International Journal of Child-Computer Interaction 31 (2022) 100355

Contents lists available at ScienceDirect

International Journal of Child-Computer Interaction

journal homepage: www.elsevier.com/locate/ijcci

Children's play and problem-solving in motion-based learning technologies using a multi-modal mixed methods approach

Serena Lee-Cultura *, Kshitij Sharma, Michail Giannakos

Department of Computer Science, Norwegian University of Science and Technology, 7491, Trondheim, Norway

ARTICLE INFO

Article history:
Received 1 November 2020
Received in revised form 10 July 2021
Accepted 11 July 2021
Available online 23 July 2021

Keywords:
Multimodal data
MMLA
Learning Analytics
Sensors
Cognitive Load Theory
Educational Technologies
Embodied Interaction
Multi-Modal mixed methods approach

ABSTRACT

Motion-Based Learning Technologies (MBLT) offer a promising approach for integrating play and problem-solving behaviour within children's learning. The proliferation of sensor technology has driven the field of learning technology towards the development of tools and methods that may benefit from the produced Multi-Modal Data (MMD). Such data can be used to uncover cognitive, affective and physiological processes during learning activities. Combining MMD with more traditionally exercised assessment tools, such as video content analysis, provides a more holistic understanding of children's learning experiences and has the potential to enable the design of educational technologies capable of harmonising children's cognitive, affective and physiological processes, while promoting appropriately balanced play and problem-solving efforts. However, the use of an MMD mixed methods approach that combines qualitative and MMD data to understand children's behaviours during engagement with MBLT is rather unexplored. We present an in-situ study where 26 children, ages 10–12, solved a motion-based sorting task for learning geometry. We continuously and unobtrusively monitored children's learning experiences using MMD collection via eye-trackers, wristbands, Kinect joint tracking, and a web camera. We devised SP3, a novel observational scheme that can be used to understand children's solo interactions with MBLT, and applied it to identify and extract children's evoked play and problem-solving behaviour. Collective analysis of the MMD and video codes provided explanations of children's task performance through consideration of their holistic learning experience. Lastly, we applied predictive modelling to identify the synergies between various MMD measurements and children's play and problem-solving behaviours. This research sheds light on the opportunities offered in the confluence of video coding (a traditional method in learning sciences) and MMD (an emerging method that leverages sensors proliferation) for investigating children's behaviour with MBLT.

1. Introduction

Play can be characterised as "an expression of children's actions and their own creative meaning" (Pramling Samuelsson & Johansson, 2006) and is generally regarded as a cornerstone of children's behaviour. When children play, they surrender to imaginative thinking and lose themselves in a multi-sensory experience derived on their own volition (Rieber, 1996). Research shows that in the context of learning, play behaviour promotes valuable outcomes, such as increased enjoyment (Rubens et al., 2020), elevated engagement (Ofer, Erel, David, Hitron, & Zuckerman, 2018; Yu, Zheng, Tamashiro, Gonzalez-millan, & Roque, 2020), and amplified motivation (Fotaris, Pellas, Kazanidis, & Smith, 2017; Liu et al., 2019; Radu, 2014); thus contributing to healthy childhood development. Moreover, children's learning activities are often more effective when they are playfully framed (Cohen, 2007), with certain types and applications of play being more efficient in assisting learning and children's development (Hainey, Connolly, Boyle, Wilson, & Razak, 2016; Lai, Ang, Por, & Liew, 2018; Lillard et al., 2013; Wilkinson, Taylor, & Readman, 2018).

One such genre, made possible by recent technological advancements, is Motion-Based Learning Technologies (MBLT). MBLT utilise sensing technologies (e.g., Microsoft Kinect) to capture, map and interpret, players' body movements as an input to the learning task (Bartoli, Corradi, Garzotto, & Valoriani, 2013). In absence of tangible input devices, these "touchless" technologies provide a more fluid (Grandhi, Joue, & Mittelberg, 2011; Nielsen, Störring, Moeslund, & Granum, 2003; Villaroman, Rowe, & Swan, 2011), and engaging learner experience (Hsu, 2011). As such, MBLT are being investigated as a promising interaction paradigm in children's learning experiences (where children's learning experiences are defined as discrete segments of time, encapsulating their personal interactions with the learning environment Schmidt, Tawfik, Jahnke, & Earnshaw, 2020).

* Corresponding author.
E-mail address: serena.leecultura@ntnu.no (S. Lee-Cultura).

https://doi.org/10.1016/j.ijcci.2021.100355

Previous research (Blikstein & Worsley, 2016; Sharma & Giannakos, 2020) illuminates the value and usefulness of wearable and ubiquitous data sources for collecting MMD, including eye-tracking, skeletal coordinates and skin-conductance data. In particular, sensing technologies and the resulting MMD, afford us real-time access to information regarding participants' cognitive, affective, and physiological, processes, which are not always directly observable (Thorson, West, & Mendes, 2018). Moreover, contemporary research depicts that wearable and ubiquitous data sources (e.g., Lee-Cultura, Sharma, Papavlasopoulou, Retalis and Giannakos, 2020; Papavlasopoulou, Sharma, Giannakos, & Jaccheri, 2017; Sharma, Papavlasopoulou and Giannakos, 2019; Worsley et al., 2016) possess the capacity to uncover information that is pivotal in explaining the various aspects of children's problem-solving behaviour and predicting aspects of student's learning experiences (Cukurova, Giannakos, & Martinez-Maldonado, 2020; Giannakos, Sharma, Papavlasopoulou, Pappas and Kostakos, 2020). Therefore, augmenting children's MBLT experiences with MMD capture and analytics may also afford researchers with an enriched view of children's interactions with MBLT and deeper insights into the processes (e.g., cognitive load, stress, arousal) that children undergo during MBLT.

The goal of this study is to identify the connections between children's behaviours (specifically, play and problem-solving), and the produced MMD, during children's interactions with MBLT. Formally, we define play and problem-solving behaviours as the different ways that children express themselves when interacting with the MBLT. As such, behaviours are encompassed within the learning experience. Further, we aim to determine how the relationships between children's expressions (play and problem-solving) and the produced MMD might provide insights that inform the instructional design and use of MBLT. Specifically, in this paper we ask the following research questions (RQ):

RQ1: How do play and problem-solving behaviours manifest in MBLT?

RQ2: How can MMD inform on different elements of play and problem-solving behaviours in MBLT?

RQ3: How can the synergy between MMD and elements of play and problem-solving behaviours help us support the design and instruction of MBLT?

To address these questions, we conducted an in-situ study in which 26 children, aged 10–12, solved a motion-based sorting task targeting geometry skill development. We automatically, continuously and unobtrusively collected children's MMD (i.e., video from webcam, gaze from eye-tracking glasses, physiological from wristbands, and skeletal from Microsoft Kinect). To investigate the relationship between children's MMD and their behaviours (play and problem-solving) during children's motion-based interactions, we employed a multi-step mixed methods approach. First, we employed qualitative video coding to understand how children manifest play and problem-solving behaviours, and to identify additional characteristics (e.g., physical activity level, social interaction) that may contribute to shaping their MBLT experience (RQ1). Next, we used statistical analyses to examine the relationship between the human-labelled video coding and the MMD measurements that were automatically extracted (RQ2). Lastly, we employed predictive modelling techniques to identify the relationships which most accurately forecast children's task performance (as measured by correctness of their responses). This mixed-methods approach allows us to understand the synergies between the MMD and the different elements of play and problem-solving behaviour in MBLT (RQ3).

2. Related work and background theories

The theoretical underpinning of our work is driven by Piaget's learning theory on cognitive development, in particular, his model on schema construction. We draw on research heavily influenced by Piagetian themes, namely Cognitive Load Theory (CLT) and Cognitive Theory of Multimedia Learning (CTML), and Embodied Cognition. Additionally, we build upon prior studies in play and MMD in CCI. In this section, we offer an overview of the relevant works that ground and guide our research.

2.1. Piaget's theory of cognitive development, cognitive load theory, and multimedia learning

Piaget (2003)'s theory of cognitive development provides a theoretical model for understanding how children acquire and develop knowledge. Piaget (2003) hypothesised that children's cognitive growth evolves across four stages differentiated by qualitative developmental differences. Of particular interest to our work, is the Concrete Operational stage (i.e., ages 7–11 years old), wherein children begin to develop the capacity for logical, organised thought, and inductive reasoning, and advance their understanding of concepts such as transitivity and object classification (Piaget, 2013a). Abstract thought emerges in the Formal Operational stage (i.e., ages 12 and older), wherein children become more confident grappling with deductive reasoning and hypothetical scenarios.

A central tenet of Piaget (2003)'s theory finds that children are born with basic mental structures by which knowledge and learning is developed. These structures, namely *schemas*, which Piaget (2003) regarded as building blocks of intelligent behaviour, represent mental and physical actions derived from our experiences. Formally, a schema is defined as "a cohesive, repeatable action sequence possessing component actions that are tightly interconnected and governed by a core meaning" (Piaget & Cook, 1952). As a schema matures, its level of complexity evolves, and it may become connected to, or embedded with, more primitive schemas. Piaget introduced that knowledge creation is an active process, stimulating the assimilation and accommodation of schemas. *Assimilation* describes the act of relating the new information to preexisting schemas that were developed in a previous context. *Accommodation* is the process of reorganising or revising an existing schema, or creating a new schema, to better integrate the new information (Lefrançois, 2000). The continual interweaving of these two concepts results in *adaptation* of the child's schemas (i.e., what the child knows) (Lefrançois, 2000). Moreover, Piaget argues that children's cognitive development primarily progresses through the ongoing process of establishing *equilibrium* between assimilation and accommodation progresses and claims that in its absence, "the schemas will not be altered as the situation requires and subjects will fail to solve the problem" (Lefrançois, 2000).

Sweller (1988, 1989) introduced CLT, which concerns the focus and distribution of cognitive resources throughout learning and problem-solving. It recognises three types of cognitive load: intrinsic, extraneous and germane. Each of these is associated with different aspects of the cognitive process of knowledge acquisition and manipulation. *Intrinsic* cognitive load relates to the innate complexity of the given learning material. This is determined by the number of novel elements (i.e., schemas), and corresponding connections, which must be considered and stored in the learner's working memory at a given instant. A greater number of elements requiring concurrent storage in working memory, yields a higher intrinsic load. This indicates that educational materials which minimise the amount of information simultaneously presented, may benefit the learner by implicitly helping to manage their levels of intrinsic cognitive load. *Extraneous* cognitive load stems from the learning material's presentation, and is thus influenced by a task's instructional design (Sweller, 1989). Specifically, distracting and unnecessary elements that are integrated into the presentation of learning material, impose on the learner's mental capacity by occupying cognitive resources (i.e., activating superfluous schemas) that would better serve other aspects of the learning and problem-solving process (i.e., the management of new incoming information and integration of information into long-term memory via assimilation and accommodation). Lastly, *germane* cognitive load derives from the construction and automation of the learning content held in working memory, into mental representations (i.e., schemas), which are required to integrate the information in long-term memory for permanent storage. As we develop our capacity of what we store in long term memory regarding a specific topic, future learning of related content becomes easier due to an increase in preexisting and readily available relevant schemas.

The CTML offered by Mayer (2002), is an extension of Sweller's CLT (Plass, Moreno, & Brünken, 2010) and describes how learners accumulate their knowledge through images and words according to three core assumptions. The *dual channel assumption* asserts the existence of two separate channels used to acquire, represent and manipulate learning content. The auditory-verbal channel is responsible for managing words (spoken or printed text), and the visual-pictorial channel processes images (and other visuals such as animations, video, drawings, photographs). By the *limited capacity assumption* (Baddeley & Logie, 1999), the amount of information that each channel can process at one time is constrained, and exceeding these upper thresholds results in cognitive overload of the associated channel. The *active processing assumption* suggests that the synthesis of new knowledge occurs as learners actively participate in cognitive processes within these channels, such as identifying relevant information, constructing and automating pictorial and verbal schemas, and integrating newly developed schemas with preexisting knowledge in long-term memory. Lastly, while Piaget's theory of Cognitive Development describes the advancement of knowledge production, both Sweller's CLT and Mayer's CTML, are focused on tailoring best practices for presenting information to support learning.

2.2. Play in CCI

Play encompasses activities carried out for pleasure, utilises information already possessed by the child, but does not contribute to the construction of new knowledge (Piaget, 2013b). Rather, play results from disequilibrium caused by the dominance of assimilation (conversely, changes in the child's developmental progress, i.e., learning, correspond to the supremacy of accommodation Lefrançois, 2000). When engaged with a learning task, children play by stepping away from the task's structure and objective, to freely explore and interact with various elements within the task's learning environment. From a Piagetian perspective, ongoing play translates to the repeated assimilation of the involved schemas, which reinforces these them, and increases their accessibility for subsequent learning opportunities (Lefrançois, 2000).

Over the past decades, technological advancements and ease of device accessibility have contributed to a rapid growth in the study of children's technology-enhanced play in CCI research (Giannakos, Papamitsiou, Markopoulos, Read and Hourcade, 2020). Play has been explored from various perspectives, and in an extensive range of contexts. Researchers have proposed guidelines for designing playful interactions to facilitate outdoor play (Bekker, Sturm, & Eggen, 2010), developed theoretical measurement tools for the assessment of fun (Markopoulos, Read, & Giannakos, 2021; Tisza, Gollerizo, & Markopoulos, 2019) and observational schemes for the systematic evaluation of children's play qualities (Soute, Bakker, Magielse, & Markopoulos, 2013). Children enjoy individual play and repetitive sensorimotor actions (Markopoulos et al., 2021), which stabilises their existing schemas through practice (Lefrançois, 2000). Moreover, play during learning is well regarded as a catalyst for enjoyment (Rubens et al., 2020), engagement (Jakobsen et al., 2016; Ofer et al., 2018; Yu et al., 2020), and motivation (Fotaris et al., 2017; Liu et al., 2019; Radu, 2014). Consequently, play has seen much traction as a driving force used to promote children's learning across numerous domains. This especially applies to science, technology, engineering, and mathematics (STEM), where educational and CCI design researchers have explored the relationship (and benefits) between learning and play (Ofer et al., 2018), and focused heavily on the integration of play as a support vehicle for developing children's maths (Arroyo et al., 2017; Li, Van der Spek, Hu, & Feijs, 2019) and computational thinking skills (Brooks & Sjöberg, 2020; Yu et al., 2020). These studies and a recent mapping of CCI research (Giannakos, Papamitsiou et al., 2020) demonstrate that play is a growing and fertile topic, ripe for exploring its potential in CCI.

In this paper, we attempt to further our knowledge on children's play by augmenting video observation with MMD. By exploring the ways that MMD support and enrich our understanding of children's playful experiences in the context MBLT, we enable developers to leverage playfulness and educators to employ MBLT appropriately.

2.3. MMD in CCI

MMD's key affordances, such as temporality and direct access to new indicators of cognitive and affective processes, offer new forms of information to the field of learning sciences and CCI (Lee-Cultura, Sharma, Papavlasopoulou, Retalis, 2020). There is an ongoing and growing discussion on MMD's potential to contribute to theories about human behaviours in learning contexts (Cukurova et al., 2020). Recent studies such as Lee-Cultura, Sharma, Papavlasopoulou, Retalis (2020) and Worsley and Blikstein (2018) argued that the existing strategies for analysing MMD could provide more meaningful insights into complex learning processes than traditional approaches.

In recent years, the CCI community has engaged in discussions regarding the promised benefits and ethical impositions of utilising MMD to tailor children's experiences, and further CCI research (Hourcade et al., 2018; Sharma & Giannakos, 2021). Previous works advocate the use of MMD to analyse the complex interactions exchanged between children and the systems they employ (Black et al., 2011; Loh, 2012). This endorsement is collectively driven by different data stream's capacities to inform on the unique qualities of children's behaviour (Black et al., 2011; Loh, 2012) and contribute to a holistic understanding of their experiences. For example, system logs combined with the kinaesthetic data were used to assess and support children's short term memory in an MBLT (Kosmas, Ioannou, & Retalis, 2018). Kinaesthetic data were combined with the electrodermal activity (EDA) and video data to predict children's learning performance in a construction task (Worsley & Blikstein, 2015). Additionally, hand movements and video data were used to explain children's understanding levels of learning material (Andrade, Delandshere, & Danish, 2016). Further improvement in the MMD range of gesture/movement and video with speech and eye-tracking were also used explain different activities that the children engaged with (Andrade, 2017).

