

Generative AI and multimodal data for educational feedback: Insights from embodied math learning

Giulia Cosentino¹  | Jacqueline Anton^{2,3} | Kshitij Sharma¹  |
Mirko Gelsomini^{1,4} | Michail Giannakos¹ | Dor Abrahamson^{2,5} 

¹Department of Computer Science,
Norwegian University of Science and
Technology, Trondheim, Norway

²Berkeley School of Education, University
of California Berkeley, Berkeley, California,
USA

³Department of Special Education, San
Francisco State University, San Francisco,
California, USA

⁴Department of Innovative Technologies,
Scuola Universitaria Professionale della
Svizzera Italiana, Lugano, Switzerland

⁵College of Education, Korea University,
Seoul, South Korea

Correspondence

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

Abstract: This study explores the role of generative AI (GenAI) in providing formative feedback in children's digital learning experiences, specifically in the context of mathematics education. Using multimodal data, the research compares AI-generated feedback with feedback from human instructors, focusing on its impact on children's learning outcomes. Children engaged with a digital body-scale number line to learn addition and subtraction of positive and negative integers through embodied interaction. The study followed a between-group design, with one group receiving feedback from a human instructor and the other from GenAI. Eye-tracking data and system logs were used to evaluate student's information processing behaviour and cognitive load. The results revealed that while task-based performance did not differ significantly between conditions, the GenAI feedback condition demonstrated lower cognitive load and students show different visual information processing strategies among the two conditions. The findings provide empirical support for the potential of GenAI to complement traditional teaching by providing structured and adaptive feedback that supports efficient learning. The study underscores the importance of hybrid intelligence approaches that integrate human and AI feedback to enhance learning through synergistic feedback. This research offers valuable insights for educators, developers and researchers aiming to design hybrid AI-human educational environments that promote effective learning outcomes.

KEYWORDS

embodied learning, formative feedback, generative AI, hybrid intelligence, teacher-AI collaboration

Practitioner notes

What is already known about this topic?

- Embodied learning approaches have been shown to facilitate deeper cognitive processing by engaging students physically with learning materials, which is especially beneficial in abstract subjects like mathematics.
- GenAI has the potential to enhance educational experiences through personalized feedback, making it crucial for fostering student understanding and engagement.
- Previous research indicates that hybrid intelligence that combines AI with human instructors can contribute to improved educational outcomes.

What this paper adds?

- This study empirically examines the effectiveness of GenAI-generated feedback when compared to human instructor feedback in the context of a multisensory environment (MSE) for math learning.
- Findings from system logs and eye-tracking analysis reveal that GenAI feedback can support learning effectively, particularly in helping students manage their cognitive load.
- The research uncovers that GenAI and teacher feedback lead to different information processing strategies. These findings provide actionable insights into how feedback modality influences cognitive engagement.

Implications for practice and/or policy

- The integration of GenAI into educational settings presents an opportunity to enhance traditional teaching methods, enabling an adaptive learning environment that leverages the strengths of both AI and human feedback.
- Future educational practices should explore hybrid models that incorporate both AI and human feedback to create inclusive and effective learning experiences, adapting to the diverse needs of learners.
- Policymakers should establish guidelines and frameworks to facilitate the ethical and equitable adoption of GenAI technologies for learning. This includes addressing issues of trust, transparency and accessibility to ensure that GenAI systems are effectively supporting, rather than replacing, human instructors.

INTRODUCTION

Embodied learning, based on the theory of embodied cognition, highlights the intrinsic link between cognitive processes and physical interactions with the environment (Shapiro, 2019; Sheets-Johnstone, 2011). This approach has proven particularly effective in education, where physically engaging with learning contents can deepen conceptual understanding, especially in abstract subjects such as mathematics (Bofferding & Hoffman, 2014; Varma & Schwartz, 2011). Multisensory environments (MSEs), which integrate visual, auditory

and kinesthetic stimuli, further amplify the benefits of embodied learning by providing immersive and interactive learning experiences to students (Cosentino et al., 2023). Previous research has demonstrated that incorporating multisensory environments into educational contexts can significantly enhance engagement and learning outcomes (Moyer-Packenham et al., 2015; Mutlu-Bayraktar et al., 2022). However, while researchers have started to investigate the pedagogical value of embodied learning and multisensory environments, the integration of advanced technologies like Generative AI (GenAI) within these frameworks remains underexplored. GenAI, with its promise to accurately detect student challenges and potential misconceptions, as well as use this information to generate appropriate feedback (Suraworachet et al., 2024), holds the potential to further enrich multisensory educational environments (Akata et al., 2020). GenAI can equip systems with personalized feedback capabilities addressing learners' cognitive and attentional needs (Ke et al., 2024). Such systems not only augment traditional instructional methods but also open new opportunities for addressing challenges like cognitive overload and attention management in real time (Holstein et al., 2020; Roschelle et al., 2020).

This study investigates the impact of GenAI-produced formative feedback in the context of learning mathematics with an MSE. The study focuses on learning positive and negative integer operations, where children interacted with a digital number line (NL) system through embodied interaction. To empirically test the effects of GenAI feedback, this study utilized a between-group design. In the control group, children received feedback from a human instructor, while in the experimental group, feedback was generated by a large language model (LLM) based on the students' movement data. This set-up allowed for a comparison between human and AI-generated feedback. Insights were collected using eye-tracking technology to capture students' areas of interest (AOI) and their transitions, alongside motion data system logs, providing comprehensive insights into the students' learning.

Specifically, this research addresses the following questions:

- RQ1: How does feedback from GenAI compare to that from human instructors in terms of learning outcomes and cognitive engagement in embodied math education?
- RQ2: How did students engage with the feedback from GenAI compared to human instructor, and what do these patterns reveal about challenges and opportunities of integrating AI-generated feedback within MSEs?

Previous studies have shown that hybrid intelligence (HI), where AI supports human instructors, holds promise for improving educational outcomes Roschelle et al. (2020), but the impact of AI-driven formative feedback in multisensory settings remains underexplored. This study provides empirical evidence on the benefits and limitations of using GenAI to deliver feedback, offering insights that can inform future designs of AI systems aimed at enhancing learning experiences. By examining the effectiveness of GenAI in providing formative feedback, this research contributes to the broader discourse on GenAI's role in education, offering valuable implications for educators, researchers and developers in designing and using AI-enhanced learning environments.

