Children's Play and Problem Solving in Motion-Based Educational Games: Synergies between Human Annotations and Multi-Modal Data

Serena Lee-Cultura
serena.leecultura@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

Kshitij Sharma
kshitij.sharma@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

Giulia Cosentino
giulia.cosentino@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

Sofia Papavlasopoulou
spapav@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

Michail Giannakos
michailg@ntnu.no
Norwegian University of Science and Technology
Trondheim, Norway

ABSTRACT

Identifying and supporting children's play and problem solving behaviour is important for designing educational technologies. This can inform feedback mechanisms to scaffold learning (provide hints or progress information), and assist facilitators (teachers, parents) in supporting children. Traditionally, researchers manually code video to dissect children's nuanced play and problem solving behaviour. Advancements in sensing technologies and their respective Multi-Modal Data (MMD), afford observation of invisible states (cognitive, affective, physiological), and provide opportunities to inspect internal processes experienced during learning and play. However, limited research combines traditional video annotations and MMD to understand children's behaviour as they interact with educational technology. To address this concern, we collected data from webcam, wristband, eye-trackers, and Kinect, as 26 children, aged 10-12, played a Motion-Based Educational Games (MBEG). Results showed significant differences in children's experience during play and problem solving episodes, and motivate design considerations aimed to facilitate children's interactions with MBEG.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in interaction design**; **Empirical studies in HCI**; • **Applied computing → Interactive learning environments**.

KEYWORDS

play, problem solving, motion-based games, sensors, education, learning, multi-modal data

IDC '21, June 24–30, 2021, Athens, Greece
Published version can be found:
https://doi.org/10.1145/3459990.3460702

ACM Reference Format:
Serena Lee-Cultura, Kshitij Sharma, Giulia Cosentino, Sofia Papavlasopoulou, and Michail Giannakos. 2021. Children's Play and Problem Solving in Motion-Based Educational Games: Synergies between Human Annotations and Multi-Modal Data. In *Interaction Design and Children (IDC '21), June 24–30, 2021, Athens, Greece.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3459990.3460702

1 INTRODUCTION & MOTIVATION

Play can be described as the manifestation of children's actions and their own creative meaning [73], and is traditionally attributed as a substrate for children's individual understanding and experiences of the surrounding world. During play, children submit to imaginative thinking and are captivated by a rich multi-sensory experience derived on their own volition [75]. Research shows that integrating play into children's problem solving scenarios, yields a multitude of valuable outcomes, such as increased engagement [63], elevated enjoyment [77], and amplified motivation [29]. Accordingly, playfully framed problem solving experiences may be more meaningful [15] for children's learning. Children demonstrate several different problem solving strategies and behaviours that oftentimes need different level or forms of support [13, 100]. Identifying and properly facilitating these behaviours has the potential to inform the design of technological affordances and equip children's support sphere (i.e., learning facilitators, such as, teachers, parents, and therapist). This support can guide appropriate feedback delivery to children (e.g., via the system or the facilitator), with the goal of nurturing children's learning experience.

The last decade has given rise to an enormous penetration of wearable technologies in children's lives. Due to their ubiquitous nature, these technologies have become more readily accessible for use with young children; supporting their play, communication, education and other endeavours. One such application is the development of Motion-Based Educational Games (MBEG), which utilise sensing devices (e.g., Microsoft Kinect) to capture, map and interpret children's full-bodied movements as game input [7]. Within these games, playful expression emerges as children appropriate the game context [82] by creatively interpreting, and engaging with, learning content, and at times challenging the game's confines (i.e., rules, story line, and objectives) through freedom of movement. Moreover, due to their "touchless" nature, MBEG have been shown to provide children with a more natural [36] and engaging learning experience [43]. The inherent richness of investigating the natural interactions exhibited by children during learning, highlights MBEG as a potentially valuable untapped resource for better understanding children's learning behaviours.

A core element of MBEG is the use of sensing technologies (e.g., motion sensors), which enable the automatic, continuous, and unobtrusive collection of Multi-Modal Data (MMD), such as physiological, gaze, and skeletal data. Such data collections empower us to transcend the limits of human observation, by accessing real-time information on children's seemingly "invisible" cognitive, affective and physiological states [92]. Accordingly, sensing technologies are gaining traction as useful, reliable means of investigative practice for understanding multi-faceted problem solving phenomena and supporting learning in-situ [11, 19], specifically in the domain of children's problem solving behaviours during interactive learning experiences [33, 51, 52, 71]. Additionally, sensing technologies and their respective MMD, allow us to closely monitor and understand children's play and problem solving behaviours, leveraging the key affordances of MMD (e.g., temporality and direct access to indicators of children's cognitive and affective processes [19]). However, despite this potential, limited research has attempted to interlace these ideas by using MMD from sensing technologies to investigate children's play and problem solving behaviours. To bridge this gap, our research attempts to address the following research question (RQ).

RQ1 How do children experience play and problem solving during their interaction with MBEG?

RQ2 How do children experience guessing and informed problem solving during their interaction with MBEG?

To explore these questions, we conducted an empirical study in which 26 children, aged 10 – 12, played three games of a geometry-focused MBEG, called Marvy Learns. We recorded children's gameplay and employed various sensing devices which allowed us to capture children's MMD (e.g., gaze from eye-tracking glasses, physiological from wristband, and skeletal from motion sensor). We manually annotated the video recordings, and used a mixed-methods approach to investigate how children experience play and problem solving. We present the following contributions:

- We offer insights from an in-situ experiment where children, aged 10 – 12 years, were monitored by wearables and sensing technologies, as they played a geometry-focused MBEG.
- We outline the differences in children's play and problem solving behaviours during their interaction with MBEG.
- We elaborate on how our findings can be used to scaffold children's learning through provision of individually tailored feedback mechanisms concerning student's cognitive, affective and physiological states.

2 RELATED WORK

We draw inspiration from the domains of embodied interaction, MBEG, MMD, and the confluence of play and problem solving in Child-Computer Interaction (CCI). Here, we offer a overview of the groundwork that directs our research.

2.1 Embodied Interaction and Motion-Based Educational Games

Recent advancements in wearables and high precision motion sensing devices (e.g., Microsoft Kinect) demonstrate the capacity to play an instrumental role in the design and implementation of learning experiences [25]. These interactive technologies encourage the coupling of the mind and body, and have given rise to the concept of embodied interaction. Embodied interaction describes the relationship between one's physical actions and mental faculties, combined with their social and environmental context, and its influence on the sharing, creation, and manipulation of knowledge through natural, meaningful interactions (i.e., gesture, full bodied movements, facial expression) with technology [25].

MBEG allow children interact with learning resources in a natural and playful way, by using their body to improve cognitive skills [104]. These games have gained traction as a powerful pedagogical strategy to promote children's learning [7, 43]. Concerning proficiency in maths, MBEG have traversed a variety of sub-domains, such as calculus [67], algebra [45], arithmetic [91], and geometry [83]. Remarkable studies show that MBEG might bring benefits to a player's maths learning experience; especially concerning enhanced problem understanding [83], reduced anxiety [44], and increased academic performance [93]. Additionally, the introduction of programming languages using motion-based technology (i.e, Kinect2Scratch), may enhance students' computational thinking and problem solving skills [2].

These contributions illustrate that learning scientists, instructional designers and educational facilitators are beginning to consider MBEG as a viable approach by which to complement current education [43], specifically regarding maths. Moreover, the innate ability of MBEG to foster children's playful movement whilst learning, makes them attractive research candidates in the intersection of education and MMD. However, despite the aforementioned benefits of integrating opportunities for playful behaviour in educational contexts, research addressing how MMD generated from children's play and problem solving interactions with MBEG is lacking. Exploring this untapped wealth of potential may help researchers advance the design and development of MBEG (e.g., via games that can recognise, support, adapt, or respond, to children's cognitive, affective, physiological and behavioural states in real-time during game play sessions).

2.2 Multimodal data in child-computer interaction

MMD combines multiple data modalities both from physical and digital dimensions, allowing computational methods to access and analyse them [11]. Leveraging on the key attributes of MMD, such as temporality and possible direct access to novel measurements of cognitive, affective, and physiological processes, consists of relatively new sources of information in the domains of learning sciences and CCI [19, 52]. There is an ongoing and increasing debate on MMD opportunities to contribute to theories regarding human behaviours in learning contexts [19]. Recent studies, such as Worsley and Blikstein [102] and Lee-Cultura et al. [52], indicate that the existing strategies for analysing MMD may contribute more significant perspectives to complex learning processes than conventional methods. In a similar vein, a recent literature review on MMD for young children [16], demonstrates this advantages, while also highlighting potential ethical issues.

Large-scale MMD collection of children's affective and behavioural data is a relatively new practice in CCI research [41]. In recent years, the CCI community has argued over the potentiality and ethical impositions of tailoring children's experiences using ubiquitous technologies and their produced traces (e.g., MMD) [42]. Related works encourage the use of MMD for analysis of complex interactions between children and systems [10, 54]. This is mutually motivated by different data stream's capacities to inform on key aspects of children's behaviour [10, 54] and lead to a comprehensive understanding of their interplay. For instance, kinaesthetic data combined with system logs, have been used to evaluate and support children's short term memory during MBEG play [48]. To predict children's learning performance in a construction task, kinaesthetic data was combined with electrodermal activity (EDA) and video data [101]. Improvements in wearable sensors have bolstered research using EDA and Blood Volume Pulse (BVP) [62, 72]. Moreover, integration of EDA and BVP data with system logs have explained differences between various task performance levels in construction-based activities [11]. Furthermore, hand movement and video data supported the interpretation of children's understanding of learning material [4]. Further, range of gesture/movement, combined with video, speech and eye tracking, was used to explain children's engagement with different activities [3]. By analysing gaze tracking, facial expressions and speech, it is also possible to automatically recognise real-time social signals, and understand affective states, including children's basic and complex emotions (i.e., happiness, sadness, confusion, frustration) [23, 61].