Off-the-shelf access to MMD devices has increased their use to understand and/or explain children's problem-solving behaviour

(Alepis, 2011; Clabaugh, Sha, Ragusa, & Mataric, 2015; Sridhar, Chan, & Nanayakkara, 2018) and playfulness (Crowell, 2018; Crowell, Sayis, Benitez, & Pares, 2019; Ohnishi, Saito, Terada, & Tsukamoto, 2017). For example, audio and video recordings were use to detect children's affective state while they solved sorting and pattern recognition problems (Yildirim & Narayanan, 2009). Analysis of log files and facial video have also contributed to the creation of constructive user-friendly interaction to facilitate fun and learning in acquisition of programming skills (Alepis, 2011). These studies provide a few examples on how MMD might enable researchers to better understand children's experiences while interacting with technology.

2.4. Embodied cognition and motion-based learning technologies

Piaget (2003) hypothesised that from an early age, a human's understanding of their surrounding world is shaped by their senses and bodily movements (Lefrançois, 2000). In line with this, several recent empirical studies suggest that movement influences learning and problem-solving processes, particularly in the domains of gesture and mathematics (Broaders, Cook, Mitchell, & Goldin-Meadow, 2007; Cook, Mitchell, & Goldin-Meadow, 2008). For example, Cook et al. (2008) showed that subsequent to hearing teacher's explanation of mathematics concepts, children who were instructed to produce specific gestures while solving maths problems outperformed children who were restricted to only speaking. CLT cannot account for such learning outcomes, and in response to the growing body of studies demonstrating similar results (Alibali & Nathan, 2012; Goldin-Meadow, 2005; Goldin-Meadow et al., 2012; Goldin-Meadow, Nusbaum, Kelly, & Wagner, 2001), researchers have derived alternate and expanded models, theories and guidelines (Hornecker & Buur, 2006; Hurtienne & Israel, 2007; Skulmowski, Pradel, Kühnert, Brunnett, & Rey, 2016; Wilson, 2002), to explain and support the role of bodily movement in the cognitive processes that learners undertake during learning and problem-solving activities.

Embodied cognition emphasises the importance of the connection between the brain and body, and its relationship to the surrounding environment in the acquisition, development and understanding of knowledge (Wilson, 2002). *Embodied schemas* (Lakoff & Johnson, 2008) result from the unconscious development of metaphors which arise from "experiential gestalts relating to the body's movement, orientation in space, and its interaction with objects" (Antle, Corness, & Droumeva, 2009). Due to their ability to represent abstract concepts via metaphor, embodied schemas are regarded by some as "common primitives of thought" (Hurtienne & Israel, 2007). *Embodied metaphors* (Antle et al., 2009), which describe metaphorical extensions grounded by embodied schemas, underpin the conceptualisation of information by forming connections between concrete actions and abstract ideas. Goldin-Meadow et al. (2001) found that when gesturing through maths problems, participants were able to express the same ideas as if using speech. They claim that performing gestures subconsciously transfers some cognitive load on to additional cognitive systems (i.e., motor-gestural system), through the creation of visualspatial representation formats, which may enrich the information encoding process. This reduces demands on a learner's working memory by relinquishing resources associated with the verbal-auditory and pictorial-visual channels, to assist in the development of "higher quality schemas in long term memory". Moreover, McNeill (1992)'s topology of gestures identifies *iconic gestures*, where the movement depicts semantic content directly via the motion trajectory of the hands (i.e., motion to grab an object as a way of selecting it), and *metaphorical gestures*, where the gesture depicts semantic content by way of metaphor (i.e., motioning a line between an object and its name to convey that they represent the same shape). Alibali and Nathan (2012) posits that these *representational gestures* (i.e., the grouping of iconic and metaphorical gestures) manifest mental simulations of action and perception when produced, and that the relationship between perception and action is both dependent and bi-directional. Further, they argue that via gesture production, mathematical cognition is embodied, and they reinforce that integrating bodily movement into educational environments may lessen demands on working memory in the context of mathematical learning and problem-solving.

Although advancements in motion sensing technologies (e.g., Microsoft Kinect) have only recently facilitated the development of MBLT, researchers are beginning to consider movement, particularly play (Bartoli et al., 2013; Lindgren & Moshell, 2011), as a powerful pedagogical tool to promote children's learning (Hsu, 2011). In MBLT, children use their body to naturally, and playfully, interact with learning material to develop cognitive skills (Yap, Zheng, Tay, Yen, & Do, 2015) using representational gestures.

Research centred on MBLT for proficiency in maths, permeates numerous divisions, such as calculus (Orona, Maldonado, & Martínez, 2015), algebra (Johnson, Pavleas, & Chang, 2013), arithmetic (Thakkar, Shah, Thakkar, Joshi, & Mendjoge, 2012), and geometry (Rosenbaum, Kaur, & Abrahamson, 2020; Smith, King, & Hoyte, 2014). Noteworthy studies demonstrate that MBLT may yield advantages to player's maths learning experience; particularly concerning enhanced problem understanding (Angotti & Bayo, 2012; Smith et al., 2014), reduced anxiety (Isbister, Karlesky, Frye, & Rao, 2012), and increased academic performance (Kourakli et al., 2017; Retalis et al., 2014; Tsai, Kuo, Chu, & Yen, 2015). These contributions illustrate that educational researchers and instructional designers are beginning to consider MBLT as a viable approach by which to supplement current educational instruction (Hsu, 2011), specifically regarding maths (Abrahamson & Bakker, 2016). Furthermore, MBLT intrinsic ability to entice children's playful movement whilst learning, makes them an attractive candidate for research in the confluence of education and MMD from wearable and sensing devices. However, despite the aforementioned advantages of introducing opportunities for play into educational domains, research addressing how MMD generated during children's play and problem-solving interactions with MBLT is lacking.

3. Marvy learns: The MBLT

In the Marvy Learns MBLT, children assist a large blue monster, named Marvy, with sorting tasks, such that they organise a collection of six cards by moving each card into a labelled box. Sorting one card is analogous to answering a single question "what is the name of this shape?". There are two variations of cards that were used in separate sorting tasks. In the first task variant, each card displayed a shape-net (i.e., a 2D shape that can be folded to create a 3D shape). The 3D shapes included were: tetrahedron, triangular prism, cube, or cuboid. In the second task variant, each card showed a 4×4 grid with disconnected blue dots that, if connected, formed a 2D shape. The 2D shapes included were: isosceles triangle, right triangle, rectangle, square, rhombus, and trapezoid. The names of the shapes acted as the box labels.

For example, a collection of six cards that a child must sort, may consist of 2 cube shape-nets, 2 tetrahedron shape-nets, and 2 triangular prism shape-nets (see Fig. 1a); with boxes labelled Triangular Prism, Tetrahedron, and Cube. The actions a child must take include reading the box labels, deciding which shapes correspond to each box, and moving them accordingly. More specifically, to answer the question, the child must (1) examine the cards and read the box labels (i.e., see and understand the question), (2) determine which cards correspond to each box

(a) initiate a selection: child assesses item cards (2 unfolded tetrahedrons, 2 unfolded cubes, and 2 unfolded triangular prisms), for sorting into labelled boxes.

(b) selection process: child chooses the pink tetrahedron card. The card is selected and movable once it has been filled with blue background.

(c) card categorisation: child has moved the pink tetrahedron card into the box labelled Tetrahedron.

Fig. 1. Marvy Learns requires a child to classify an item to a labelled box according to its attributes. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(i.e., mentally solve the problem), (3) perform a specific gesture/posture to select a card, and (4) re-locate the selected card to the labelled box by maintaining the posture and moving their body to the proper box. Marvy's arms mirror the arms of the child, so arrangement of cards takes place as the child moves their arms in physical space. In our example (see Fig. 1), the child would be expected to match the pink and red cube shape-nets to Cube, the yellow and green triangular prism shape-nets to Triangular Prism, and the red and pink tetrahedrons to Tetrahedron.

Children received positive messages (e.g., "Nice!", "Good Work!"), paired with celebratory animations (e.g., an eruption of sparkles, bubbles, or confetti) upon making a correct card-box match. Fig. 7c shows a congratulatory "Nice!" message, paired with celebratory bubbles in response to a correctly matched card-box pair. Incorrect matches prompted messages of encouragement (e.g., "Try Again!"), and the incorrectly matched card would automatically return to its original location. Children were not penalised for incorrect matches and were permitted unlimited match attempts. Additionally, there was neither a running timer, nor score displayed to the child during a sorting task, as to minimise any pressure which may originate from these element, but there was background music. Table 1, presented in Section 4.6 provides description of ways that children demonstrated play behaviour while playing Marvy Learns.

3.1. Design and learning opportunities

Marvy Learns questions are focused on developing a child's ability to visualise, identify, pattern match, and classify, geometric shapes according to the shape's properties. To accomplish this, children must extract the underlying patterns and properties associated with the geometric concepts presented (i.e. shapenet or gridnet, to shape name). Extracting patterns and properties corresponding to a shape's realisation in the next dimension (shapenet into 3D object, grid dots into 2D shape), relies on abstract thought. Mapping a specific shape image to a generalised shape name (i.e., the box label) depends on inductive reasoning. The interaction mechanism for answering Marvy Learns questions is specifically designed to provoke the creation and development of embodied metaphors through the production of representational gestures by motioning a line between a shape card and its label (or box filled with like shapes) to convey the "togetherness" of the items. Embodied metaphors are extensions of embodied schema, thus performing these representational gestures during the sorting task activates the adaptation processes (assimilation and accommodation) of the involved embodied schemas.

To clarify the roles played by schematic assimilation and accommodation consider the following example (see Fig. 1b). The child has sorted an unfolded triangular prism (yellow) and a tetrahedron (red), and they are now attempting to classify a second tetrahedron (pink) which is oriented differently than the previously classified prism. A few different scenarios may occur, of which we will discuss two. In one scenario, the child may already have thorough preexisting schemas representing tetrahedra, and instantly recognise that the unfolded pink shape they are considering is a tetrahedron. The child gestures the card to the box labelled "Tetrahedron" to represent that they are connected or the same, and the new information (i.e., associated with pink tetrahedron card) is assimilated into the child's preexisting schema. Another scenario sees that the child does not recognise the pink shape as a tetrahedron. In this case, the child may examine the previously sorted cards and labelled boxes, activating the schemas associated with each encountered element, and try to extract patterns or similarities to the pink shape card. They may note the similarities between the sorted red tetrahedron (that they have a supporting schema for) and the differently oriented unsorted pink shape card, and through transitivity, deduce that the pink shape is also a tetrahedron. Under the hood, the child has revised, or accommodated, their preexisting schema to include the new information from the differently oriented pink tetrahedron, resulting in the construction of new knowledge.

While Marvy Learns does not explicitly offer teaching (i.e., the process of instructing learning content), the target educational content (i.e., geometry concepts) is integral to the sorting task (rather than being tangential to the game's plot-line Fisch, 2005), such that engaging with the task directly exercises children's geometry knowledge. As the task progresses, and a child encounters a mix of known and unknown shapes needing to be sorted, the child will strive for equilibrium as they cognitively dance between the processes of assimilation and accommodation. In this way, Marvy Learns affords opportunities for learning through problem-solving, self-evaluation, trial and error and self correction; which all nurture children's self-awareness and -reflection.

4. Method

4.1. Context

Our study was conducted in a grade six class (10–12 year olds) at a local public school in Trondheim, Norway, in the fall of 2019. Children were provided a thorough explanation of the study by researchers and their maths instructor, after which they opted to participate on their own volition. Researchers/authors conducted the study in a room dedicated strictly to the experiment. The room was arranged to accommodate two MBLT, so two children engaged with different sorting tasks simultaneously.

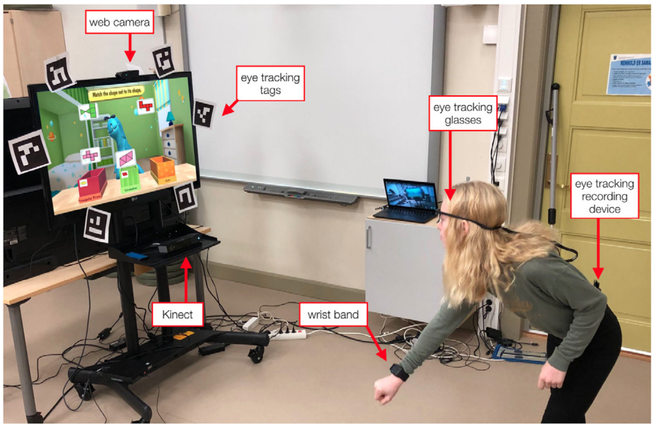


Fig. 2. The experimental set up of a child interacting with Marvy Learns MBLT. The MMD devices are indicated by the labels.

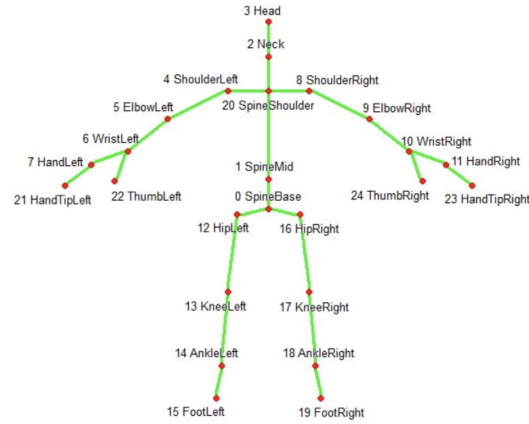


Fig. 3. The 25 joints captured by the Kinect data.

4.2. Participants

Our sample consisted of 26 typically developing children (10 F, 16 M) with an average age of 10.95 years ($SD = 0.21$ years). Children engaged in 3 Marvy Learns sorting tasks, for which they received a gift card for their participation. Prior to their participation, all children and their guardians provided verbal/written informed assent/consent, respectively. All procedures were approved by the Norwegian Centre for Research Data, (i.e., a national human research ethics organisation).

4.3. Procedure

We conducted a mixed-methods study to explore the connections between children's play and problem-solving behaviours (as identified from the video content analysis) with children's cognitive, affective and physiological processes (as derived from MMD), that occur during children's experiences with the MBLT. Children were given a pair of Tobii eye-tracking glasses, and an Empatica E4 wristband to wear. Fig. 2 shows the experimental set up of a child interacting with Marvy Learns while wearing the data collection devices. Children solved 3 consecutive Marvy Learns sorting tasks. The first task was for practice with the goal to familiarise children with the task's physical interaction mechanics (i.e., selection gestures and how to move an object on-screen using their body), and ensure that they understood the sorting task's objective and rules. Children were given an opportunity to ask questions to confirm their understanding. Recorded sorting tasks did not commence until the experimenter was confident that each child had a full understanding of these critical elements. In total, we recorded 76 sorting tasks. However, due to technical difficulties, one sorting task was discarded, resulting in 75 sorting tasks, averaging a length of 7.69 min (SD = 2.22). None of the children had prior exposure to MBLT.

4.4. Data collection

We recorded children's sorting task engagement via video camera, in addition we captured data from 3 devices: physiological data (with sensors for heart rate, blood-pressure, temperature and EDA levels) from wristbands, gaze data from eye-tracking glasses, skeleton data from Microsoft Kinect, and the event data from system logs.

Eye-tracking: We collected children's gaze data using Tobii eye-tracking glasses at 50 Hz sampling rate and one-point calibration. The child's field of view was video recorded using Tobii glass controller software and an objective camera built into the nose-bridge of the glasses. Video resolution was 1920 × 1080 at 25 Frames Per Second (FPS).

Empatica E4 wristbands: We used the Empatica E4 wristband to capture children's physiological data consisting of 4 different variables: Heart Rate Variability (HRV) (1 Hz), EDA (64 Hz), skin temperature (4 Hz), and Blood Volume Pulse (BVP) (4 Hz).

Kinect skeleton: Children's skeletal data were captured using Microsoft Kinect sensor. Data were recorded at a sampling rate of 1 Hz and consisted of the 3D position for 25 joints: head, shoulder-centre, spine and hip centre, hands, wrists, elbows, feet, ankles, knees, hips. See Fig. 3.

Video: a front facing Logitech web camera was installed to the top of the screen that displayed the sorting task, to capture facial expressions. The camera recorded HD video at 10 FPS and to target the child's face it was zoomed in to 200%.

4.5. Task performance

Children's task performance was determined for an entire sorting task, and calculated as follows:

$$performance = \frac{correct\ matches}{correct\ matches\ + incorrect\ matches} \quad (1)$$

where,
correct matches = number of times the child placed a card in the correct box.
incorrect matches = number of times the child placed a card in an incorrect box.

By this formula, the maximum task performance score attainable is **one**. This occurs only when the child matches all six cards to their correctly labelled box, without mistake (i.e, zero incorrect matches). Alternatively, a child can continually create incorrect match attempts (as we did not discourage this behaviour), meaning that there can be any number of incorrect matches. Thus, it can be deduced that there is no lower limit on the performance score.

