Unveiling cognitive processes in digital reading through behavioural cues
A hybrid intelligence (HI) approach

Lee, Yoon; Migut, Gosia; Specht, Marcus

DOI
10.1111/bjet.13551
Publication date
2025
Document Version
Final published version
Published in
British Journal of Educational Technology

Received: 11 June 2024 | Accepted: 20 November 2024

DOI: 10.1111/bjet.13551

ORIGINAL ARTICLE

British Journal of Educational Technology BERA

Unveiling cognitive processes in digital reading through behavioural cues: A hybrid intelligence (HI) approach

Yoon Lee[1,2,3] | Gosia Migut[3] | Marcus Specht[3]

[1]Industrial Design, Graduate School of Design, University of Seoul, Seoul, Republic of Korea

[2]Learning and Educational Technology (LET) Research Lab, Faculty of Education and Psychology, University of Oulu, Oulu, Finland

[3]Faculty of Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, Delft, The Netherlands

Correspondence
Yoon Lee, Industrial Design, Graduate School of Design, University of Seoul, Seoul 02504, Republic of Korea.
Email: yoonlee@uos.ac.kr

Funding information
Technische Universiteit Delft; Oulun Yliopisto; University of Oulu profiling project Profi7 Hybrid Intelligence, Grant/Award Number: 352788; Leiden-Delft-Erasmus Centre for Education and Learning (LDE-CEL)

Learner behaviours often provide critical clues about learners' cognitive processes. However, the capacity of human intelligence to comprehend and intervene in learners' cognitive processes is often constrained by the subjective nature of human evaluation and the challenges of maintaining consistency and scalability. The recent widespread AI technology has been applied to learning analytics (LA), aiming at a more accurate, consistent and scalable understanding of learning to compensate for challenges that human intelligence faces. However, machine intelligence has been criticized for lacking contextual understanding and difficulties dealing with complex human emotions and social cues. In this work, we aim to understand learners' internal cognitive processes based on the external behavioural cues of learners in a digital reading context, using a hybrid intelligence (HI) approach, bridging human and machine intelligence. Based on the behavioural frameworks and the insights from human experts, we scope specific behavioural cues that are known to be relevant to learners' attention regulation, which is highly relevant for learners' cognitive processes. We utilize the public WEDAR dataset with 30 subjects' video data, behaviour annotation and pre–post tests on multiple choice and summarization tasks. We apply the explainable AI (XAI) approach to train the machine learning model so that human evaluators can also understand which behavioural features were essential for predicting the usage of the cognitive processes (ie, higher-order thinking skills [HOTS] and lower-order thinking skills [LOTS]) of learners, providing insights for the next-round feature engineering and intervention design. The result indicates that the dominant use of attention regulation behaviours is a reliable indicator of low use of LOTS with 79.33% prediction accuracy, while reading speed is a valuable indicator for predicting the overall usage of HOTS and LOTS, ranging from 60.66% to 78.66% accuracy, highly surpassing random guess of 33.33%. Our study demonstrates how various combinations of behavioural features supported by HI can inform learners' cognitive processes accurately and interpretably, integrating human and machine intelligence.

KEYWORDS

hybrid intelligence (HI), explainable AI (XAI), behaviour analysis, cognitive processes, digital reading, attention regulation behaviours

Practitioner notes

What is already known about this topic

- Human attention is a cognitive process that allows us to choose and concentrate on relevant information, which leads to successful learning.
- In affective computing, certain behavioural cues (eg, attention regulation behaviours) are used to indicate learners' attentional states during learning.

What this paper adds

- Attention regulation behaviours during digital reading can work as predictors of different levels of cognitive processes (ie, the utilization of higher-order thinking skills [HOTS] and lower-order thinking skills [LOTS]), leveraged by computer vision and machine learning.
- By developing an explainable AI model, we can predict learners' cognitive processes, which often cannot be achieved by human observations, while understanding behavioural components that lead to such machine decisions is critical. It can provide valuable machine-driven insights into the relationship between humans' external and internal states in learning.
- Based on the frameworks spanning cognitive AI, psychology and education, expert knowledge can contribute to initial feature selection and engineering for the hybrid intelligence (HI) model development and next-round intervention design.

Implications for practice and/or policy

- Human and machine intelligence form an iterative cycle to build a HI to understand and intervene in learners' cognitive processes in digital reading, balancing each other's strengths and weaknesses in decision-making. It can eventually inform automated feedback loops in widespread e-learning, a new education norm since the COVID-19 pandemic.
- Our framework also has the potential to be extended to other scenarios with digital reading, providing concrete examples of where human intelligence and machine intelligence can contribute to building a HI. It represents more systematic supports that apply to real-life practices.

INTRODUCTION

Digital technologies have transformed how we engage with educational materials (Järvelä et al., 2021). With the increasing use of digital texts in formal and informal education (Hussain et al., 2015), understanding learners' cognitive processes in digital reading has become more critical (Shaughnessy, 2020). It is a foundation for learning analytics (LA) and designing timely and effective interventions for learners who engage in digital reading (Wang, 2018). However, sensor-based laboratory experiments often used in LA challenge evaluating learners' natural cognitive processes by changing the nature of real-life digital reading and the ecosystems with intrusive sensor implementations (Li et al., 2016) and experimental design. In this sense, our work aims to bridge learner behaviours and their cognitive states in real life, leveraged by AI technologies, with a multimodal WEDAR dataset (Lee & Specht, 2023b) that premises a real-life digital reading with a webcam-based framework.

The existing approaches to digital reading assessment on cognitive dimension have predominantly relied on eye trackers (Bixler & D'Mello, 2016; Hutt et al., 2019). It is because indicators, such as pupil dilation (Wang, 2011), fixation and saccades (Salvucci & Goldberg, 2000), work as objective and solid cues for understanding learners' cognitive states, supported by previous work. At the same time, reading is a straightforward task with regular eye movement patterns (eg, character-level fixations (Yan et al., 2022), scanning and skimming (Liu, 2012), area of interest (AOI, Popa et al., 2015), number of blinks (Roschke & Radach, 2016), re-reading (ChanLin, 2013)), making it a solid indicator of evaluating the cognitive demands in digital reading. Various multimodal indicators, such as video data (eg, valence, arousal (Wang, 2018)) and multiple layers of log data (eg, mouse dynamics (Li et al., 2016)), have been combined with other features for the more multidimensional understanding of learner states and learning. However, feature-based analysis has suffered from lacking standards for defining ideal learner features, which is often the case for digital reading analytics, too (Wang, 2018).

Based on multimodal LA, learners' cognitive states, such as mind-wandering (Bixler & D'Mello, 2016; Hutt et al., 2019), switches of internal thoughts (Huang et al., 2019), working memory (Li et al., 2016) and affects (eg, valence, arousal (Wang, 2018)) have been the target of previous analyses. In the process, self-reported data showing learners' subjective perceptions about their learning and experts' observations have often been used as ground truths for machine reasoning (Lee, Limbu, et al., 2023). In this context, different physiological patterns found in learners with and without successful learning outcomes have been targets of the machine learning model training in previous work (Liu et al., 2023).

Models from the previous framework aimed at finding critical features that are automatically learned in the model training processes in optimal ways, focusing on the accuracy of models. However, due to the non-explainable nature of black-box AI applied in the previous frameworks (Gohel et al., 2021), there is a growing need for explainable AI (XAI) in education to understand the reasoning behind the model's decision (Alonso & Casalino, 2019), which is often fundamental objectives of LA; thus, practitioners can act upon the analysis and further design and implement relevant interventions.

To address the limitations of previous research, our work aims to bridge the gap by utilizing a non-intrusive computer vision approach and developing an XAI model for cognitive process assessment based on learners' behaviours in digital reading. This approach identifies and analyses critical features to predict learners' cognitive processes. Our model offers significant support to learners by providing insights into their cognitive processes, a capability beyond the reach of human educators alone. By linking these LA results to interventions supported by human educators' insights, digital reading can benefit from a human–AI collaborative intervention loop at scale, where both human and machine intelligence contribute their unique strengths to enhance learning outcomes.

We focus on digital reading since, especially in higher education, the significance of digital reading is substantial, as learners are required to independently process, comprehend, retain and apply knowledge gained through reading as part of their regular coursework. Digital reading directly influences learning outcomes, self-efficacy and overall academic success (Lee, Migut, & Specht, 2023b). Traditionally, post hoc analysis using log data has been employed to assess digital learning interventions (Jivet et al., 2018). However, with the rise of technology-enhanced learning (TEL) approaches—supported by sensing and machine learning technologies—there have been advancements in real-time LA and personalized learning support. Despite these technological developments and the critical role of digital reading in higher education, the design and implementation of interventions for digital reading remain limited (Lee, 2024).

Also, digital reading involves distinct cognitive demands compared to traditional reading, such as multitasking, visual scanning and interactive features. These differences can affect comprehension and retention, increasing the cognitive load (Brüggemann et al., 2023). Moreover, digital environments with various platforms and peripheral devices allow for personalized feedback and behavioural tracking, enabling TEL-based interventions (Lee, Limbu, et al., 2023). In this context, our study explores how these unique aspects of digital reading impact learners' cognitive processes, specifically in utilizing higher-order thinking skills [HOTS] and lower-order thinking skills (HOTS) and lower-order thinking skills (LOTS).

Below, we articulated three research questions that we focused on in our study.

- RQ1. How can learners be clustered based on their use of cognitive processes during digital reading to inform intervention design?

By addressing the first research question, we aimed to identify how learners can be clustered based on their cognitive processes (ie, HOTS and LOTS) to guide the design of personalized, data-driven interventions tailored to their specific learning needs. We implemented and compared two clustering methods—statistical and unsupervised—that were linked to different behavioural cues and thus can inform different targeted intervention strategies for each learner segment.

- RQ2. How can an automated system evaluate learners' use of cognitive processes during digital reading while incorporating human insights for more effective interventions?

The second research question focused on exploring how machine learning models can assess learners' cognitive skills at scale, while also integrating human expertise to enhance the effectiveness and personalization of interventions. To answer this, we developed XAI models that accurately predict learner clusters with an explainable decision-making process, making it possible to scale more personalized interventions.

- RQ3. What key behavioural indicators can machines use to predict learners' cognitive processes and form an iterative cycle between human-machine intelligence?

The final research question sought to identify the key behavioural indicators that predict different cognitive processes, enabling a continuous cycle of machine-driven insights and human interventions. This approach facilitates the development of expert-informed interventions alongside automatic cognitive process predictions, enhancing the digital reading experience.

All in all, this work suggests a hybrid intelligence (HI) framework for human attention analysis in digital reading (see Figure 1). As suggested as an essential challenge for future

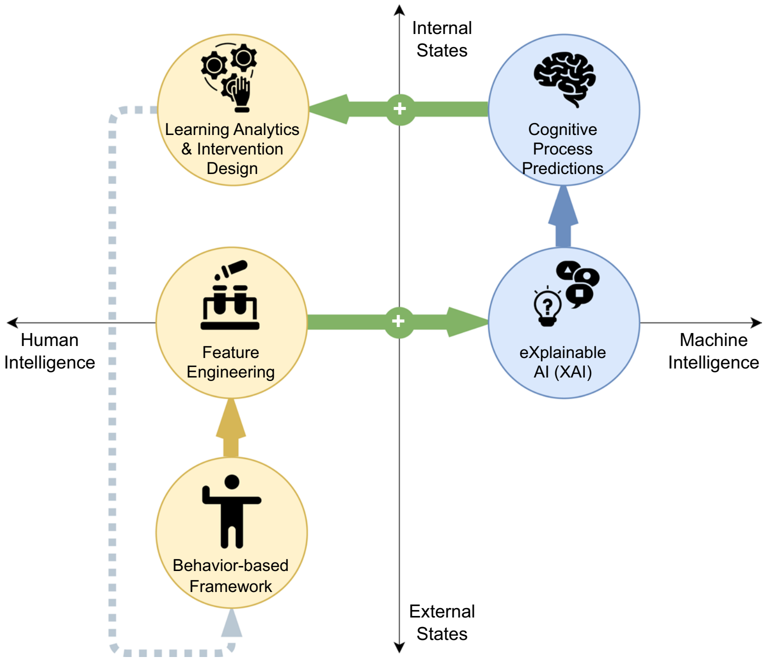


FIGURE 1 Our approach aimed at building a hybrid intelligence (HI) in digital reading by introducing a behaviour-based cognitive AI framework. Human intelligence provides the first insights for machine model training, while machines can achieve behaviour-based cognitive process prediction, which human evaluators cannot achieve. Humans investigate and interpret the explainable model to gain insights for the next-round LA and intervention design.

AI applications in education (Molenaar, 2022), we strived to balance the human insights from the experts and understand critical components for machine reasoning via the XAI approach. Our framework is (1) based on the behaviour-based frameworks. (2) Using human experts' insights, we select behavioural features that we hypothesize to correlate with human attention for the machine learning model training. (3) We take the XAI so that we can trace behavioural features for predicting learners' cognitive processes. (4) Using the model, we evaluate the internal states of learners via external behavioural cues and make the prediction automatically at scale, which has been a common challenge that human educators face in providing real-life e-learning support. Also, the results of the analysis could be linked to the next-round intervention design, where human intelligence can occur in iteration next to the machine's decisions.

THEORETICAL BACKGROUND

In this section, we investigate various behavioural indicators known to be directly and indirectly correlated to learners' attention that construct our behaviour-based framework. As illustrated in Figure 2, we mainly utilize the dataset and framework of learners' attention regulation investigated in a real-life digital reading scenario (Lee et al., 2022). We leveraged the XAI approach to understand learners' cognitive processes, especially the utilization of HOTS and LOTS on various levels during digital reading, representing distinct levels of cognitive engagement, from basic information recall to complex analysis and synthesis, reflecting how learners process and apply knowledge.