Standard access to MMD devices has increased their adoption for understanding and/or explaining children's play [17, 64] and problem solving behaviour [1, 86]. For example, audio and video recordings were used to detect children's affective states as they solved sorting and pattern recognition problems [105]. System log and facial video analysis have also led to the development of constructive user-friendly experiences to promote fun and learning in the acquisition of programming skills [1]. More recently, Sridhar et al. [86] demonstrated the use of Heart Rate Variability (HRV) and galvanic skin response to differentiate between children's cognitive-affective states as they executed tasks of variable mental effort. Physiological data has been leveraged during learning tasks to help manage cognitive load, as both overload and underload can result to weakness [14, 50]. EDA, HRV and affective states have also been assessed to understand children's level of physiological arousal [35] and stress [21], the dynamics of goal-oriented open-ended gameplay, proxemics, and to encourage group collaboration [17]. Collectively, these studies highlight the importance of using MMD to further our understanding on children's experiences during learning and play activities.

2.3 Play and Problem Solving in CCI

When a child lacks the answer to a proposed problem, they engage in a cognitive processing called problem solving, directed at determining the solution. The problem solving process has four notable characteristics [57]. It is cognitive, involves representing and manipulating knowledge, is directed by the problem itself, and is personal (i.e., the individual knowledge and skills of the child determine the problem's difficulty or ease). Problem solving is an extremely important educational goal and is a behaviour that can greatly benefits children as they find themselves in new situations [57]. It is part of our daily lives [31] and is, thus, critical in children's learning and for their future integration to society [30, 46].

Play is a core activity for children as it contributes to their well-being and development. In particular, playful technologies have been successful in motivating children "off the couch" to support learning and recreational activities [20]. Many researchers in the fields of education and child psychology have studied how play promotes emotional, cognitive, language, and physical development, and can be seen as a positive and natural means of engaging children in problem solving and knowledge building [74, 99].

Play and problem solving are coherently and well integrated into most new learning technologies. Children assume different behaviours throughout various phases of the learning process. Thus, it is advisable to know how to distinguish and interpret them to promote better educational experiences. Related studies have examined ways of detecting such learning behaviours by utilising different features derived from system-generated data which identify the various stages of children's experiences [34, 40]. In particular, it is possible to understand when a child engages in guessing behaviour [100] or informed problem solving behaviour [13]. Previous research has employed various parameters to automatically detect children's behaviours, such as response times, number of attempts to respond correctly, and the response itself [47, 55]. Nevertheless, the bulk of these studies have been based on system logs and very few studies have examined the use of motion features to investigate children's play and problem solving behaviours during their interactions. For instance, Olugbade et al. [66] explored the automatic detection of reflective thinking through motion data while children solved a maths problem in order to tailor feedback/support via technology for learning. Shin et al. [81] analysed learner's finger movements as they played KitKit School, a tablet based game, recognising guessing behaviour from solution behaviour. Despite these limited studies, the majority of CCI works utilise video annotations (i.e. manual coding) conducted by experts as the standard method for characterising children's experience (e.g., playing, problem solving) [94]. This study aims to combine human coding methods and MMD to identify how children experience play and problem solving behaviours in the context of MBEG.

3 MARVY LEARNS: THE MBEG

In our study, children played a single-player MBEG called Marvy Learns, which aims to develop their maths skills, specifically concerning geometry. Children assisted a large blue monster, named Marvy, in sorting a collection of cards by moving them into labelled boxes according to geometric shape attributes. Each game session consisted of six cards to be sorted to the correct box (each card represents a problem to be solved). The game had two different types of geometry questions (i.e., shape-grid and shape net), which were used in separate game sessions. In shape-grid questions, as shown in Figure 1, each card showed a 4x4 grid with disconnected blue points. When connected, the points formed a 2D polygon.

Figure 1: Marvy Learns requires a child to match a grid card to a labelled box according to its geometric characteristics. *left:* child reads the instruction and assesses the cards to sort into the labelled boxes. *centre:* child chooses a trapezoid card. The card is selected when it turns blue. *right:* Marvy Learns requires a child to match a grid card to a labelled box according to its geometric characteristics.

These cards included shape-grids for: rhombi, trapezoids, isosceles triangles, right triangles, rectangles, and squares. In shape-net questions, each card displayed a flattened, or unfolded, 2D representation of a 3D shape. When folded, the resulting 3D shapes included: tetrahedrons, triangular prisms, cubes, and cuboids. In both cases, children were asked to either visualise the geometric shape that results from connecting the multiple grid points to their 2D shape name, or match the flattened 2D shape to its 3D shape name. The boxes were labelled with the shape names. In this way, children learned geometry shapes (e.g., qualities and characteristics of, different dimensional representations of) and the respective terminology by associating the cards with the shape names on the boxes. Marvy Learns also fosters logical and inductive thinking through practice of arranging and classifying objects.

For example, a collection of six cards that a child must sort, may consist of 3 rhombi grids, 2 trapezoid grids, and a square grid (Figure 1, left); with boxes labelled as Rhombus, Trapezoid, and Square. To answer a question, the child must examine the cards and read the box labels (i.e., see and understand the question), visualise the shape resulting from connecting the blue points on the grid, determine its corresponding labelled box, select the card by performing a specific gesture/posture, and re-locate the selected card to the labelled box by maintaining the posture and moving their body to the proper box to place the card. The Marvy avatar mirrors the child's movement, so arrangement of cards takes place as the child moves their body in physical space. In our example (Figure 1), the child would be expected to match 3 rhombi to the red Rhombus box, 2 trapezoids to the green Trapezoid box, and 1 square to the orange Square box.

When a question was answered correctly, children received a positive reinforcement message (e.g., "Good Work!", "Nice!"), coupled with celebratory animations (e.g., eruptions of confetti or sparkles). Incorrect answers prompted messages of encouragement (e.g., "Try Again!"), and the incorrectly matched card was automatically returned to its original location. Children were permitted unlimited match attempts and not penalised for incorrect answers. Moreover, to mitigate additional pressure during the game sessions, Marvy Learns did not display a timer or running game score. As such, the game did not encourage or discourage guessing behaviour [100]. Lastly, Marvy Learns affords opportunities for play through children's creative control of selected item cards and movement synchronised direction of Marvy.

4 METHODS

4.1 Context

Our study was conducted during winter of 2019, in collaboration with a grade six class from a local Norwegian public school. Researchers and the class maths instructor, provided children with a thorough explanation of the study and children were given the opportunity to participate by their own free will. The study took place in a room dedicated to concurrently accommodate two game sessions (i.e., two children), and was specifically setup to avoid distractions.

4.2 Participants

Our sample consisted of 26 typically developing children (10 M, 16 F), with an average age of 10.95 years (SD = 0.21 years). None of the children had prior exposure to MBEG. Children engaged in three gameplay sessions lasting, with an average duration of 8.63 minutes. Prior to their participation, verbal/written informed assent/consent was obtained by children and their guardians respectively. Additionally, each child received a gift card for their time. All procedures were granted prior approval from the national human research ethics organisation.

4.3 Procedure

We conducted a mixed methods study to investigate the use of MMD to explore children's play and problem solving behaviours during their interactions with Marvy Learns, a geometry focused MBEG. Children were given an Empatica E4 wristband and pair of Tobii eye-tracking glasses to wear. Researchers then introduced the children to the Marvy Learns game, explaining the game's objective, rules, and interaction techniques. Children were given opportunity to ask questions to gain further clarity where needed, prior to commencing the game. Children played three consecutive game sessions: one practice session, during which children exercised their understanding of the games rules and objective, and 2 non-practice sessions. The 2 non-practice sessions delivered a different types of geometry questions (i.e., shape-grid and shape-net, see section

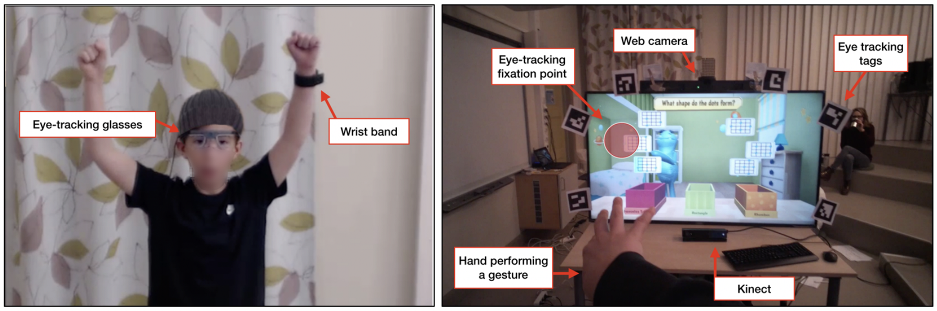

Figure 2: *left:* view from the Logitech web camera, of a child during an episode of play. They are wearing Tobii eye-tracking glasses and an Empatica E4 wristband. *right:* view from the Tobii eye-tracking glasses, showing the Logitech webcamera and Microsoft Kinect. The red dot on-screen represents the child's real-time point of focus. The child does not see this during gameplay.