4.6. Coding process and classification scheme

Method: Our initial research question (RQ1) aimed to explore how play and problem-solving behaviour manifested during children's interactions with MBLT. For video coding, the methodology proposed by Mayring (2000) was adopted. Mayring (2000) combines several techniques for systematic coding and analysis with an elaborate and systematic research process. According to Mayring (2000), the category system can be either derived

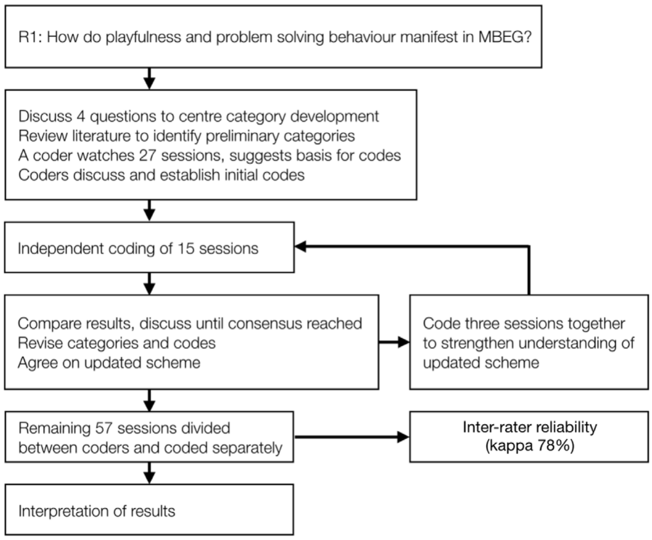


Fig. 4. The phases outlining the inductive category development process, based on Mayring (2000).

from a suitable existing theory, from related work or developed inductively from the codes during the analysis. Our approach was inductive, with support by previous works (Bakker, Markopoulos, & De Kort, 2008; Bartoli et al., 2013; Bhattacharya, Gelsomini, Pérez-Fuster, Abowd, & Rozga, 2015; Fink et al., 2014).

Inductive category development process: We selected children's behavioural events as the entities (e.g., aspects of interpretation) to be coded during the content analysis process. Prior to observing any video data, researchers discussed the following four questions: (1) what (sorting task) elements incite play behaviour?, (2) which levels of physical intensity (i.e., energy levels) do children demonstrate during their sorting task experience?, (3) which problem-solving behaviours are identifiable to the observer?, and (4) what additional factors, if any, influence children's play and problem-solving behaviour? Next, researchers engaged in an iterative inductive category development (Mayring, 2000) process, for which we describe complete details.

Coding procedure: Our inductive coding process, outlined in Fig. 4, was conducted by two coders with expertise in learning technologies and learning theories (and who are also authors), using NVivo 12. In the first phase, one coder watched 27 sorting tasks (i.e., 36% of the video data) and took observational notes on children's behavioural events, specifically in relation to the aforementioned four questions proposed to guide the development process of the observational behaviour classification system. The coders then met and discussion transpired to develop an initial set of categories and corresponding codes. Next, the two coders randomly selected 15 new sorting tasks (i.e., 20%) and coded them separately, according the initial coding criteria. After this, they reconvened to review the subset of coded sessions, compare individual coding results, discuss any discrepancies until resolutions were agreed upon, and refined the coding rules accordingly. To assist in this process, the coders watched three new sorting tasks together and discussed how they would code the children's behaviour until code assignment consensus was established. This resulted in the formation of new codes, as well as new understanding of a few of the rules defined in the initial coding criteria. The two coders then re-coded the same 15 sorting tasks separately, according to their newly re-aligned understanding. The results from the second round of individual coding were used to determine inter-rater reliability. The consensus was measured using Cohen's kappa value. The result of the Kappa was 0.78, which illustrates substantial agreement (Landis & Koch, 1977). In this way, the coders settled on four non-mutually exclusive observable behavioural events (i.e., categories), grounded in CCI literature on play and learning (Bakker et al., 2008; Bartoli et al., 2013; Bhattacharya et al., 2015; Fink et al., 2014), and 8 mutually exclusive codes. Table 1 presents the categories, codes and definitions used to classify play and problem-solving behaviour during children's interactions with MBLT. Following, the remaining sorting tasks ($n = 57$) were divided between the two coders and coded separately.

The classification scheme: The resulting category system (codes), which we call the SP3 observational scheme, was devised according to inductive coding, and inspired by previous works (Bakker et al., 2008; Bartoli et al., 2013; Bhattacharya et al., 2015; Fink et al., 2014) with a focus on children's play and problem-solving behaviour. In particular, the categories of SP3 are: **s**ocial, **p**hysical activity, **p**layfulness, and **p**roblem-solving behaviour. Each category was further reduced into mutually exclusive codes. The categories, associated codes and descriptions are presented in Table 1. Each code was applied across the duration of time over which it transpired. For example, if a child engaged in play behaviour with Marvy, the event was coded as *avatar play* from the moment the play behaviour began until the behaviour ceased. Furthermore, codes in some categories (i.e., physical activity) were always present, while others (i.e., playfulness, problem-solving, and social interaction) occurred at random durations in time.

Following, we discuss details pertaining to the codes associated with each category.

Problem — solving concerns the cognitive process that children demonstrated when determining a solution to the proposed problem (i.e., matching a card to the correctly labelled box). This category was decomposed into informed learning (Bruce, 2008), and trial and error, learning methods; resulting in two codes: *informed*, and *trial and error*.

Playfulness (Fink et al., 2014) was divided into two codes, namely *avatar play* and *card play*, based on the task element that the child actively included as their main focus for playful interaction (i.e., the avatar or a card). Playfulness is sporadically demonstrated by children during sorting task engagement, and may co-occur with children's other behaviours (e.g., physical activity).

Physical activity (Bakker et al., 2008) classifies the amount of energy associated with children's movement, into three different levels of intensity. Namely, *highly intensive*, *moderately intensive*, and *non-intensive*. This code category was applied across the complete duration of children's sorting task experiences.

Social interaction (Bakker et al., 2008) refers to children's interactions with the support person exclusively (as there were no spectators in the room, and thus no opportunities for other social behaviour). There is a single code: *social*. This event occurred at random intervals, and manifested as children asking questions to confirm their understanding of the given questions, or through children's commentary throughout the sorting task.

4.7. MMD pre-processing

To remove noise, personal and additional conditional biases from the sensor data, the following pre-processing was conducted.

Table 1
SP3 Observation scheme of children's behavioural events during their interactions with MBLT.

Category	Codes	Description
Playfulness	• card play • avatar play	• playful avatar control with card selected. Includes waving, exaggerated and exploratory movements • playful avatar control without card selected. Includes dancing, jumping, waving limbs, clapping, exaggerated, exploratory, and celebratory movements
Problem solving	• trial and error • informed	• dragging a card to every box, or different cards to the same box, until a match is found • theory or evidence-based card-box matching i.e., identifying match by card characteristics, recognising similarities between boxed cards and unboxed cards.
Physical activity	• highly intensive • moderately intensive • non-intensive	• exaggerated movements i.e., quickly swinging arms or jumping up to grab card, dancing, • walks and smooth moves • stands in place and moves so subtly that the avatar does not move
Social interaction	• social	• discussion with support personnel

Eye-tracking: the mobile eye-tracker records three separate data-streams: pupil diameter, objective video, and raw eye-tracking data. We applied a different pre-processing method to each stream.

Pupil data is recorded as the pupil diameter, which is prone to variances caused by several reasons (e.g., age, gender, time of the day, sleep quality, screen brightness). To accommodate for these personal and contextual biases, we followed the techniques described by Lee-Cultura, Sharma, Papavlasopoulou and Giannakos (2020) and Lee-Cultura, Sharma, Papavlasopoulou, Retalis (2020). Raw eye-tracking data is stored as each new position of the eye (i.e., where the child is looking), in the pixel coordinate system in the objective video's frame-of-reference. Due to extreme instability, this data must be divided into more stable moments of exploration (i.e., fixations) and long jumps from on-screen locations (i.e., saccades). Fixations and saccades were detected using the Tobii's default algorithm (Olsen, 2012). A filter was also applied to remove raw gaze points that classified as blinks during this process.

Finally, we pre-processed the mobile eye-tracking settings. The objective video is different for each user, which makes the comparison and calculation of eye-tracking measurements difficult. We mapped the fixations and saccades derived from the objective video's frame-of-reference to the screen-recording of the sorting task, as this is a more stable and comparable frame of reference across participants. Using fiducial markers which were taped to the frame of the screen (Fig. 2), we performed a computer-vision operation termed *homography* (see Appendix B). This operation finds corresponding points in different images from two different frames-of-reference. We computed the correspondences between the video from the eye-tracker's objective camera (i.e., objective video) and the screen recording video (i.e., ground-truth). The outcome of the process is summarised in Fig. 5a.

Empatica E4 Wristband: HR and EDA can have subjective and contextual biases as pupil dilation. Thus, we employed the same pre-processing techniques as with the pupil diameter data, to remove these biases from the HR and EDA data.

Kinect Skeleton: No pre-processing was required.

Video recordings: No pre-processing was required.

4.8. Measurements computed from MMD

From the MMD, we computed the following measurements, used in contemporary multimodal research for education and problem-solving (Andrade, Danish, & Maltese, 2017; Blikstein & Worsley, 2016; Lee-Cultura, Sharma, Papavlasopoulou, Retalis, 2020; Sharma, Papamitsiou and Giannakos, 2019; Worsley & Blikstein, 2018). Each MMD measurement was normalised between zero and one.

Cognitive load is related to the amount of mental effort invested when solving a given problem. We used eye-tracking data to compute cognitive load as a function of pupilary activity (Duchowski et al., 2018). Below, we provide the steps for computing cognitive load for each sorting task. Complete details can be found in Duchowski et al. (2018).[1]

Transitions among Areas Of Interest (AOIs): We defined three AOIs on screen: (1) the cards, (2) the correct box, and (3) the wrong boxes. Fig. 5b shows an example of the AOIs positions for one of the questions used in our study. For each card, we matched the respective correct box and the respective pairs of incorrect boxes. Once the AOIs on the stimulus screen were defined, we computed the proportion of gaze transitions from one AOI to another. We only computed the transitions between two distinct AOIs, and they were defined as three types of transitions: (1) between question and correct answer; (2) between question and incorrect answers; and (3) between correct answer and incorrect answers. Each transition was labelled based on the actual card-box pair. For example, if gaze from rectangle was followed by box other than one marked as rectangle, this transition was labelled as "transition to wrong", otherwise "transition to correct".

Information Processing Index (IPI): is the ratio of the global to local processing. Global processing is defined as a series of short fixations and long saccades. While the local processing is defined by the series of long fixations and short saccades (Poole & Ball, 2006).

Saccade speed: is the speed of the saccades. This measurement has been found to be associated with the perceived difficulty of a problem-solving task (Bronstein & Kennard, 1987; Smit & Van Gisbergen, 1989).

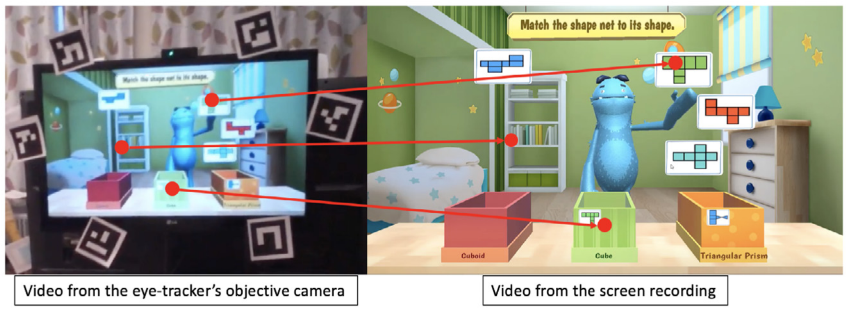
Mean HR: corresponds to the mean HR of the child per second. Increase in HR is often related to stressful situations (Harada, 2002; Herborn et al., 2015).

Number of EDA peaks: is the computed using the same method as described by Santini (Di Lascio, Gashi, & Santini, 2018) and is associated with the physiological arousal (Hasson, Furman, Clark, Dudai, & Davachi, 2008; Leiner, Fahr, & Früh, 2012).

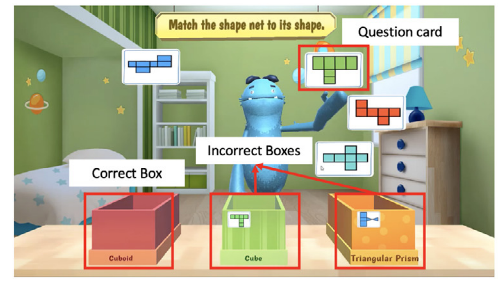
Phasic EDA level: EDA signal is comprised of two parts: the tonic and phasic components. The tonic component of the EDA signal is the one with slow evolving patterns. The phasic component of the EDA signal is the one with rapid changes and is found to be related to physiological arousal (Di Lascio et al., 2018). In this paper, we consider only the mean phasic EDA component as a measure of physiological arousal.

Motion stability: is computed using the average distance between the joints (Fig. 3) over two consecutive seconds. A shorter distance indicates a more stable posture. To achieve a positive

[1] This is a standard wavelet form that is implemented as one of the basic options in both the Python and Matlab signal processing libraries/toolboxes.



(a) Homography example. *left:* video from the eye-tracking glasses' objective camera. *right:* video from the screen recording.



(b) AOI example. The areas of interest are labelled and encapsulated in red squares.

Fig. 5. *left:* A homography example showing the synchronisation between the video data from the eye-tracking glasses' objective camera and the screen recording. *right:* AOI example showing the areas of interest.

scale, we inverted the average distance $(1+(1/distance))$ between the joints to compute the posture stability.

Motion Entropy: is computed using the Shannon Entropy of the movement of each joint detected in the Kinect data stream.

Motion Jerk: is computed using the time derivative of the acceleration of the joint's movement, and represents the average jerk of all of the joints. Jerk is shown to be inverse of the energy spent (Guigon, Baraduc, & Desmurget, 2007). In other words, the smoother is the movement (low jerk), the lower the energy spent should be.

Amount of Movement in specific body parts: we also computed the average distance travelled by the different body parts using the specific joints. Torso (joints: 0, 1, 20, 4, 8), left hand (joints: 5, 6, 7, 21, 22), right hand (joints: 9, 10, 11, 23, 24), left leg (joints: 12–15), left leg (joints: 16–19). This is calculated to capture the localised body movement. The total body movement could be misleading for certain movements, e.g., cases where only arms are moving and no other body parts are moving.

Match select time difference: We logged the timestamp when the child selected a card, and when the selected card was matched to a box. We computed the time difference between these two timestamps to show the time difference between selection and response times.

4.9. Data analysis

To answer RQ1 we used the video coding process explained in Section 4.6, Coding Process and Classification Scheme, with coding protocol shown in Table 1. Based on the results of our coding we present the content of each code and extract qualitative insights.

To answer RQ2 we adopt a two-step process that allows us to identify which MMD measurements (commonly used in MMLA) are associated with children's play and problem-solving behaviours (as indicated by the SP3 categories). First, we used Exploratory Factor Analysis (EFA) to identify the underlying relationship between the qualitative coding and the MMD measurements. The steps followed for the rotational factor analysis are outlined below.

(i) For the validity of the scales, factorial analysis with principal components and varimax rotation for the MMD features of each code category, was performed.[2]
(ii) For the convergent validity we computed the loadings (of the new dimensions) for the new rotated space with respect to all the quantitative vectors.
(iii) We then use the Cronbach's Alpha to compute the internal validity scores of the MMD measurements, to determine how they consistently explain a given video code. This is done in a backward elimination manner.

We chose EFA to analyse data for **RQ2** since this analysis allows us to understand shared variance of measured variables that is believed to be attributable to a factor or latent construct (Howard, 2016) and due to a lack of clear theoretical hypotheses in examining the relationship between the MMD measurements and the video codes. The studies that provided the inspiration for the current video coding scheme (Bakker et al., 2008; Bartoli et al., 2013; Bruce, 2008; Fink et al., 2014), are purely qualitative in nature. Furthermore, most studies related to the MMD measurements are either quantitative (Andrade et al., 2017; Giannakos, Sharma et al., 2020; Giannakos, Sharma, Pappas, Kostakos, & Velloso, 2019; Sharma, Papamitsiou et al., 2019) or examine qualitative codes from a narrow point of view (Andrade et al., 2016; Worsley & Blikstein, 2014, 2018). For this reason, and to explain the relationship between the *qualitative video codes* (the third column in Fig. 6), *quantitative MMD measurements* (the fourth column in the Fig. 6), we selected an exploratory method (i.e., EFA) which has the provision of incorporating both types of measurements and is commonly used in HCI and educational measurements research (Gaskin & Happell, 2014; Howard, 2016). With rotational factor analysis, we aimed to achieve a balance between the exploratory nature of the factor analysis and the statistical confidence of accurate factor loadings, while we considered the video codes as new "factors". Using this method, we were be able to find the many-to-many mappings between the video codes and MMD measurements. These mappings not only contain information about the association of the video subcodes and the MMD measurements, but also the positive/negative direction of these associations.