RELATED WORK AND BACKGROUND THEORIES

Embodied learning and generative AI in education

Embodied cognition theory emphasizes that intelligence and human reasoning evolve from bodily experience (Abrahamson & Bakker, 2016; Dourish, 2001; Piaget, 1970). Studies that relate to the embodied cognition theory have demonstrated that physical actions, including

gestures, movements and spatial manipulation, play a pivotal role in facilitating learning. These activities provide students with bodily experiences, bridging the gap between abstract concepts and practical understanding (Giannakos et al., 2020; Moyer-Packenham et al., 2015; Riconscente, 2013). Previous studies highlighted that embodied learning approaches can significantly improve students' grasp of mathematical concepts, as they visualize and physically manipulate these ideas (Barrocas et al., 2023; Moyer-Packenham et al., 2015; Riconscente, 2013). Technological advancements like MSEs and interactive interfaces have opened up novel opportunities for incorporating embodied learning into educational practices (Giannakos & Cukurova, 2023). Tools like multi-touch applications, motion-tracking systems and projection-based technologies provide children with dynamic and immersive ways to interact with and explore educational content (Chen, 2022; Cosentino & Giannakos, 2023; Sinclair & Heyd-Metzuyanim, 2014).

The integration of embodied learning strategies with GenAI technologies presents a unique opportunity to enrich educational experiences further. GenAI systems can analyse students' physical interactions with the learning contents, allowing for real-time adaptation and personalized feedback based on observed behaviours. This synergy fosters dynamic learning environments where AI responds to students' physical cues, enhancing engagement and understanding (Druga et al., 2019; Yang et al., 2024). By tailoring feedback to individual movements, these systems can bridge the gap between physical and cognitive learning, promoting a more comprehensive educational experience.

GenAI's capacity to process multimodal data complements embodied learning by facilitating personalized educational experiences. These technologies can integrate several inputs to create a rich feedback mechanism aligned with diverse learning contexts and pedagogical needs (Huang et al., 2019; Järvelä et al., 2022). The combined approach of embodied learning and GenAI supports the creation of interactive educational tools that respond to both cognitive and physical aspects of learning (Yang, 2022). As education continues to evolve, AI plays a transformative role through processes like co-teaching (Holstein et al., 2020) and co-learning (Huang et al., 2019), making it essential for human instruction and AI systems to complement and intertwine with each other. Human–AI HI approaches have demonstrated considerable promise in advancing educational theories and practices (Cosentino et al., 2024; Järvelä et al., 2023; Roschelle et al., 2020). For instance, Holstein et al. (2020) proposed a conceptual framework that highlights various dimensions of human–AI hybrid adaptivity, detailing how humans and AI can mutually augment their capabilities. Despite these advancements, further research is needed to deepen our understanding of HI's role in education and to refine its application to effectively support learning with advanced technologies such as MSEs.

Eye tracking for information processing

Eye tracking is increasingly used in digital environments to monitor how learners interact with different types of online content, such as videos, simulations and interactive tools. In embodied learning scenarios, where children move and interact with several digital elements, analysing eye movements such as fixations and saccades allows researchers to interpret cognitive load (CL) (Abrahamson et al., 2016; Lee-Cultura et al., 2022). CL refers to the level of mental effort involved in solving a given task. It is captured using pupil diameter (Duchowski et al., 2018) and is related to performance and different phases of learning (Paas & Van Merriënboer, 1994; Yang et al., 2018). Longer fixation durations typically indicate deeper cognitive processing, while rapid saccades suggest information searching, especially when users are trying to integrate multiple information sources, like text and visuals, in educational platforms (Ke et al., 2024; Van Gog & Jarodzka, 2013).

Eye-tracking metrics such as fixation duration and saccadic patterns provide valuable insights into cognitive engagement by revealing how learners allocate their attention and process complex information across varied digital interfaces. Research indicates that deeper cognitive engagement, often marked by longer fixations and fewer saccades, correlates with better learning outcomes, as it reflects sustained attention and effortful processing of educational materials (Ke et al., 2024). Cognitive engagement, which refers to the extent of learners' active participation in processing information, is closely linked to learning outcomes, as it shapes how effectively individuals integrate and apply new knowledge (Chi & Wylie, 2014). Conversely, disengagement can be identified through erratic eye movements or reduced focus on key informational elements, highlighting areas where instructional design might fail to capture or maintain learners' attention. Understanding these dynamics enables the creation of digital learning environments that foster meaningful engagement and improve outcomes by aligning cognitive demands with learners' processing capabilities (?). These insights enable digital technologies to adapt and customize learning experiences by analysing where learners direct their attention and the duration of their engagement with specific interface elements.

In the context of eye-tracking research, the Information Processing Index (IPI)¹ is a metric used to assess the balance between global and local processing strategies during visual tasks (Poole & Ball, 2006). In this way, information processing refers to the mechanisms by which humans perceive, interpret, store and retrieve information. Global processing is characterized by a pattern of short fixations and long saccades, indicating rapid scanning of a scene to grasp the overall structure or locate specific items. Conversely, local processing involves longer fixations and shorter saccades, suggesting detailed examination of specific areas or objects (Lee-Cultura et al., 2022). These eye movement patterns provide insights into cognitive processes and attentional focus (Ting & Gluth, 2024). Eye-tracking studies reveal that high-performing learners tend to adopt more efficient information processing strategies, such as focusing on key areas of a digital interface, while lower performing learners often exhibit scattered fixation patterns, suggesting less organized cognitive processing (Ke et al., 2024). This distinction helps educators design more effective digital tools by optimizing how information is presented. Moreover, eye-tracking data have been crucial in addressing the split-attention effect, a cognitive load theory phenomenon where learners divide their focus between different types of content, often leading to inefficiency. Research has demonstrated that poorly designed digital interfaces exacerbate this issue, forcing learners to switch frequently between different sources, such as textual descriptions and corresponding images, resulting in higher cognitive load (Deng & Gao, 2023). Eye-tracking data reveal how learners allocate their visual attention across digital interfaces, helping identify where overload occurs and how to reduce it by integrating or simplifying the presentation of contents. Studies confirm that reducing unnecessary cognitive load enhances learning outcomes, especially when digital content is presented in formats that minimize the need for constant attention switching (Xue et al., 2024).

Immersive technologies such as MSEs hold significant potential for keeping users' cognitive load at levels where they can effectively process the information provided to them by presenting information in a more integrated and engaging manner (Lee-Cultura et al., 2022). However, managing the rich range of immersive elements within these environments can also pose challenges for learners, as they may struggle to focus effectively when faced with multiple sensory stimuli. While researchers have started to investigate the effectiveness of immersive environments such as MSEs, augmented reality (AR) and virtual reality (VR) (Andrade, 2017; Moreno-Arjonilla et al., 2024), the introduction of GenAI adds another layer of complexity calling for research that considers their interplay. GenAI systems often rely on digital cues, such as textual, auditory or visual feedback, which could further split learners' attention. This makes it crucial to optimize the presentation of information and reduce the

potential for cognitive overload. Eye tracking can play a crucial role in understanding and mitigating these issues by providing insights into how learners interact with AI-driven cues and digital elements. These data can inform the design of AI systems to deliver feedback that aligns with learners' focus, fostering a seamless integration of GenAI and MSEs to create learner-centred educational experiences.