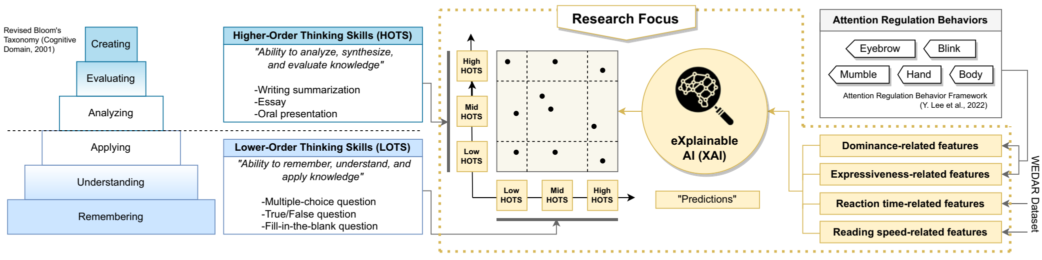


FIGURE 2 Our framework is based on the revised Bloom's taxonomy (Krathwohl, 2002), which has HOTS and LOTS as components of learners' cognitive processing. Using decision trees, we strived to predict learners' HOTS and LOTS based on attention regulation behaviours (Lee et al., 2022). We tried to understand the reasoning behind the model's decision to find the critical behavioural components for predicting different levels and combinations of learners' cognitive processing. HOTS, higher-order thinking skills; LOTS: lower-order thinking skills.

Current XAI approaches in AI in education (AIED)

XAI is an emerging topic, especially in areas where the reasoning behind decision-making is especially critical (eg, healthcare, law, autonomous driving (Alonso & Casalino, 2019; Richmond et al., 2024)). In education, LA and educational data mining (EDM) are two areas where AI-driven approaches commonly take place for various stakeholders (eg, teachers, tutors, students and managers (Khosravi et al., 2022)) for collecting, processing, exploiting and reporting the learning data (Alonso & Casalino, 2019). However, while machine learning models can successfully assist and complement humans in tasks with classification, regression, clustering, transferring and optimization capabilities, black-box AI has a limitation in that researchers do not directly understand the reasoning behind the models' decisions (Cukurova et al., 2020). Therefore, understanding the specific task in feature engineering and result interpretation from human experts has been considered critical (Khosravi et al., 2022) while the limited explainability of models still raises ethical and trustworthiness issues for educational applications, lacking transparency, trust and fairness in machine-based decision-making (Gohel et al., 2021). Therefore, various XAI frameworks have been introduced in educational research, focusing on revealing the feature dominance, correlation among features used for the training, the reasoning behind predictions (Gohel et al., 2021) and sources of noise in the decision-making (Khosravi et al., 2022).

Khosravi et al. (2022) have suggested a framework of XAI in Education (XAI-ED) that aligns the needs of stakeholders, interfaces and AI models. Various XAI approaches in education, such as the generalized additive model (GAM) with a linear relationship between independent and dependent variables (Dikaya et al., 2021), the decision tree model with a hierarchical structure, a rule-based model with conditional statements, the clustering method with specific data patterns and natural language processing with data cross-validations among learning data have been introduced. Gohel et al. (2021) introduced various XAI approaches and made baseline comparisons of different state-of-the-art methods with multiple modalities and features through a survey. In the work, XAI applications have been explained as transparent methods (eg, Bayesian model, decision trees, linear regression, fuzzy inference systems), explaining superficial relationships among features, while post hoc methods (eg, LIME, perturbation, LRP, SHAP) for the task with higher data complexity. Alonso and Casalino (2019) have suggested the GUI web-based ExpliClas, which provides text descriptions and a dashboard with data visualizations regarding the feature use and recommendations. Cukurova et al. (2020) implemented a decision tree to find critical features among learners' listening, watching, making, and speaking behaviours to predict and understand collaborative problem-solving competencies.

All in all, the general focus has been finding XAI implementation opportunities in education with model comparisons and platform suggestions. However, according to our best knowledge, neither XAI in behaviour analysis (ie, learners' external states) for understanding learners' cognitive processes (ie, learners' internal states) nor XAI for digital reading applications has yet to be attempted. It is essential for the rapidly growing necessity of LA and feedback loop design for real-life digital reading, supported by widespread computer vision, which we fundamentally aim to foster in line with hybrid human and machine intelligence.

Our contributions to the HI in education

HI in education is an emerging area that aims to integrate human capabilities—such as adaptability, collective productivity, socio-emotional skills and self-regulation throughout learning—with AI (Järvelä et al., 2023). Cukurova (2024) highlighted several limitations in the current AIED field, including (1) an excessive focus on the effectiveness of specific AI models, which fails to capture the complexity of educational processes; (2) the potential for dehumanizing education through automation and prioritizing data collection; (3) reduced student motivation when interacting with AI tools, which limits their benefits; and (4) a lack of attention to social and cultural dimensions, as AI introduces a new educational ecosystem. In response, the HI approach has been proposed to address AI shortcomings in educational contexts (Cukurova, 2024).

While existing AI models have limitations in fully explaining learning processes, HI holds promise for bridging the gap between LA and key educational areas like feedback, motivation, awareness and learner contributions (Cukurova, 2024). Also, Molenaar (2022) emphasized developing collaborative, adaptive, responsible and explainable HI, which bridges human cognition and AIED. To strengthen the theoretical framework, we build on existing HI frameworks, behavioural LA and learners' cognitive dimensions, aiming to address gaps in the current literature. Building on the work of Järvelä et al. (2023), who developed an HI model to understand cognitive self-regulation in collaborative learning through a hybrid human-AI shared regulation in learning (HASRL) model, our study extends this understanding to the mental processes and self-regulation of learners in independent digital reading scenarios, understood via behavioural cues. All in all, by utilizing the HI approach, our potential contributions to the AIED in digital reading are listed below.

1. *Simple understanding of learners' cognitive processes through behavioural cues*: Our study represents the first attempt to apply XAI to understand learners' cognitive processes in digital reading. We can grasp complicated cognitive processes via combinations of observable learner behaviours and simple webcam implementation. By using our behaviour-based cognitive process predictions with the XAI approach, we can identify critical behavioural indicators for the machine predictions. Traditionally, LA on learners' cognition has been done via dedicated biosensors, such as an eye tracker (Liu et al., 2023). It has challenged educational researchers with complicated hardware implementations, hindering learning processes in real life with intrusiveness. Moreover, combinations of multimodal data streams with different granularity and black-box models often challenged the alignment of the LA and intervention design due to complex and uninterpretable machine decisions. However, our webcam-based behaviour analysis with XAI directly reveals the relationship between semantically understandable behavioural cues and learners' hidden cognitive processes in learning without destructiveness.
2. *Future extension of the framework to intervention strategies and interaction design*: Using the XAI approach, we strived for a semantic understanding of the influential features of

machine reasoning with objectivity and scalability. Understanding prediction mechanisms related to different cognitive processes of individuals provides valuable insights to instructional designers for more concrete and adaptive intervention plans (Khosravi et al., 2022). For instance, the framework can extend to diverse learning strategies and interventions for digital reading. Our framework can connect diverse interfaces with various feedback strategies (eg, conversational agents), successfully closing the feedback loop with learning's behavioural, cognitive and affective enhancements.

Understanding learners' cognitive processes: Use of HOTS and LOTS in digital reading

Understanding how learners utilize different thinking skills is crucial for gaining insight into their cognitive processes (Krathwohl, 2002). These skills shape the pace (Tanujaya et al., 2017) and effectiveness (Heong et al., 2011) of learning. Krathwohl's (2002) revised Bloom's taxonomy identifies HOTS and LOTS as essential cognitive objectives. LOTS include remembering, understanding and applying knowledge, while HOTS involve analysing, evaluating and creating (Qasrawi & BeniAbdelrahman, 2020). Despite their critical role in understanding cognitive processes, there remains a gap in research that links these cognitive dimensions to HI using behavioural frameworks, especially within digital reading contexts. Addressing this gap serves as the main motivation for our study, which has never been attempted, according to our best knowledge.

We hypothesize that the distinct nature of HOTS and LOTS can be observed through behavioural cues, such as attention regulation, which correlate with distraction and attention management (Lee et al., 2022; Lee & Specht, 2023a). Through an Explainable AI (XAI) approach that emphasizes behaviour-based prediction, we aim to identify critical behavioural markers—such as the prominence of attention regulation behaviours, reaction time, and reading speed—that could serve as observable indicators of HOTS and LOTS during digital reading. This understanding can guide practitioners in designing targeted interventions based on learners' cognitive processes and their external behaviours.

Our study uses Krathwohl's framework to categorize thinking skills into LOTS (ie, remembering, understanding, applying) and HOTS (ie, analysing, evaluating, creating). LOTS support foundational cognitive processes, often relying on short-term memory, and are assessed through simple question formats (Abosalem, 2016; Narayanan & Adithan, 2015). HOTS, in contrast, involves deeper analysis and synthesis, enabling long-term retention and knowledge transfer (Mainali, 2012; Qasrawi & BeniAbdelrahman, 2020). Understanding learners' cognitive processes is complex, as some students may excel in LOTS but struggle with HOTS, while others may exhibit the opposite pattern (Abosalem, 2016; Qasrawi & BeniAbdelrahman, 2020). Therefore, recognizing the varied use and interplay of HOTS and LOTS is crucial for researchers and educators, as it enables the design of tailored interventions that address different cognitive needs, fostering personalized learning experiences that better support each learner's development.

Behaviour-based framework for evaluating learners' HOTS and LOTS in digital reading

This study identifies critical features from existing behavioural frameworks that we hypothesize are correlated with learners' HOTS and LOTS, using human expertise and domain knowledge. Our primary focus is on learners' behavioural traits that can inform the design of subsequent interventions, specifically: (1) the dominance of attention regulation behaviours,

(2) the expressiveness of attention regulation behaviours, (3) reaction time to secondary blur stimuli and (4) reading speed. These indicators have previously been explored as direct and indirect cues for understanding learners' cognition in digital reading contexts.

Attention regulation behaviours

Lee et al. (2022) focused on behaviours that have been defined as voluntary and spontaneous actions of learners to regain attention, especially during digital reading context: behaviours in eyebrows (eg, raising, bringing together), blinks (eg, blink flurries, voluntary prolonged blink), mumble (eg, mumble reading), hand (eg, touching body and or face) and body (eg, adjusting position and or angle of torso, arm). Those behaviours were correlated with self-aware distractions in digital reading, which provides essential context for the automatic attention regulation behaviour recognition leveraged by video-based distraction recognition (Lee, Migut, & Specht, 2023a). Also, attention regulation behaviours were understood as help-seeking behaviour of self-regulated learning (SRL) that involves phases of goal-setting, self-monitoring, help-seeking and self-evaluating, where intervention can greatly assist learners with strategic approaches (Järvelä et al., 2023; Zimmerman & Moylan, 2009). Based on the framework of Lee et al. (2022) as a foundational framework and the public WEDAR dataset, we develop an XAI model to find which specific attention regulation behaviours inform learners' usage of HOTS and LOTS in digital reading. Our primary aim was to fill the knowledge gap in LA and intervention design. We implemented a behaviour-based XAI approach because traditional black-box models do not clearly explain the reasoning behind their decisions, which is essential for informing intervention strategies. By identifying specific behavioural patterns linked to cognitive processes, we enable human educators to design more effective instructional strategies and integrate them with automatic feedback systems.

Please note that in the feature engineering process, we focused on behavioural patterns identified in the literature, hypothesizing that these patterns correlate with cognitive processes during digital reading. While it is still unclear whether attention regulation behaviours cause or result from attention loss, they are consistently linked to self-reported attention lapses, making them valuable cues for post hoc intervention timing. We consider these behaviours as indicators of 'perceived' cognitive states that signal the need for intervention based on previous work (Lee et al., 2022). By assuming that attention regulation affects cognitive engagement differently, we aim to refine our interventions through iterative post hoc analysis to address these cognitive states better.

Dominance and expressiveness of attention regulation behaviours

Contextual features, such as individual and cultural factors (Greenaway et al., 2018), are known to highly influence human behaviours' frequency and expressiveness (Zuckerman, 1994). Such individual differences in behaviours often challenge generalized behaviour-based LA (Zunino et al., 2017). In this study, we aimed to investigate such differences in line with the usage of attention regulation behaviours and learners' cognitive processes.

Reaction time to the screen blur at randomized timing during digital reading

Reaction time has long been a reliable indicator of learners' arousal and attention during task performances (Huang et al., 2019). Fast reaction time is commonly associated with efficient attentional control and vigilance (van Kempen et al., 2019), indicating the ability to maintain focus and allocate cognitive resources effectively. Conversely, slow reaction time has been suggested as disengagement from the task and challenges sustaining an optimal attentional state amidst distractions (Huang et al., 2019). The influence of affective states, including arousal and engagement, is known to shape individuals' reaction time differently (Bless & Fiedler, 1995), representing its potential correlations to cognitive processes in learning. Given the suggested insights, we hypothesized that reaction time to the screen blur could work as a feature that robustly predicts the utilization of learners' HOTS and LOTS during their digital reading.

Reading speed

Reading speed provides valuable insights into learners' cognitive load and information-processing capabilities (Brysbaert, 2019). Reading speed is often influenced by the complexity of the material (ie, intrinsic cognitive load) and the way the information is presented (ie, extraneous cognitive load). Faster reading speeds might indicate lower cognitive load when the material is simple or familiar. In contrast, slower speeds could signal an increase in cognitive effort, either due to challenging content (eg, intrinsic load) or poor presentation design (eg, extraneous load; Orru & Longo, 2019). Though faster reading does not guarantee better learning, it is often associated with more rapid information gain and reduced cognitive load compared to slower readers (Wirth et al., 2020). Moreover, fast readers of screen-based reading are known to experience fewer distractions (Dyson & Haselgrove, 2001), which supports our attempt to predict cognitive processes based on learners' reading speed in digital reading. In this regard, we hypothesized that higher attention and faster reading speed would enhance HOTS and LOTS during digital reading.

METHODS

This chapter introduces how we preprocessed the multimodal WEDAR dataset to train an XAI model to predict various cognitive learner clusters based on behaviours. We specifically applied the Bidirectional Encoder Representations from Transformers (BERT) method to the WEDAR dataset for the HOTS evaluation. Also, in this chapter, we derived various combinations of LOTS and HOTS, using an unsupervised *k*-means clustering method and statistical quartile analysis to define the target features (ie, various cognitive processes) of our behaviour-based prediction model.