3), and provided researchers the opportunity to collect more data per child. As well, previous experience informed us that children enjoyed the game and typically wanted to play several sessions. The experimental setup of a child, wearing the data collection devices, while playing Marvy Learns, is shown in Figure 2.

4.4 Data Collection

Children's game sessions were recorded using a Logitech video camera and three additional sensor devices: Tobii eye-tracking glasses, Empatica E4 wristbands, and Microsoft Kinect. System logs were also collected to observe event data and game analytics (e.g., response times, game score).

4.4.1 Logitech video camera. Children's interactions were captured using a front facing Logitech web camera, which was fastened to the top of the game play screen. Because the children's play space was located approximately 1.5-2 metres from the camera, the HD recording set to a zoom level of 200% and at 10 Frames Per Second (FPS).

4.4.2 Tobii eye-tracking glasses. Children's gaze data was collected using Tobii eye-tracking glasses with a 50Hz sampling rate and one-point calibration. The glasses contain an objective camera built into the nose-bridge, which was used, in conjunction with Tobii glass controller software, to capture children's field of view. Video resolution was 1920x1080 at 25 FPS.

4.4.3 Empatica E4 wristbands. Children's wrist data was captured using the Empatica E4 wristband, which collects four different variables: HRV (1Hz), EDA (64Hz), skin temperature (4Hz), and BVP (4Hz).

4.4.4 Kinect Skeleton. Children's skeletal data was collected using the Kinect sensor, which recorded at a sampling rate of 1Hz. This data represented the 3D position of 20 joints: head, shoulder-centre, spine and hip-centre, hand, wrist, elbow, shoulder, feet, ankle, knee, and hip (both left and right for the last 8).

4.5 Data Pre-processing

4.5.1 Video Coding Procedure. To identify and annotate children's naturally expressed behaviours, two researchers with expertise in learning theories and technologies (and who are authors), coded the video data by adopting an iterative inductive coding approach [58], as outlined in Figure 3. First, one coder viewed 27 game sessions (i.e., 36% of video data) and made observational notes on children's innate behaviours. The coders discussed the findings and, with support from previous relevant works in CCI [6, 28], categorised children's actions as *play*, *problem solving*, and *else* codes. Coders selected 15 new game sessions (i.e., 20% of video data) for separate coding, according the initial coding criteria, after which comparison of individual assignments commenced. To resolve discrepancies, the coders parallel coded three new game sessions, and engaged in discussion until consensus was reached. This resulted in revision of the original coding criteria. The original 15 game sessions were then re-coded separately according to the new re-aligned understanding and the results from these individual codings were used to determine inter-rater reliability. The consensus was measured using Cohen's kappa value. The result of the Kappa was 0.84, which illustrates substantial agreement [49]. The remaining 57 game sessions were divided and coded separately. In this way, the coders settled on three non-mutually exclusive observable codes (i.e., play, problem solving [sub-divided into informed [13] and guessing [8, 100]], and else), grounded in CCI literature on play and learning [6, 7, 28]. These behaviours manifested by children naturally and were not intentions of the game. *Play* [28] concerned interactions that were primarily directed at invoking fun, rather than solving Marvy Learns tasks. *Problem solving* described the process that children exhibited when attempting to match a card to the correct box (i.e., determining a solution to Marvy Learns tasks). For informed problem solving, children actively seeked to determine the *correct* solution to a given question by considering the information presented, and learning objective (i.e., determine the correct name of the geometric shape), and through deductive reasoning and logical thought, they attempted to construct a correct answer. Guessing behaviour [8, 100] was evident when a child "chose multiple incorrect answers before the correct answer" [8] (e.g., engaged in an exhaustive search). This was supported by children's response time Beal and Cohen [8]. Specifically, when children responded (i.e., grabbed a card and attempted to match it to a box) so quickly that their response time was less than the minimum time required to read the question and consider the options presented [100], children's behaviour was coded as guessing. The else code

Table 1: Coding scheme of children's play and problem solving (guessing, informed) behaviours during their interaction with Marvy Learns.

Code	Description
Play	Playing with the card included waving, exaggerated and exploratory movements. Playing with the avatar included dancing, jumping, waving limbs, clapping, exaggerated, exploratory, celebratory movements.
Guessing	Dragging a card to every box to find a match.
Informed	Theory or evidence-based card-box matching included matching based on card characteristics, and recognising similarities between boxed cards and unboxed cards.
Else	Moments where the child did not exhibit any play or problem solving behaviours, such as: tying their shoe, speaking to the experimenter.

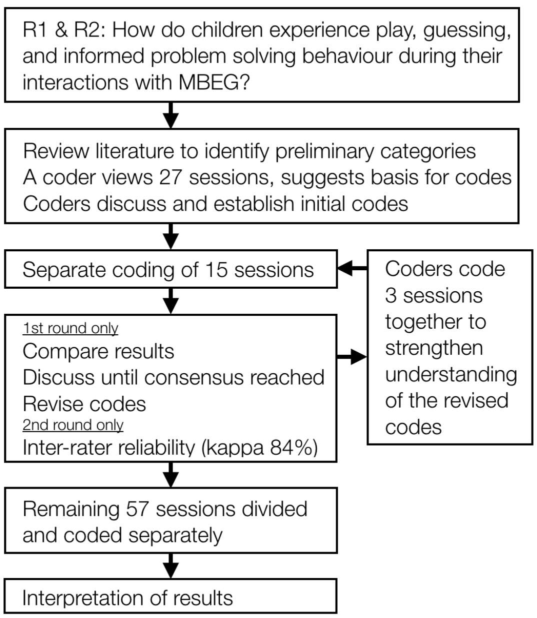

Figure 3: The phases outlining the inductive category development process, based on Mayring [58]

was applied to behaviour that could not be categorised as play or problem solving (e.g., asking experimenters a question, stopping to tie shoelace, etc). Examples of the different behaviours associated with the play and problem solving codes are presented in Table 1.

4.5.2 Tobii eye-tracking glasses. Fixations and saccades were detected using the Tobii's default algorithm [65]. We also removed blinks from the raw data before detecting the fixations and saccades by using a speed based filter. To remove the noise and subjective variance in the pupil diameter caused by factors such as brightness of screen, time of day, the child's gender, age, amount of sleep, and contextual biases, we considered the first 30 seconds of eye tracking data to normalise pupil dilation.

4.5.3 Empatica E4 wristbands. Similar to pupil dilation, HRV and EDA can have subjective and contextual biases. Correspondingly, we used the same pre-processing techniques to remove these biases from the HRV and EDA data, which weree used with the pupil diameter data (except the brightness-control step).

4.5.4 Kinect Skeleton. No pre-processing was required.

4.6 Measurements

We extracted the following six variables from the collected MMD: cognitive load, perceived difficulty, physiological stress, physiological engagement, emotional regulation, and fatigue (Table 2).

4.7 Data Analysis

Though our coding process included non-mutually exclusive codes (*play* and *problem solving*, Figure 4), the resulting data revealed that children's interactions with MBEG contained minimal overlaps. To accommodate for the few overlapping episodes, we adhered to the following guidelines. Overlaps occurring at the transition between the two codes (Figure 4, left), were discarded. Additionally, for short duration overlaps with one code completely embedded within the other, we discarded the minority episode and the dominant code prevailed. This resulted in removal of 7%, 2.3% (Figure 4, centre), 1.8% (Figure 4, right) of the data, respectively. Furthermore, the "else" category was also found to be negligible (1.7 % of total time).

To address the cognitive and affective differences that children experienced during the play and problem solving episodes (RQs), we conducted repeated-measure Analysis of Variance (ANOVA). For ANOVA, we use the student ID as a grouping factor. We checked the preconditions of ANOVA, i.e., normality and homoscedasticity, using Shapiro-Wilk's test [76] and Breusch-Pagan test [12], respectively. Furthermore, we applied Bonferroni corrections for p-values for multiple tests. Lastly, during a post-hoc analysis, we compared play episodes against informed problem solving and guessing episodes, separately. For this purpose, we used an ANOVA in a similar way as for our two main research questions, RQ1 and RQ2.

5 RESULTS

Initially, we verified the lack of age and gender bias for each of the six MMD measurements. Additionally, we did not find any relation between children's time-on-task effect and MMD measurements.

To address RQ1, we compared the MMD measurements between episodes of play and problem solving behaviour (see Figure 5). All statistical analyses reported were conducted with a significance

Table 2: Descriptions of the multimodal measurements used in our study, including their device source.

Measurement	Data source	Explanation and relevant citations
Cognitive load	Eye-tracker	This is the level of mental processing involved to solve a given problem. It is captured using pupil diameter [26], and is related to performance and different phases of learning [68, 103]
Perceived difficulty	Eye-tracker	This is captured using saccades from the eye-tracking glasses, and is computed as the saccade velocity in a given temporal window [22]. Perceived difficulty has been used in various problem solving and educational contexts to differentiate between learning performance levels [79, 80]
Physiological Stress	Wristband	This is computed as heart rate's increasing slope. The more positive the slope of the heart rate is in a given time window, the higher the stress is [90]. The heart rate has been use to measure stress in educational [79] and problem solving [59] contexts.
Physiological engagement	Wristband	This is computed as a linear combination of EDA's increasing slope and the arrival rate of EDA peaks. The more positive the slope of the EDA and the higher the rate of arrival of peaks in a given time window is, the higher the engagement is [39, 53].
Emotional regulation	Wristband	This measurement is directly computed from HRV, captured by the Empatica E4 wristband. The index of emotional regulation was computed as the rate of arrival of HRV peaks as suggested by [9, 97]. The lower the arrival rate of HRV peaks the higher is the emotional regulation.
Fatigue	Kinnect	Fatigue is proportional to the Jerk in the movement. Jerk is computed as the time derivative of the acceleration of the joint's movement (also known as the fourth derivative of displacement), and represents the average jerk of all of the joints. It is shown to be inverse of the energy spent [37].