To answer RQ3 we used a predictive modelling approach that allows us to identify the synergies between the various MMD measurements and the elements of play and problem-solving behaviour. In other words, to identify which of the MMD measurements can best predict children's play and problem-solving behaviours (as indicated by the SP3 categories) and use these predictors to support the design and instruction of MBLT. To achieve this, we first divide the entire sorting task into individual video code episodes. We then compute all MMD measurements (described in Section 4.8) for each video code episode. We use the MMD features computed for the episodes to predict children's task performance.

[2] The new space was rotated in a way that with every rotation one principal component aligns with the nearest code. The distance with the code is computed using the angle between the new components and the MMD measurement vectors).

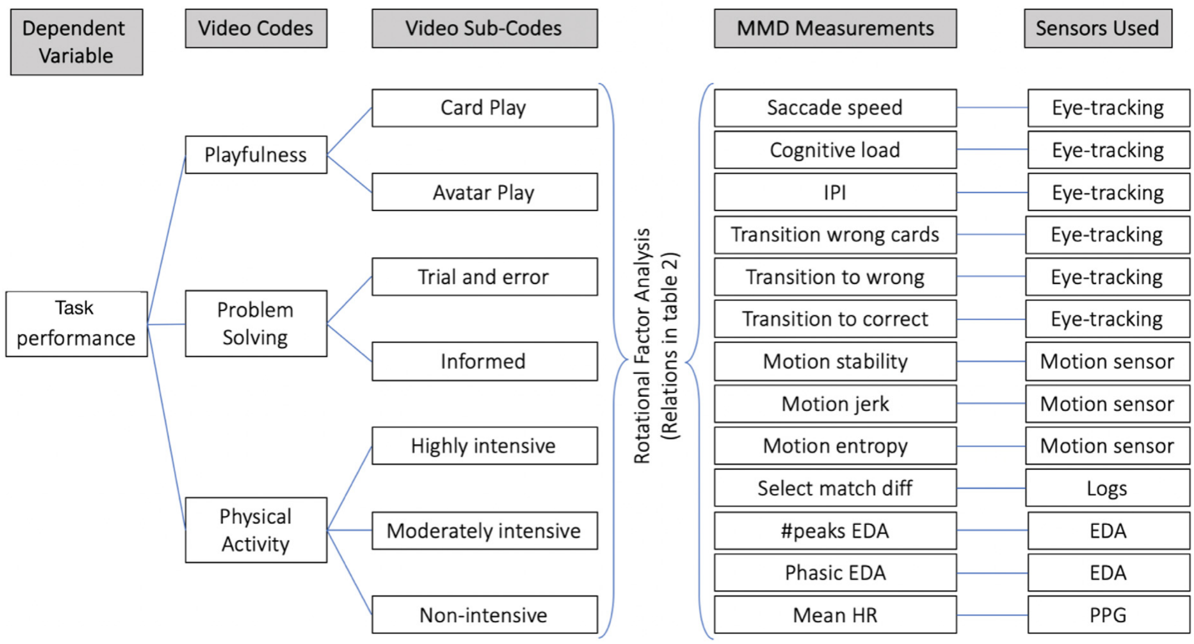


Fig. 6. The relation between the video codes and the MMD measurements.

This predictive analysis technique allows us to employ MMD measurements that have established relationships with problem-solving and other learning related behaviours (e.g., cognitive load, information processing index, heart rate) and identify the "most important measurements" with respect to those behaviours (also known as the most important *features* in machine learning terminology) (Sharma, Papamitsiou et al., 2019). Therefore, this predictive analysis technique can inform us on the set of MMD measurements (e.g., cognitive load and IPI during informed problem-solving) during children's play and problem-solving behaviours and their importance (or not) in predicting children's task performance.

These combinations, as the set of most important features from predictive modelling, can later be used to support design guidelines for a scaffolding tool which supports children during play and problem-solving activities. Specifically, we followed these predictive modelling steps:

(i) Divide the time into windows based on the duration of the video sub-codes.
(ii) Compute all MMD measurements from Section 4.8 for each of these windows.
(iii) Compute the average of the MMD measurements for windows with the same video sub-code label. For example, compute the average cognitive load for all informed problem-solving episodes. This step resulted in 130 features per child, per sorting task.
(iv) Use these feature for predictive modelling.

For prediction we used a random forest predictor with the normalised root mean squared error (NRMSE) as an evaluation measurement. Random forest predictions offer easy extraction of feature importance, and have been found to be a top performing algorithm in a large comparative study (Fernández-Delgado, Cernadas, Barro, & Amorim, 2014). For evaluation, NRMSE is the proposed metric for student models (Pelánek, 2015), and is used widely in learning technologies (Moreno-Marcos, Alario-Hoyos, Muñoz-Merino, & Kloos, 2018) for measuring the accuracy of learning prediction. For the computation of NRMSE, see Appendix C.

To remove sampling bias we performed **out-of-sampling testing** (i.e., leave-one-participant-out), and divided the dataset into 3 subsets: (1) training, (2) validation, and (3) testing. We kept the testing set aside (10% based on participant ID). The dataset is split based on participant identifiers. Each model is trained and validated using the training and validation sets with a cross validation. The cross-validation is performed using leave-one-participant-out.

5. Results

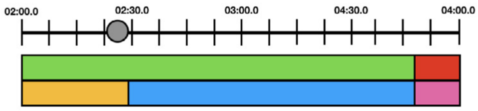
5.1. Manifestations of play and problem-solving in MBLT

We explored children's play and problem-solving behaviour through application of the SP3 categories to video data collected during their engagement with MBLT. Below we share insights on how each category was realised.

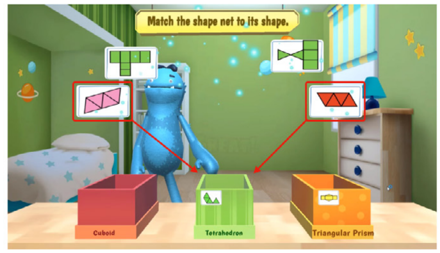
"*problem-solving – Finding the correct card-box pair*"

Various *trial and error* behaviours were identified across children's sorting tasks. Some children employed a strictly brute force approach in which they selected a card, and then systematically dragged the card to each box (typically from left to right), until a correct match was found. Alternatively, in a different trial and error behaviour, children would make two incorrect matches with the same card, and then change their focus to a different card. In these instances, two different scenarios occurred upon returning to the original card. Some children would immediately drag the card to the correct box, demonstrating that they had remembered the results from their previous trial and error behaviour. While other children would begin their brute force trial and error tactics from scratch, repeating incorrect matches that they had already encountered.

Informed problem-solving manifested in three different manners during children's interactions with Marvy Learns. Some children front-loaded their decision making by examining the cards and boxes, and selecting a card only after they had decided on the corresponding box (Fig. 9b). Other children selected a card, moved it closer to the boxes by lowering it on screen, and then



(a) Code distribution during a two minute segment of a child's sorting task. Initially, the child applied trial and error (i.e., orange strip) to determine a correct match between a green tetrahedron card and the tetrahedron box. When the match was found (denoted by the grey circle), the child switched to using an informed problem-solving behaviour (i.e., blue strip). Other colour codes green = moderate intensity, red = highly intensity, pink = avatar play.



(b) The sorting task view during the transition between trial and error, and informed problem-solving processes. During informed problem-solving, the child used the information learned from previously using trial and error behaviour, to identify the two remaining tetrahedron cards (circled in red) to be matched to the tetrahedron box.

(c) The sorting task view at the end of the informed problem-solving. During this phase, the child successively correctly matched all tetrahedron cards to the tetrahedron box using information learned during the trial and error phase.

Fig. 7. Overview of the events that occurred during a two minute segment of a child's sorting task where the child first applied the trial and error followed by the informed problem-solving behaviour.

held their selection gesture while they examined the boxes and made their decision (Fig. 9c). Lastly, some children displayed a short segment of trial and error, followed by a long duration of informed problem-solving behaviour. This describes the scenario when children first use trial and error to find a correct match, and then use the information learned to identify and match all cards of the same genre to the involved box (see Fig. 7).

In sum, we observed three different problem-solving behaviours that children employed during their interactions with MBLT. They were: informed, trial and error for informed, and brute force trial and error.

"*Social interaction*"

All children engaged in *social* interaction during their sessions, though most children demonstrated this behaviour at the beginning of each sorting task to establish or confirm their understanding of the proposed questions. The majority of children solved the problems in complete silence, however, throughout the sessions, a small number of children continued to verbalise their thoughts, either to themselves or the support person. Social interaction was used as a tool to acquire knowledge about how to proceed with the questions, and so it was strongly aligned with instances of informed problem-solving.

"*Playfulness – Constituent factors of fun*"

Children exhibited play through interactions with the avatar or a selected card. Each of these behaviours took on multiple forms, ranging from small to large scale movements. *Avatar play* was typically demonstrated as exaggerated actions which occurred when children first became aware of the high fidelity movement congruency they shared with Marvy. Primarily dancing, but also running on the spot, and uncontrollable arm waving (Fig. 8c), were frequently observed behaviours. Upon achieving a correct match, several children also celebrated by flapping their arms in a flying motion (Fig. 8a) or by shooting their arms up into a victory v-shape pose (Fig. 8b). A few children stepped side to side at varying speeds, as if testing the response latency of Marvy. More subtly, one child kept her body still while she bobbed her head back and forth in a playful manner while grinning from ear to ear, complete fixated on Marvy. Play behaviours with the cards (i.e., *card play*) were less frequent than avatar play. For example, we observed children demonstrating zestful and dramatic actions while dragging a selected card towards a box. One child consistently used both hands to quickly slam a card into a box, despite knowing that the second hand was not motion tracked, and therefore played no role in card movement (Fig. 9a). Another playful tactic observed was children bouncing a selected card between box tops. Children would drag a selected card towards a box until it was hovering directly above a box, and then jump the card to the adjacent box, and back again, as if to playfully test how close they could place the card to the box before the task considered a match attempt. This behaviour indicated elements of risk taking in their playfulness.

"*Physical activity – Intensity of children's actions*"

Children exhibited a wide spectrum of intensity regarding their physical activity during sorting tasks. *Non-intensive* physical activity was primarily observed during informed problem-solving, and when socially engaged with the support person. During informed problem-solving, non-intensive activity manifested as children determined a card-box match. Several children remained primarily motionless (Fig. 9b) or displayed only subtle head nods, as they compared a cards to boxes (and their content) prior to making a card selection. Other children swayed slowly from side to side. Children also frequently selected a card before determining its corresponding box, and would then either maintain the selection posture while holding their body completely still (Fig. 9c) or lower the selected card slowly, as they examined each box in search of the match.

On the other hand, the majority of *highly intensive* physical activity occurred as children played with the avatar (i.e., *avatar play*). A few instances were also derived from children's interaction with a selected card (i.e., *card play*). Upon card selection, one child quickly swung their arm in a large circular motion, tracing out several circles before attempting to deselect the card at exact location of a box. Celebratory behaviour, such as dancing, waving arms erratically in the air (Fig. 8c), flying (Fig. 8a), and shooting arms upward into a victory pose (see Fig. 8b), also attributed for much of children's high physical intensity.

5.2. Understanding play and problem-solving in MBLT with MMD

EFA, Principal Components Analysis (PCA) with varimax rotation, was conducted in order to identify orthogonally aligned


(a) Child exercises avatar play by making peace signs on both hands as he flaps his arms in a flying motion to celebrate a correct card-box match.


(b) Child strikes a victory pose as he plays with the avatar to celebrate task completion, demonstrating avatar play.


(c) Child erratically waving arms during an episode of avatar play.

Fig. 8. Different play behaviours exhibited by children during sorting tasks.


(a) Child uses both hands to control a card in a playful manner during episode of card play.


(b) Child stands completely still as he makes decisions prior to selecting card.


(c) Child stands still, maintaining posture on a selected card, as she determines a card-box match, during informed problem-solving.

Fig. 9. Children's different play and problem-solving behaviours with the Marvy Learns MBLT.

scales (MMD measurements) and codes (factors) (Mertler & Vannatta, 2005). The EFA scree plots (see Figs. A.1 and A.2 in the appendix) indicate that the MMD measurements can represent the respective codes, based on the scree plot of factors/codes with EigenValues (EV) greater than one. According to Stevens (2012)'s suggestion that constructs the sharp descent of the graph, only the EV more than one, should be retained. Thus, we identified the MMD measurements that can adequately represent each code. In this section, we provide two detailed examples to clarify the exploratory analysis process for the two problem-solving codes: informed, and trial and error. For the remaining codes, we offer a summary of the results (Table 2) from the exploratory analysis. The exploratory analysis figure for each code is in Appendix A.

First, we conducted the PCA step of EFA using the informed problem-solving code as the qualitative variable and the MMD measurements as the quantitative variables. There were four EV greater than one, therefore we used a varimax rotation, with four PCs and rotated the Principal Component (PC) such that the most significant PC aligned with the video code (i.e., informed problem-solving). Next, we used the absolute loadings values of the MMD measurements with respect to the first PC. Finally, we report the MMD measurements and their association with the video code based on their internal reliability. The MMD measurements were selected using a backward elimination method with the objective of keeping the reliability metric greater than, or equal, to 0.7.

In the case of informed problem-solving, we observed that the most positively associated MMD measurements were the two different transitions between AOIs (transitions between cards and wrong boxes, and transition between cards and correct box), cognitive load and IPI. Additionally, the informed problem-solving code is not negatively associated with any of the MMD measurements. We repeated the same process using the trial and error problem-solving code as the qualitative variable. Results revealed one positive association with motion entropy and four negative associated: cognitive load, phasic EDA, and IPI. Lastly, Table 2 presents the remaining strong associations identified by the rotational factor analysis procedure (discussed in Section 4.9). Associations were considered strong provided they had an internal validity exceeding 0.7. This was determined using the same processed described above.

5.3. Synergies between playfulness, problem-solving and MMD measures

While addressing the previous two RQs (i.e., how play and problem-solving behaviour is manifested, and the MMD features explaining those manifestations), we established that the information from the MMD measurements holds significant explanatory power. However, certain aspects of children's interaction with MBLT could not be explained by MMD. This demonstrates complementarity between the two sources of information. RQ3 investigates how we can leverage this complementary information from qualitative behaviour and quantitative MMD to scaffold children's task performance in MBLT. For this purpose we present the performance prediction results using video codes and MMD.

Table 2
Results from the rotational factor analysis to understand the relation between the video coding category and the MMD measurements. **EV** = Number eigenvalues ≥ 1; **AVE** = Average variance extracted. α = Cronbach's alpha. For all codes, the exploratory analysis figures are in Appendix A.

Code	EV (AVE)	Positive associations	Negative associations	α
Trial and error	4 (0.70)	Motion entropy	Cognitive load, IPI, phasic EDA	0.79
Informed	4 (0.73)	Cognitive load, transitions to wrong transition to correct, IPI		0.84
High intensity	3 (0.68)	Motion entropy, phasic EDA	Motion stability	0.73
Moderate intensity	2 (0.62)		Motion jerk, phasic EDA	0.70
Non-intensive	2 (0.63)		Motion entropy, motion jerk	0.71
Card play	4 (0.69)	Motion stability, IPI	Saccade speed, select match diff	0.74
Avatar play	4 (0.70)	Cognitive load, #peaks EDA,	Motion jerk, transitions wrong cards	0.76

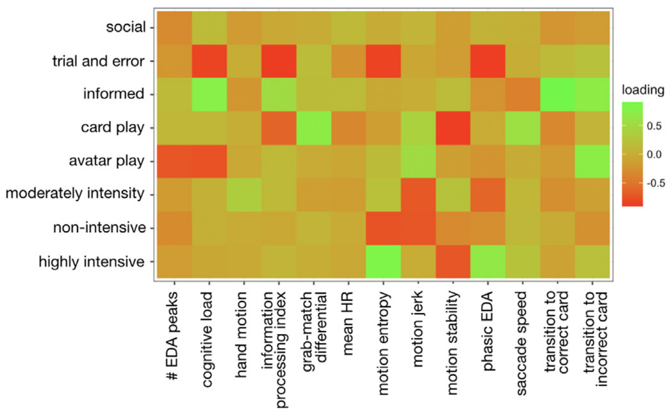


Fig. 10. Overall relation between video codes and the MMD.