Multimodal WEDAR dataset

This study used the WEDAR dataset (see Figure 3) collected from 30 higher education learners during computer screen-based digital reading. The dataset includes learning results assessed via multiple-choice questions and text summarization, moment-to-moment self-reported distractions, learners' reaction time to the randomized screen blur and attention regulation behaviours annotated every second from video samples from 30 subjects, approximately 8.7 hours long.

Please note that this study only utilized post hoc features because HOTS and LOTS were not collected in real time. Predicting post hoc targets (ie, HOTS and LOTS) based on real-time behavioural features could be misleading. This is because cognitive processes like

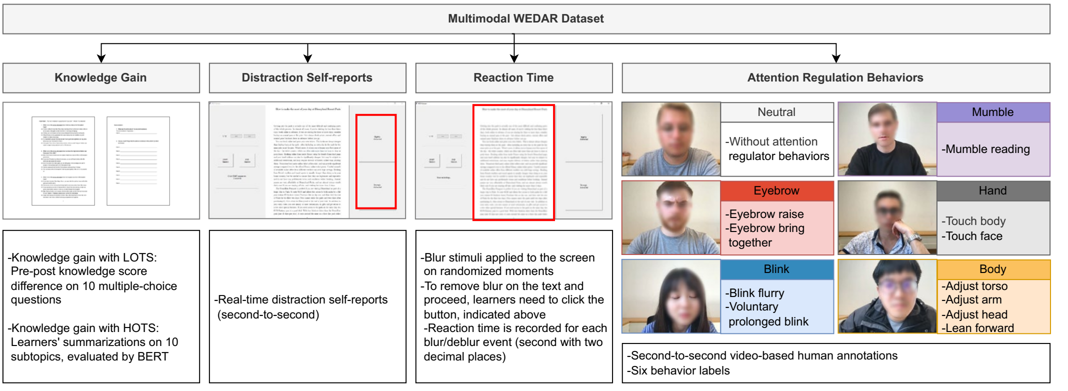


FIGURE 3 Our work utilized the knowledge gained for assessing HOTS and LOTS, distraction self-reports, reaction time to screen blur stimuli, and attention regulation behaviours of learners from a multimodal WEDAR dataset.

HOTS and LOTS were evaluated post hoc as part of a summative assessment. The utilization of these cognitive processes requires a sufficient amount of input, which makes them unsuitable for formative, real-time assessment.

Furthermore, collecting data for learner assessment in real-time could alter the nature of the e-reading activity itself, which was not the intention of our study. Our goal was to capture the natural use of cognitive processes without influencing learners' behaviour during the task.

Understanding different learners based on their cognitive processes

This section answers '*RQ1. How can learners be clustered based on their use of cognitive processes during digital reading to inform intervention design?*' First, we introduce how LOTS and HOTS of different learners were assessed using the difference between the pre–post scores of the multiple-choice questionnaire and summarization tasks. Second, we introduce learner segmentation methods (1) with an unsupervised learning model with *k*-means clustering and (2) with statistical quartile analysis. Please refer to the WEDAR dataset (Lee & Specht, 2023b) for details on the reading materials used in the digital reading tasks. See Appendix A for the pre–post questionnaire design.

Evaluating LOTS: Pre–post multiple-choice questionnaire

In the WEDAR, 10 multiple-choice questions related to the reading materials were given before and after the reading (pretest, posttest) to evaluate LOTS. We calculated the LOTS by subtracting the pretest score from the posttest score, making the final LOTS ranging from a scale of 0 to 10;

$$\text{Score}_{\text{LOTS}} = \sum_{i=1}^{Npost} S_i^{post} - \sum_{i=1}^{Npre} S_i^{pre}, \tag{1}$$

where S_i^{post} is the posttest score (0 or 1) while S_i^{pre} is the pretest score (0 or 1) for question i. Note that the pretest and posttest multiple-choice questionnaire content were the same.

Evaluating HOTS: BERT applied to the text summarization

A summarization questionnaire has been given only in the posttest to evaluate learners' HOTS. We found that evaluating summarized text by human evaluators can be subjective; thus, we utilized the automatic evaluation technique, the BERT, a type of natural language processing model (NLP, Devlin et al., 2018). BERT is a widely used language model that can handle various language tasks while considering contexts. It is especially relevant to understanding the similarity of learners' summarization (ie, inputs for evaluation) and the original text (ie, ground truth) that represents learners' ability to reconstruct the contents they read.

$$R_{\mathrm{BERT}} = \frac{1}{|X|} \sum_{x_i \in x} \max_{\hat{x}_j \in \hat{x}} \mathbf{x_i}^{\top} \hat{\mathbf{x}}_{\mathbf{j}}, \tag{2}$$

$$P_{\mathrm{BERT}} = \frac{1}{\left|\hat{X}\right|} \sum_{\hat{x}_j \in \hat{x}} \max_{x_i \in x} \mathbf{x_i}^{\top} \mathbf{x_j}, \tag{3}$$

$$F_{\mathrm{BERT}} = 2 \frac{P_{\mathrm{BERT}} \cdot R_{\mathrm{BERT}}}{P_{\mathrm{BERT}} + R_{\mathrm{BERT}}}, \tag{4}$$

Based on BERT, we evaluated participants' summaries using precision (P_{BERT}), recall (R_{BERT}) and F1 scores (F_{BERT}), ranging from 0 to 1. The ground truth summaries (x) were compared with participant summaries ($\hat{x}$). Above, X represents token vectors from the ground truth, while $\hat{X}$ represents token vectors from participant summaries. x_i and $\hat{x}_j$ are vectors within X and $\hat{X}$, respectively, with the dot products ($\mathbf{x_i}^{\top}\hat{\mathbf{x}}_{\mathbf{j}}$ and $\mathbf{x_i}^{\top}\mathbf{x_j}$) used to calculate the average maximum similarity between tokens. We used the F_{BERT} score to measure each learner's HOTS, as it balances precision and recall for a more comprehensive assessment.

Initially, we considered using ChatGPT to evaluate the quality of summaries due to its powerful text generation capabilities supported by Generative AI. However, we ultimately adopted the BERT model because it allows us to calculate the similarity between the embeddings of the original text and the student's summary, providing a fine-grained analysis of how closely the summary reflects the original meaning. In contrast, ChatGPT may generate plausible but not necessarily factually accurate text when used for tasks it was not explicitly trained for, such as detailed text comparison and summarization evaluation. Additionally, our BERT-based model was fine-tuned for key phrase extraction, enabling it to determine whether the student's summary captures the essential information from the original text without introducing irrelevant details (Zhong et al., 2023).

Note that we acknowledge that the choice of language, specific wording and potential typos (Zhong et al., 2023) could influence the similarity score when evaluating HOTS. To mitigate these concerns, the original dataset used in this study already included corrections for typos and minor grammatical errors. This was done to compensate for disadvantages faced by second-language learners, ensuring a fairer assessment for all participants since most learners in this experiment were non-native English speakers and the educational environment used English as the primary language.

Learner segmentation based on combinations of HOTS and LOTS

As suggested in the previous work (Abosalem, 2016; Qasrawi & BeniAbdelrahman, 2020), we implemented a machine learning method and statistical analysis method, respectively.

It was to explore different levels and combinations of learners' HOTS and LOTS: (1) with *k*-means clustering for machine-based clustering and (2) quartile analysis to define thresholds for high (first quartile: Q1), mid (second quartile: Q2) and low (third quartile: Q3) ranges of HOTS and LOTS. These categorizations provide insight into how learners can be divided using HOTS and LOTS and their diverse combinations, which can further be combined with various feedback strategies for different feedback objectives.

Learner segmentation based on unsupervised method: k-means clustering

As can be seen from Figure 4, we performed *k*-means clustering (Trivedi & Patel, 2020) using HOTS and LOTS as feature vectors to segment the learners in an unsupervised way. We determined *k* as 3 based on the elbow method and made three clusters. The clustering results helped to define one group of learners (C2) with a relatively low LOTS range and two groups (C1 and C0) with a comparatively higher LOTS range. One of the groups with high LOTS (C1) demonstrated a higher HOTS, while the other (C0) exhibited a lower HOTS. The sample consisted of 12 learners in C0, 11 learners in C1 and 7 learners in C2, respectively. Note that we standardized HOTS and LOTS by mean-max scaling (Shanker et al., 1996) for both segmentations, subtracting and scaling the mean to unit variance for a fair comparison of HOTS and LOTS with different data ranges of 0–1 and 0–10, respectively.

Learner segmentation based on quartile analysis: Defining the high, mid and low ranges of HOTS and LOTS

As can be seen from Figure 5, we also conducted a quartiles analysis (Goffin et al., 2009) to define thresholds for high, mid and low ranges of HOTS and LOTS. By doing so, we can understand who achieved how in HOTS and LOTS in each of the three levels, which could be fundamental for developing different feedback. Based on quartile analysis, learners were grouped into three performance categories: the top 25% (Q1) were classified as high achievers, the middle 25–75% (Q2) as mid-range and the bottom 25% (Q3) as low achievers. This categorization applied to both HOTS and LOTS. By clustering learners according to their quartile rankings, we identified nine distinct segments (3 HOTS * 3 LOTS) based on their cognitive processes. Additionally, each quartile label was one-hot encoded to determine if the model could predict whether a learner belonged to a specific quartile (eg, high HOTS was coded as 1, while learners without high HOTS were coded as 0).

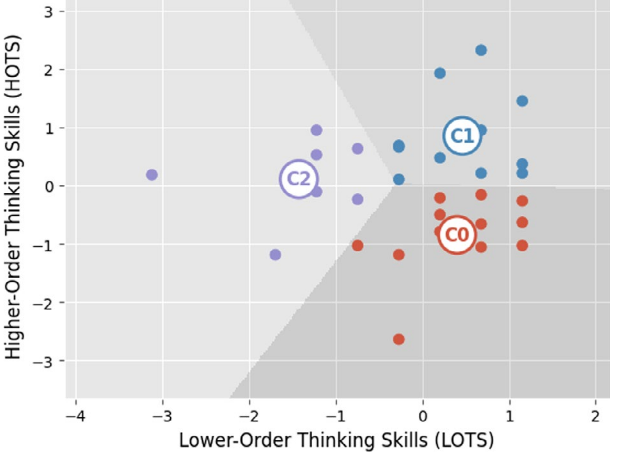


FIGURE 4 Learner segmentation based on unsupervised method: *k*-means clustering.

FIGURE 5 Learner segmentation based on quartile analysis: Defining the high, mid and low ranges of HOTS and LOTS.

Feature engineering for the behaviour-related features for the XAI model training

To answer '*RQ2. How can an automated system evaluate learners' use of cognitive processes during digital reading while incorporating human insights for more effective interventions?*', we list various features that are used independently and in combinations to build XAI models in effectively predicting the usage of learners' cognitive processes (see Table 1). We extracted 19 features across five categories from the multimodal WEDAR dataset. These categories included dominance-related, expressiveness-related, reaction time-related and reading speed-related features, which served as predictors of HOTS and LOTS, the prediction targets, since we hypothesize that the first four categories make semantic correlations with HOTS and LOTS.

Model training protocols

For the XAI model development, we implemented the decision tree model for its simple implementation and straightforward result interpretation. Also, the decision tree has advantages in that it can automatically exclude irrelevant features and include only influential features by calculating the Gini impurity in its training process while achieving robust prediction as a classical machine learning model (Cukurova et al., 2020). The prediction target has been: (1) multi-class prediction of clusters derived from *k*-means and (2) binary prediction of high, mid and low usage of HOTS and LOTS, based on the thresholds derived from the quartile analysis.

It is important to highlight that we used different target levels in prediction (ie, multi-class and binary) because each approach is better suited to the specific characteristics of its target. The clusters derived from the *k*-means method can change dynamically as new data points are added (Shafeeq & Hareesha, 2012). Therefore, it is more meaningful to focus on understanding the features that distinguish these clusters rather than the clusters themselves, making multi-level classification a more appropriate approach.

On the other hand, quartiles represent fixed, interpretable and distinctive data segments—low, mid and high (Brown et al., 2008). This stable structure makes binary classification more

TABLE 1 Preprocessed multimodal WEDAR dataset, which has been used for the XAI model training.

Feature categories	#	Feature names	Feature description	Categorical/nominal
Dominance-related	F1	Behaviour_eyebrow	Number of eyebrow behaviours/total number of attention regulation behaviours	Continuous (0–1)
(Attention regulation behaviour)	F2	Behaviour_blink	Number of blink behaviours/total number of attention regulation behaviours	Continuous (0–1)
	F3	Behaviour_mumble	Number of mumble behaviours/total number of attention regulation behaviours	Continuous (0–1)
	F4	Behaviour_hand	Number of hand behaviours/total number of attention regulation behaviours	Continuous (0–10)
	F5	Behaviour_body	Number of body behaviours/total number of attention regulation behaviours	Continuous (0–1)
	F6	First_behaviour (one-hot encoded)	Occurrences of having the most dominant attention regulation behaviours	5-classes (0, 1)
	F7	Second_behaviour (one-hot encoded)	Occurrences of having the second dominant attention regulation behaviours	5-classes (0, 1)
	F8	Third_behaviour (one-hot encoded)	Occurrences of having the third dominant attention regulation behaviours	5-classes (0, 1)
Expressiveness-related	F9	Expressiveness	Number of attention regulation behaviours/duration of the video	Continuous (0–1)
(Attention regulation behaviour)	F10	Exp_level (one-hot encoded)	Low (Q1), mid (Q2), high (Q3) expressiveness levels of each participant	3-classes (0, 1)
Reaction time-related	F11	Indiv_reaction_average	Reaction time average of each participant	Continuous values
	F12	Reaction_time (one-hot encoded)	Fast (Q1), mid (Q2), slow (Q3) reaction time levels of each participant	3-classes (0, 1)
Reading speed-related	F13	Individ_reading_speed	Reading speed average of each participant (word/duration of reading)	Continuous values
	F14	Reading_speed (one-hot encoded)	Fast (Q1), mid (Q2), slow (Q3) reading speed levels of each participant	3-classes (0, 1)

TABLE 1 (Continued)

Feature categories	#	Feature names	Feature description	Categorical/nominal
HOTS & LOTS	F15	LOTS	Posttest score–pretest score (multiple choice, full score:10)	Continuous (0–10)
	F16	HOTS	BERTScore calculated based on written summarizations	Continuous (0–1)
	F17	LOTS_ (high, mid, low, one-hot encoded)	Low (Q1), mid (Q2), high (Q3) LOTS of each participant	3-classes (0, 1)
	F18	HOTS_ (high, mid, low, one-hot encoded)	Low (Q1), mid (Q2), high (Q3) HOTS of each participant	3-classes (0, 1)
	F19	Clusters (C0, C1, C2)	3 clusters derived from *k*-means with LOTS and HOTS as feature vectors	3-classes (0, 1)

effective, allowing for a clear and straightforward analysis of whether specific traits contribute to predictions based on defined thresholds. Therefore, our methodological choices are carefully tailored to the unique properties of each data type, ensuring we leverage the strengths of each classification approach.