METHODS

The design

In this study, we developed a technology referred to as MOVES. This system is a portable MSE-enabling platform designed to address the hardwired limitations of conventional environments (Cosentino et al., 2024).

In the MOVES educational design, students participate in an activity that involves walking along a body-scale number line (NL) to solve integer arithmetic problems. The NL serves as an effective pedagogical tool for teaching integer arithmetic, as it represents integers in an ordinal and spatially organized manner, with negative integers serving as reflections of their positive counterparts (Varma & Schwartz, 2011).

The basic instructions for how to solve integer arithmetic problems by walking on the NL are (see Figure 1):

- a. start by standing on the first number in the problem;
- b. turn to the right (positive side of the NL) for addition problems, or turn to the left (negative side of the NL) for subtraction; and

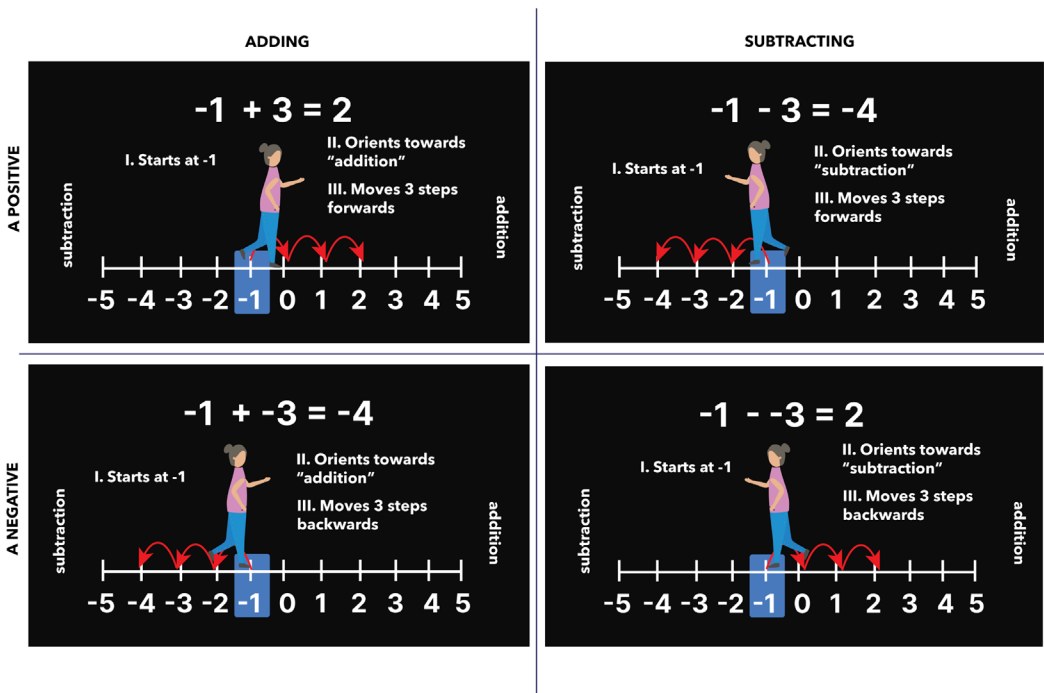


FIGURE 1 Representation on how a student walks through the NL under the four different possible combination schemes of adding or subtracting positive or negative integers.

- c. walk the number of steps indicated by the second number in the problem (forwards if the number is positive, backward if the number is negative).

Figure 1 depicts a child's movements for the four potential combination schemes of adding or subtracting positive or negative integers on the Walking NL, as reflected by the three-move instructions above.

The SENSEi software (Gelsomini, 2023) enables the operation of dual projectors, one displaying on the wall and the other on the floor, to track and monitor students' movements, positions and orientations as they interact with the body-scale NL.

For instance (see Figure 2), to solve the equation $-1 + 2$, the student begins by standing on the -1 hash mark on the NL. As they step onto -1 , the number beneath their feet turns blue, while the -1 displayed on the wall in front changes to green, accompanied by a sound. Next, the student turns to the right to face the addition direction. Upon turning, the addition sign on the wall also turns green, and another sound plays. Finally, the student takes two steps forward and raises their hands above their head to indicate that they have reached the solution. If correct, the entire problem, along with the solution on the wall, turns green, and a congratulatory sound plays. If incorrect, the solution remains unchanged, and no sound plays, as negative feedback is avoided to prevent demotivating or discouraging students (Van Duijvenvoorde et al., 2008).

GenAI integration

For this study, we introduced the integration of the GPT-4 language model into a Node.js web server, aimed at enhancing interactive educational systems that rely on body movement and dynamic user input. Upon receiving feedback on the student's actions, the system communicates with OpenAI's GPT-4 API to generate personalized hints or suggestions, creating an adaptive learning experience. The core system is built on a Node.js server, responsible for handling real-time interactions between the students and the GPT-4 model via OpenAI's API. As students engage with the physical number line, their responses, whether correct or incorrect, are detected and processed by the system's sensors. These responses are

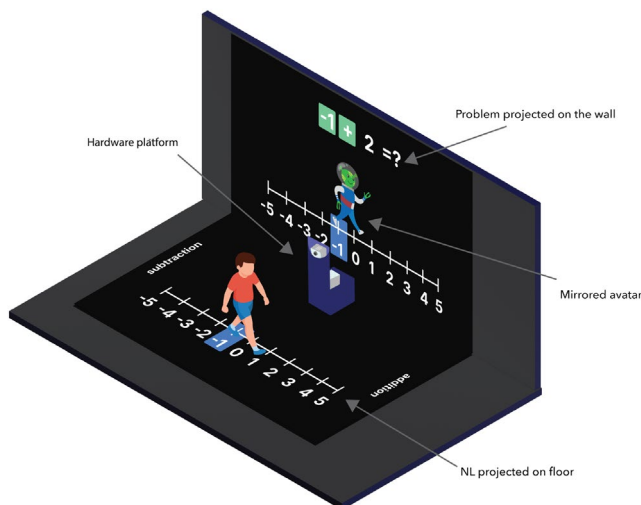


FIGURE 2 Students walk a floor-projected interactive number line, and a screen-based avatar is introduced that mimics their whole-body movements.

then sent to the web server, which interacts with GPT-4 to retrieve context-aware suggestions or hints that support the student's learning process. The integration of asynchronous communication between the Node.js server and the API ensures minimal latency, while error-handling protocols are in place to manage system failures and optimize API usage. Furthermore, the system incorporates caching mechanisms to reduce redundant API calls and improve response efficiency.