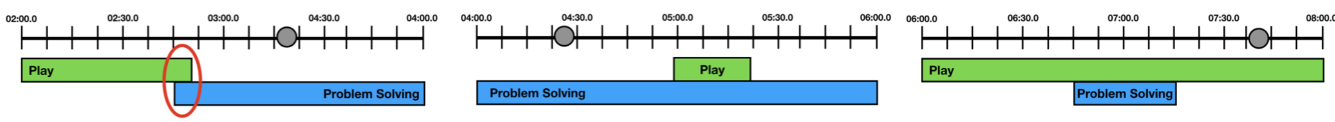

Figure 4: Annotations with overlapping episodes. *left:* episodes with discarded overlap.*centre:* episode considered as problem-solving. *right:* episode considered as play.

level of 0.05. Results from the ANOVA revealed that physiological stress in play episodes was significantly lower than physiological stress in problem solving episodes (F[1,73] = 52.14, p < .0001). As well, emotional regulation during play behaviour, was lower than emotional regulation in problem solving episodes (F[1,73] = 38.14, p < .0001). However, we found no significant difference between physiological engagement during play and problem solving episodes (F[1,73] = 3.80, p > .05). Cognitive load, on the other hand, was significantly lower during play episodes, than during children's problem solving (F[1,73] = 31.97, p < .0001). Similarly, perceived difficulty was also lower during play episodes than during problem solving episodes (F[1,73] = 9.25, p < .01). Finally, fatigue was higher during play episodes than during problem solving episodes (F[1,73] = 10.39, p < .01).

To address RQ2, we compared the MMD between children's guessing and informed problem solving episodes. We observed that children's physiological stress during informed behaviour was higher than their physiological stress during guessing behaviour (F[1,73] = 19.30, p < .0001). Similarly, emotional regulation was also higher during informed behaviour than during guessing behaviour (F[1,73] = 24.29, p < .0001). In line with the comparisons between play and problem solving, children's physiological engagement did not warrant significant differences between informed and guessing behaviour (F[1,73] = 2.95, p > .05). Considering the eye-tracking measurements, cognitive load was higher during informed behaviour than during guessing behaviour (F[1,73] = 24.29, p < .0001); while perceived difficulty for informed behaviour was lower than for guessing behaviour (F[1,73] = 4.11, p < .05). Finally, we did not encounter a significant difference between children's fatigue during informed and guessing episodes (F[1,73] = 1.39, p > .05).

After identifying the differences between children's play and problem solving episodes (RQ1), and their guessing and informed problem solving (RQ2) episodes, we performed a two part post-hoc analysis during which we compared the MMD measurements between children's play and informed problem solving episodes. We observed that children's physiological stress (F[1,73] = 22.70, p < 0.0001), emotional regulation (F[1,73] = 29.95, p < 0.00001), and cognitive load (F[1,73] = 22.14, p < 0.00001), were higher during informed problem solving episodes than during play episodes; however fatigue was lower for the informed problem solving (F[1,73] = 6.14, p < 0.001); and no significant difference was observed for engagement (F[1,73] = 1.80, p > 0.05) or perceived difficulty (F[1,73] = 1.47, p > 0.05). During the second set of post-hoc comparisons, we compared the MMD measurements between children's play and guessing episodes. We observed that physiological stress (F[1,73] = 15.47, p < 0.0001), emotional regulation (F[1,73] = 7.40, p < 0.01), cognitive load (F[1,73] = 5.01, p < 0.05), and perceived difficulty (F[1,73] = 8.78, p < 0.01) were higher during children's guessing than during play episodes; fatigue was lower when children guessed, as opposed to when they demonstrated play behaviour (F[1,73] = 18.54, p < 0.0001); and no difference between guessing and play episodes was observed for engagement (F[1,73] = 0.03, p > 0.05).

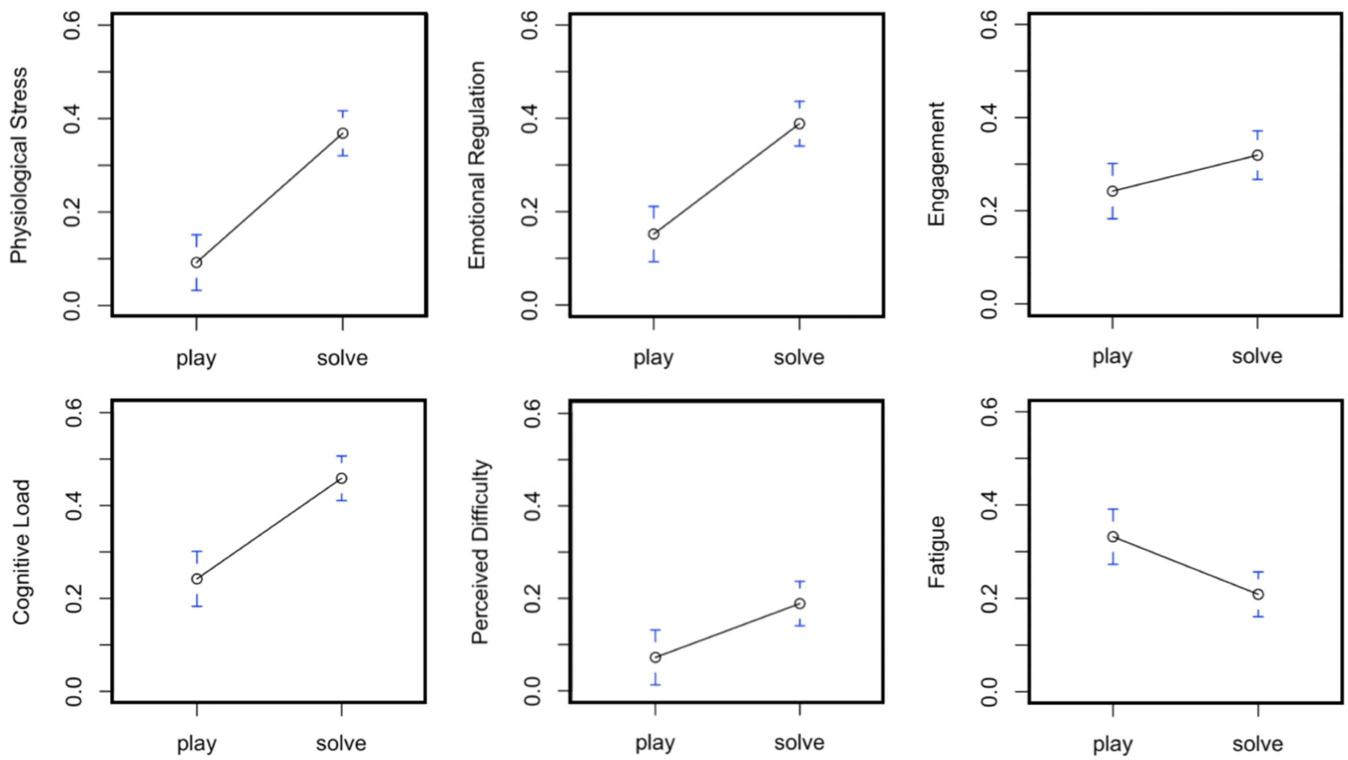

Figure 5: Differences between play and problem solving episodes (RQ1). The vertical bars show the 95% confidence interval.

6 DISCUSSION

6.1 Interpretation of Results

In the results section, we offer findings that address each RQ separately. However, our discussion is primarily centred on the post hoc analysis, as it provides the most detailed comparison/account of children's different play and problem solving behaviours (informed, guessing) exhibited during their interactions with MBEG through the lens of MMD.

During informed problem solving episodes, we observed that children's physiological stress, cognitive load, and emotional regulation were the highest. When children are presented a problem to solve, they may feel under pressure (external or self-imposed [60]) to answer the question correctly. An example of such pressure, is the "fear of evaluation or fear to perform" which is characteristic of examination anxiety [96]. During our study, children interacted with a MBEG; however, despite the intended "fun factor" that typically accompanies games, the pressure to academically perform (i.e., correctly match a card-box pair) may have elevated children's stress levels [38, 87]. This may explain why children's stress levels peaked during episodes of informed problem solving. In a similar vein, increased levels of cognitive load during informed problem solving may be directly linked to the mental effort that children expended as they reasoned through problems [88]. Lastly, emotional regulation relates to children's HRV [9, 97]. A plausible reason for observing the highest levels of emotional regulation during informed problem solving might be due to the immediate feedback that children received directly after they attempted to make a card-box match. The anticipation of the MBEG assessment/evaluation may have influenced children's heart rate, causing high levels of variability as children invested themselves in informed problem solving. Thus, in accordance with prior research [32], we hypothesise that the feedback in general, may have triggered cognitive and affective responses which affected learning, particularly during this ongoing tasks (i.e., a collection of questions asked in series).