Due to the limited sample size for predicting the post hoc results, we employed the Synthetic Minority Oversam-pling TEchnique (SMOTE) to generate extra samples and match the number of samples to the largest class (Chawla et al., 2002). We applied a 2:1 ratio to make the sample size twice bigger than the original dataset to compensate for the shortage of the number of samples. Following the standard sampling method, we divided the training and testing sets into 80% and 20%, respectively. To further enhance model performance, we applied a pruning method to remove branches that provided little information during the decision model training process. We used fivefold cross-validation to train and evaluate the model, averaging the results across folds to account for the limited sample size. This approach helped reduce the variance caused by different data splits in each fold and mitigated the risk of overfitting potentially introduced by the oversampling process. We used one-hot-encoded LOTS (F17) and HOTS (F18) as the training targets to achieve the quartile prediction. For cluster prediction, we set *k*-means-driven clusters (F19) as the training target. We used dominance-related (F1–F8), expressiveness-related (F9 and F10), reaction time-related (F11 and F12) and reading speed-related (F13 and F14) features as predictors of HOTS and LOTS, in combinations and independently. Please refer to Table 2 for the accuracy of predictions. We further conducted the feature importance analysis in the later section to understand critical behavioural components that are used for predicting the cognitive processes by machines.

RESULTS

Accuracy of the model prediction with different feature categories

As shown in Table 2, using all feature categories led to the best prediction performances for predicting the three cognitive process clusters derived from *k*-means, achieving an accuracy of 72.00%, highly surpassing the random guess made by 33.33%. When examining the prediction results from individual feature categories, accuracy ranged from 27.99% for reaction-time-related features to 62.00% for dominance-related features. This suggests that dominance-related features heavily influence the overall prediction performance. Also, dominance-related features achieved the highest accuracy for the quartile prediction of HOTS and LOTS. Dominance-related features achieved the highest accuracy, at 79.33% for predicting the low LOTS and 72.66% for predicting the low HOTS. The result indicates that understanding the usage of attention regulation behaviours works as robust cues of learners' low utilization of HOTS and LOTS. This is particularly significant given that it is nearly impossible to accurately determine the type of thinking skills used through human observation alone, where random guessing would only yield a prediction accuracy between 33.33% and 50.00%, depending on the specific type of prediction being made. Reading speed-related features also worked as robust predictors for predicting overall quartiles of HOTS and LOTS, with accuracy ranging from 60.66% to 78.66%. On the other hand, expressiveness-related features from attention regulation behaviours were only valuable for predicting low and high LOTS levels, with an accuracy of 71.33% each, while showing limitations in predicting HOTS. Similarly, reaction time-related features have shown sound prediction results for high (67.99%) and low (75.33%) LOTS. However, comparatively poor prediction results of 44.00% were shown for HOTS and mid-range LOTS. We assume that learners with low and high ranges of HOTS and LOTS exhibit distinctions that can inform

TABLE 2 Accuracies for predicting clusters and quartiles based on the decision tree.

	Prediction objectives						
	Thinking_skills_clusters (*k*-means)	LOTS_level_ high	LOTS_level_ mid	LOTS_level_ low	HOTS_level_ high	HOTS_level_ mid	HOTS_ level_low
Features	**(F19)**	**(F17)**	**(F17)**	**(F17)**	**(F18)**	**(F18)**	**(F18)**
Random guess	33.33	50.00	50.00	50.00	50.00	50.00	50.00
All *(F1–F14)*	**72.00**	67.33	49.33	67.33	**70.66**	<u>66.66</u>	<u>72.00</u>
Dominance-related *(F1–F8)*	<u>62.00</u>	64.66	<u>63.99</u>	**79.33**	52.66	58.00	**72.66**
Expressiveness-related *(F9 and F10)*	36,66	<u>71.33</u>	52.66	71.33	38.66	49.33	40.00
Reaction time-related *(F11 and F12)*	27.99	67.99	44.00	75.33	56.66	57.33	64.00
Reading speed-related *(F13 and F14)*	42.66	**78.66**	**65.33**	<u>78.66</u>	<u>60.66</u>	**72.00**	70.66

Note: The best performances are **bolded**. The second best performances are <u>underlined</u>.

machine reasoning. In contrast, learners with mid-range HOTS and LOTS did not show consistent behaviour patterns, particularly in terms of attention regulation behaviours and reaction time to the screen blur events. Note that we set the decision tree model's maximum depth to 10 to ensure simpler interpretability and prevent possible overfitting.

Identifying significant behavioural predictors for understanding learners' cognitive processes

In this section, we examine plot trees (see Appendix B) and feature the importance of models in identifying the essential behavioural features for predicting learners' cognitive processes. By doing so, we aimed to answer '*RQ3. What key behavioral indicators can machines use to predict learners' cognitive processes and form an iterative cycle between human-machine intelligence?*'

Feature importance analysis for predicting thinking skill clusters derived from *k*-means clustering (F19)

In Figure 6, we listed features used for the model training and ranked their feature importance from the tree model. The result shows that hand behaviours (F4) are usually the dominant feature for predicting the thinking skill clusters derived by *k*-means clustering (F19). Not only the dominance of the hand behaviour (F4, 28.32%) but also hand behaviours as the first dominant (F4, 7.15%) and the second dominant (F7, 17.93%) behavioural features contributed to making decisions for differentiating thinking skill clusters (F19). The expressiveness of the learner's behaviour (F9, 25.32%) was the second most significant feature used to predict the thinking skill clusters (F19). Additionally, individual reading speed (F13, 8.94%), the dominance of eyebrow movements (F1, 6.97%), and body movements (F5, 5.36%) among attention regulation behaviours were also used as indicators for the thinking skill clusters (F19) prediction.

Feature importance analysis for predicting high, mid and low HOTS and LOTS derived from quartile analysis (F17 and F18)

Figure 7 shows that different behavioural features are essential in predicting LOTS and HOTS. The dominance of hand behaviours (F4, 40.08%), mumble reading (F3, 39.77%) and behavioural expressiveness (F9, 20.15%) have been identified as the most critical features for predicting high LOTS. For predicting mid LOTS, the dominance of movements in hand (F4, 25.52%), eyebrow (F1, 14.91%) and body (F5, 11.67%) have been used as significant indicators. Low LOTS have been predicted through features such as the dominance of blink behaviours (F2, 40.17%) and movements in eyebrows (F1, 14.91%). The dominance of body movements (F5, 12.15%) as the first dominant behaviour and hand movements (F4, 22.18%) as the second dominant behaviour (F7, 17.50%) have also been considered meaningful in understanding mid LOTS. Behavioural expressiveness of attention regulation behaviour (F9, 5.82%) has been identified as a critical predictor of low LOTS. In general, the dominance of diverse attention regulation behaviours (F1–F5) and behavioural expressiveness (F9) have been utilized to predict LOTS. For high HOTS, individual reaction time average (F11, 26.94%) towards the screen blur stimuli has been identified as the most critical feature. The dominance of mumbling (F3, 23.33%), movements from the body (F5, 10.61%) and hand (F4, 13.92%) have been considered

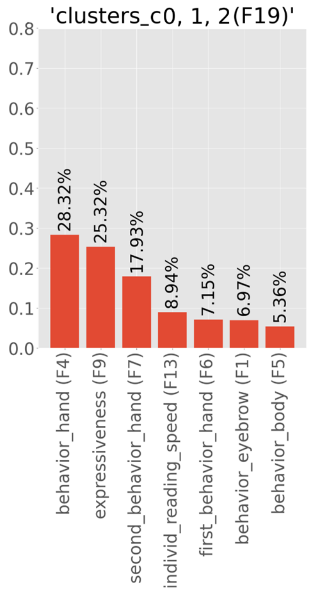


FIGURE 6 The feature importance for predicting three clusters of c0, c1 and c2 derived from *k*-means clustering.

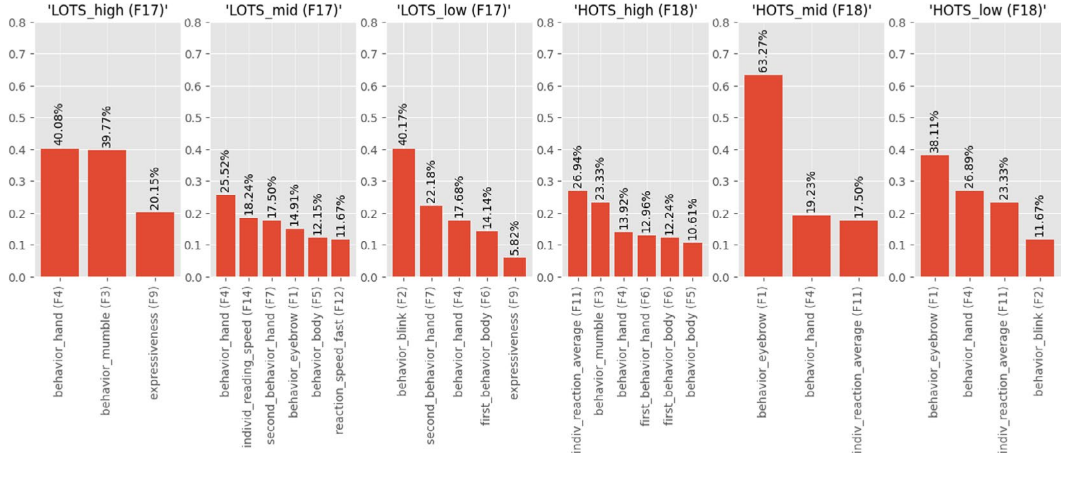


FIGURE 7 The feature importance for predicting high, mid and low range of HOTS and LOTS, derived from quartile analysis.

essential for predicting high HOTS. For learners with mid HOTS, behavioural dominance of eyebrow movements (F1, 63.27%), hand movements (F4, 19.23%) as well as individual reaction time average (F11, 17.50%) have been identified as critical features. For predicting low HOTS, the dominance of the eyebrow (F1, 38.11%), hand (F4, 26.89%) and blink (F2, 11.67%) have been used, along with individual reaction time average (F11, 23.33%) towards the secondary blur stimuli. All in all, the dominant movements from the eyes (ie, eyebrows (F1), blinks (F2)) were commonly used for predicting both low LOTS and HOTS. Behavioural expressiveness of attention regulation behaviours (F9) has been used for predicting LOTS. At the same time, learners' reaction time (F11, ie, arousal) has been identified as a critical feature for predicting HOTS. Contrary to our hypothesis, reading speed (F13) was not considered more important than other behavioural feature categories for predicting the overall LOTS and HOTS.

DISCUSSION

Reaction time as a predictor of HOTS and expressiveness as a predictor of thinking skill clusters, low and high LOTS

The result indicates that reaction time (F11) has only been used for predicting HOTS, while it has not been considered for making judgements for LOTS and thinking skill clusters (F19). It might indicate that more arousal, observed from fast reaction time, is related to learners' HOTS. On the other hand, expressiveness (F9) is used for predicting high and low LOTS (F17) and thinking skill clusters (F19). As higher expressiveness indicates more attention regulation behaviour during learners' reading, more distractions likely led to low LOTS (F17). Also, fewer attention regulation behaviours have been interpreted as cues to predict high LOTS (F17). All in all, we assume that learners' arousal has been targeted for predicting HOTS (F18) in machine-driven decision-making, while more self-aware distractions (ie, attention regulation behaviours) have been used for predicting LOTS (F17).

Behaviour-based human-machine HI for future intervention loops

The findings emphasize the critical role of attention regulation as a consistent predictor of cognitive performance, guiding the development of more targeted interventions and feedback mechanisms in digital learning environments (Cukurova, 2024). This focus on behavioural indicators provides a more nuanced understanding that goes beyond traditional self-reports and knowledge-gain metrics. Integrating these behavioural measures with cognitive assessments strengthens the model's theoretical foundation, making it a robust tool for educational research. The iterative combination of human insights and machine intelligence demonstrates how HI can be applied in both theory and practice.

Potential intervention strategies for learners with different HOTS and LOTS

HOTS has been known to be encouraged by: asking open-ended questions (Sofyan et al., 2024), implementing problem-based (Jailani et al., 2017) and inquiry-based learning (Mubarok et al., 2019), facilitating collaborative learning (Poudel, 2020) and peer teaching (Zaid et al., 2018), and fostering reflection (Jarvis & Baloyi, 2020). Conversely, LOTS can be strengthened through activities such as repetition and drills (Larsen-Freeman, 2012), demonstrations, practice with worksheets (Hayikaleng et al., 2016), tasks involving classification and categorization, and regular review of key concepts and facts (Wegerif, 2002). Such different learning strategies can inform future instructional design and feedback strategies, in line with our unsupervised (ie, *k*-means clustering) and statistical (ie, quartile analysis) cluster prediction, leveraged by machine intelligence.