For the prediction results, we obtained an NRMSE value of 15.82% on the testing set. This performance out performs the random baseline NRMSE of 41.26%. We also computed the feature importance (normalised between 0 and 100) from the final training model and selected the features with feature importance greater than 75 (the most important quartile). These features are shown in Fig. 11. For informed problem-solving, the most important features were cognitive load (rank 2), IPI (rank 1), and grab match differential (rank 4). The most important features for trial and error were saccade speed (rank 6), mean HR (rank 5), IPI (rank 7) and grab match differential (rank 10). Jerk and phasic EDA were the most important features during the card play. Whereas, the most important feature during avatar play was phasic EDA.

6. Discussion

This study explored elements of children's play and problem-solving behaviour during their interactions with MBLT. In RQ1, we used an inductive category development approach (Mayring, 2000) to devise the SP3 observation scheme, and extract qualitative data on children's behavioural events in the four non-mutually exclusive categories that emerged: namely, play, problem-solving, physical activity, and social interaction. The inductive process provided valuable insights into the visible (i.e. externally observable) processes that children undergo during their interactions with MBLT. For example, we were able to quickly identify patterns of behavioural co-occurrences and dependencies, recognise when social interactions were transpiring, based on observing the direction of their movements, reactions, and facial expressions (discussed in Section 5.1). In sum, addressing RQ1 by qualitatively *looking on to* the data as such (i.e., coders based on SP3) helped us understand the main play and problem-solving behaviour which manifested during children's interactions with MBLT.

In RQ2, we used EFA and PCA to identify the most prominent relationships (negative and positive) between the qualitative video codes and a set of quantitatively computed MMD measurements which are widely used in LA (see Fig. 10). This connected children's cognitive, affective, and physiological processes to the human-observed behavioural events (as annotated in the video codes).

Children's interactions with MBLT are complex and multifaceted events which require them to engage in mental activity while simultaneously moving their bodies into different postures at various speeds. During this time, children demonstrate numerous deliberate behaviours and undergo many involuntary processes, each with varying degrees of visibility to the external observer (i.e., support personnel, researchers). Previous research on children's play interactions in the education and CCI domains, highlight qualitative observation (primarily video analysis) as the gold standard for assessing children's interactions with MBLT. These techniques provide a richness of contextual detail that MMD measures lack. For example, by analysing the video data, we were able to differentiate between occurrences where children made incorrect card-box matches due to lack of conceptual understanding, versus accidentally placing the card in the wrong box. However, the video coding process lacks the capacity to report on the unobservable processes (i.e., cognitive, affective and physiological), which are critical to understanding children's experience and learning with MBLT. Furthermore, expert annotation processes requires extreme resource allocation (i.e., human-hours) and can only be executed only as post processing. Consequently, some researchers (Soute et al., 2013) are calling for new methods and tools of exploration to support their work. On the other hand, MMD from sensors, supplement the shortcomings of video coding with their real-time ability to "see the invisible", but provide details only pertaining to deliberately selected and highly specific aspects of the researcher's interest and are incapable of discerning the contextual and nuanced behaviour detected by human observation. Ergo, the complementarity of these two analysis techniques (video coding and MMD capture) stems from their observational quoins of vantage, and when harnessed in tandem, the bidirectional benefits of their union are undeniably simple: the weaknesses and limitations of one method are supplemented by the strengths of the other, and visa versa. Aligning these perspectives constructs a (more) holistic representation of children's experiences, amplifying researchers ability to increase their understanding of children's learning experience with MBLT.

6.1. MMD based explanations of play and problem-solving behaviour

Our second research question addresses the insights that MMD can provide into children's play and problem-solving behaviour while children engaged with MBLT. Table 2 shows the most prominent associations (positive and negative) between the MMD and the qualitative video codes. Overall, we found a high degree of explainability regarding the problem-solving codes (trial and error, and informed), play codes (card play, avatar play) and physical activity (high, moderate, and non-intensive). We found

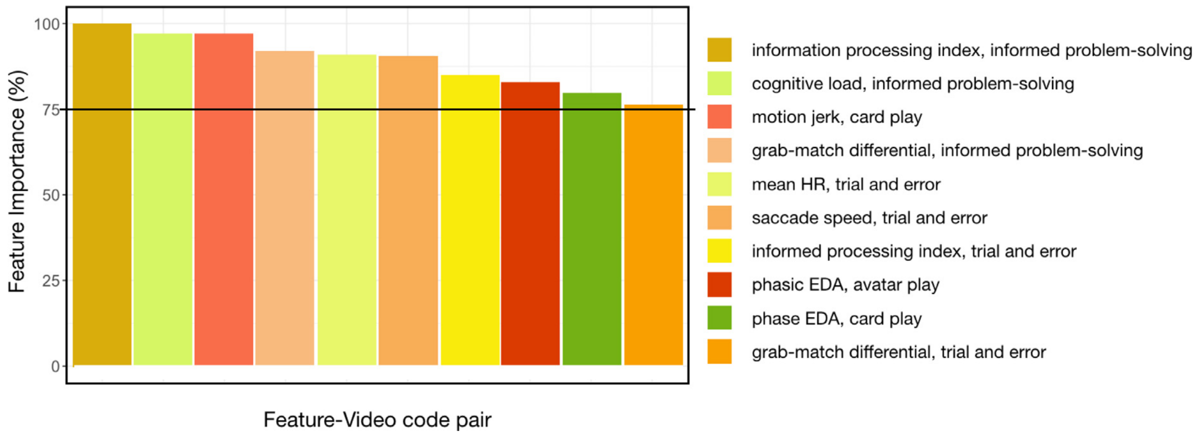


Fig. 11. Variable importance for predicting performance. This will guide the design to improve performance.

no strong associations between the MMD measurements and the social interaction codes.

Regarding the problem-solving codes, cognitive load and IPI were positively associated with informed problem-solving, and negatively associated with trial and error. This indicates that children are mentally invested (e.g., have high cognitive load) while solving problems in an informed manner; conversely, they display opposite behaviour for episodes of trial and error. IPI is also high during informed problem-solving, which demonstrates that children carry out more global than local processing of the on-screen information. This is further corroborated by two additional positive associations with informed problem-solving behaviour, namely transition to wrong, and transition to correct. These transitions might contribute to higher global information processing and, in turn, an increased IPI. On the other hand, during trial and error episodes, the transitions recorded all support hand–eye coordination (e.g., select a card and match it to a box), thus decreasing the IPI. We also found that motion entropy is positively associated with trial and error problem-solving, whereas, phasic EDA level is negatively associated. The motion entropy might be elevated during trial and error episodes due to frequent selection and matching of the cards and boxes. This might also have a negative impact on the engagement with the MBLT as suggested by lower levels of phasic EDA (Levine, Conradt, Goodwin, Sheinkopf, & Lester, 2014).

During episodes of play, card play (which occurred only during the physically actionable stage of problem-solving, Section 5.1), mainly involved hand movements. Since hand joints contribute to less than 10% of the total joint motion (other joints remain relatively stable), this actually attributes to an increase in overall motion stability. Further, once children began performing extreme and exaggerated movements with the card (e.g., quickly tracing out large circle patterns with card in hand), all visual focus was on the card, resulting in smaller saccades and shorter fixations. This attributed to a decreased saccade speed and decreased amount of local information processing (increasing IPI). Finally, since the "match select time difference" timestamp starts as soon as the card is selected, the "card play" duration adds to the time difference between the card selection and card-to-box matching timestamps. On the other hand, during avatar play children might be investing mental effort to understand how their movement patterns are mirrored by the avatar, thus increasing their cognitive load. This engagement also increases their physiological arousal (Lee-Cultura, Sharma, Papavlasopoulou, Retalis, 2020) (increasing peaks in EDA). As well, avatar play consisted primarily of celebratory dance moves (Section 5.1), which lead to increased smoothness in motion (assuming they are fluid dancers!).

We found that the majority of motion-based measurements and the phasic EDA were associated with physical activity. The first set of MMD measurements could be regarded as a direct measurement of the physical activity (entropy — randomness, stability — postural control, and jerk — smoothness). Moreover, phasic EDA captures physiological arousal and could be used as a proxy for engagement with MBLT (Levine et al., 2014). Different physical activity levels encompass different forms of children's physical expression and interaction (Section 5.1). For example, avatar play consists primarily of specific dance moves, which increase motion entropy and decrease motion stability. Similar interactions also depict a high level of engagement with the interface, increasing the phasic EDA levels (Di Lascio et al., 2018). During moderate and non-intensive physical activity, children's motion was mostly smooth and without sudden outbursts of large physical actions. This indicates a lower motion jerk, as motion jerk is observed to be negatively associated with both moderate and non-intensive physical activities. Further, for non-intensive physical activity, the low amount of motion attributed to lower motion entropy. Additionally, physiological arousal may also be lower, as when children are engaged in problem-solving (moderately intensive actions occur mostly during the core problem-solving episodes, see Section 5.1) their momentarily phasic EDA momentarily decreases.

In sum, the results from the exploratory analysis (Section 5.2) show that a significant range of play and problem-solving behaviours could be explained using MMD measurements. However, there are certain aspects of the MBLT interaction that current MMD measurements are unable to explain (e.g., social). We claim that there exists a complementarity of information between the qualitative video coding and the quantitative MMD measurements. The synergistic nature of information from these separate sources can be used to design advanced interaction support systems for augmenting learning outcomes and experiences such as, engagement, learning gain, and emotions. RQ3 presents a case study in which we show how the synergy between video coding and MMD can be used to design support mechanisms for children's task performance.

6.2. Implications to support children's task performance with MBLT

Our final RQ asks how the synergy between MMD and elements of play and problem-solving behaviour can assist us in supporting children's learning through the design and instruction of MBLT. In this section, we discuss how we might nurture a collection of the 10 feature-importance pairs between video codes and MMD for predicting task performance (see Section 5.3), by offering a collection of analysis driven design guidelines.

6.2.1. Cognitive feedback

Cognitive load and informed problem-solving: During informed problem-solving, cognitive load is positively associated with children's task performance. Results from the analysis in RQ2 show that cognitive load was significantly higher when children exhibited informed, relative to trial and error, problem-solving behaviour (Fig. A.8). Cognitive load is not negative, per say, as this mental effort derives from the process of accommodation, which is required to progress learning (Piaget, 2003). However, situations in which children's cognitive load becomes unmanageable (i.e., cognitive overload) may result in children making errors during problem-solving (Harbluk, Noy, Trbovich, & Eizenman, 2007; Liu, Lai, & Chuang, 2011) or negatively impact children's desire to continue with the MBLT (e.g., boredom, disengagement, frustration). Thus, if a sudden steep increase in cognitive load is observed, it is advised to intervene with feedback in the form of a content related support to ease the accommodation process.

A generalisable tactic for helping alleviate cognitive overload, is the removal or reduction of superfluous information caused by unnecessary elements. According to CTML, such elements can hinder a learner's cognitive capacity by introducing unnecessary strain onto children's processing channels, thereby challenging the collective capacity of these channels (i.e., dual channel assumption, and limited capacity assumption). We offer two methods to realise this suggestion. First, the temporary removal of select item cards to reduce the number of options simultaneously presented to the child. When a child's cognitive load skyrockets as they demonstrate informed problem-solving behaviour, the system could remove (i.e., make temporarily invisible) cards which display shapes that differ from the current (if a card is selected), or previous (if no card is selected) match attempt. This would reduce the total information presented to the child at one time, causing a reduction in intrinsic cognitive load, as there would be less novel elements and corresponding connections to consider and store in working memory concurrently. The second method considers synchronising manipulations of the background music (e.g., mute, decrease volume, or adjust tempo) to the child's cognitive load levels. From a CTML perspective, background music acts as redundant information which places additional strain onto the audio-verbal processing channel, thereby attributing extraneous cognitive load. On the other hand, some studies suggest high motivational potential of background music (Linek, Marte, & Albert, 2011), as well as its ability to facilitate an immersion in video games (Fu, 2015; Jennett et al., 2008). Thus, the complete removal of background music may not holistically support a child's learning experience. Therefore, we suggest adjusting the background music to correspond with children's levels of cognitive load. For instance, reducing volume or muting music when load is detected as high, and then restoring the background music once the cognitive load has stabilised to normal levels. Such a feedback mechanism may be capable of appropriately managing cognitive overload, while also preserving the added benefits of background music in multimedia MBLT.

Saccade speed and trial and error: Saccade speed is negatively associated with children's task performance during episodes of trail and error. High saccade speed indicates that children perceive the presented problem as difficult (Bronstein & Kennard, 1987; Smit & Van Gisbergen, 1989). Moreover, this perceived difficulty may be the root cause of children's trial and error behaviour. In Marvy Learns, when an incorrect match was formed, the task responded with a motivational message, "Try Again!". Although such messages may act to encourage continued engagement, they neglect to address the reason for children's incorrect response: the child's inability to integrate the newly introduced task-based information into their schematic library via the process of adaptation. If children are observed in this state, content related support which assists them in understanding the target concept (by stimulating schema adaptation), should be provided to encourage them towards using informed problem-solving behaviours. Thus, instead of merely providing awareness regarding the correctness of their answers, such feedback should scaffold children towards discovering the answers themselves (Fisch, 2005), thereby cultivating a learning opportunity. Rather than simply stating "Try Again!", Marvy Learns could visually highlight the last box that the child tried to match, and then offer critical information purposed to (1) activate the related schemas, or (2) supplement their schematic deficiency, to help the child identify the correct cards. For example, if the last box was labelled *cube*, the task could offer a clue (via narration would employ the auditory-verbal channel and impose less load, according to CTML's Modality Principal Mayer, 2002) with the message "Find all of the unfolded cubes. They are made up of 6 squares". In this way, the child would gain valuable information which they can apply to isolate the (correct) remaining cube shapenet cards. In the same vein, the system can also provide further support via alternations of the options, for instance providing semi-folded cubes, that can enable children in finding the solution of half-based problems. Hints such as these reduce task complexity (and associated intrinsic cognitive load, and schematic complexity required to solve the problem) by minimising the amount of information to be considered by the child. Furthermore, they scaffold the construction of mental representations and schema development (which are tied to germane load).

6.2.2. Behavioural feedback

Information processing index and informed problem-solving: Fig. 11 indicates that during informed problem-solving, IPI is positively associated with children's task performance. In Marvy Learns, IPI manifested from children moving their gaze between the cards widely spread across the top of the screen and the boxes positioned at the bottom of the screen, as they compared characteristics of the card images to the box labels (or previously correctly matched cards within the boxes) in search of a correct match. Specifically, the distance between the question components attributed to long saccades, whereas the frequent switching between question components, lead to short fixations. Essentially, elevated global information processing indicates that children are assessing and comparing the different match options prior to selecting their answer, which is desired problem-solving behaviour. It follows that in matching and multiple choice style questions, where problem elements (e.g., question and answers) are distributed across the screen, global information processing should be encouraged. Thus, if global information processing is detected to be low, aids could be used to guide children's vision back and forth between the question (e.g., a labelled box) and potential answers (e.g., the cards) as a means to invite them to engage in more comparing and contrasting decision making. By CTML's Modality Principal (Mayer, 2002), such aids should be introduced as narration (as opposed to on-screen text), as this will employ children's auditory-verbal channel, thereby distributing the additionally imposed load required to process the hint, across both channels (e.g., auditory-verbal, visual-pictorial), to remain bounded by the upper threshold of each channel (i.e., limited capacity assumption).

6.2.3. Affective feedback

Mean HR and trial and error problem-solving: Mean HR during trial and error is negatively associated with children's task performance. Elevated HR is an indicator that the child is experiencing stress (Herborn et al., 2015). Conversely, an exceptionally low HR may indicate a lack of engagement with the task at hand (Monkaresi, Bosch, Calvo, & D'Mello, 2016). From the video coding, we observed that trial and error behaviour transpired under 2 circumstances: (1) in short bursts to set up a scenario that supports informed learning, and (2) as the child's primary problem-solving behaviour. The former illustrates a strategic use of trial and error behaviour, and thus a low mean HR in this context is of little consequence. However, low mean HR during extended periods of trial and error may indicate that the child is not meaningfully engaged with the sorting task (e.g., making continual mistakes and feeling little stress over the matter) and that they are neglecting to extrapolate and apply the knowledge that becomes available from their successful and unsuccessful attempts. In these cases, intervention is necessary to encourage the child to reflect on the information that follows from their mistakes during the trial and error process, in order to encourage accommodation resulting in the production of schemas that may be used in sorting of subsequent cards (i.e., nudge them to scenario (1) define above). Such intervention may suggest the child compare the incorrectly match card to the remaining cards to identify similarities (or differences) and make future matches based on same or transitive (or unlike) characteristics.