To optimize the effectiveness of the feedback provided by GenAI, we collaborated closely with the teacher in the co-design of the feedback prompts. By reiterating and refining the prompts, we ensured that the AI-generated feedback was tailored to support the specific learning objectives set by the teacher. This iterative process involved aligning the AI's responses with the pedagogical approach, ensuring that the feedback complemented the teacher's instructional strategies and met the students' needs. Through this co-design process, we aimed to create a seamless integration of AI-driven feedback with human guidance, fostering a more adaptive and personalized learning environment for students. By aligning the AI's prompts with the teacher's expertise, we aimed to deliver contextually relevant, constructive and pedagogical feedback that would enhance the learning experience and promote student engagement. Below the prompt we gave to GPT-4, and examples of feedback it delivered to students from the system logs.

PROMPT CODE:

```
function feedbackUser()  
  var systemPrompt =  
  
  'You are an intelligent tutor helping a student learn mathematical operations with negative numbers using the number line approach. The student's number line is projected on the floor, with numbers ranging from -5 to +5. Here are the rules the student must follow.  
  
  O1 = Operand1 = First operand  
  Operator = The symbol between the operands  
  O2 = Operand2 = Second operand  
  
  1. Positive O1 + Positive O2  
  Walk to the position of O1 on the number line.  
  Rotate to the right (for +).  
  Walk forward O2 steps.  
  
  2. Positive O1 - Positive O2  
  Walk to the position of O1 on the number line.  
  Rotate to the left (for -).  
  Walk forward O2 steps.  
  
  3. Positive O1 + Negative O2  
  Walk to the position of O1 on the number line.  
  Rotate to the right (for +).  
  Walk backward the absolute value of O2 steps.  
  
  4. Positive O1 - Negative O2  
  Walk to the position of O1 on the number line.
```

Rotate to the left (for -).
Walk backward the absolute value of 02 steps.

5. Negative 01 + Positive 02
Walk to the position of 01 on the number line.
Rotate to the right (for +).
Walk forward 02 steps.

6. Negative 01 - Positive 02
Walk to the position of 01 on the number line.
Rotate to the left (for -).
Walk forward 02 steps.

7. Negative 01 + Negative 02
Walk to the position of 01 on the number line.
Rotate to the right (for +).
Walk backward the absolute value of 02 steps.

8. Negative 01 - Negative 02
Walk to the position of 01 on the number line.
Rotate to the left (for -).
Walk backward the absolute value of 02 steps.

Summarize the steps taken by the student using a past tense and in second person, then provide feedback.
Use no more than 30 words.;

```
var userMessage = convertHistoryToHumanForm();
console.log(userMessage, systemPrompt);
callGPT(userMessage, systemPrompt);
}
```

GenAI FEEDBACK EXAMPLES FROM LOGS:

EXAMPLE 1:

```
NLS:operationSelected|"1|+|-4"
```

```
NLS:LLM-UserMessage|"
01: +1
Operator: +
02: -4
Result: -3
The student walked on the correct number: 1
The student correctly rotated his body to the right
The student correctly walked backward"
```

```
The student walked on the correct number: -3
Provide feedback."
```

```
NLS:LLM-Response|"You correctly walked to 1, rotated right, and walked back 4 steps to -3. Well done following the operation rules with negative numbers on the number line!"
```

EXAMPLE 2:

```
NLS:operationSelected|"2|-|-1"
```

```
NLS:LLM-UserMessage|"
01: +2
Operator: -
02: -1
Result: 3
The student walked on the correct number: 2
The student correctly rotated his body to the left
The student erroneously walked forward"
```

instead of backward\nThe student correctly walked backward\nThe student walked on the correct number: 3\nProvide feedback.”
NLS:LLM-Response|“You walked correctly to 2, turned left, but walked forward instead of backward. Remember, with -1, you should walk backward on the number line.”

Context and procedures

In collaboration with a school in Trondheim, Norway, the study lasted 2 weeks in May 2024. The teacher provided us with a list of participants for each day, and one of the researchers randomly called the participating students from the class list one by one; on average, each student's session lasted 40–50 minutes. The study aimed to assess the impact of adaptive formative feedback within the MOVES-NL educational tool. Participants were divided into two groups:

- Control Group ($N=16$): Received formative feedback from a teacher.
- Experimental Group ($N=14$): Received formative feedback through GenAI.

Students were randomly assigned to one of two conditions, differentiated by the source of formative feedback: teacher feedback or AI-generated feedback. The feedback was formative, as it was provided after the student solved each arithmetic problem and the solution was revealed. In the teacher feedback condition, the first author, an experienced mathematics teacher with over 5 years of teaching experience, conducted task-based, semistructured clinical interviews (as per (Ginsburg, 1997)) to help students ground mathematical procedures in physical actions. Following the completion of each problem, the teacher provided formative feedback in a dialogic manner, co-constructing ideas interactively with the student. For example, the teacher often asked students ‘does this answer make sense?’ The teacher employed gestures and full-body movements to pose questions and receive student responses, engaging in a high-frequency, multimodal and interactive communication style that was naturalistic and rich in detail.

The study room was set up in a dedicated classroom inside the school to avoid external distractions. Each study session consisted of the phases:

1. Facilitator's introduction covering our identity, planned activities and the data collection process (including camera recording and eye-tracking set-up).
2. Introduction to the Walking NL: The facilitator showed students how to move to solve problems.
3. Walking NL: The student walks the NL in order to solve arithmetic problems while an avatar mirrors movements on the screen.
4. Formative feedback: The student get feedback on their task based on the group setting.

During the sessions, children worked on various difficulty levels of math problems, with specific success criteria outlined for each level.

Participants

Our sample consisted of 34 students (18 females) between 11 and 13 years old, selected based on the curriculum timing when students begin learning about negative numbers and

their applications in mathematical concepts. Prior to their participation, written informed consent was obtained from their legal guardians. All the ethical procedures were approved by the national human research ethics organization. Students' participation and data collection were conducted after approval from the Norwegian Agency for Shared Services in Education and Research (Sikt) and with Institutional Review Board approval from the University of California, Berkeley (protocol #2022-10-15703), following all the regulations and recommendations for research with students. Notably, none of the participants had prior experience with digital body-scale number line systems and GenAI technologies. This lack of familiarity was considered in the experimental design and analysis, as it could influence their interaction with the tools and their embodied math learning outcomes.

Data collection

We recorded students' interactions with MOVES-NL and employed sensing devices which allowed us to capture students' experience via multimodal data. The decision to use these data collection techniques was also influenced by the fact that they account for (to some extent) students' embodied learning and their importance in multisensory systems (eg, students externalize their actions with the use of their body/skeleton). The sensing devices and their respective multimodal data allowed us to closely monitor and understand how students experienced the received support, leveraging the key affordances of multimodal data (eg, temporality and direct access to indicators of students' cognitive and affective processes (Cukurova et al., 2020)). Students' activity sessions were recorded using two mobile cameras and one additional sensor device: gaze data from eye-tracking glasses. We collected children's gaze data using Tobii eye-tracking glasses at a 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). Due to errors in data collection, we had to discard four students. These participants were excluded because of incomplete data and this exclusion did not affect the overall group sizes because their data were not factored into the final analysis.