When applying *k*-means clustering and generating groups of three, the relatively high-performing group (C1) represented high HOTS and LOTS simultaneously, while group C0 showed similar patterns of LOTS with C1 but a lower range of HOTS. Therefore, in our digital reading scenario, suggesting more learning practices for HOTS after the session would benefit learners classified in C1. Likewise, feedback strategies for learners assigned to C2 with relatively low LOTS and HOTS could focus more on activities for fostering LOTS and HOTS.

Iterative HI for adaptive learning with AI

Incorporating an iterative intervention loop appears to be effective as LA and interventions can be applied post hoc to observe shifts in the relative use of HOTS and LOTS among peer learners. With multiple rounds of intervention, learning thresholds are expected to progressively increase over time. However, the current analysis primarily reflects learners' relative performance compared to their peers and forms clusters based on this data. After several rounds of iteration, it would be advantageous to incorporate data samples from previous sessions and discontinue the interventions once an upward levelling of learning outcomes is observed. This approach helps avoid providing interventions beyond the point of diminishing returns.

The need for understanding the system's impact in real-world settings

While the model has shown promising results in a simple webcam-based setup, its application in real-world educational environments remains underexplored. Assessing AI-based systems in real-world contexts is particularly important for AIED implementations, where learners' perceived effectiveness and willingness to engage with AI-driven interactions play a significant role (Cukurova, 2024). Additionally, educational environments are dynamic, with diverse learners and unpredictable challenges that may affect the system's effectiveness. Therefore, comprehensive real-world testing is essential to evaluate the model's adaptability and long-term impact on actual learning outcomes.

Limitations due to small sample size

Although compensatory methods such as SMOTE, pruning and fivefold cross-validation were employed to address the challenges coming from the limited sample size, this remains a notable limitation. While the results offer valuable insights into the relationship between cognitive processes and observable behavioural cues, supporting an empirical HI approach in education, the small number of data points may not fully represent the broader population or context, potentially affecting the dynamics of the analysis. Future studies with larger samples could enhance the robustness of these findings or open up new directions for further research.

Varied focuses and expertise in human insights

A key limitation of this work is the variability in human expertise within the HI approach. Experts from diverse fields may interpret learners' cognitive processes differently, leading to inconsistencies in selecting behavioural indicators or designing interventions. These differences can affect the model's outcomes and limit its scalability across different educational contexts. To address this, future work could leverage advanced collaboration tools, such as consensus algorithms or machine-learning models that can integrate and harmonize multiple expert inputs. Additionally, continuous feedback loops between experts and AI systems could help refine the decision-making process, aligning varied human perspectives while maintaining flexibility for different educational contexts and supporting the advancement of HI.

CONCLUSION

This study focused on developing behaviour-based XAI models in digital reading to predict learners' cognitive processes, especially learners' utilization of HOTS and LOTS. Using the public multimodal WEDAR dataset, we extracted behavioural features related to learners' attention, including dominance and expressiveness of attention regulation behaviours, reaction time to secondary blur stimuli and reading speed. We hypothesized that these features could serve as predictors of HOTS and LOTS. We adopted an unsupervised clustering method and statistical quartile analysis to define targeted learners' cognitive processes in various levels and combinations for predictions. To achieve better explainability, we employed decision tree models with maximum depths of 10, suitable for small datasets with fewer feature categories.

The prediction results for thinking skill clusters and each high, mid and low level of HOTS and LOTS demonstrate robust accuracies ranging from 65.33% to 78.66% across different behavioural features and their combinations. The feature importance analysis reveals that attention regulation behaviour is consistently a strong predictor for all types of HOTS and LOTS. According to the following critical component analysis of training features, individual reading speed was found to be relevant only in predicting thinking skill clusters. At the same time, behavioural expressiveness played an essential role in predicting thinking skill clusters and LOTS. Individual reaction time to secondary stimuli was utilized only to predict HOTS.

In conclusion, our study successfully developed XAI models for behaviour-based prediction of learners' cognitive processes with HOTS and LOTS in digital reading, leveraged by the HI approach of combining human and machine intelligence. The findings highlight the significance of attention regulation behaviour as a consistent predictor across different cognitive processes with various levels of HOTS and LOTS. At the same time, we found that behavioural expressiveness worked as a critical component for predicting the thinking skill clusters, high and low LOTS, which seems to be related to learners' self-aware distractions (ie, attention regulation behaviours). On the other hand, reaction time was used for predicting HOTS, which we found to be related to learners' arousal, which needs further validation. The results contribute to understanding various behavioural factors to predict learners' HOTS and LOTS in digital reading, providing valuable insights for educators. Expanding the framework in line with specific feedback strategies can assist instructional designers with real-life digital reading practices.

All in all, our study provides actionable insights into designing personalized interventions for learners with different cognitive needs regarding HOTS and LOTS through machine intelligence. By leveraging clustering techniques, such as *k*-means clustering and quartile analysis, we segmented learners based on their performance, allowing educators to tailor interventions more effectively. For example, learners in higher-performing groups can receive strategies to improve HOTS, while lower-performing learners can focus on strengthening both HOTS and LOTS. Additionally, we propose an iterative intervention loop informed by LA to continuously monitor learners' progress and optimize interventions over time, helping to avoid intervention fatigue once the upward levelling of learning outcomes is observed in practice.

Our study also theoretically advances the field of HI by demonstrating how human insights and machine intelligence can be integrated to refine educational strategies. Specifically, we show how behavioural indicators such as attention regulation play a critical role in cognitive performance. This approach goes beyond traditional self-report methods and knowledge-gain metrics, offering a robust framework for using cognitive assessments to inform targeted interventions and feedback mechanisms.

ACKNOWLEDGEMENTS

This work was supported by the University of Oulu profiling project Profi7 Hybrid Intelligence, Grant/Award Number: 352788. Also, this work was supported by the Leiden-Delft-Erasmus Centre for Education and Learning (LDE-CEL).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

We utilize the data that support the findings of this study, which are openly available in 4TU.ResearchData at https://doi.org/10.4121/8f730aa3-ad04-4419-8a5b-325415d2294b.v1.

ETHICS STATEMENT

Approval of all ethical and experimental procedures and protocols was granted by the TU Delft Human Research Ethics Committee under Application No. 1760 and performed in line with the General Data Protection Regulation (GDPR).

ORCID

Marcus Specht https://orcid.org/0000-0002-6086-8480

REFERENCES

Abosalem, Y. (2016). Assessment techniques and students' higher-order thinking skills. *International Journal of Secondary Education*, *4*(1), 1–11. https://doi.org/10.11648/j.ijsedu.20160401.11

Alonso, J. M., & Casalino, G. (2019). Explainable artificial intelligence for human-centric data analysis in virtual learning environments. In D. Burgos, M. Cimitile, P. Ducange, R. Pecori, P. Picerno, P. Raviolo, & C. M. Stracke (Eds.), *Higher education learning methodologies and technologies online. HELMeTO 2019. Communications in Computer and Information Science* (Vol. *1091*, pp. 125–138). Springer. https://doi.org/10.1007/978-3-030-31284-8_10

Bixler, R., & D'Mello, S. (2016). Automatic gaze-based user-independent detection of mind wandering during computerized reading. *User Modeling and User-Adapted Interaction*, *26*, 33–68. https://doi.org/10.1007/s11257-015-9167-1

Bless, H., & Fiedler, K. (1995). Affective states and the influence of activated general knowledge. *Personality and Social Psychology Bulletin*, *21*(7), 766–778. https://doi.org/10.1177/0146167295217010

Brown, D. S., Aydin, C. E., & Donaldson, N. (2008). Quartile dashboards: Translating large data sets into performance improvement priorities. *The Journal for Healthcare Quality (JHQ)*, *30*(6), 18–30. https://doi.org/10.1111/j.1945-1474.2008.tb01166.x

Brüggemann, T., Ludewig, U., Lorenz, R., & McElvany, N. (2023). Effects of mode and medium in reading comprehension tests on cognitive load. *Computers & Education*, *192*, 104649. https://doi.org/10.1016/j.compedu.2022.104649

Brysbaert, M. (2019). How many words do we read per minute? A review and meta-analysis of reading rate. *Journal of Memory and Language*, *109*, 104047. https://doi.org/10.1016/j.jml.2019.104047

ChanLin, L.-J. (2013). Reading strategy and the need of e-book features. *The Electronic Library*, *31*(3), 329–344. https://doi.org/10.1108/EL-08-2011-0127

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Cukurova, M. (2024). The interplay of learning, analytics and artificial intelligence in education: A vision for hybridintelligence. *British Journal of Educational Technology*, 1–20. https://doi.org/10.1111/bjet.13514

Cukurova, M., Zhou, Q., Spikol, D., & Landolfi, L. (2020). Modelling collaborative problem-solving competence with transparent learning analytics: Is video data enough? In *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge*, Frankfurt, Germany (pp. 270–275). https://doi.org/10.1145/3375462.337548

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*. https://doi.org/10.48550/arXiv.1810.0480

Dikaya, L. A., Avanesian, G., Dikiy, I. S., Kirik, V. A., & Egorova, V. A. (2021). How personality traits are related to the attitudes toward forced remote learning during covid-19: Predictive analysis using generalized additive modeling. *Frontiers in Education*, *6*, 629213. https://doi.org/10.3389/feduc.2021.629213

Dyson, M. C., & Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. *International Journal of Human-Computer Studies*, *54*(4), 585–612. https://doi.org/10.1006/ijhc.2001.0458

Goffin, R. D., Jelley, R. B., Powell, D. M., & Johnston, N. G. (2009). Taking advantage of social comparisons in performance appraisal: The relative percentile method. *Human Resource Management: Published in Cooperation with the School of Business Administration, the University of Michigan and in Alliance with the Society of Human Resources Management*, *48*, 251–268. https://doi.org/10.1002/hrm.20278

Gohel, P., Singh, P., & Mohanty, M. (2021). Explainable ai: current status and future directions. *arXiv preprint arXiv:2107.07045*.

Greenaway, K. H., Kalokerinos, E. K., & Williams, L. A. (2018). Context is everything (in emotion research). *Social and Personality Psychology Compass*, *12*, e12393. https://doi.org/10.48550/arXiv.2107.07045

Hayikaleng, N., Nair, S. M., & Krishnasamy, H. N. (2016). Thai students' L2 reading comprehension level for lower order thinking skills and higher order thinking skills questions. *Journal of Applied Linguistics and Language Research*, *3*(5), 83–91.

Heong, Y. M., Othman, W. B., Yunos, J. B. M., Kiong, T. T., Hassan, R. B., & Mohamad, M. M. B. (2011). The level of marzano higher order thinking skills among technical education students. *International Journal of Social Science and Humanity*, *1*(2), 121–125. https://doi.org/10.7763/IJSSH.2011.V1.20

Huang, M. X., Li, J., Ngai, G., Leong, H. V., & Bulling, A. (2019, October). Moment-to-moment detection of internal thought during video viewing from eye vergence behavior. In *Proceedings of the 27th ACM International Conference on Multimedia*, Nice, France (pp. 2254–2262). https://doi.org/10.1145/3343031.3350573

Hussain, S., Minaz, M., Ahmad, N., & Idris, M. (2015). The effect of e-reading and printed document reading on student's comprehension and retention power. *International Conference on Computational and Social Sciences*, *25*(8), 408–414.

Hutt, S., Krasich, K., Mills, C., Bosch, N., White, S., Brockmole, J. R., & D'Mello, S. K. (2019). Automated gaze-based mind wandering detection during computerized learning in classrooms. *User Modeling and User-Adapted Interaction*, *29*, 821–867. https://doi.org/10.1007/s11257-019-09228-5

Jailani, J., Sugiman, S., & Apino, E. (2017). Implementing the problem-based learning in order to improve the students' HOTS and characters. *Jurnal Riset Pendidikan Matematika*, *4*(2), 247–259. https://doi.org/10.21831/jrpm.v4i2.17674

Järvelä, S., Malmberg, J., Haataja, E., Sobocinski, M., & Kirschner, P. A. (2021). What multimodal data can tell us about the students' regulation of their learning process? *Learning and Instruction*, *72*, 101203. https://doi.org/10.1016/j.learninstruc.2019.04.004

Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, *54*(5), 1057–1076. https://doi.org/10.1111/bjet.13325

Jarvis, M. A., & Baloyi, O. B. (2020). Scaffolding in reflective journaling: A means to develop higher order thinking skills in undergraduate learners. *International Journal of Africa Nursing Sciences*, *12*, 100195. https://doi.org/10.1016/j.ijans.2020.100195

Jivet, I., Scheffel, M., Specht, M., & Drachsler, H. (2018, March). License to evaluate: Preparing learning analytics dashboards for educational practice. In *Proceedings of the 8th International Conference on Learning Analytics and Knowledge*, Sydney, Australia (pp. 31–40). https://doi.org/10.1145/3170358.3170421

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y.-S., Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., & Gašević, D. (2022). Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *3*, 100074. https://doi.org/10.1016/j.caeai.2022.100074

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory Into Practice*, *41*, 212–218. https://doi.org/10.1207/s15430421tip4104_2

Larsen-Freeman, D. (2012). On the roles of repetition in language teaching and learning. *Applied Linguistics Review*, *3*(2), 195–210. https://doi.org/10.1515/applirev-2012-0009

Lee, Y. (2024). *Interactive intelligence: Multimodal AI for real-time interaction loop towards attentive e-reading*. [Dissertation]. (TU Delft), Delft University of Technology. https://doi.org/10.4233/uuid:764408e4-72c1-4cf9-8bff-1ce20b8944b2

Lee, Y., Chen, H., Zhao, G., & Specht, M. (2022). Wedar: Webcam-based attention analysis via attention regulator behavior recognition with a novel e-reading dataset. In *Proceedings of the 2022 International Conference on Multimodal Interaction*, Bengaluru, India (pp. 319–328). https://doi.org/10.1145/3536221.3556619

Lee, Y., Limbu, B., Rusak, Z., & Specht, M. (2023). Role of multimodal learning systems in technology-enhanced learning (TEL): A scoping review. In *European Conference on Technology Enhanced Learning* (pp. 164–182). Springer. https://doi.org/10.1007/978-3-031-42682-7_12

Lee, Y., Migut, G., & Specht, M. (2023a). Behavior-based feedback loop for attentive e-reading (BFLAe): A real-time computer vision approach. In *Micro-gesture analysis for hidden emotion understanding 2023*, (Vol. *12*). CEUR-WS.