Contrary to previous research [69, 84], our results did not indicate a connection between guessing behaviour and children's lack of engagement during their interactions with the MBEG (Figure 6, top right). As such, we propose that during episodes of guessing, children experienced some degree of external and/or self-imposed pressures to determine answers correctly (as during informed problem solving). Aside lack of engagement, it has been suggested that children also exercise guessing behaviour when a question's level of difficulty exceeds the scope of their prior domain knowledge and sits outside their problem solving ability [18]. We argue that under these particular circumstances (i.e, children do not possess the knowledge or problem solving skills required to reason through the task at hand, but have not disengaged), children must receive certain support (e.g., hints offered by the system or the educational facilitator). Moreover, such scenarios can be prevented, or greatly reduced, by adopting adaptive MBEG, which scale questions based

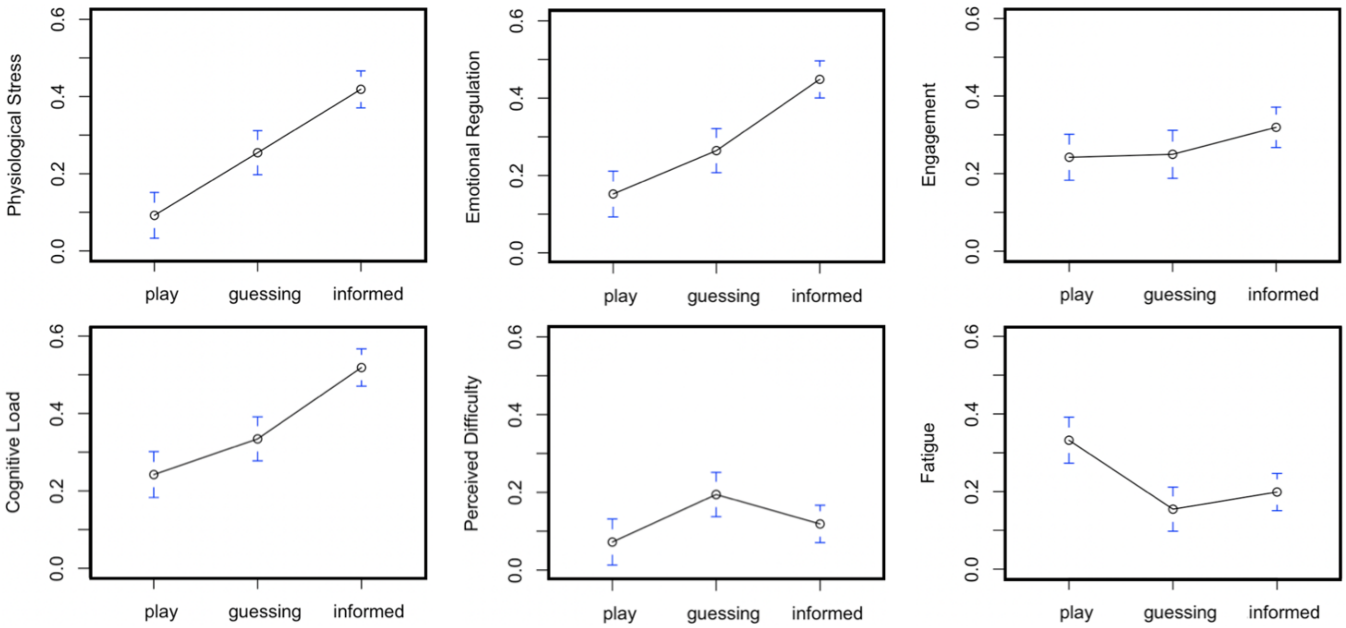

Figure 6: Post-hoc analysis of measurements showing significant differences between the playful, guessing and informed behaviours.

on their level of difficulty to appropriately match children's prior knowledge and demonstrated abilities (e.g., ongoing performance).

Children's experienced the lowest stress, cognitive load, and emotional regulation, during episodes of play. *Playing is freedom* [82], and as children exhibited play behaviour, we observed that they temporarily suspended their investment in Marvy Learns' rules and game objective, to freely and creatively engage in autotelic [82] exploratory movements. Within play episodes, children were not focused on solving the given task (i.e., making a card-box match), and lacked (external and self-imposed [60]) pressures to academically perform. Rather, play episodes were comprised of exploration into game aspects, such as its embodied affordances, and avatar synchronicity. This finding potentially indicates a connection between play behaviour and low levels of stress, emotional regulation and cognitive load. Moreover, children's fatigue was significantly higher during play, than during problem solving. Play encompassed children's behaviours which were directed at invoking fun, rather than making card-box matches. These behaviours manifested as physical interactions with on-screen content (e.g., Marvy and card), and resulting in frequent, quick, directional changes resulting in high amounts of movement (as confirmed by video data). On the other hand, it was observed that on several occurrences during problem solving episodes, children only moved once they knew which box they planned to match the card to (either after reasoning it through in the case of informed problem solving, or according to a brute force approach if guessing). In this way, during the problem solving episodes, children's directional change in movement (aka, fatigue) was reduced, as they already knew the exactly where they planned to put the card once selected, and executed this movement swiftly and purposefully.

6.2 Implications for Research

The bulk of contemporary research on children's interactive learning technologies centres on data collection from surveys [104], interviews [24] and observational practices (e.g., video data [5, 7]). We recognise the merits of such practices, as they afford rich contextual details that MMD measures, derived from wearable and ubiquitous sensors, lack. However, these techniques confine researcher to children's visible external behaviours (e.g., participant observation [27]) and subjectively reported states (e.g., motivation [93, 104], enjoyment [93]). Furthermore, manual video coding and compilation of survey and interview data, requires a considerable investment of resources (e.g., person hours) and does not warrant real-time analysis. As a result, researchers [85] in CCI have emphasised the need for more feasible (i.e., resource friendly) methods to provide additional perspectives (i.e., triangulation of data) and facilitate their work. On the other hand, MMD devices strengthen researcher's arsenal with which to understand children's invisible internal states (e.g., cognitive load, stress, perceived difficulty) through the objective, real-time investigation of children's interactive learning experiences (i.e. MBEG). Recently, CCI researchers on the cusp of MMD and education [33, 71, 101] have demonstrated the value of applying wearable and sensing devices to learning environments. However, MMD measures only report on specifically predefined metrics selected by researchers, and cannot discern children's nuanced behaviours.

Our research contributes to the detection of children's "invisible" states, and showcases how sensing technology can augment the results of experts' annotations. This opens avenues for synergies between sensor data and automated processes with traditional research methods (e.g., manual annotation) for deepening our understanding of children's behaviours during their interactions with technology. Moreover, by demonstrating the feasibility of using MMD to identify differences in children's play, guessing and informed problem solving behaviours, future work can leverage these differences to develop techniques which support human annotations. Consequently, this may reduce the number of human hours dedicated to video annotation, and enable a more elegant management of larger data sets (e.g., more participants, longer data capture sessions). As well, this MMD provides additional means for researchers to better understand and properly annotate challenging episodes by augmenting researcher's abilities and comprehension capacities via examination of both sensor and video data. Thereupon, and in accordance with a newly emerging discourse [51, 52], we highlight, and advocate, the complimentary of (human-centred) traditional practices (e.g., observation, interviews) and wearable and ubiquitous sensors, particularly when constructing a holistic understanding of children's learning experiences.

6.3 Implications for Design

Currently, feedback mechanisms integrated with children's educational technologies are primarily based on performance metrics (i.e., correctness of answers [89], response time [70]). However, the added immediacy of sensor data leading to the identification of children's needs and different behaviours during learning activities (e.g., lack of move off-task behaviour), may afford researchers and designers new opportunities for the creation and streamlining of educational technologies. Augmenting children's performance metrics with the real-time assessment of their cognitive, affective, physiological states, *and* learning behaviours (play, guessing, informed problem solving) from MMD, could be used to identify moments where children might benefit from additional assistance. For instance, an educational technology could notify the child's learning facilitator (i.e., teachers, parents, therapist) so that they can use their contextualised awareness of the child's current interactions, combined with knowledge of the child's learner profile to determine appropriate and timely means of feedback by which to direct the child through their playful learning experience.

In this study, children's behaviour was annotated as play, guessing, and informed problem solving. Though some may criticise play behaviour as a distraction from learning (e.g., steering children away from learning and problem solving),CCI literature recognises play as integral to the ways in which children learn [95], and celebrates the many successful roles of play, such as supporting children's social and emotional skills, and assisting children with special abilities [98]). On the other hand, one main purpose of MBEG is to facilitate the acquisition of knowledge through logical and inductive thinking; a characteristic of informed problem solving but not guessing. With this in mind, our underlying motivation for design recommendations centres on moving children away from guessing behaviour.

Our results found that a children's perceived difficulty was significantly higher when they exercised guessing behaviour, rather than during informed problem solving. It follows that once children's saccade speed exceeds a given threshold, children are most likely employing a guessing tactic. Therefore, it seems plausible to identify guessing behaviour from children's gaze. In hopes to encourage children towards an informed problem solving approach (and away from guessing behaviour), in situations where children demonstrate excessive saccade speed, we recommend offering problem solving hints (e.g., via the system or educational facilitator). Additionally, we can take preventive measures to reduce children guessing. Adapting content based on its difficulty is a common approach, but to achieve this, it is important to develop the necessary infrastructure (e.g., question bank ranked on difficulty) and intelligence in the system (i.e., learner model). Specifically, for games that do not afford immediate content adaptation (e.g., when problems are presented in tandem, such as in Marvy Learns), offering hints may help children identify the next logical steps to pursue. On the other hand, if the game content is malleable, then offering a slightly easier problem may benefit children by redirect the child towards informed problem solving. Additionally, re-engaging children with informed problem solving by scaling the content difficulty to a more manageable level, might also provide children with more opportunities to regain self-confidence using the informed strategy [18]. Scaling the content difficulty would then occur as children's problem solving capabilities increase.