Task performance

Children's task performance was determined for an entire session and calculated as follows. We counted the number of attempts a child took to provide the correct answer and then counted the number of correct answers. The task-based performance was computed as the ratio of correct_answers and the sum of correct_answers and total_attempts.

Data collection and data pre-processing

The mobile eye tracker provides three sources of information. First, the objective video from the participants' view and second the gaze location with respect to the objective video; it also provides the pupil diameter. There are two different preprocessing steps for the gaze location and the pupil diameter.

We computed the correspondences between the video from the eye tracker's objective camera (objective video) and the (ground-truth). These correspondences are called homographies. Regarding eye-tracking data, the field of view for each student was completely different. This particularly pertained to all dyads due to the use of mobile eye trackers. We

corrected this issue using a computer vision operation called homography. We applied the method used in Sharma et al. (2022) for computing this step.

Pupil dilation is highly susceptible to personal and contextual biases, for example, brightness of the screen, time of day, pre-existing physical health conditions, a student's gender, age, and amount of sleep. We used the first 30 seconds of eye-tracking data to normalize pupil dilation, effectively removing subjective and contextual biases. Further normalization was obtained using the darkest (ie, put to a maximum value of 1) and brightest (ie, put to a minimum value of 0) screenshots obtained from a student's complete interaction to account for screen brightness.

Measurements

Cognitive load

Cognitive load is a gaze-based proxy to the mental effort invested when solving a given problem (Joseph & Murugesu, 2020; Palinko et al., 2010). We used eye-tracking data to compute cognitive load as a function of pupillary activity (Duchowski et al., 2018). We give a brief explanation of the process of computing cognitive load or mental effort from pupil diameter. We capture the pupil diameter at 250 Hz. Since pupil diameter can be affected by various factors (eg, ambient light, blinks), preprocessing steps are necessary to remove noise. This includes two steps. Correcting for lighting changes: Light affects pupil size significantly, so experimental conditions typically control for consistent lighting. Blink removal: Data points where the pupil is obscured (during blinks) are interpolated to avoid skewing the results. We focused on small, rapid fluctuations in pupil diameter, known as microscale pupillary oscillations. These oscillations occur even when the overall pupil size is stable. We specifically look at the amplitude and frequency of these oscillations, as greater fluctuations in these oscillations are often linked to higher cognitive load or stress. This is carried out using a wavelet transform. The continuous pupil diameter signal is divided into short time windows (eg, 1-second intervals). The variance or standard deviation of pupil diameter within each window is calculated. Higher variability indicates more frequent and significant pupil oscillations. Across all time windows, the variances are either summed or averaged to get a final value.

Time spent on areas of interest (AOI)

We defined five AOIs (Figures 3 and 4), namely the feedback provider, task text, correct option, incorrect option and the number line projected on the ground. These AOIs were selected because they represent the key informational and interactive components that underpin the learning process within the experimental set-up. The feedback provider (teacher or GenAI) delivered the formative feedback to the students; the task text provided context for solving the problem; the correct and incorrect options reflect the decision-making components of the task, highlighting how participants allocate their attention between possible choices; lastly, the number line projected on the ground is central to the embodied learning process, as it integrates the physical interaction with the mathematical concepts. This selection is rooted in well-established principles of embodied cognition and multimodal interaction, which emphasize the importance of visual and interactive focus points in learning environments (Abrahamson & Bakker, 2016; Giannakos et al., 2020; Piaget, 1970). This approach ensures that our findings are directly tied to the behavioural mechanisms driving student engagement with the feedback.

Then, we computed the proportion of time each participant spent looking at them. The proportion of time spent on these AOIs indicates how learners distribute their attention

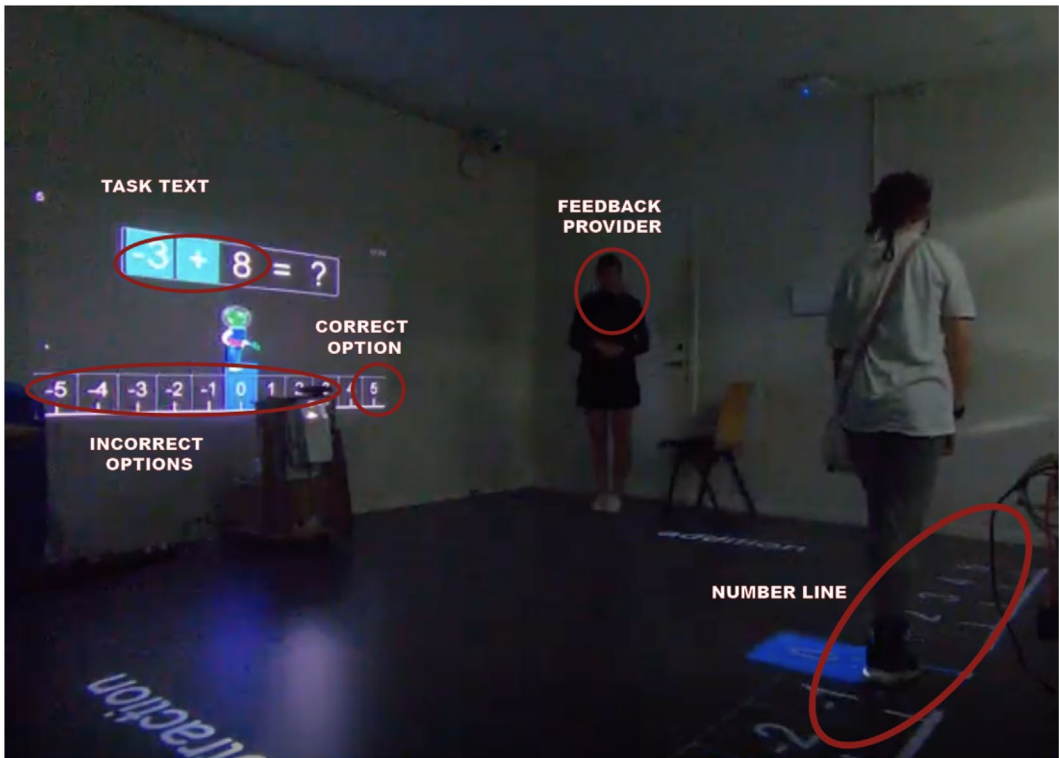


FIGURE 3 AOIs in the control group condition.

across relevant elements, serving as an indirect measure of cognitive engagement. It is important to point out here that we normalized the time spent using the area of the total viewing space. This is because, in the teacher condition, the viewing space had a higher area consisting of the teacher, the screen and the projected area on the ground. Whereas the viewing space in the Gen-AI condition contains only the screen and the projected area on the ground. Moreover, we have also taken into account the time to read the feedback in the GenAI condition by subtracting the time that the participants took to linearly read the text.