Lee, Y., Migut, G., & Specht, M. (2023b). What attention regulation behaviors tell us about learners in e-reading?: Adaptive data-driven persona development and application based on unsupervised learning. *IEEE Access*, *11*, 118890–118906. https://doi.org/10.1109/ACCESS.2023.3327334

Lee, Y., & Specht, M. (2023a). Can we empower attentive e-reading with a social robot? An introductory study with a novel multimodal dataset and deep learning approaches. In *LAK 2023 Conference Proceedings-Towards Trustworthy Learning Analytics-13th International Conference on Learning Analytics and Knowledge. Association for Computing Machinery (ACM)*, Texas, USA (pp. 520–530). https://doi.org/10.1145/3576050.3576122

Lee, Y., & Specht, M. (2023b). Multimodal WEDAR dataset for attention regulation behaviors, self-reported distractions, reaction time, and knowledge gain in e-reading. (Version 1) [Data set]. 4TU.ResearchData. https://doi.org/10.4121/8f730aa3-ad04-4419-8a5b-325415d2294b.v1

Li, J., Ngai, G., Leong, H. V., & Chan, S. C. (2016). Multimodal human attention detection for reading from facial expression, eye gaze, and mouse dynamics. *ACM SIGAPP Applied Computing Review*, *16*, 37–49. https://doi.org/10.1145/3015297.3015301

Liu, X., Ma, J., & Wang, Q. (2023). A social robot as your reading companion: Exploring the relationships between gaze patterns and knowledge gains. *Journal on Multimodal User Interfaces*, *18*, 21–41. https://doi.org/10.1007/s12193-023-00418-5

Liu, Z. (2012). Digital reading. *Chinese Journal of Library and Information Science (English Edition)*, *5*, 85–94.

Mainali, B. P. (2012). Higher order thinking in education. *Academic Voices: A Multidisciplinary Journal*, *2*, 5–10. https://doi.org/10.3126/av.v2i1.8277

Molenaar, I. (2022). Towards hybrid human-AI learning technologies. *European Journal of Education*, *57*, 632–645. https://doi.org/10.1111/ejed.12527

Mubarok, H., Suprapto, N., & Adam, A. S. (2019, February). Using inquiry-based laboratory to improve students higher order thinking skills (HOTS). *Journal of Physics: Conference Series*, *1171*(1), 012040. https://doi.org/10.1088/1742-6596/1171/1/012040

Narayanan, S., & Adithan, M. (2015). Analysis of question papers in engineering courses with respect to hots (higher order thinking skills). *American Journal of Engineering Education (AJEE)*, *6*(1), 1–10. https://doi.org/10.19030/ajee.v6i1.9247

Orru, G., & Longo, L. (2019). The evolution of cognitive load theory and the measurement of its intrinsic, extraneous and germane loads: A review. In *Human Mental Workload: Models and Applications: Second International Symposium, H-WORKLOAD 2018, Amsterdam, The Netherlands, September 20–21, 2018, Revised Selected Papers 2* (pp. 23–48). Springer International Publishing. https://doi.org/10.1007/978-3-030-14273-5_3

Popa, L., Selejan, O., Scott, A., Mureşanu, D. F., Balea, M., & Rafila, A. (2015). Reading beyond the glance: Eye tracking in neurosciences. *Neurological Sciences*, *36*, 683–688. https://doi.org/10.1007/s10072-015-2076-6

Poudel, A. P. (2020). Computer supported collaborative learning for developing higher order thinking skill in ELT. *International Journal of Multidisciplinary Perspectives in Higher Education*, *5*(1), 34–51. https://doi.org/10.32674/jimphe.v5i1.2536

Qasrawi, R., & BeniAbdelrahman, A. (2020). The higher and lower-order thinking skills (hots and lots) in unlock English textbooks (1st and 2nd editions) based on bloom's taxonomy: An analysis study. *International Online Journal of Education and Teaching*, *7*, 744–758.

Richmond, K. M., Muddamsetty, S. M., Gammeltoft-Hansen, T., Olsen, H. P., & Moeslund, T. B. (2024). Explainable AI and law: An evidential survey. *Digital Society*, *3*(1), 1. https://doi.org/10.31235/osf.io/dypn2

Roschke, K., & Radach, R. (2016). Perception, reading, and digital media. In *The cognitive development of reading and reading comprehension* (Vol. *1*, pp. 33–52). Routledge. https://doi.org/10.4324/9781315694429

Salvucci, D. D., & Goldberg, J. H. (2000). Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, Florida, USA (pp. 71–78). https://doi.org/10.1145/355017.355028

Shafeeq, A., & Hareesha, K. (2012). Dynamic clustering of data with modified k-means algorithm. In *Proceedings of the 2012 Conference on Information and Computer Networks*, Singapore (pp. 221–225). https://doi.org/10.13140/2.1.4972.3840

Shanker, M., Hu, M. Y., & Hung, M. S. (1996). Effect of data standardization on neural network training. *Omega*, *24*, 385–397. https://doi.org/10.1016/0305-0483(96)00010-2

Shaughnessy, M. F. (2020). An interview with miriam schcolnik: Reading, e-reading and writing and their assessment. *European Scientific Journal*, *16*(28), 1–22. https://doi.org/10.19044/esj.2020.v16n28p1

Sofyan, F. A., Sartono, E., Badaruddin, K., Fauzi, M., Syarnubi, S., Oviyanti, F., Soraya, N., & Sukirman, S. (2024, April). Analysis of higher-order thinking skill (HOTS) of Madrasah Ibtidaiyah students in solving open-ended mathematics problems. *AIP Conference Proceedings*, *3058*(1), 060026. https://doi.org/10.1063/5.0201104

Tanujaya, B., Mumu, J., & Margono, G. (2017). The relationship between higher order thinking skills and academic performance of student in mathematics instruction. *International Education Studies*, *10*(11), 78–85. https://doi.org/10.5539/ies.v10n11p78

Trivedi, S., & Patel, N. (2020). Clustering students based on virtual learning engagement, digital skills, and e-learning infrastructure: Applications of k-means, dbscan, hierarchical, and affinity propagation clustering. *Sage Science Review of Educational Technology*, *3*, 1–13.

van Kempen, J., Loughnane, G. M., Newman, D. P., Kelly, S. P., Thiele, A., O'Connell, R. G., & Bellgrove, M. A. (2019). Behavioural and neural signatures of perceptual decision-making are modulated by pupil-linked arousal. *eLife*, *8*, e42541. https://doi.org/10.7554/eLife.42541

Wang, J. T.-Y. (2011). Pupil dilation and eye tracking. In *A handbook of process tracing methods for decision research: A critical review and user's guide* (Vol. *1*, pp. 185–204). Psychology Press. https://doi.org/10.4324/9780203875292

Wang, L. (2018). Attention decrease detection based on video analysis in e-learning. *Transactions on Edutainment XIV*, *10790*, 166–179.

Wegerif, R. (2002). Literature review in thinking skills. *Technology and Learning*, *1*, 1–44. https://doi.org/10.1007/978-3-662-56689-3_14

Wirth, J., Stebner, F., Trypke, M., Schuster, C., & Leutner, D. (2020). An interactive layers model of self-regulated learning and cognitive load. *Educational Psychology Review*, *32*, 1127–1149. https://doi.org/10.1007/s10648-020-09568-4

Yan, S., Hahn, M., & Keller, F. (2022). Modeling fixation behavior in reading with character-level neural attention. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, Toronto, Canada (Vol. *44*, pp. 2171–2177).

Zaid, N. M., Yaacob, F., Shukor, N. A., Said, M. M., Musta'amal, A., & Rahmatina, D. (2018). Integration of peer instruction in online social network to enhance higher order thinking skills. *International Journal of Interactive Mobile Technologies*, *12*(8), 30–40. https://doi.org/10.3991/ijim.v12i8.9672

Zhong, Q., Ding, L., Liu, J., Du, B., & Tao, D. (2023). Can ChatGPT understand too? A comparative study on ChatGPT and fine-tuned BERT. *arXiv preprint arXiv:2302.10198*. https://doi.org/10.48550/arXiv.2302.10198

Zimmerman, B. J., & Moylan, A. R. (2009). Self-regulation: Where metacognition and motivation intersect. In *Handbook of metacognition in education* (Vol. *1*, pp. 299–315). Routledge.

Zuckerman, M. (1994). *Behavioral expressions and biosocial bases of sensation seeking*. Cambridge University Press.

Zunino, A., Cavazza, J., & Murino, V. (2017). Revisiting human action recognition: Personalization vs. generalization. In *Image Analysis and Processing-ICIAP 2017: 19th International Conference*, Catania, Italy (pp. 469–480). https://doi.org/10.1007/978-3-319-68560-1_42

How to cite this article: Lee, Y., Migut, G., & Specht, M. (2025). Unveiling cognitive processes in digital reading through behavioural cues: A hybrid intelligence (HI) approach. *British Journal of Educational Technology*, *00*, 1–34. https://doi.org/10.1111/bjet.13551

APPENDIX A

A.1 | Multiple-choice and summarization questionnaires from the WEDAR dataset

The questionnaire from the WEDAR dataset used for both the pretest and posttest is described below. The pretest consisted of 10 multiple-choice questions, while the posttest used the same questionnaire with an additional summarization question (see Figures A1 and A2).

A.1.1. | Pretest questionnaire: multiple-choice

A.1.2. | Posttest questionnaire: multiple-choice and summarization

APPENDIX B

B.1 | Predicting three thinking skill clusters derived from *k*-means clustering (F19)

Decision tree models provide great interpretability with the plot tree. Figure B1 illustrates the model's depth-by-depth decision-making process for predicting the three-level thinking skill clusters. The tree uses Gini impurity to understand the quality of the split of groups based on the condition, having 0 as the best purity with the best distinctions in the decision. At the same time, 1 indicates the impurity, which requires another round of decision-making. Values in the bracket indicate the possibility that each condition is classified as C0, C1 and C2, respectively.

B.1.1. | Plot tree analysis for predicting thinking skill clusters derived from *k*-means (F19)

As can be seen from Figure B1, the first depth of the model considers the dominance of hand behaviour (F4) as the most influential feature in the decision-making: It informs that if the feature marks less than 0.082, samples are classified as C1, making decisions for 21.4% of the samples. For the remaining 78.6% of samples, the condition in the second depth, expressiveness of the attention regulation behaviour (F9) of 0.342, has been used. On the right branch, 17.9% of the samples were classified as C0, having less than or the same as 1.5 as individual reading speed (F13) as the condition.

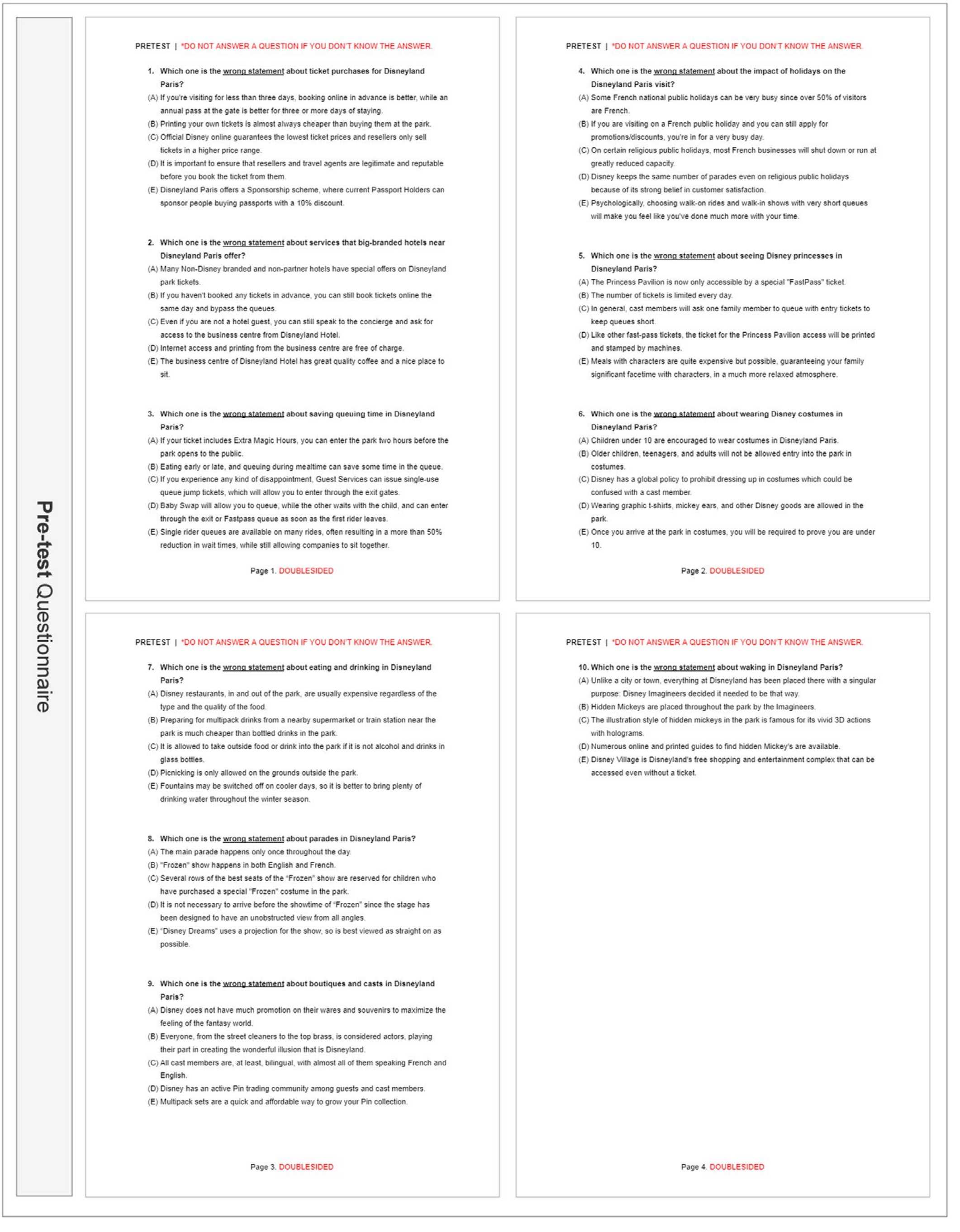
The following condition of an individual reaction time average (F11) of less than or equal to 0.793 classified 3.6% of the samples as C0. Other 7.1% of the samples were classified as C1, with an individual reaction time average (F11) of more than 0.793. From the left branch, 3.6% of the samples were classified as C1, with a dominant hand behaviour (F4) of more than 0.211. In the left branch, a second dominant hand behaviour (F7) of 0.5 was the following condition, and 32.1% of the samples were classified as C2. The following condition of having dominance of body behaviour (F5) of less than 0.398 classified 7.1% of the samples as C0. Finally, the last condition classified 3.6% of the samples as C0, having less than or equal to 0.017 as eye behaviour dominance (F1). In contrast, 3.6% were classified as C2, with more than 0.017 as eye behaviour dominance.