6.2.4. Design considerations

The aforementioned paragraphs presented contextualised *feedback mechanisms* to facilitate the child's experience in real time. However, opportunities exist at the task design level as well, concerning interaction mechanisms and motivated by our results from the cognitive and behavioural processes.

First, we acknowledge that the interaction mechanisms and corresponding affordances offered by Marvy Learns might influence children's propensity towards play behaviour, and facilitate their problem-solving abilities. The grids and shapenets displayed on the cards in Marvy Learns were static, and so interacting with them was limited as children could only move the card on-screen. However, including additional representational and metaphorical gestures that offer greater interaction with the shapes (e.g., gestures which translate to connecting the grid dots into their 2D shape, or fold the shapenets into their 3D shape), may afford possibilities for either play or problem-solving behaviour. For example, by gesturing to fold a shapenet into its 3D representation, a child will experience a more embodied interaction furthering the construction of embodied schemas, which may assist them to better understand, identify, and learn the geometric shape's properties, while facilitating their problem-solving capacity.

Concerning design ideas which cater to children's cognitive processes, we emphasise that by CTML, decomposing the question into smaller sub-problems to be sequentially solved, should incur less cognitive strain on the child as they engage in the task, due to a reduced number of concerns to be simultaneously managed across the child's cognitive channels and within working memory (i.e., limited capacity assumption). Presently, a task requires the application of pattern matching, transitivity, ability to visualise, *and* applying knowledge of geometric characteristics pertaining to each box label's corresponding shape. Orchestrating all of these concerns in tandem, while working towards task completion, may exceed/exhaust the child's cognitive resources. A practical example of breaking the sorting problem down into 2 sub-problems may be as follows. For the first sub-problem, the question would be presented without box labels and the child asked to group, or *stack*, all cards representing the same shape, into piles. This process would require only recognition of patterns and visual similarities, and there would be no need for geometry knowledge at this stage, which makes the solution more accessible to the child. For example, a child does not need to understand the geometric characteristics of a cube, but may still be able to use pattern matching strategies to group all cards with shapes composed of 6 squares into a single pile. To stack cards into a pile, the child would perform a selection gesture to grab a card, drag it overtop of a second card representing the same shape, and then deselect the card to *stack* it on top of the like card. The child would create piles by performing stacking gestures for all cards, until each was associated with the other cards of the same shape genre. Once all cards have been stacked into pile, the box labels appear. The second sub-problem would require the child to gesture an entire stack of cards into a labelled box. This sub-problem requires knowledge of geometric characteristics of each box label shape. As the child's success (via completion time and attempts) at solving the 2-phase problem stabilises, the system may scale the problem difficulty by presenting the integrated problem presentation to the child (i.e., the present version of the question, without stacking sub problem).

Lastly, regarding children's behavioural processes, encouraging children to engage in visual comparisons prior to making match attempts will contribute to an informed (rather than trial and error) problem-solving behaviour. To facilitate increased visual comparisons, we suggest introducing a maximum number of "free" match attempts for each card to be sorted. Once the child exceeds the maximum attempt thresholds for a given card, the card would disappear. Disappearing the card has 2 purposes. First, it simplifies the task by reducing the problem space in which the child is struggling. The child is left with less cards to simultaneously consider, which incurs less strain on their cognitive channels (limited capacity assumption). Second, removing the card acts as a task penalty. Although Marvy Learns does not display the child's score (i.e., number of correct and incorrect match attempts) during task engagement, the score is presented on screen at task completion. Removing a card creates a situation where the child is unable to attain a perfect score (i.e., 6 correct and 0 incorrect matches). In the event that the child does not care about their score, an alternate reason to remove a card may be that this prevents the child from completely solving the problem.

7. Limitations and ethical concerns

The participants' age may impact the results. The children in our study were ages 10–12 years old, representing the cusp on the Concrete and Formal Operational stages in Piaget (2003)'s theory of Cognitive Development. This stage of cognitive maturity is appropriate for our work, as this demographic can engage in abstract thought and perform in/deductive reasoning required by the Marvy Learns questions. Additionally, different age groups may inherently demonstrate greater or lesser propensity towards play (or problem-solving behaviour), producing slightly different results. For example, younger children may naturally play more (i.e., engage in pretend play, which is a defining quality of the Preoperational stage of cognitive development), while older children may naturally play less, despite efforts to preserve the appropriateness of question difficulty.

With sensing technologies becoming more accessible to CCI researchers and practitioners, there is growth in both the number of publications that use sensors to investigate children's behaviour while they interact with specific technologies, and also in the ethical consideration related to data collection, analysis and use (Sharma & Giannakos, 2021). Therefore, it is important to inform the children and support-sphere (parents, teachers, caregivers) about the potential benefits and risks of using such technology. Furthermore, it is also important to implement "ethics-by-design" in such studies (d'Aquin et al., 2018; Dignum et al., 2018; Nelson & Afonso, 2019; Urquhart & Craigon, 2021). Inspired by "privacy-by-design" initiative (Gürses, Troncoso, & Diaz, 2011), ethics-by-design provides certain guidelines that should be considered while using sensing technologies in CCI contexts.

For example, the participants and facilitators should be made aware of issues (d'Aquin et al., 2018) such as, what effects will the research have on them? What are the potential biases and how they are mitigated? How can the results be interpreted and used (or in what ways might they be abused)? Besides informing the children and support-sphere, the researchers should be able to understand how these results would support changes and which of the changed aspects can be measurable (Dignum et al., 2018). Making sure that there is a certain level of common ground between the children, parents, teachers, caregivers, and researchers about aforementioned aspects of technology use, we can strengthen trust from sensing technologies users, compared to the current scenarios (Nelson & Afonso, 2019). With the recent developments of practical tools, such as Moral-IT Deck, the efforts towards the transparent and responsible use of the sensor data can be ensured (Urquhart & Craigon, 2021).

Lastly, as observed in our study, children demonstrate a multitude of behaviours while engaged in MBLT, and at times, these behaviours (e.g., play and problem-solving behaviour) compete for children's resources (e.g., attention). A primary function of learning technologies is to foster the construction of new knowledge and skills regarding specific educational content. Thus, it may seem tempting to integrate design considerations (and feedback mechanisms) which take every available opportunity to encourage children *away* from play and *towards* problem-solving behaviour. However, play serves as an important role in the course of children's cognitive development. According to Piaget, the repeated assimilation involved in ongoing play, reinforces the associated schemas, and increases its accessibility for subsequent learning (Lefrançois, 2000). Design of children's learning experiences should cater to the quality of children's complete experience, rather than focusing only on the outcome of the experience (i.e., whether or not the target educational material was learned).

8. Conclusion

In this paper, we explore the application of MMD to understand children's play and problem-solving behaviour in the context of MBLT. Our work exemplifies how the confluence of MMD and video coding can go further than data triangulation, and contribute to a holistic understanding of children's play and problem-solving behaviours during their interactions with MBLT, by enabling researchers and designers the capacity to cater to children's cognitive, affective and physiological processes to support learning through use of MBLT. To the best of our knowledge, there are no previous studies that use MMD from wearable and ubiquitous sensors (e.g., eye tracking glasses, wristbands and skeletal tracking) to investigate children's behaviours in this context. In particular, this work makes the following contributions, in the context of MBLT:

- We present insights on children's play and problem-solving behaviour, drawn from an in-situ experiment with 10–12 years old children engaged with a MBLT.
- Drawing from previous, well-established protocols and related works, we developed the SP3 observation scheme, that allowed us to code children's social, physical activity, playfulness, and problem-solving behaviour.
- We collected 75 videos of children's interactions with MBLT. Those videos were analysed through content analysis and provide insights on children's play and problem-solving behaviour.
- We employed a wide range of MMD and explored their relationship with the moments of social, physical, play and problem-solving activities.

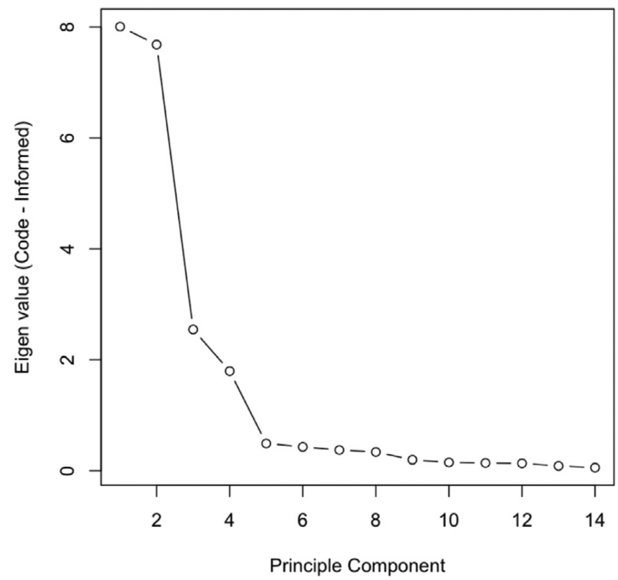


Fig. A.1. Eigenvalues of the different principle components for informed video code.

- Through a mixed-methods approach, we used the synergy (instead of the antagonism, that is oftentimes seen in the literature) between the SP3 moments; and identified implications to support the design and instruction to improve children's learning experiences, specifically, their task performance.

9. Selection and participation

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.

Declaration of competing interest

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

Appendix A. Rotational factor analysis results

A.1. Scree plots for different video codes

See Figs. A.1–A.7.

A.2. PCA variable graphs for the different video codes

See Figs. A.8–A.14.

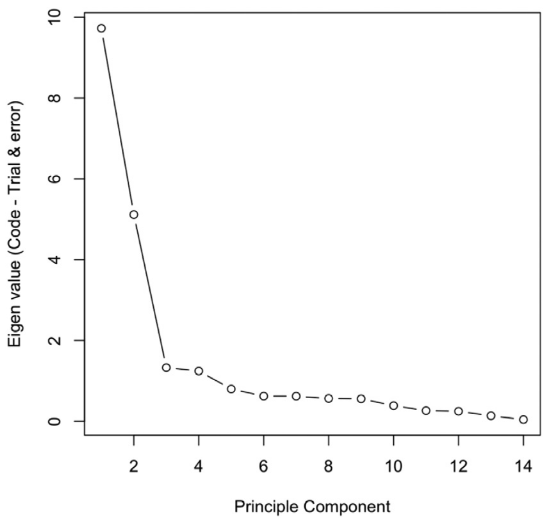


Fig. A.2. Eigenvalues corresponding to the different principle components for trail and error video code.

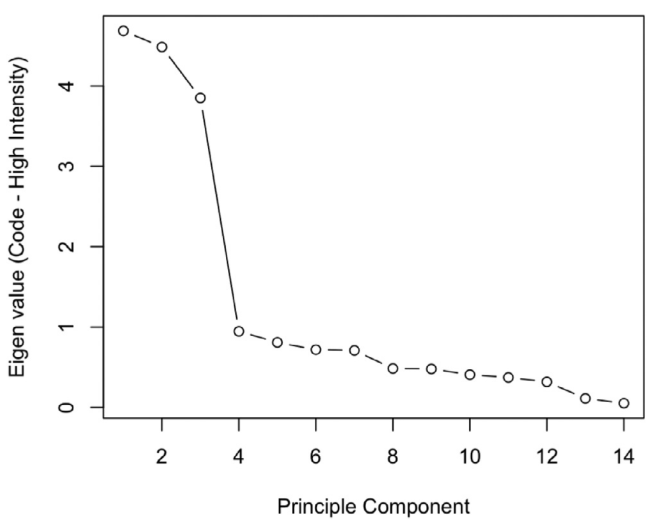


Fig. A.3. Eigenvalues corresponding to the different principle components for high intensity video code.

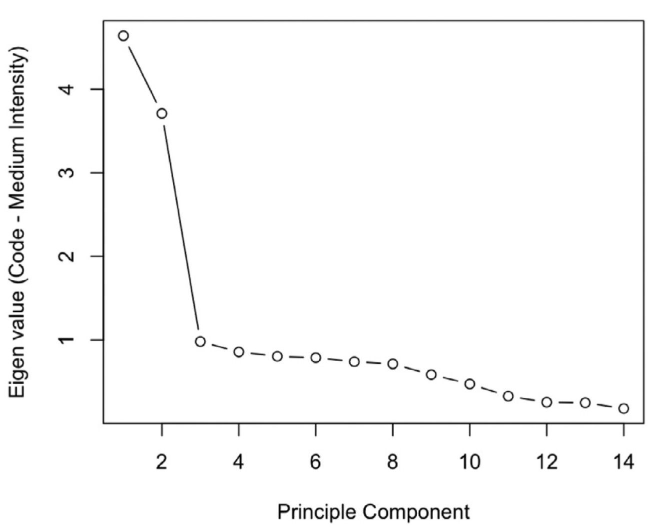


Fig. A.4. Eigenvalues corresponding to the different principle components for medium intensity video code.

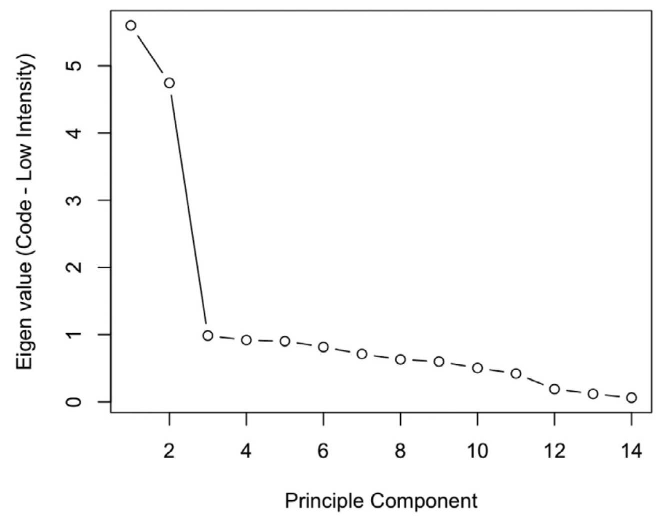


Fig. A.5. Eigenvalues corresponding to the different principle components for low intensity video code.

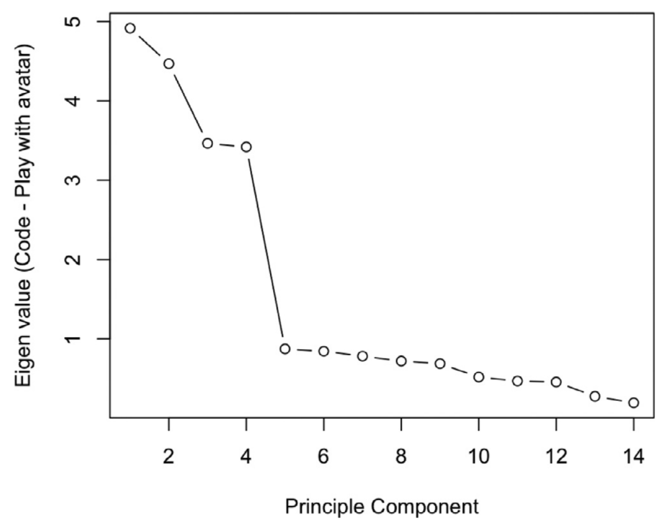


Fig. A.6. Eigenvalues corresponding to the different principle components for play with avatar video code.

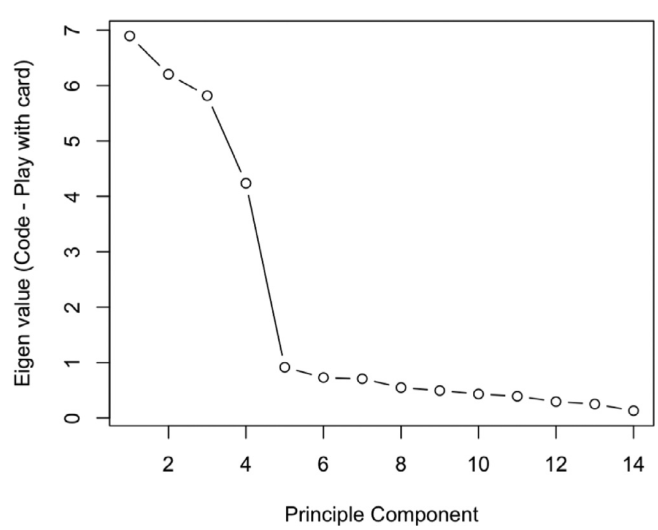


Fig. A.7. Eigenvalues corresponding to the different principle components for play with card video code.