Transition among the AOIs

We computed the proportion of times the gaze of the participants shifts from one AOI to another. These transitions capture the attention shifts between the feedback provider, questions, correct/incorrect answers and the projected number line on the ground. This measure complements the proportion of time spent on individual AOIs by adding the temporal aspect in the gaze measurements.

Information processing index (IPI)

Information processing index (IPI) is the ratio of global and local processing. The global processing is identified as a series of short fixations and long saccades. The local processing is characterized by a series of long fixations and short saccades. Small saccades typically occur during tasks like reading or focusing on detailed objects (Poole & Ball, 2006). Larger

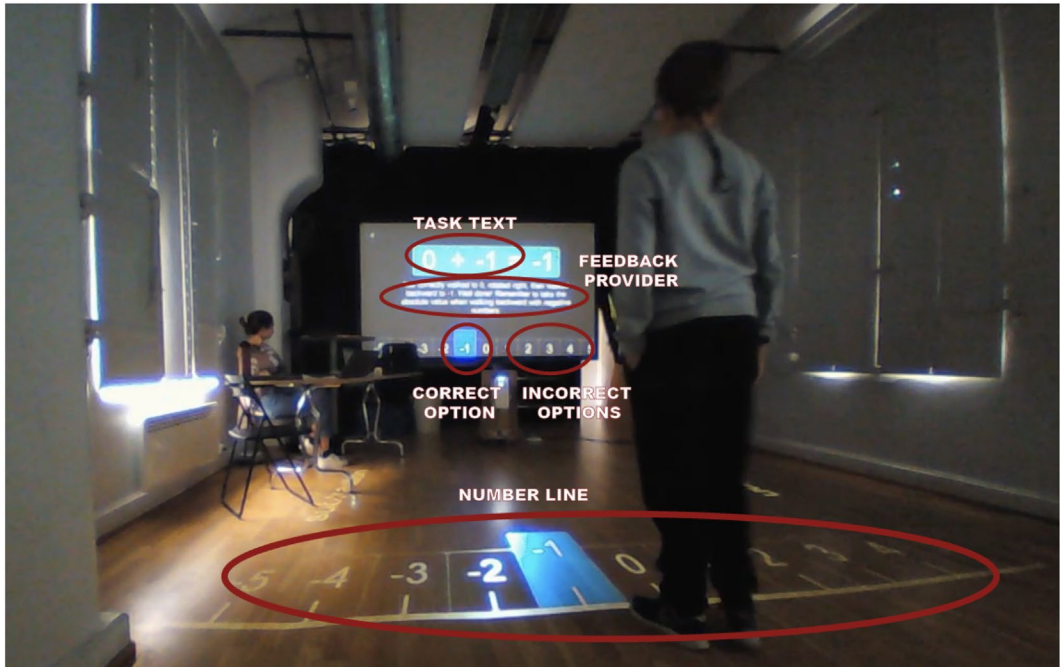


FIGURE 4 AOIs in the experimental group condition.

saccades are common when quickly scanning a scene, looking for objects or navigating visually across a large display. Longer fixation durations may suggest that the object or text being viewed is undergoing more careful processing. Shorter fixations can imply a lack of interest. For computing the two types of information processing (global and local), the same normalizations were carried out as in the case of time spent on AOIs.

Task-based performance

Task-based performance is simply the number of correct solutions divided by the number of attempts made to submit the solution. This measure is related to learning gain for few aspects: first, task-based performance measures how the participants applied the newly acquired knowledge to complete tasks accurately, which can result in the improved understanding of concepts and principles and thus having a higher learning gain. Second, during the tasks, practical experience gained through learning directly improves the ability to perform tasks. This enhances procedural knowledge or practical skills through hands-on learning or task-specific training, resulting in higher learning gains. Third, during solving the questions, participants perform complex tasks that require critical thinking and adaptability that increases their cognitive abilities and aids in the development of problem-solving, decision-making and analytical skills through training or guided learning experiences which increases the learning gain.

By examining these measures in conjunction, we can better understand the interplay between cognitive engagement and learning outcomes. Cognitive load, time spent on AOIs, AOI transitions and IPI collectively provide a multidimensional view of how participants process and engage with information during learning tasks. These measures, combined with task-based performance, offer insights into how engagement translates into measurable educational gains.

Data analysis

To address the research questions, the study employed quantitative data analysis techniques across the two conditions (human instructor and GenAI). We first conducted a *t*-test between the learning gain and conditions as the dependent and independent variables, respectively. Next, we conduct *t*-tests between the eye-tracking measurements and the conditions as the dependent and independent variables, respectively. Prior to the *t*-test, we check the normality assumption using the Shapiro–Wilk test, and upon encountering a non-normally distributed variable, we normalize it by computing the z-score. We also used Welch correction to compute the effective degrees of freedom for unequal variances due to the different number of participants in the two groups.

RESULTS

The analysis for RQ1 focused on assessing learning outcomes and cognitive engagement metrics across the two feedback modalities.

Learning gain

We observe that there is no significant difference in the task-based performance between the two conditions ($t[31.63]=0.73, p < 0.05$) (Figure 5).

Cognitive engagement

In this study, cognitive engagement is defined as the mental effort and attentional focus participants direct towards understanding and solving tasks. It reflects the depth of cognitive processing required to integrate new information and make informed decisions (Joseph & Muruges, 2020; Poole & Ball, 2006). We observe a significant difference in the cognitive load between the two conditions ($t[26.12]=2.89, p < 0.01$). The cognitive load

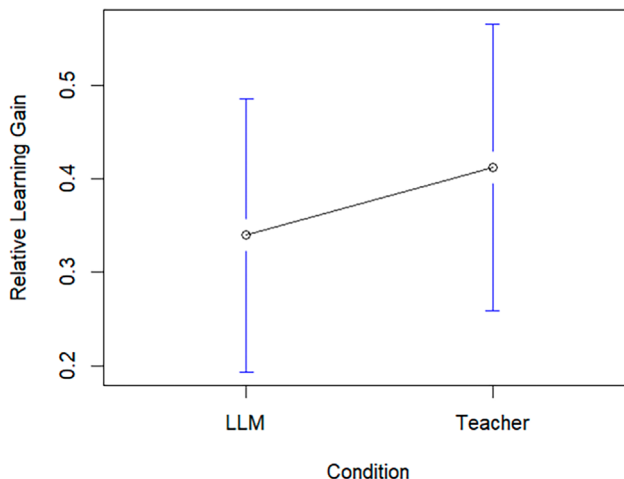


FIGURE 5 Relative learning gain across two conditions.

in the control condition is significantly higher than that in the experimental condition. We also observe a significant difference in IPI between the two conditions ($t[27.56]=5.16$, $p < 0.001$). The IPI in the control condition is significantly higher than that in the experimental condition. This indicated that the participants in the control condition predominantly performed global information processing while the participants in the experimental condition had a more balanced information processing behaviour between the global and local processing (Figures 6 and 7).