6.4 Limitations

The findings of this paper tackle the differences on children's cognitive and behavioural states during their play and problem solving episodes. Our findings are subject to certain limitations, the age of the children represent an adequate population for the indented tasks (i.e., children who are grown enough to be able to read and perform the needed motor movements, but not too old for the indented content). However, younger or older children might produce slightly different results (e.g., more difficult to be stressed or to experience mental overload). It was inherent part on our research design to conduct an in-situ study; meaning that the produced data exhibit high ecological validity, but are vulnerable to potential disruptions and noise. During our study we did not experience any long disruption since we had an isolated space in a school, in turn, the data-quality was very high. Sensor based measurements involve inference, and inference within complex psycho-physiological constructs (such as ours) involves a degree of error. In our study we selected time tested sensing devices (e.g., Tobii, Empatica), and data streams that have been used to infer various learning and user experience-related constructs in previous works (e.g., [79]). Thus, although different methodological decisions (e.g., different variables) might have had a slight impact on the results, our general approach followed valid and time-tested devices and variables. Lastly, we recognise that our work constitutes a single study, and future longitudinal studies, are needed to determine how our findings maintain and further unfold over longer time (e.g., regular and everyday experience of play and problem solving episodes).

6.5 Ethical aspects

Sensing technology is becoming more and more prevalent in CCI research due to their inherent benefits (e.g., automatic, pervasive, temporal insights), but also on their ability to be employed with more traditional research methods and complement them [78]. From a practical standpoint, preparations of studies using sensing technologies require additional time and special attention to the ethics of data collection [56]. From a recent literature review in CCI [78], we can see the growth of sensing technologies in CCI research but also clear recognition to the need for thorough consideration of the ethical underpinnings. Most of the children and the adult facilitators (e.g., teachers, parents) will be new to some of the technologies, therefore, when planning to utilise MMD in your research, it is not enough to just describe the details in the consent form. Rather, it is extremely important to engage in discussion with the children and parents to explaining the rationale and added value of such data collections [78].

Specifically in this study, children were curious about what it looked like ("can I see my heart rate?"), and inquired as to our motivations for collecting it ("what does my heart rate tell you about how I learn/feel?"). For example, when shown the real-time tracing of gaze movement from the eye-tracking glasses, one child became excited that we could "see through his eyes" and then asked why this was important ("why do you care where I am looking?"). We replied that it was interesting to see how much time the child spent looking at different parts of the screen. This lead to more questions ("why does it matter how long I look at the boxes or the cards for?", and "What if I look at the monster instead?"). This anecdote illustrates the need to prepare child-friendly demonstrations and explanations of how the wearable and ubiquitous sensing technologies are used, the data collection process and reason for use. Children's are naturally curious, and, as with any other user-group, the indented use of the data collected from them needs be clarifying prior to enrolling them in research. Consequently, given that it is unlikely for written consents to capture all the potential questions from children, researchers must anticipate these inherent needs by allocating enough time to engage in meaningful child-friendly discussion and demonstration when obtaining children's assent, as well as throughout the study. Parallel to this is the need to ensure that parents feel comfortable with the data being collected. On a few occasions, children where excited to wear the devices, but parents needed additional assurance on the safety of the employed devices and our ability and intention to anonymise children's sensitive data.

7 CONCLUSION & FUTURE DIRECTIONS

Leveraging sensing technologies to explore children's experience while interacting with systems is a valuable evaluation method [52, 56], but, rather like an observation without dialogue, the data is enhanced in terms of its usefulness when the evaluation triangulates with other methods [56] including observations, interviews, verbalization methods and video annotation by human experts (in our case). In this study we employed sensing technology (i.e., eye-trackers, wristbands, Kinect motion sensor) to explore children's play and problem solving behaviour, and to provide academic stakeholders (i.e., children, parents, teachers) with educational support, during children's interactions with MBEG. To do so we first apply an inductive coding scheme to video data, to classify children's behaviours as play, guessing, and informed problem solving. Then, we explored potential differences on children's cognitive, affective and physiological states within these episodes, with the use of MMD data produced during children's interaction. The results demonstrate significant differences between children's cognitive, affective and physiological states during children's play, guessing and informed problem solving behaviours. We provide insights to support those episodes in educational games and help designers and facilitators to focus on children's experience. Future work should focus on developing consistent ways to account for children's important but sometimes "invisible" states thought either technological affordances (e.g., intelligent agent, visualisations) and/or by adapting the facilitation processes (e.g., revising the best-practices).

8 SELECTION AND PARTICIPATION OF CHILDREN

All the study's participants were students from public schools in Trondheim, Norway. The study took place at a science museum (Vitensenteret) and a primary school, in rooms strictly designated to the experimental setup. Data related to the study were collected after approval from the national Data Protection Official for Research (Norsk Senter for Forskningsdata), following all the regulations and recommendations for research with children. A researcher contacted the teacher and legal guardian of each child to obtain written consent permitting the data collection. Children were informed about the data collection process and their participation in the study was completely voluntary. In addition, children were able to withdraw their consent for the data collection at any time without affecting their participation in the activity.

ACKNOWLEDGMENTS

We would like to thank Kinems Inc. for providing for free user licenses of its educational gaming platform as well as access to the skeletal data.

REFERENCES

[1] Efthimios Alepis. 2011. AFOL: Towards a new intelligent interactive programming language for children. In *Intelligent Interactive Multimedia Systems and Services*. Springer, 199–208.
[2] Ioannis Altanis and Symeon Retalis. [n.d.]. A multifaceted students' performance assessment framework for motion-based game-making projects with Scratch. *Educational Media International* ([n. d.]).
[3] Alejandro Andrade. 2017. Understanding student learning trajectories using multimodal learning analytics within an embodied-interaction learning environment. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*. 70–79.
[4] Alejandro Andrade, Ginette Delandshere, and Joshua A Danish. 2016. Using Multimodal Learning Analytics to Model Student Behaviour: A Systematic Analysis of Behavioural Framing. *Journal of Learning Analytics* 3, 2 (2016), 282–306.
[5] Alissa N Antle. 2013. Exploring how children use their hands to think: An embodied interactional analysis. *Behaviour & Information Technology* 32, 9 (2013), 938–954.
[6] Saskia Bakker, Panos Markopoulos, and Yvonne De Kort. 2008. OPOS: an observation scheme for evaluating head-up play. In *Proceedings of the 5th Nordic conference on Human-computer interaction: building bridges*. 33–42.
[7] Laura Bartoli, Clara Corradi, Franca Garzotto, and Matteo Valoriani. 2013. Exploring motion-based touchless games for autistic children's learning. In *Proceedings of the 12th international conference on interaction design and children*. ACM, 102–111.
[8] Carole R Beal and Paul R Cohen. 2008. Temporal data mining for educational applications. In *Pacific Rim International Conference on Artificial Intelligence*. Springer, 66–77.
[9] Gary G Berntson and John T Cacioppo. 2004. Heart rate variability: Stress and psychiatric conditions. *Dynamic electrocardiography* 41, 2 (2004), 57–64.
[10] Matthew P Black, Daniel Bone, Marian E Williams, Phillip Gorrindo, Pat Levitt, and Shrikanth Narayanan. 2011. The usc care corpus: Child-psychologist interactions of children with autism spectrum disorders. In *Twelfth Annual Conference*