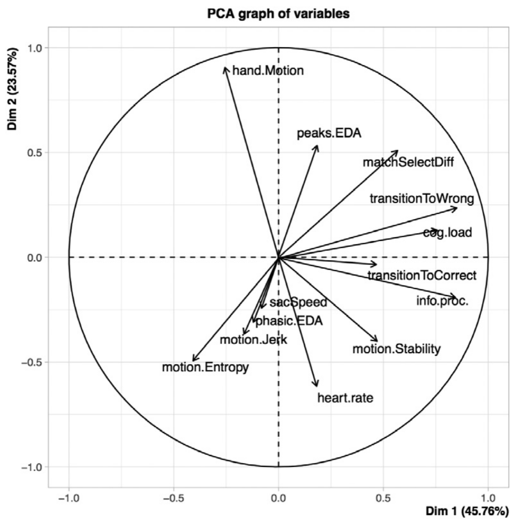


Fig. A.8. Top two principle components, for informed video code. The numbers in the parenthesis represent the explained variance of the data by the rotated PC.

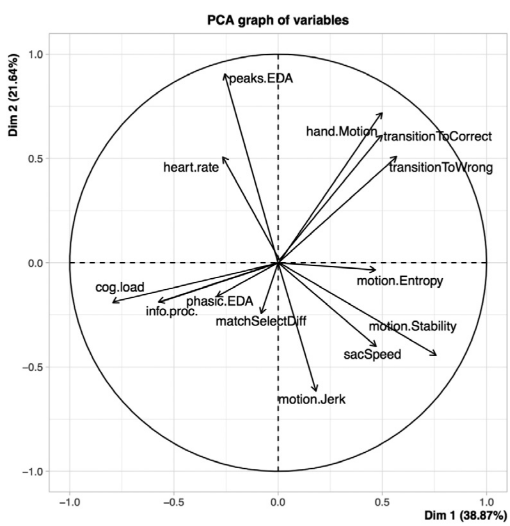


Fig. A.9. Top two principle components for informed video code. The numbers in the parenthesis represent the explained variance of the data by the rotated PC.

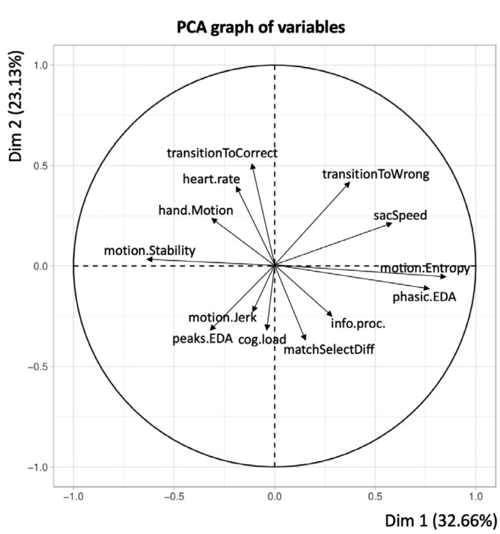


Fig. A.10. Top two principle components for high intensity code. The numbers in the parenthesis represent the explained variance of the data by the rotated PC.

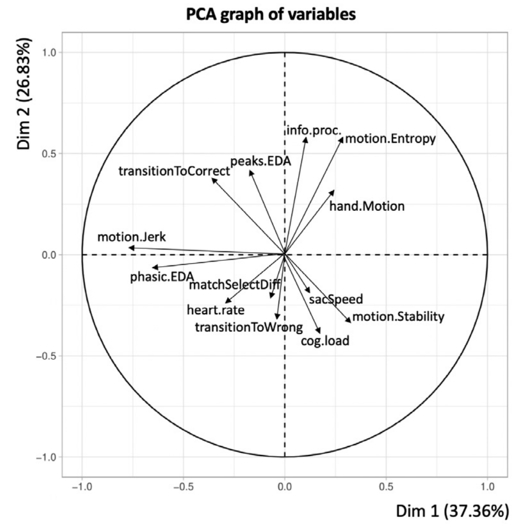


Fig. A.11. Top two principle components for medium intensity code. The numbers in the parenthesis represent the explained variance of the data by the rotated PC.

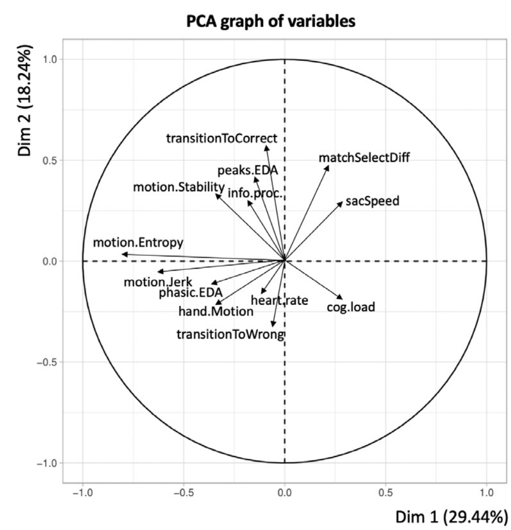


Fig. A.12. Top two principle components for low intensity code. The numbers in the parenthesis represent the explained variance of the data by the rotated PC.

Appendix B. Homography calculation

The main functionality of homography is to find correspondences between the two images of overlapping fields-of-view. For each child, we had the their eye-tracking data in the objective-video-frame (left panel of Fig. B.1. We created a ground-truth of all the screen recordings with the fiducial markers around the frame of the TV screen. These were the two images input for the homography operator. We knew the positions of every fiducial marker in both the ground-truth video and each objective-video-frame. We compute the homography between the objective-video-frame and the ground-truth. The output of the homography calculation is a matrix that encodes the correspondence from the ground-truth to the objective-video-frame. Since we also knew the position of gaze-pointer (where the child is looking at) in the objective-video-frame, we can compute the position of the gaze-pointer on the ground truth by using the inverse of the correspondence matrix (the process is shown in Fig. B.1).

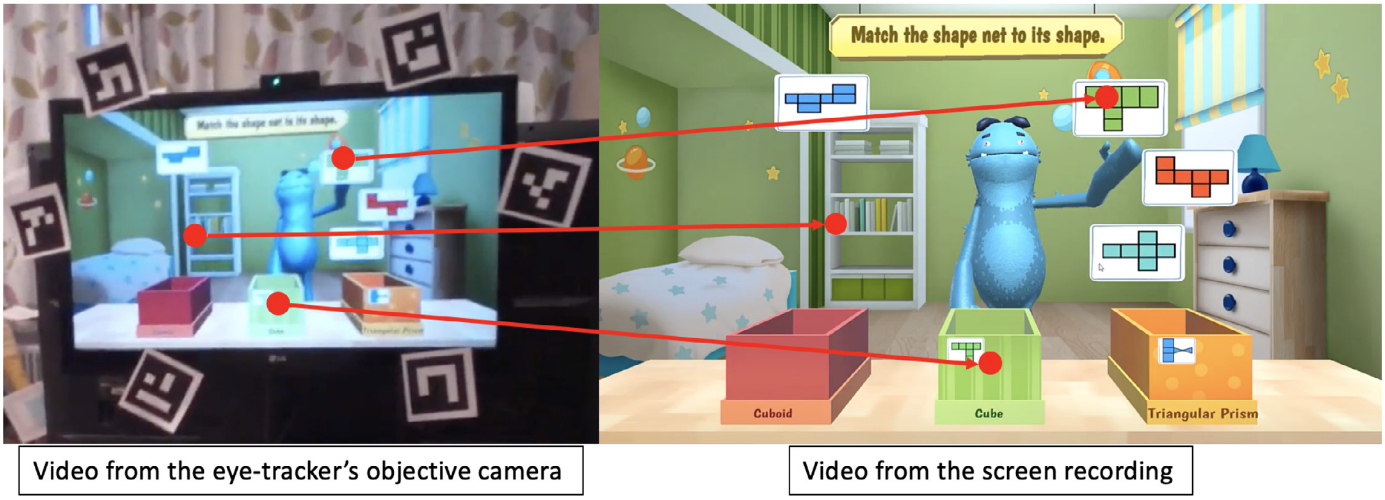


Fig. B.1. Homography example. *left:* video from the eye-tracking glasses' objective camera. *right:* video from the screen recording.

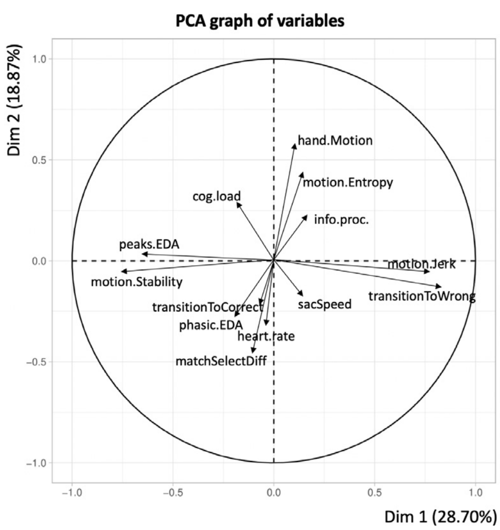


Fig. A.13. Top two principle components for play with avatar code. The numbers in the parenthesis represent the explained variance of the data by the rotated PC.

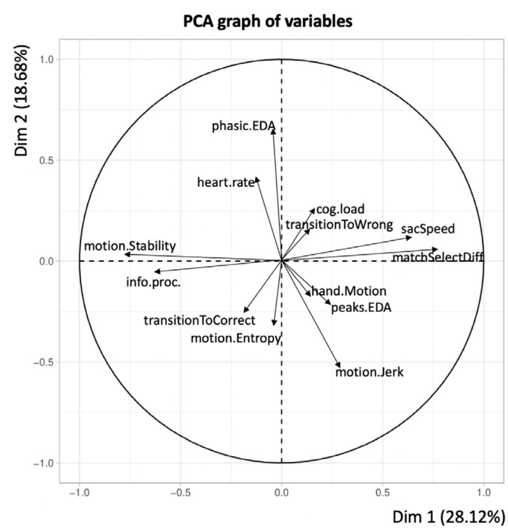


Fig. A.14. Top two principle components for play with card code. The numbers in the parenthesis represent the explained variance of the data by the rotated PC.

Appendix C. Normalised Root Mean Squared Error (NRMSE)

The Root Mean Squared Error (RMSE) is calculated using the following formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{Number\ of\ samples}(predicted_i - original_i)^2}{Number\ of\ samples}} \tag{C.1}$$

Once we have calculated the RMSE, we normalise it to obtain NRMSE using the following formula:

$$NRMSE(\%) = 100 \times \frac{RMSE}{original_{max} - original_{min}} \tag{C.2}$$

References

Abrahamson, D., & Bakker, A. (2016). Making sense of movement in embodied design for mathematics learning. *Cognitive Research: Principles and Implications*, *1*(1), 1–13.

Alepis, E. (2011). Afol: Towards a new intelligent interactive programming language for children. In *Intelligent interactive multimedia systems and services* (pp. 199–208). Springer.

Alibali, M. W., & Nathan, M. J. (2012). Embodiment in mathematics teaching and learning: Evidence from learners' and teachers' gestures. *Journal of the Learning Sciences*, *21*(2), 247–286.

Andrade, A. (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* (pp. 70–79).

Andrade, A., Danish, J. A., & Maltese, A. V. (2017). A measurement model of gestures in an embodied learning environment: Accounting for temporal dependencies. *Journal of Learning Analytics*, *4*(3), 18–46.

Andrade, A., Delandshere, G., & Danish, J. A. (2016). Using multimodal learning analytics to model student behaviour: A systematic analysis of behavioural framing.. *Journal of Learning Analytics*, *3*(2), 282–306.

Angotti, R., & Bayo, I. (2012). Making kinections: Using video game technology to teach math. In *Prato CIRN community informatics conference, Vol. 2012*.

Antle, A. N., Corness, G., & Droumeva, M. (2009). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Interactive Computing*, *21*(1–2), 66–75.

Arroyo, I., Micciollo, M., Casano, J., Ottmar, E., Hulse, T., & Rodrigo, M. M. (2017). Wearable learning: multiplayer embodied games for math. In *Proceedings of the annual symposium on computer-human interaction in play* (pp. 205–216).

Baddeley, A. D., & Logie, R. H. (1999). Working memory: The multiple-component model. In *Models of working memory: mechanisms of active maintenance and executive control* (pp. 28–61). Cambridge University Press.

Bakker, S., Markopoulos, P., & De Kort, Y. (2008). Opos: an observation scheme for evaluating head-up play. In *Proceedings of the 5th nordic conference on human-computer interaction: building bridges* (pp. 33–42).

Bartoli, L., Corradi, C., Garzotto, F., & Valoriani, M. (2013). Exploring motion-based touchless games for autistic children's learning. In *Proceedings of the 12th international conference on interaction design and children* (pp. 102–111). ACM.

Bekker, T., Sturm, J., & Eggen, B. (2010). Designing playful interactions for social interaction and physical play. *Personal and Ubiquitous Computing*, *14*(5), 385–396.

Bhattacharya, A., Gelsomini, M., Pérez-Fuster, P., Abowd, G. D., & Rozga, A. (2015). Designing motion-based activities to engage students with autism in classroom settings. In *Proceedings of the 14th international conference on interaction design and children* (pp. 69–78). ACM.

Black, M. P., Bone, D., Williams, M. E., Gorrindo, P., Levitt, P., & Narayanan, S. (2011). The usc care corpus: Child-psychologist interactions of children with autism spectrum disorders. In *Twelfth Annual conference of the international speech communication association*.

Blikstein, P., & Worsley, M. (2016). Multimodal learning analytics and education data mining: Using computational technologies to measure complex learning tasks. *Journal of Learning Analytics*, *3*(2), 220–238.

Broaders, S. C., Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2007). Making children gesture brings out implicit knowledge and leads to learning.. *Journal of Experimental Psychology: General*, *136*(4), 539.

Bronstein, A., & Kennard, C. (1987). Predictive eye saccades are different from visually triggered saccades. *Vision Research*, *27*(4), 517–520.

Brooks, E., & Sjöberg, J. (2020). A designerly approach as a foundation for school children's computational thinking skills while developing digital games. In *Proceedings of the interaction design and children conference* (pp. 87–95).

Bruce, C. (2008). *Informed learning*. Assoc of Cllge & Rsrch Libr.

Clabaugh, C., Sha, F., Ragusa, G., & Mataric, M. (2015). Towards a personalized model of number concepts learning in preschool children. In *Proceedings of the icra workshop on machine learning for social robotics*, Seattle, WA, USA (pp. 26–30).

Cohen, D. (2007). *The development of play*. Routledge.

Cook, S. W., Mitchell, Z., & Goldin-Meadow, S. (2008). Gesturing makes learning last. *Cognition*, *106*(2), 1047–1058.

Crowell, C. (2018). Analysis of interaction design and evaluation methods in full-body interaction for special needs. In *23rd international conference on intelligent user interfaces* (pp. 673–674). ACM.

Crowell, C., Sayis, B., Benitez, J. P., & Pares, N. (2019). Mixed reality, full-body interactive experience to encourage social initiation for autism: Comparison with a control nondigital intervention. *Cyberpsychology, Behavior, and Social Networking*.

Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics.

d'Aquin, M., Troullinou, P., O'Connor, N. E., Cullen, A., Faller, G., & Holden, L. (2018). Towards an" ethics by design" methodology for ai research projects. In *Proceedings of the 2018 AAAI/ACM conference on ai, ethics, and society* (pp. 54–59).

Di Lascio, E., Gashi, S., & Santini, S. (2018). Unobtrusive assessment of students' emotional engagement during lectures using electrodermal activity sensors. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *2*(3), 103.

Dignum, V., Baldoni, M., Baroglio, C., Caon, M., Chatila, R., Dennis, L., et al. (2018). Ethics by design: Necessity or curse?. In *Proceedings of the 2018 AAAI/ACM conference on ai, ethics, and society* (pp. 60–66).

Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., et al. (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* (pp. 1–13).

Fernández-Delgado, M., Cernadas, E., Barro, S., & Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research*, *15*(1), 3133–3181.

Fink, J., Lemaignan, S., Dillenbourg, P., Rétornaz, P., Vaussard, F., Berthoud, A., et al. (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* (pp. 439–446).

Fisch, S. M. (2005). Making educational computer games" educational". In *Proceedings of the 2005 conference on interaction design and children* (pp. 56–61).