The analysis for RQ2 focused on how students engaged with feedback from GenAI and human instructors by examining transitions between various areas of interest (AOIs) and the proportion of time spent on each. This approach highlights key differences in how participants interacted with feedback in the experimental (GenAI) and control (human feedback) conditions (Figure 8). We observe the following:

- participants in control condition shift more often between the feedback provider and task text than the participants in the experimental condition ($t[27.96]=2.55$, $p < 0.05$),
- participants in the control condition shift less often between the feedback provider and the correct option than the participants in the experimental condition ($t[27.99]=2.84$, $p < 0.01$),
- participants in the control condition shift more often between the feedback provider and the incorrect options than the participants in the experimental condition ($t[26.37]=3.47$, $p < 0.01$),
- participants in the control condition shift more often between the task text and the incorrect options than the participants in the experimental condition ($t[26.60]=2.38$, $p < 0.05$),
- participants in the control condition shift less often between the task text and the correct option than the participants in the experimental condition ($t[22.54]=2.47$, $p < 0.05$),
- there are no other transitions that show any significant differences between the two conditions.

When we look at the proportion of time spent on various AOIs (Figure 9), we observe the following:

- the participants looked longer at the feedback provider in the control condition than in the experimental condition ($t[26.83]=3.61$, $p < 0.01$),

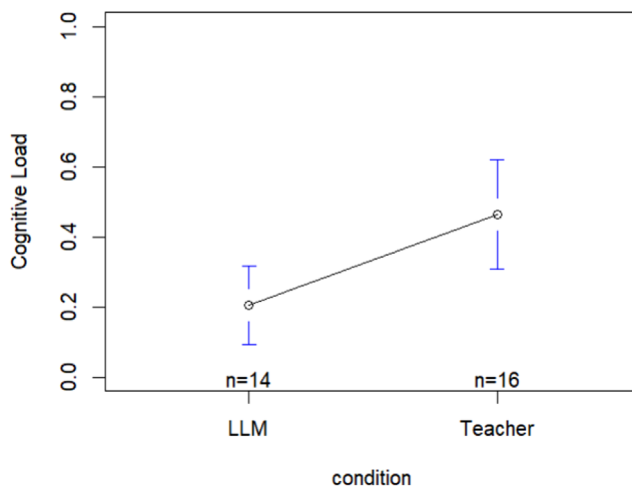


FIGURE 6 Cognitive load across two conditions.

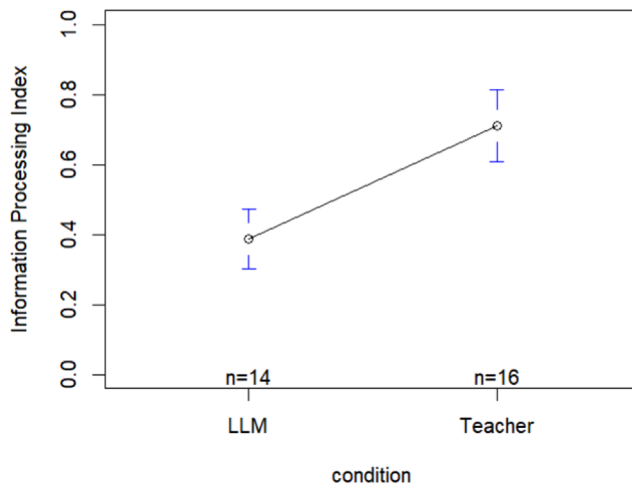


FIGURE 7 Information processing index across two conditions.

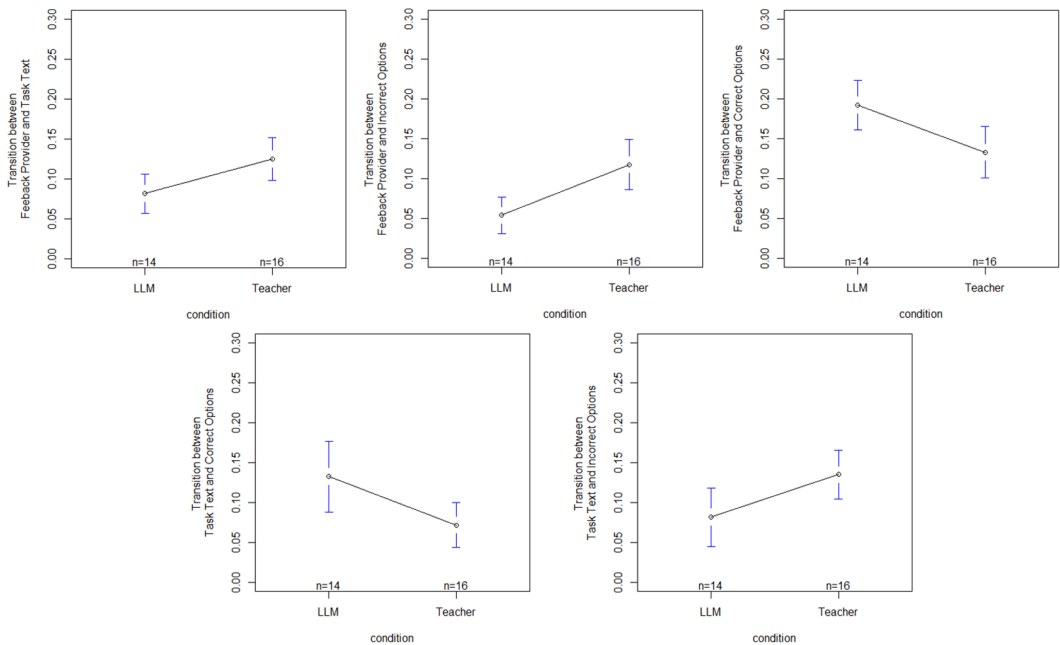


FIGURE 8 Proportion of time spent while shifting between different AOIs across two conditions.

- the participants looked longer at the task text in the control condition than in the experimental condition ($t[22.64]=5.06, p < 0.001$),
- the participants looked shorter at the correct option in the control condition than in the experimental condition ($t[23.30]=3.11, p < 0.01$),
- the participants looked shorter at the number line projected in the control condition than in the experimental condition ($t[25.03]=2.68, p < 0.05$),
- the participants looked longer at the incorrect options in the control condition than in the experimental condition ($t[21.87]=3.82, p < 0.001$).