of the International Speech Communication Association.
[11] Paulo Blikstein and Marcelo Worsley. 2016. Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics* 3, 2 (2016), 220–238.
[12] Trevor S Breusch and Adrian R Pagan. 1979. A simple test for heteroscedasticity and random coefficient variation. *Econometrica: Journal of the Econometric Society* (1979), 1287–1294.
[13] Christine Bruce. 2008. *Informed learning.* Assoc of Cllge & Rsrch Libr.
[14] Fang Chen, Jianlong Zhou, Yang Wang, Kun Yu, Syed Z Arshad, Ahmad Khawaji, and Dan Conway. 2016. *Robust multimodal cognitive load measurement.* Springer.
[15] David Cohen. 2007. *The development of play.* Routledge.
[16] Lucrezia Crescenzi-Lanna. 2020. Multimodal Learning Analytics research with young children: A systematic review. *British Journal of Educational Technology* 51, 5 (2020), 1485–1504.
[17] Ciera Crowell. 2018. Analysis of Interaction Design and Evaluation Methods in Full-Body Interaction for Special Needs. In *23rd International Conference on Intelligent User Interfaces.* ACM, 673–674.
[18] Mihaly Csikszentmihalyi. 2020. *Finding flow: The psychology of engagement with everyday life.* Hachette UK.
[19] Mutlu Cukurova, Michail Giannakos, and Roberto Martinez-Maldonado. 2020. The promise and challenges of multimodal learning analytics.
[20] Bronwyn J Cumbo and Ole Sejer Iversen. 2020. CCI in the wild: designing for environmental stewardship through children's nature-play. In *Proceedings of the Interaction Design and Children Conference.* 335–348.
[21] Sabrine DHAOUADI and Mohamed Moncef BEN KHELIFA. 2020. A multimodal Physiological-Based Stress Recognition: Deep Learning Models' Evaluation in Gamers' Monitoring Application. In *2020 5th International Conference on Advanced Technologies for Signal and Image Processing (ATSIP).* IEEE, 1–6.
[22] Leandro L Di Stasi, Andrés Catena, José J Canas, Stephen L Macknik, and Susana Martinez-Conde. 2013. Saccadic velocity as an arousal index in naturalistic tasks. *Neuroscience & Biobehavioral Reviews* 37, 5 (2013), 968–975.
[23] Sidney D'Mello. 2013. A selective meta-analysis on the relative incidence of discrete affective states during learning with technology. *Journal of Educational Psychology* 105, 4 (2013), 1082.
[24] Afke Donker and Panos Markopoulos. 2002. A comparison of think-aloud, questionnaires and interviews for testing usability with children. In *People and computers XVI-Memorable yet invisible.* Springer, 305–316.
[25] Paul Dourish. 2004. *Where the action is: the foundations of embodied interaction.* MIT press.
[26] Andrew T Duchowski, Krzysztof Krejtz, Izabela Krejtz, Cezary Biele, Anna Niedzielska, Peter Kiefer, Martin Raubal, and Ioannis Giannopoulos. 2018. The index of pupillary activity: Measuring cognitive load vis-à-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems.* 1–13. https://doi.org/10.1145/3173574.3173856
[27] Adelmo Eloy, Danielle Cruz, Kavindya Thennakoon, and Wayne Grant. 2020. Buildagram: a constructionist environment for spatial reasoning. In *Proceedings of the 2020 ACM Interaction Design and Children Conference: Extended Abstracts.* 280–283.
[28] Julia Fink, Séverin Lemaignan, Pierre Dillenbourg, Philippe Rétornaz, Florian Vaussard, Alain Berthoud, Francesco Mondada, Florian Wille, and Karmen Franinović. 2014. Which robot behavior can motivate children to tidy up their toys? Design and Evaluation of" Ranger". In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction.* 439–446.
[29] Panagiotis Fotaris, Nikolaos Pellas, Ioannis Kazanidis, and Paul Smith. 2017. A systematic review of Augmented Reality game-based applications in primary education. In *Memorias del XI Congreso Europeo en Aprendizaje Basado en el Juego Graz.* 181–191.
[30] Robert S Friedman and Fadi P Deek. 2002. The integration of problem-based learning and problem-solving tools to support distributed education environments. In *32nd Annual Frontiers in Education*, Vol. 2. IEEE, F3E–F3E.
[31] Joachim Funke and Peter Frensch. 1995. Complex problem solving research in North America and Europe: An integrative review. *Foreign Psychology* 5 (1995), 42–47.
[32] Emily R Fyfe and Bethany Rittle-Johnson. 2016. Feedback both helps and hinders learning: The causal role of prior knowledge. *Journal of Educational Psychology* 108, 1 (2016), 82.
[33] M. N. Giannakos, S. Papavlasopoulou, and K. Sharma. 2020. Monitoring Children's Learning Through Wearable Eye-Tracking: The Case of a Making-Based Coding Activity. *IEEE Pervasive Computing* (2020), 1–12. https://doi.org/10.1109/MPRV.2019.2941929
[34] Michail N Giannakos, Kshitij Sharma, Sofia Papavlasopoulou, Ilias O Pappas, and Vassilis Kostakos. 2020. Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human-Computer Studies* 136 (2020), 102384.
[35] Robyn M Gillies, Annemaree Carroll, Ross Cunnington, Mary Rafter, Kelsey Palghat, Jeff Bednark, and Amanda Bourgeois. 2016. Multimodal representations during an inquiry problem-solving activity in a Year 6 science class: A case study investigating cooperation, physiological arousal and belief states. *Australian Journal of Education* 60, 2 (2016), 111–127.
[36] Sukeshini A Grandhi, Gina Joue, and Irene Mittelberg. 2011. Understanding naturalness and intuitiveness in gesture production: insights for touchless gestural interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* 821–824.
[37] Emmanuel Guigon, Pierre Baraduc, and Michel Desmurget. 2007. Computational motor control: redundancy and invariance. *Journal of neurophysiology* 97, 1 (2007), 331–347.
[38] Gary Hackbarth, Varun Grover, and Y Yi Mun. 2003. Computer playfulness and anxiety: positive and negative mediators of the system experience effect on perceived ease of use. *Information & management* 40, 3 (2003), 221–232.
[39] Uri Hasson, Orit Furman, Dav Clark, Yadin Dudai, and Lila Davachi. 2008. Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron* 57, 3 (2008), 452–462.
[40] Curtis R Henrie, Lisa R Halverson, and Charles R Graham. 2015. Measuring student engagement in technology-mediated learning: A review. *Computers & Education* 90 (2015), 36–53.
[41] Juan Pablo Hourcade, Alissa Antle, Lisa Anthony, Jerry Fails, O Iversen, Elisa Rubegni, Mikael Skov, Petr Slovak, Greg Walsh, Anja Zeising, et al. 2018. Child-computer interaction, ubiquitous technologies, and big data. *interactions* 25, 6 (2018), 78–81.
[42] Juan Pablo Hourcade, Anja Zeising, Ole Sejer Iversen, Mikael B Skov, Alissa N Antle, Lisa Anthony, Jerry Alan Fails, and Greg Walsh. 2018. Child-Computer Interaction SIG: Ubiquity and Big Data–A Changing Technology Landscape for Children. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems.* ACM, SIG07.
[43] Hui-mei Justina Hsu. 2011. The potential of kinect in education. *International Journal of Information and Education Technology* 1, 5 (2011), 365.
[44] Katherine Isbister, Mike Karlesky, Jonathan Frye, and Rahul Rao. 2012. Scoop!: a movement-based math game designed to reduce math anxiety. In *CHI'12 extended abstracts on human factors in computing systems.* ACM, 1075–1078.
[45] Keri Johnson, Jebediah Pavleas, and Jack Chang. 2013. Kinecting to mathematics through embodied interactions. *Computer* 46, 10 (2013), 101–104.
[46] David H Jonassen. 2000. Toward a design theory of problem solving. *Educational technology research and development* 48, 4 (2000), 63–85.
[47] E Joseph. 2005. Engagement tracing: using response times to model student disengagement. *Artificial intelligence in education: Supporting learning through intelligent and socially informed technology* 125 (2005), 88.
[48] Panagiotis Kosmas, Andri Ioannou, and Symeon Retalis. 2018. Moving bodies to moving minds: a study of the use of motion-based games in special education. *TechTrends* 62, 6 (2018), 594–601.
[49] J Richard Landis and Gary G Koch. 1977. The measurement of observer agreement for categorical data. *biometrics* (1977), 159–174.
[50] Charlotte Larmuseau, Jan Cornelis, Luigi Lancieri, Piet Desmet, and Fien Depaepe. 2020. Multimodal learning analytics to investigate cognitive load during online problem solving. *British Journal of Educational Technology* 51, 5 (2020), 1548–1562.
[51] Serena Lee-Cultura, Kshitij Sharma, Sofia Papavlasopoulou, and Michail Giannakos. 2020. Motion-Based Educational Games: Using Multi-Modal Data to Predict Player's Performance. In *2020 IEEE Conference on Games (CoG).* IEEE, 17–24.
[52] Serena Lee-Cultura, Kshitij Sharma, Sofia Papavlasopoulou, Symeon Retalis, and Michail Giannakos. 2020. Using sensing technologies to explain children's self-representation in motion-based educational games. In *Proceedings of the Interaction Design and Children Conference.* 541–555.
[53] Dominik Leiner, Andreas Fahr, and Hannah Früh. 2012. EDA positive change: A simple algorithm for electrodermal activity to measure general audience arousal during media exposure. *Communication Methods and Measures* 6, 4 (2012), 237–250.
[54] Christian S Loh. 2012. Information trails: In-process assessment of game-based learning. In *Assessment in game-based learning.* Springer, 123–144.
[55] Jing Lu, Chun Wang, Jiwei Zhang, and Jian Tao. 2020. A mixture model for responses and response times with a higher-order ability structure to detect rapid guessing behaviour. *Brit. J. Math. Statist. Psych.* 73, 2 (2020), 261–288.
[56] Panos Markopoulos, Janet Read, and Michail Giannakos. 2020. (in press) Design of Digital Technologies for Children. *Handbook of Human Factors and Ergonomics 5th ed* (2020).
[57] Richard E Mayer and Merlin C Wittrock. 2006. Problem solving. *Handbook of educational psychology* 2 (2006), 287–303.
[58] Philipp Mayring. 2000. Qualitative Content Analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [On-line Journal], http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm* 1 (06 2000).
[59] Shayan Mirjafari, Kizito Masaba, Ted Grover, Weichen Wang, Pino Audia, Andrew T Campbell, Nitesh V Chawla, Vedant Das Swain, Munmun De Choudhury, Anind K Dey, et al. 2019. Differentiating Higher and Lower Job Performers in the Workplace Using Mobile Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3, 2 (2019), 37.

[60] Ranjita Misra, Michelle McKean, Sarah West, and Tony Russo. 2000. Academic stress of college students: Comparison of student and faculty perceptions. *College Student Journal* 34, 2 (2000).

[61] Behnaz Nojavanasghari, Charles E Hughes, and Louis-Philippe Morency. 2017. Exceptionally social: Design of an avatar-mediated interactive system for promoting social skills in children with autism. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. ACM, 1932–1939.

[62] Omid Noroozi, Iman Alikhani, Sanna Järvelä, Paul A Kirschner, Ilkka Juuso, and Tapio Seppänen. 2019. Multimodal data to design visual learning analytics for understanding regulation of learning. *Computers in Human Behavior* 100 (2019), 298–304.

[63] Netta Ofer, Hadas Erel, Idan David, Tom Hitron, and Oren Zuckerman. 2018. A little bit of coding goes a long way: effects of coding on outdoor play. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 599–604.