All in all, by conducting the feature analysis, we aimed to grasp how the model made the decision. Having those procedures aligned is especially insightful for education researchers and instructional designers, who work with the same sets of indicators and parameters. By having such standards, they can take a more systematic approach to learning analytics and subsequent intervention design, especially with learning behaviours.

B.2 | PREDICTING HIGH, MID AND LOW HOTS AND LOTS (F17 AND F18)

B.2.1. | Plot tree analysis for high, mid and low HOTS and LOTS (F17 and F18)

In our comparative analysis, three decision tree models were developed to predict the high, mid and low levels of LOTS and HOTS, respectively. The three trees for predicting LOTS share a consistent set of predictors, utilizing a variety of dominant hand (F4), mumble (F3)



Pre-test Questionnaire

PRETEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

1. **Which one is the wrong statement about ticket purchases for Disneyland Paris?**
(A) If you're visiting for less than three days, booking online in advance is better, while an annual pass at the gate is better for three or more days of staying.
(B) Printing your own tickets is almost always cheaper than buying them at the park.
(C) Official Disney online guarantees the lowest ticket prices and resellers only sell tickets in a higher price range.
(D) It is important to ensure that resellers and travel agents are legitimate and reputable before you book the ticket from them.
(E) Disneyland Paris offers a Sponsorship scheme, where current Passport Holders can sponsor people buying passports with a 10% discount.

2. **Which one is the wrong statement about services that big-branded hotels near Disneyland Paris offer?**
(A) Many Non-Disney branded and non-partner hotels have special offers on Disneyland park tickets.
(B) If you haven't booked any tickets in advance, you can still book tickets online the same day and bypass the queues.
(C) Even if you are not a hotel guest, you can still speak to the concierge and ask for access to the business centre from Disneyland Hotel.
(D) Internet access and printing from the business centre are free of charge.
(E) The business centre of Disneyland Hotel has great quality coffee and a nice place to sit.

3. **Which one is the wrong statement about saving queuing time in Disneyland Paris?**
(A) If your ticket includes Extra Magic Hours, you can enter the park two hours before the park opens to the public.
(B) Eating early or late, and queuing during mealtime can save some time in the queue.
(C) If you experience any kind of disappointment, Guest Services can issue single-use queue jump tickets, which will allow you to enter through the exit gates.
(D) Baby Swap will allow you to queue, while the other waits with the child, and can enter through the exit or Fastpass queue as soon as the first rider leaves.
(E) Single rider queues are available on many rides, often resulting in a more than 50% reduction in wait times, while still allowing companies to sit together.

Page 1. DOUBLESIDED

PRETEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

4. **Which one is the wrong statement about the impact of holidays on the Disneyland Paris visit?**
(A) Some French national public holidays can be very busy since over 50% of visitors are French.
(B) If you are visiting on a French public holiday and you can still apply for promotions/discounts, you're in for a very busy day.
(C) On certain religious public holidays, most French businesses will shut down or run at greatly reduced capacity.
(D) Disney keeps the same number of parades even on religious public holidays because of its strong belief in customer satisfaction.
(E) Psychologically, choosing walk-on rides and walk-in shows with very short queues will make you feel like you've done much more with your time.

5. **Which one is the wrong statement about seeing Disney princesses in Disneyland Paris?**
(A) The Princess Pavilion is now only accessible by a special "FastPass" ticket.
(B) The number of tickets is limited every day.
(C) In general, cast members will ask one family member to queue with entry tickets to keep queues short.
(D) Like other fast-pass tickets, the ticket for the Princess Pavilion access will be printed and stamped by machines.
(E) Meals with characters are quite expensive but possible, guaranteeing your family significant facetime with characters, in a much more relaxed atmosphere.

6. **Which one is the wrong statement about wearing Disney costumes in Disneyland Paris?**
(A) Children under 10 are encouraged to wear costumes in Disneyland Paris.
(B) Older children, teenagers, and adults will not be allowed entry into the park in costumes.
(C) Disney has a global policy to prohibit dressing up in costumes which could be confused with a cast member.
(D) Wearing graphic t-shirts, mickey ears, and other Disney goods are allowed in the park.
(E) Once you arrive at the park in costumes, you will be required to prove you are under 10.

Page 2. DOUBLESIDED

PRETEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

7. **Which one is the wrong statement about eating and drinking in Disneyland Paris?**
(A) Disney restaurants, in and out of the park, are usually expensive regardless of the type and the quality of the food.
(B) Preparing for multipack drinks from a nearby supermarket or train station near the park is much cheaper than bottled drinks in the park.
(C) It is allowed to take outside food or drink into the park if it is not alcohol and drinks in glass bottles.
(D) Picnicking is only allowed on the grounds outside the park.
(E) Fountains may be switched off on cooler days, so it is better to bring plenty of drinking water throughout the winter season.

8. **Which one is the wrong statement about parades in Disneyland Paris?**
(A) The main parade happens only once throughout the day.
(B) "Frozen" show happens in both English and French.
(C) Several rows of the best seats of the "Frozen" show are reserved for children who have purchased a special "Frozen" costume in the park.
(D) It is not necessary to arrive before the showtime of "Frozen" since the stage has been designed to have an unobstructed view from all angles.
(E) "Disney Dreams" uses a projection for the show, so is best viewed as straight on as possible.

9. **Which one is the wrong statement about boutiques and casts in Disneyland Paris?**
(A) Disney does not have much promotion on their wares and souvenirs to maximize the feeling of the fantasy world.
(B) Everyone, from the street cleaners to the top brass, is considered actors, playing their part in creating the wonderful illusion that is Disneyland.
(C) All cast members are, at least, bilingual, with almost all of them speaking French and English.
(D) Disney has an active Pin trading community among guests and cast members.
(E) Multipack sets are a quick and affordable way to grow your Pin collection.

Page 3. DOUBLESIDED

PRETEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

10. **Which one is the wrong statement about waking in Disneyland Paris?**
(A) Unlike a city or town, everything at Disneyland has been placed there with a singular purpose: Disney Imagineers decided it needed to be that way.
(B) Hidden Mickeys are placed throughout the park by the Imagineers.
(C) The illustration style of hidden mickeys in the park is famous for its vivid 3D actions with holograms.
(D) Numerous online and printed guides to find hidden Mickey's are available.
(E) Disney Village is Disneyland's free shopping and entertainment complex that can be accessed even without a ticket.

Page 4. DOUBLESIDED

FIGURE A1 The pretest questionnaire consists of 10 multiple-choice questions.

behaviours, and behavioural expressiveness (F9) play a significant role in the prediction across all LOTS levels. Figure B2 initiates the split with dominant hand behaviours (F4), suggesting its strong influence for predicting the high level of LOTS. Subsequent splits on

Post-test Questionnaire

POSTTEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

1. **Which one is the <u>wrong statement</u> about ticket purchases for Disneyland Paris?**
(A) If you're visiting for less than three days, booking online in advance is better, while an annual pass at the gate is better for three or more days of staying.
(B) Printing your own tickets is almost always cheaper than buying them at the park.
(C) Official Disney online guarantees the lowest ticket prices and resellers only sell tickets in a higher price range.
(D) It is important to ensure that resellers and travel agents are legitimate and reputable before you book the ticket from them.
(E) Disneyland Paris offers a Sponsorship scheme, where current Passport Holders can sponsor people buying passports with a 10% discount.

2. **Which one is the <u>wrong statement</u> about services that big-branded hotels near Disneyland Paris offer?**
(A) Many Non-Disney branded and non-partner hotels have special offers on Disneyland park tickets.
(B) If you haven't booked any tickets in advance, you can still book tickets online the same day and bypass the queues.
(C) Even if you are not a hotel guest, you can still speak to the concierge and ask for access to the business centre from Disneyland Hotel.
(D) Internet access and printing from the business centre are free of charge.
(E) The business centre of Disneyland Hotel has great quality coffee and a nice place to sit.

3. **Which one is the <u>wrong statement</u> about saving queuing time in Disneyland Paris?**
(A) If your ticket includes Extra Magic Hours, you can enter the park two hours before the park opens to the public.
(B) Eating early or late, and queuing during mealtime can save some time in the queue.
(C) If you experience any kind of disappointment, Guest Services can issue single-use queue jump tickets, which will allow you to enter through the exit gates.
(D) Baby Swap will allow you to queue, while the other waits with the child, and can enter through the exit or Fastpass queue as soon as the first rider leaves.
(E) Single rider queues are available on many rides, often resulting in a more than 50% reduction in wait times, while still allowing companies to sit together.

Page 1. DOUBLESIDED

POSTTEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

4. **Which one is the <u>wrong statement</u> about the impact of holidays on the Disneyland Paris visit?**
(A) Some French national public holidays can be very busy since over 50% of visitors are French.
(B) If you are visiting on a French public holiday and you can still apply for promotions/discounts, you're in for a very busy day.
(C) On certain religious public holidays, most French businesses will shut down or run at greatly reduced capacity.
(D) Disney keeps the same number of parades even on religious public holidays because of its strong belief in customer satisfaction.
(E) Psychologically, choosing walk-on rides and walk-in shows with very short queues will make you feel like you've done much more with your time.

5. **Which one is the <u>wrong statement</u> about seeing Disney princesses in Disneyland Paris?**
(A) The Princess Pavilion is now only accessible by a special "FastPass" ticket.
(B) The number of tickets is limited every day.
(C) In general, cast members will ask one family member to queue with entry tickets to keep queues short.
(D) Like other fast-pass tickets, the ticket for the Princess Pavilion access will be printed and stamped by machines.
(E) Meals with characters are quite expensive but possible, guaranteeing your family significant facetime with characters, in a much more relaxed atmosphere.

6. **Which one is the <u>wrong statement</u> about wearing Disney costumes in Disneyland Paris?**
(A) Children under 10 are encouraged to wear costumes in Disneyland Paris.
(B) Older children, teenagers, and adults will not be allowed entry into the park in costumes.
(C) Disney has a global policy to prohibit dressing up in costumes which could be confused with a cast member.
(D) Wearing graphic t-shirts, mickey ears, and other Disney goods are allowed in the park.
(E) Once you arrive at the park in costumes, you will be required to prove you are under 10.

Page 2. DOUBLESIDED

POSTTEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

7. **Which one is the <u>wrong statement</u> about eating and drinking in Disneyland Paris?**
(A) Disney restaurants, in and out of the park, are usually expensive regardless of the type and the quality of the food.
(B) Preparing for multipack drinks from a nearby supermarket or train station near the park is much cheaper than bottled drinks in the park.
(C) It is allowed to take outside food or drink into the park if it is not alcohol and drinks in glass bottles.
(D) Picnicking is only allowed on the grounds outside the park.
(E) Fountains may be switched off on cooler days, so it is better to bring plenty of drinking water throughout the winter season.

8. **Which one is the <u>wrong statement</u> about parades in Disneyland Paris?**
(A) The main parade happens only once throughout the day.
(B) "Frozen" show happens in both English and French.
(C) Several rows of the best seats of the "Frozen" show are reserved for children who have purchased a special "Frozen" costume in the park.
(D) It is not necessary to arrive before the showtime of "Frozen" since the stage has been designed to have an unobstructed view from all angles.
(E) "Disney Dreams" uses a projection for the show, so is best viewed as straight on as possible.

9. **Which one is the <u>wrong statement</u> about boutiques and casts in Disneyland Paris?**
(A) Disney does not have much promotion on their wares and souvenirs to maximize the feeling of the fantasy world.
(B) Everyone, from the street cleaners to the top brass, is considered actors, playing their part in creating the wonderful illusion that is Disneyland.
(C) All cast members are, at least, bilingual, with almost all of them speaking French and English.
(D) Disney has an active Pin trading community among guests and cast members.
(E) Multipack sets are a quick and affordable way to grow your Pin collection.

Page 3. DOUBLESIDED

POSTTEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

10. **Which one is the <u>wrong statement</u> about waking in Disneyland Paris?**
(A) Unlike a city or town, everything at Disneyland has been placed there with a singular purpose: Disney Imagineers decided it needed to be that way.
(B) Hidden Mickeys are placed throughout the park by the Imagineers.
(C) The illustration style of hidden mickeys in the park is famous for its vivid 3D actions with holograms.
(D) Numerous online and printed guides to find hidden Mickey's are available.
(E) Disney Village is Disneyland's free shopping and entertainment complex that can be accessed even without a ticket.

Page 4. DOUBLESIDED

POSTTEST | *DO NOT ANSWER A QUESTION IF YOU DON'T KNOW THE ANSWER.

Summarization

What was the article about? (in one short sentence)

This article delivers 10 tips about . . .

How to 1. ____________________.

Can you recall 10 tips that the authors mentioned in the article? (in short sentences)

How to 2. ____________________.

How to 3. ____________________.

How to 4. ____________________.

How to 5. ____________________.

How to 6. ____________________.

How to 7. ____________________.

How to 8. ____________________.