Fotaris, P., Pellas, N., Kazanidis, I., & Smith, P. (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* (pp. 181–191).

Fu, J. Z. X. (2015). The influence of background music of video games on immersion. *Journal of Psychology & Psychotherapy*, *5*(4).

Gaskin, C. J., & Happell, B. (2014). On exploratory factor analysis: A review of recent evidence, an assessment of current practice, and recommendations for future use. *International Journal of Nursing Studies*, *51*(3), 511–521.

Giannakos, M., Papamitsiou, Z., Markopoulos, P., Read, J., & Hourcade, J. P. (2020). Mapping child–computer interaction research through co-word analysis. *International Journal of Child-Computer Interaction*, Article 100165.

Giannakos, M. N., Sharma, K., Papavlasopoulou, S., Pappas, I. O., & Kostakos, V. (2020). Fitbit for learning: Towards capturing the learning experience using wearable sensing. *International Journal of Human-Computer Studies*, *136*, Article 102384.

Giannakos, M. N., Sharma, K., Pappas, I. O., Kostakos, V., & Velloso, E. (2019). Multimodal data as a means to understand the learning experience. *International Journal of Information Management*, *48*, 108–119.

Goldin-Meadow, S. (2005). *Hearing gesture: how our hands help us think*. Harvard University Press.

Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science*, *15*(6), 876–884.

Goldin-Meadow, S., Nusbaum, H., Kelly, S. D., & Wagner, S. (2001). Explaining math: Gesturing lightens the load. *Psychological Science*, *12*(6), 516–522.

Grandhi, S. A., Joue, G., & Mittelberg, I. (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* (pp. 821–824).

Guigon, E., Baraduc, P., & Desmurget, M. (2007). Computational motor control: redundancy and invariance. *Journal of Neurophysiology*, *97*(1), 331–347.

Gürses, S., Troncoso, C., & Diaz, C. (2011). Engineering privacy by design. *Computers, Privacy & Data Protection*, *14*(3), 25.

Hainey, T., Connolly, T. M., Boyle, E. A., Wilson, A., & Razak, A. (2016). A systematic literature review of games-based learning empirical evidence in primary education. *Computers & Education*, *102*, 202–223.

Harada, N. (2002). Cold-stress tests involving finger skin temperature measurement for evaluation of vascular disorders in hand-arm vibration syndrome: review of the literature. *International Archives of Occupational and Environmental Health*, *75*(1–2), 14–19.

Harbluk, J. L., Noy, Y. I., Trbovich, P. L., & Eizenman, M. (2007). An on-road assessment of cognitive distraction: Impacts on drivers' visual behavior and braking performance. *Accident Analysis and Prevention*, *39*(2), 372–379.

Hasson, U., Furman, O., Clark, D., Dudai, Y., & Davachi, L. (2008). Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. *Neuron*, *57*(3), 452–462.

Herborn, K. A., Graves, J. L., Jerem, P., Evans, N. P., Nager, R., McCafferty, D. J., et al. (2015). Skin temperature reveals the intensity of acute stress. *Physiology & Behavior*, *152*, 225–230.

Hornecker, E., & Buur, J. (2006). Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the sigchi conference on human factors in computing systems* (pp. 437–446).

Hourcade, J. P., Zeising, A., Iversen, O. S., Skov, M. B., Antle, A. N., Anthony, L., et al. (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* (p. SIG07). ACM.

Howard, M. C. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, *32*(1), 51–62.

Hsu, H.-m. J. (2011). The potential of kinect in education. *International Journal of Information and Education Technology*, *1*(5), 365.

Hurtienne, J., & Israel, J. H. (2007). Image schemas and their metaphorical extensions: intuitive patterns for tangible interaction. In *Proceedings of the 1st international conference on tangible and embedded interaction* (pp. 127–134).

Isbister, K., Karlesky, M., Frye, J., & Rao, R. (2012). Scoop!: a movement-based math game designed to reduce math anxiety. In *CHI'12 extended abstracts on human factors in computing systems* (pp. 1075–1078). ACM.

Jakobsen, K. B., Stougaard, J., Petersen, M. G., Winge, J., Grønbæk, J. E., & Rasmussen, M. K. (2016). Expressivity in open-ended constructive play: Building and playing musical lego instruments. In *Proceedings of the the 15th international conference on interaction design and children* (pp. 46–57).

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661.

Johnson, K., Pavleas, J., & Chang, J. (2013). Kinecting to mathematics through embodied interactions. *Computer*, *46*(10), 101–104.

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: a study of the use of motion-based games in special education. *TechTrends*, *62*(6), 594–601.

Kourakli, M., Altanis, I., Retalis, S., Boloudakis, M., Zbainos, D., & Antonopoulou, K. (2017). Towards the improvement of the cognitive, motoric and academic skills of students with special educational needs using kinect learning games. *International Journal of Child-Computer Interaction*, *11*, 28–39.

Lai, N. K., Ang, T. F., Por, L. Y., & Liew, C. S. (2018). The impact of play on child development-a literature review. *European Early Childhood Education Research Journal*, *26*(5), 625–643.

Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. University of Chicago press.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 159–174.

Lee-Cultura, S., Sharma, K., Papavlasopoulou, S., & Giannakos, M. (2020). Motion-based educational games: using multi-modal data to predict player's performance. In *2020 IEEE conference on games (cog)* (pp. 17–24). IEEE.

Lee-Cultura, S., Sharma, K., Papavlasopoulou, S., Retalis, S., & Giannakos, M. (2020). Using sensing technologies to explain children's self-representation in motion-based educational games. In *Proceedings of the interaction design and children conference* (pp. 541–555).

Lefrançois, G. R. (2000). *Theories of human learning: what the old man said*. Wadsworth/Thomson Learning.

Leiner, D., Fahr, A., & Früh, H. (2012). Eda positive change: A simple algorithm for electrodermal activity to measure general audience arousal during media exposure. *Communication Methods and Measures*, *6*(4), 237–250.

Levine, T. P., Conradt, E., Goodwin, M. S., Sheinkopf, S. J., & Lester, B. (2014). Psychophysiological arousal to social stress in autism spectrum disorders. *Comprehensive Guide to Autism*, 1177–1193.

Li, J., Van der Spek, E. D., Hu, J., & Feijs, L. (2019). Turning your book into a game: improving motivation through tangible interaction and diegetic feedback in an AR mathematics game for children. In *Proceedings of the annual symposium on computer-human interaction in play* (pp. 73–85).

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. M. (2013). The impact of pretend play on children's development: A review of the evidence.. *Psychological Bulletin*, *139*(1), 1.

Lindgren, R., & Moshell, J. M. (2011). Supporting children's learning with body-based metaphors in a mixed reality environment. In *Proceedings of the 10th international conference on interaction design and children* (pp. 177–180). ACM.

Linek, S. B., Marte, B., & Albert, D. (2011). Background music in educational games: Motivational appeal and cognitive impact. *International Journal of Game-Based Learning (IJGBL)*, *1*(3), 53–64.

Liu, H.-C., Lai, M.-L., & Chuang, H.-H. (2011). Using eye-tracking technology to investigate the redundant effect of multimedia web pages on viewers' cognitive processes. *Computers in Human Behavior*, *27*(6), 2410–2417.

Liu, R., Stamper, J., Davenport, J., Crossley, S., McNamara, D., Nzinga, K., et al. (2019). Learning linkages: Integrating data streams of multiple modalities and timescales. *Journal of Computer Assisted Learning*, *35*(1), 99–109.

Loh, C. S. (2012). Information trails: In-process assessment of game-based learning. In *Assessment in game-based learning* (pp. 123–144). Springer.

Markopoulos, P., Read, J., & Giannakos, M. (2021). Design of digital technologies for children. In *Handbook of human factors and ergonomics* (5th ed.).

Mayer, R. E. (2002). Cognitive theory and the design of multimedia instruction: an example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning*, *2002*(89), 55–71.

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research [on-Line Journal]*, *1*, http://qualitative-research.net/fqs/fqs-e/2-00inhalt-e.htm.

McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. University of Chicago press.

Mertler, C., & Vannatta, R. (2005). Multiple regression. *Advanced and Multivariate Statistical Methods*, 165–198.

Monkaresi, H., Bosch, N., Calvo, R. A., & D'Mello, S. K. (2016). Automated detection of engagement using video-based estimation of facial expressions and heart rate. *IEEE Transactions on Affective Computing*, *8*(1), 15–28.

Moreno-Marcos, P. M., Alario-Hoyos, C., Muñoz-Merino, P. J., & Kloos, C. D. (2018). Prediction in moocs: A review and future research directions. *IEEE Transactions on Learning Technologies*.

Nelson, K. L., & Afonso, W. B. (2019). Ethics by design: The impact of form of government on municipal corruption. *Public Administration Review*, *79*(4), 591–600.

Nielsen, M., Störring, M., Moeslund, T. B., & Granum, E. (2003). A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In *International gesture workshop* (pp. 409–420). Springer.

Ofer, N., Erel, H., David, I., Hitron, T., & Zuckerman, O. (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* (pp. 599–604).

Ohnishi, A., Saito, K., Terada, T., & Tsukamoto, M. (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* (pp. 353–363). Springer.

Olsen, A. (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.

Orona, H. M. O., Maldonado, G. S., & Martínez, N. P. S. (2015). Kinect team: Kinesthetic learning applied to mathematics using kinect. *Procedia Computer Science*, *75*, 169–172.

Papavlasopoulou, S., Sharma, K., Giannakos, M., & Jaccheri, L. (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* (pp. 171–181). ACM.

Pelánek, R. (2015). Metrics for evaluation of student models. *Journal of Educational Data Mining*, *7*(2), 1–19.

Piaget, J. (2003). Part i: Cognitive development in children–piaget development and learning.. *Journal of Research in Science Teaching*, *40*.

Piaget, J. (2013a). *Child's conception of number: selected works Vol. 2*. Routledge, volume 2.

Piaget, J. (2013b). *Play, dreams and imitation in childhood, Vol. 25*. Routledge.

Piaget, J., & Cook, M. T. (1952). The origins of intelligence in children.

Plass, J. L., Moreno, R., & Brünken, R. (2010). Cognitive load theory.

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. In *Encyclopedia of human computer interaction* (pp. 211–219). IGI Global.

Pramling Samuelsson, I., & Johansson, E. (2006). Play and learning—inseparable dimensions in preschool practice. *Early Child Development and Care*, *176*(1), 47–65.

Radu, I. (2014). Augmented reality in education: a meta-review and cross-media analysis. *Personal and Ubiquitous Computing*, *18*(6), 1533–1543.

Retalis, S., Korpa, T., Skaloumpakas, C., Boloudakis, M., Kourakli, M., Altanis, I., et al. (2014). Empowering children with ADHD learning disabilities with the kinems kinect learning games. In *European conference on games based learning, Vol. 2* (p. 469). Academic Conferences International Limited.

Rieber, L. P. (1996). Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games. *Educational Technology Research and Development*, *44*(2), 43–58.

Rosenbaum, L. F., Kaur, J., & Abrahamson, D. (2020). Shaping perception: Designing for participatory facilitation of collaborative geometry. *Digital Experiences in Mathematics Education*, *6*(2), 191–212.

Rubens, C., Braley, S., Torpegaard, J., Lind, N., Vertegaal, R., & Merritt, T. (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* (pp. 193–205).

Schmidt, M., Tawfik, A. A., Jahnke, I., & Earnshaw, Y. (2020). Learner and user experience research.

Sharma, K., & Giannakos, M. (2020). Multimodal data capabilities for learning: What can multimodal data tell us about learning? *British Journal of Educational Technology*, *51*(5), 1450–1484.

Sharma, K., & Giannakos, M. (2021). Sensing technologies and child-computer interaction: Opportunities, challenges and ethical considerations. *International Journal of Child-Computer Interaction*, Article 100331.

Sharma, K., Papamitsiou, Z., & Giannakos, M. (2019). Building pipelines for educational data using AI and multimodal analytics: A "grey-box" approach. *British Journal of Educational Technology*.

Sharma, K., Papavlasopoulou, S., & Giannakos, M. (2019). Joint emotional state of children and perceived collaborative experience in coding activities. In *Proceedings of the 18th ACM international conference on interaction design and children* (pp. 133–145). ACM.

Skulmowski, A., Pradel, S., Kühnert, T., Brunnett, G., & Rey, G. D. (2016). Embodied learning using a tangible user interface: The effects of haptic perception and selective pointing on a spatial learning task. *Computers & Education*, *92*, 64–75.

Smit, A., & Van Gisbergen, J. (1989). A short-latency transition in saccade dynamics during square-wave tracking and its significance for the differentiation of visually-guided and predictive saccades. *Experimental Brain Research*, *76*(1), 64–74.

Smith, C. P., King, B., & Hoyte, J. (2014). Learning angles through movement: Critical actions for developing understanding in an embodied activity. *The Journal of Mathematical Behavior*, *36*, 95–108.

Soute, I., Bakker, S., Magielse, R., & Markopoulos, P. (2013). Evaluating player experience for children's outdoor pervasive games. *Entertainment Computing*, *4*(1), 25–38.

Sridhar, P. K., Chan, S. W., & Nanayakkara, S. (2018). Going beyond performance scores: understanding cognitive-affective states in kindergarteners. In *Proceedings of the 17th acm conference on interaction design and children* (pp. 253–265). ACM.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. Routledge.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*(2), 257–285.

Sweller, J. (1989). Cognitive technology: Some procedures for facilitating learning and problem solving in mathematics and science. *Journal of Educational Psychology*, *81*(4), 457.

Thakkar, V., Shah, A., Thakkar, M., Joshi, A., & Mendjoge, N. (2012). Learning math using gesture. In *International conference on education and e-learning innovations* (pp. 1–3). IEEE.

Thorson, K. R., West, T. V., & Mendes, W. B. (2018). Measuring physiological influence in dyads: A guide to designing, implementing, and analyzing dyadic physiological studies.. *Psychological Methods*, *23*(4), 595.

Tisza, G., Gollerizo, A., & Markopoulos, P. (2019). Measuring fun with adolescents: introducing the spanish and dutch adaptation of the funq. In *Extended abstracts of the annual symposium on computer-human interaction in play companion extended abstracts* (pp. 715–722).

Tsai, C.-H., Kuo, Y.-H., Chu, K.-C., & Yen, J.-C. (2015). Development and evaluation of game-based learning system using the microsoft kinect sensor. *International Journal of Distributed Sensor Networks*, *11*(7), Article 498560.
Urquhart, L. D., & Craigon, P. J. (2021). The moral-it deck: a tool for ethics by design. *Journal of Responsible Innovation*, 1–33.
Villaroman, N., Rowe, D., & Swan, B. (2011). Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor. In *Proceedings of the 2011 Conference on Information Technology Education* (pp. 227–232).
Wilkinson, P., Taylor, J., & Readman, M. (2018). Mediating family play: Exploring the expectations of digital media through a mobile application designed to facilitate real-world child–parent play. *International Journal of Child-Computer Interaction*, *18*, 90–99.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin & Review*, *9*(4), 625–636.
Worsley, M., Abrahamson, D., Blikstein, P., Grover, S., Schneider, B., & Tissenbaum, M. (2016). Situating multimodal learning analytics. In *Proceedings of international conference of the learning sciences, ICLS, Vol. 2* (pp. 1346–1349).
Worsley, M., & Blikstein, P. (2014). Deciphering the practices and affordances of different reasoning strategies through multimodal learning analytics. In *Proceedings of the 2014 ACM workshop on multimodal learning analytics workshop and grand challenge* (pp. 21–27).
Worsley, M., & Blikstein, P. (2015). Leveraging multimodal learning analytics to differentiate student learning strategies. In *Proceedings of the fifth international conference on learning analytics and knowledge* (pp. 360–367). ACM.
Worsley, M., & Blikstein, P. (2018). A multimodal analysis of making. *International Journal of Artificial Intelligence in Education*, *28*(3), 385–419.
Yap, K., Zheng, C., Tay, A., Yen, C.-C., & Do, E. Y.-L. (2015). Word out!: learning the alphabet through full body interactions. In *Proceedings of the 6th augmented human international conference* (pp. 101–108). ACM.
Yildirim, S., & Narayanan, S. (2009). Recognizing child's emotional state in problem-solving child-machine interactions. In *Proceedings of the 2nd workshop on child, computer and interaction* (p. 14). ACM.
Yu, J., Zheng, C., Tamashiro, M. A., Gonzalez-millan, C., & Roque, R. (2020). CodeAttach: Engaging children in computational thinking through physical play activities. In *Proceedings of the fourteenth international conference on tangible, embedded, and embodied interaction* (pp. 453–459).