[64] Ayumi Ohnishi, Kaoru Saito, Tsutomu Terada, and Masahiko Tsukamoto. 2017. Toward Interest Estimation from Head Motion Using Wearable Sensors: A Case Study in Story Time for Children. In *International Conference on Human-Computer Interaction*. Springer, 353–363.

[65] A Olsen. 2012. The tobii i-vt fixation filter: Algorithm description [white paper]. *Retrieved from Tobii Technology from http://www. tobiipro. com/siteassets/tobiipro/learn-and-support/analyze/how-do-we-classify-eyemovements/tobii-pro-i-vtfixation-filter. pdf* (2012).

[66] Temitayo Olugbade, Joseph Newbold, Rose Johnson, Erica Volta, Paolo Alborno, Radoslaw Niewiadomski, Max Dillon, Gualtiero Volpe, and Nadia Bianchi-Berthouze. 2020. Automatic Detection of Reflective Thinking in Mathematical Problem Solving based on Unconstrained Bodily Exploration. *IEEE Transactions on Affective Computing* (2020).

[67] Héctor Manuel Ocampo Orona, Guillermo Silva Maldonado, and Norma Patricia Salinas Martínez. 2015. Kinect team: Kinesthetic learning applied to mathematics using kinect. *Procedia Computer Science* 75 (2015), 169–172.

[68] Fred GWC Paas and Jeroen JG Van Merriënboer. 1994. Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of educational psychology* 86, 1 (1994), 122.

[69] Zacharoula Papamitsiou and Anastasios A Economides. 2016. Process mining of interactions during computer-based testing for detecting and modelling guessing behavior. In *International Conference on Learning and Collaboration Technologies*. Springer, 437–449.

[70] Zacharoula K Papamitsiou and Anastasios A Economides. 2013. Towards the alignment of computer-based assessment outcome with learning goals: the LAERS architecture. In *2013 IEEE Conference on e-Learning, e-Management and e-Services*. IEEE, 13–17.

[71] Sofia Papavlasopoulou, Kshitij Sharma, Michail Giannakos, and Letizia Jaccheri. 2017. Using eye-tracking to unveil differences between kids and teens in coding activities. In *Proceedings of the 2017 Conference on Interaction Design and Children*. ACM, 171–181.

[72] Héctor J Pijeira-Díaz, Hendrik Drachsler, Paul A Kirschner, and Sanna Järvelä. 2018. Profiling sympathetic arousal in a physics course: How active are students? *Journal of Computer Assisted Learning* 34, 4 (2018), 397–408.

[73] Ingrid Pramling Samuelsson and Eva Johansson. 2006. Play and learning—inseparable dimensions in preschool practice. *Early child development and care* 176, 1 (2006), 47–65.

[74] Mary Reilly. 1974. *Play as exploratory learning: Studies of curiosity behavior.* Sage Publications, Inc.

[75] Lloyd P Rieber. 1996. Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational technology research and development* 44, 2 (1996), 43–58.

[76] JP Royston. 1982. Algorithm AS 181: the W test for normality. *Applied Statistics* (1982), 176–180.

[77] Calvin Rubens, Sean Braley, Julie Torpegaard, Nicklas Lind, Roel Vertegaal, and Timothy Merritt. 2020. Flying LEGO Bricks: Observations of Children Constructing and Playing with Programmable Matter. In *Proceedings of the Fourteenth International Conference on Tangible, Embedded, and Embodied Interaction*. 193–205.

[78] Kshitij Sharma and Michail Giannakos. 2021. (forthcoming) Sensing Technologies and Child-Computer Interaction: Opportunities, Challenges and Ethical Considerations. *International Journal of Child Computer Interaction* (2021).

[79] Kshitij Sharma, Zacharoula Papamitsiou, and Michail Giannakos. 2019. Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology* (2019).

[80] Kshitij Sharma, Zacharoula Papamitsiou, and Michail N Giannakos. 2019. Modelling Learners' Behaviour: A Novel Approach Using GARCH with Multimodal Data. In *European Conference on Technology Enhanced Learning*. Springer, 450–465.

[81] Hyunjin Shin, Bugeun Kim, and Gahgene Gweon. 2020. Guessing or Solving? Exploring the Use of Motion Features from Educational Game Logs. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[82] Miguel Sicart. 2014. *Play matters.* mit Press.

[83] Carmen Petrick Smith, Barbara King, and Jennifer Hoyte. 2014. Learning angles through movement: Critical actions for developing understanding in an embodied activity. *The Journal of Mathematical Behavior* 36 (2014), 95–108.

[84] James Soland, Nate Jensen, Tran D Keys, Sharon Z Bi, and Emily Wolk. 2019. Are test and academic disengagement related? Implications for measurement and practice. *Educational Assessment* 24, 2 (2019), 119–134.

[85] Iris Soute, Saskia Bakker, Remco Magielse, and Panos Markopoulos. 2013. Evaluating player experience for children's outdoor pervasive games. *Entertainment Computing* 4, 1 (2013), 25–38.

[86] Priyashri K Sridhar, Samantha WT Chan, and Suranga Nanayakkara. 2018. Going beyond performance scores: understanding cognitive-affective states in kindergarteners. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. ACM, 253–265.

[87] Marianne B Staempfli. 2007. Adolescent playfulness, stress perception, coping and well being. *Journal of Leisure Research* 39, 3 (2007), 393–412.

[88] John Sweller. 2011. Cognitive load theory. In *Psychology of learning and motivation*. Vol. 55. Elsevier, 37–76.

[89] Sietske Tacoma, Paul Drijvers, and Johan Jeuring. 2020. Combined inner and outer loop feedback in an intelligent tutoring system for statistics in higher education. *Journal of Computer Assisted Learning* (2020).

[90] Joachim Taelman, Steven Vandeput, Arthur Spaepen, and Sabine Van Huffel. 2009. Influence of mental stress on heart rate and heart rate variability. In *4th European conference of the international federation for medical and biological engineering*. Springer, 1366–1369.

[91] Varun Thakkar, Adeet Shah, Mohini Thakkar, Abhijit Joshi, and Neha Mendjoge. 2012. Learning math using gesture. In *International Conference on Education and e-Learning Innovations*. IEEE, 1–3.

[92] Katherine R Thorson, Tessa V West, and Wendy Berry Mendes. 2018. Measuring physiological influence in dyads: A guide to designing, implementing, and analyzing dyadic physiological studies. *Psychological methods* 23, 4 (2018), 595.

[93] Chih-Hsiao Tsai, Yin-Hao Kuo, Kuo-Chung Chu, and Jung-Chuan Yen. 2015. Development and evaluation of game-based learning system using the Microsoft Kinect sensor. *International Journal of Distributed Sensor Networks* 11, 7 (2015), 498560.

[94] Damyanka Tsvyatkova and Cristiano Storni. 2019. A review of selected methods, techniques and tools in Child–Computer Interaction (CCI) developed/adapted to support children's involvement in technology development. *International Journal of Child-Computer Interaction* 22 (2019), 100148.

[95] Lev S Vygotsky. 1978. Mind in society: The development of higher mental processes (E. Rice, Ed. & Trans.).

[96] Justin W Weeks and Ashley N Howell. 2012. The bivalent fear of evaluation model of social anxiety: Further integrating findings on fears of positive and negative evaluation. *Cognitive Behaviour Therapy* 41, 2 (2012), 83–95.

[97] DeWayne P Williams, Claudia Cash, Cameron Rankin, Anthony Bernardi, Julian Koenig, and Julian F Thayer. 2015. Resting heart rate variability predicts self-reported difficulties in emotion regulation: a focus on different facets of emotion regulation. *Frontiers in psychology* 6 (2015), 261.

[98] Cara Wilson, Margot Brereton, Bernd Ploderer, and Laurianne Sitbon. 2018. MyWord: Enhancing engagement, interaction and self-expression with minimally-verbal children on the autism spectrum through a personal audio-visual dictionary. In *Proceedings of the 17th ACM Conference on Interaction Design and Children*. 106–118.

[99] Margaret Wilson. 2002. Psychonomic bulletin & review. *Six views of embodied cognition* 9 (2002), 625–636.

[100] Steven L Wise and Xiaojing Kong. 2005. Response time effort: A new measure of examinee motivation in computer-based tests. *Applied Measurement in Education* 18, 2 (2005), 163–183.

[101] Marcelo Worsley and Paulo Blikstein. 2015. Leveraging multimodal learning analytics to differentiate student learning strategies. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*. ACM, 360–367.

[102] Marcelo Worsley and Paulo Blikstein. 2018. A multimodal analysis of making. *International Journal of Artificial Intelligence in Education* 28, 3 (2018), 385–419.

[103] Chi Yang, Chun-Hui Jen, Chun-Yen Chang, and Ting-Kuang Yeh. 2018. Comparison of animation and static-picture based instruction: Effects on performance and cognitive load for learning genetics. *Journal of Educational Technology & Society* 21, 4 (2018), 1–11.

[104] Kelly Yap, Clement Zheng, Angela Tay, Ching-Chiuan Yen, and Ellen Yi-Luen Do. 2015. Word out!: learning the alphabet through full body interactions. In *Proceedings of the 6th Augmented Human International Conference*. ACM, 101–108.

[105] Serdar Yildirim and Shrikanth Narayanan. 2009. Recognizing child's emotional state in problem-solving child-machine interactions. In *Proceedings of the 2nd Workshop on Child, Computer and Interaction*. ACM, 14.