How to 9. ____________________.

How to 10. ____________________.

How to 11. ____________________.

Page 5. DOUBLESIDED

FIGURE A2 The posttest questionnaire consists of 10 multiple-choice questions and a summarization task covering one main topic and 10 subtopics.

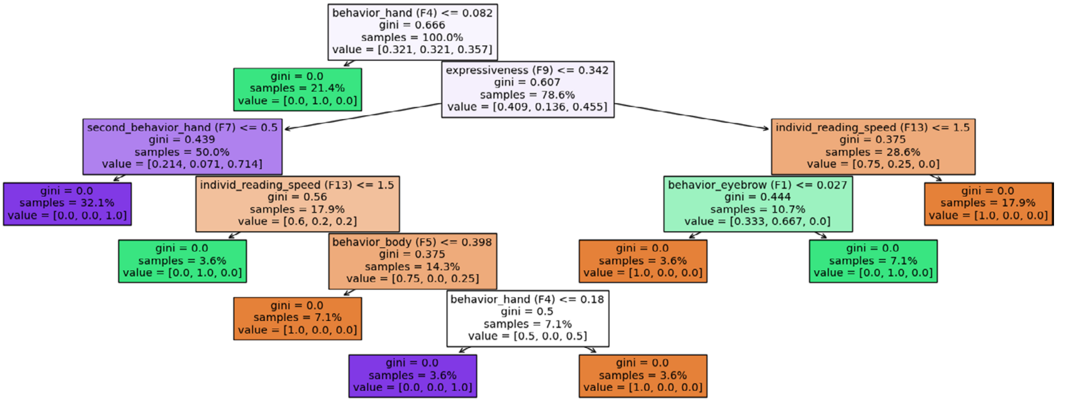


FIGURE B1 A plot tree to explain the model built upon the decision tree for predicting the unsupervised k-means clusters.

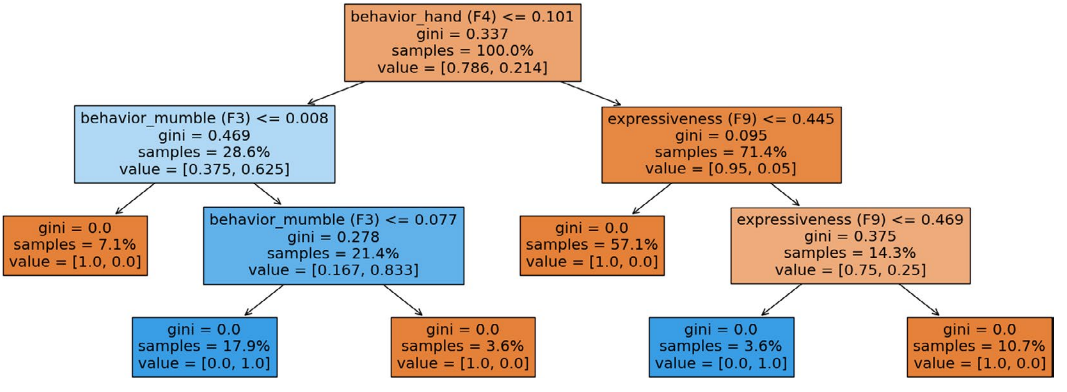


FIGURE B2 A plot tree to explain the model built upon the decision tree for predicting the high LOTS.

dominant mumble behaviours and behavioural expressiveness (F9) illustrate a focus on nuanced behaviours to refine the prediction. The tree presents a balanced path with splits occurring at both the left and right nodes, indicating diverse sample distributions. In Figure B3, aiming at the mid-level LOTS prediction, individual reading speed (F13) extends to greater depths, signalling a more complex decision-making process with multiple behavioural and reaction time-related features such as dominant blink behaviours (F2) and reaction time quartiles (F12), reflecting the intricate nature of predicting mid-range outcomes. Figure B4 predicts the low LOTS level, revealing a notable difference by starting with behavioural expressiveness (F9) as the primary split. It indicates that expressive behaviours determine lower learning outcomes in LOTS evaluation. Unlike the previous models, Figure B4 simplifies the decision process with fewer splits, potentially revealing more apparent distinctions among lower LOTS levels based on expressiveness alone.

On the other hand, all models for predicting HOTS have commonly used the dominance of hand (F4), body (F5) and mumble (5) as critical features for prediction, indicating the universal applicability of such features to different HOTS levels. To predict high levels of HOTS 12, body behaviours as the most common attention regulation behaviours (F6) have been used as the root node, suggesting that initial body language plays a significant role in predicting

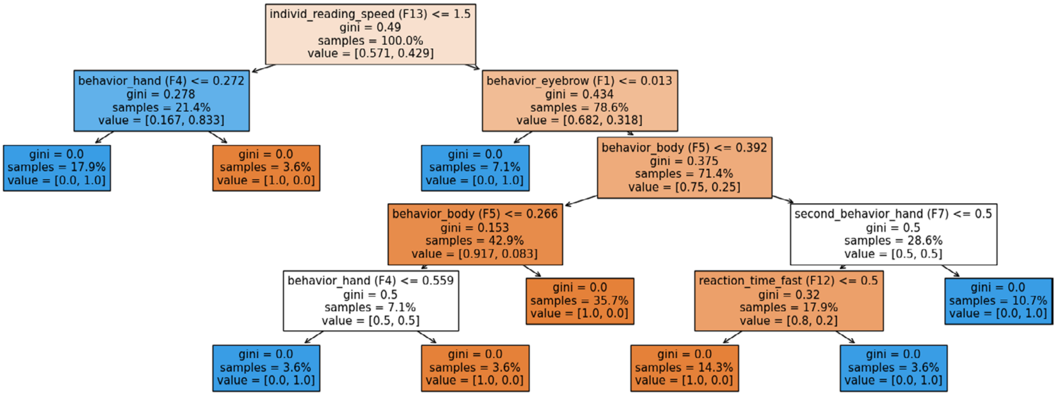


FIGURE B3 A plot tree to explain the model built upon the decision tree for predicting the mid LOTS.

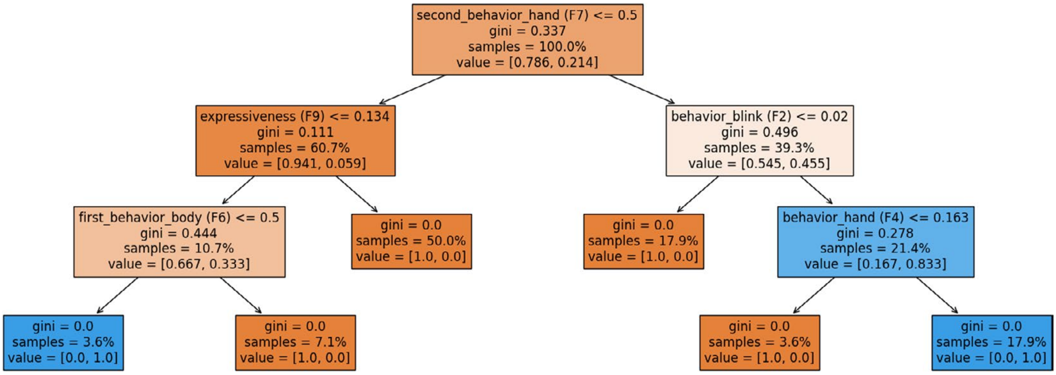


FIGURE B4 A plot tree to explain the model built upon the decision tree for predicting the low LOTS.

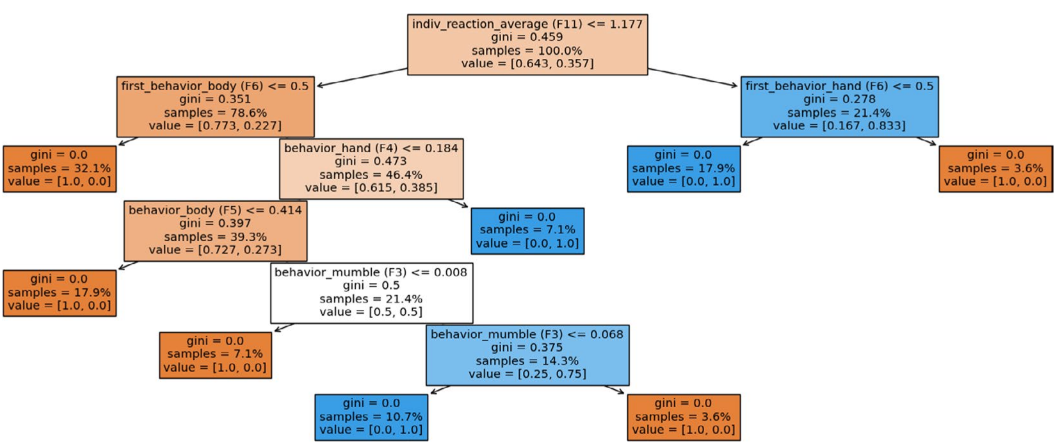


FIGURE B5 A plot tree to explain the model built upon the decision tree for predicting the high HOTS.

higher cognitive skills. Figure B5 is less complex, with fewer splits, showing a more straightforward relationship between observable behaviours and high HOTS. Figure B6 focuses on describing the mid-level HOTS, starting with the individual reaction average (F11), indicating

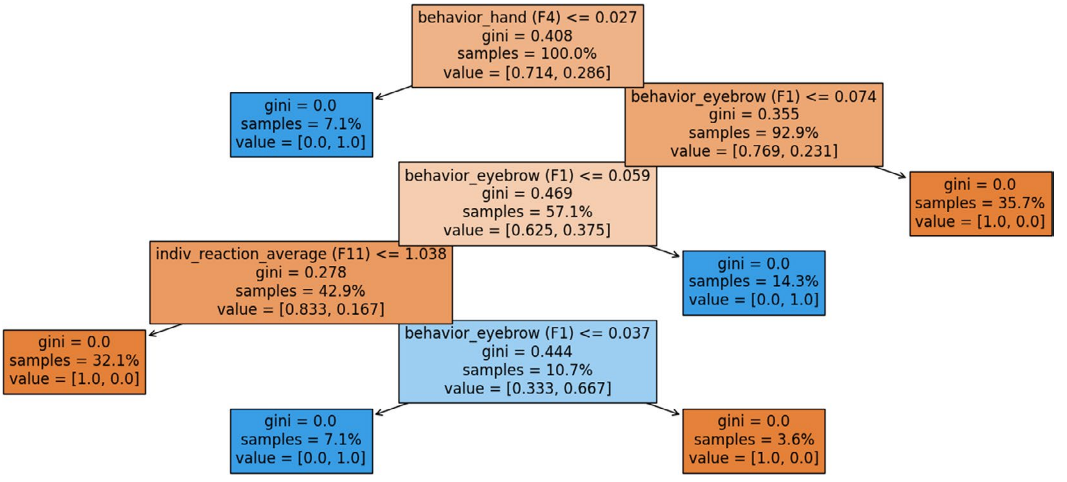


FIGURE B6 A plot tree to explain the model built upon the decision tree for predicting the mid HOTS.

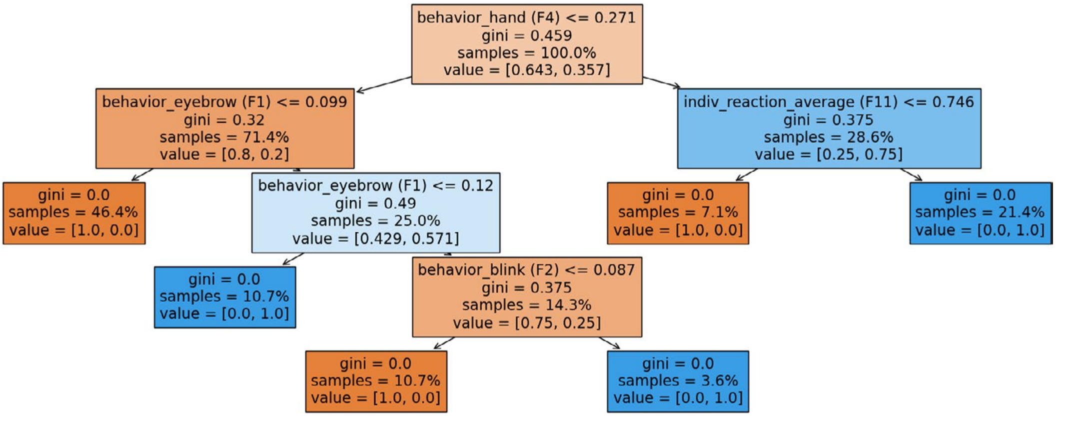


FIGURE B7 A plot tree to explain the model built upon the decision tree for predicting the low HOTS.

that mid-level HOTS may be more closely linked to the arousal levels of each individual. This model branches out into more levels of depth, requiring a deeper analysis to achieve accurate predictions. To describe low HOTS levels (Figure B7), dominant eyebrow behaviours (F1) was used as the root node, having an intermediate complexity between the high- and mid-level models, showing a balance between behaviours and individual traits in determining lower HOTS.

Both LOTS and HOTS models utilized features related to the dominance of hand (F4), body (F5), and mumble (F3) behaviours. It represents the dominance of attention regulation behaviours as predictors of thinking skill levels (F19). The LOTS models often use dominant hand behaviours (F4) as the root node. In contrast, the HOTS models vary, with the root node being body behaviours as the most dominant attention regulation behaviours (F6) for predicting high HOTS, individual reaction time average (F11) for mid HOTS, and dominant eyebrow behaviours (F1) for low HOTS. It suggests that different aspects of behaviour and individual traits are considered for predicting different thinking skill levels (F19). The HOTS models exhibited varying complexities, indicating that the prediction of HOTS levels may be more complex and require a deeper understanding of the interplay between different predictors. In contrast, the LOTS models appear more balanced, suggesting a more uniform distribution of features across varying levels of LOTS. The mid HOTS model stands out, using an individual cognitive metric as the root, whereas the LOTS and other HOTS models tend to prioritize behavioural indicators. It implies that individual cognitive metrics are more predictive of mid-level HOTS, while observable behaviours are more indicative of the extreme levels of both LOTS and HOTS.