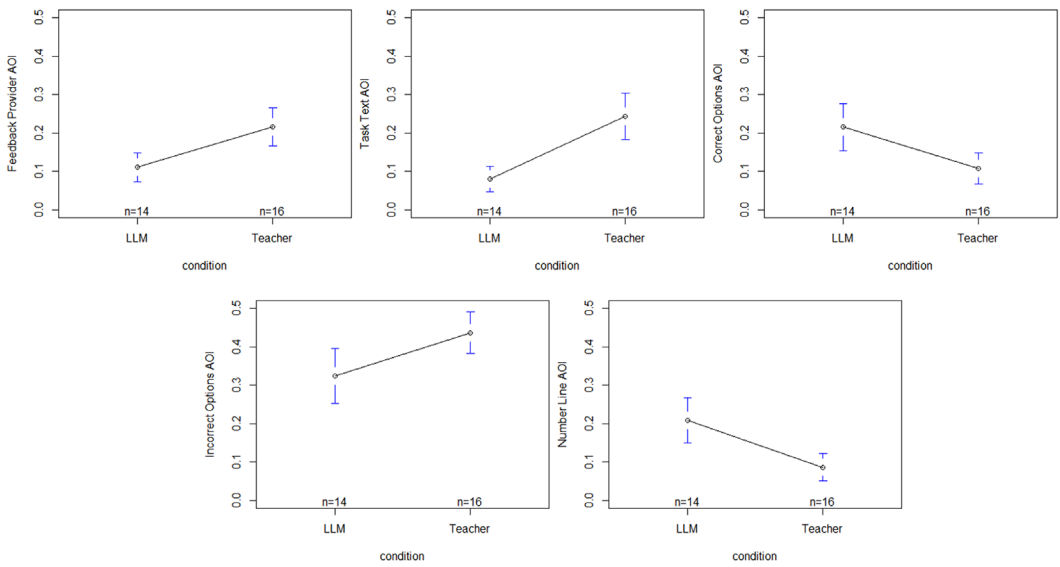


FIGURE 9 Proportion of time spent on the different AOIs across two conditions.

The results revealed that GenAI feedback led to reduced cognitive load and more balanced information processing compared to human feedback, as evidenced by significant differences in Information Processing Index (IPI) and cognitive engagement metrics. While task-based performance was similar across both conditions, participants in the GenAI condition exhibited more efficient transitions between task elements and spent more time focusing on correct responses and key areas, such as the number line projection. In contrast, participants in the human feedback condition showed higher cognitive load and more frequent shifts towards incorrect options, indicating potential engagement with the problem-solving process or less confidence in their initial answers, as discussed later. These findings highlight the potential of integrating AI-generated feedback in multisensory environments to enhance focus and streamline learning processes, offering a scalable and effective tool to complement human feedback.

DISCUSSION AND IMPLICATIONS

The findings of this study underscore the complementary roles that GenAI and human teachers can play in enhancing children's learning experiences, particularly in embodied mathematics learning. This integration is significant as it aligns with contemporary educational strategies that aim to enhance learning while leveraging technology to address diverse student needs (Cosentino et al., 2023; Shute et al., 2021; Vesin et al., 2018).

RQ1: How does feedback from GenAI compare to that from human instructors in terms of learning outcomes and cognitive engagement in embodied math education?

In examining the feedback mechanisms, while students' performance was not significantly different between the two conditions, the eye-tracking results indicate distinct information processing strategies employed by students in the teacher condition compared to the GenAI condition. The higher cognitive load observed in the teacher condition suggests that students engaged in more demanding mental activities when receiving feedback from a human instructor. This aligns with the idea that teachers often encourage deeper reflection and critical analysis by directing students to actively identify and address errors

(Csikszentmihalyi, 2020; Luckin & Cukurova, 2019). Such an approach, while valuable for fostering metacognitive skills, may also increase the mental effort required to process the feedback with the task. This elevated cognitive load might stem from the multifaceted nature of teacher feedback, which often integrates verbal, nonverbal and contextual cues, requiring students to process information globally rather than focusing on specific task components. Higher cognitive load can indeed enhance children's learning, particularly when they are engaging with complex concepts that require deeper reflection and critical analysis. This level of cognitive engagement is often essential for building a robust understanding of challenging material and promoting long-term retention (Csikszentmihalyi, 1990; Paas et al., 2003). In contrast, the lower cognitive load in the GenAI condition indicates that AI feedback may simplify the cognitive demands of the learning process. GenAI systems tend to present feedback in a more structured and consistent manner, focusing on specific task-related elements without the additional interpretive effort often required with human feedback. However, this does not imply that GenAI feedback is universally 'easier' to process; rather, its structured format aligns well with the specific needs of students navigating tasks in this context (Zhai et al., 2024). Additionally, participants in the GenAI condition demonstrated a more balanced approach to information processing (global and local), while those in the teacher condition predominantly employed global processing strategies. These differences highlight how the format and delivery of feedback influence cognitive engagement, with AI enabling more focused and task-specific interaction.

RQ2: How did students engage with the feedback from GenAI compared to human instructor, and what do these patterns reveal about challenges and opportunities of integrating AI-generated feedback within MSEs?

In our analysis of the time spent on AOs and their transitions, students in the control condition appeared to exhibit a strategy that we referred to as elimination, in which they indeed spent less time looking at correct answers and the projected number line while directing their attention more towards incorrect responses. This behaviour suggests that human teachers encourage students to critically evaluate their work and identify errors actively. Teachers are adept at guiding students to reflect on their mistakes, prompting deeper cognitive processing and understanding of the mathematical concepts involved. This strategy fosters an educational setting where students learn to self-correct and enhance their problem-solving skills. In contrast, students receiving feedback from GenAI appeared to demonstrate another processing strategy that we called verification. In the AI condition, students allocated longer periods to analyse correct answers and the number line while efficiently verifying their understanding of incorrect responses. This shift in strategy may be explained, at least in part, by the differing levels of trust students place in human teachers versus AI systems. Children are naturally more familiar with and trusting of human instructors, whose authority and expertise are well established in traditional educational settings (Baker, 2016; Thompson, 2013). As a result, they may feel more confident relying on the feedback provided by teachers, focusing their attention on correcting errors rather than questioning the validity of the feedback. Students interacting with GenAI might approach the feedback with a degree of scepticism, driven by a lack of trust in the AI's ability to provide accurate and reliable guidance (Wojcik et al., 2022). Moreover, our participants had no prior experience with GenAI technologies, which likely influenced their initial perceptions and interactions. A lack of familiarity with such systems can lead to uncertainty or scepticism, particularly when students are required to rely on AI for guidance in learning complex concepts. This unfamiliarity may have amplified their reliance on traditional teacher-centred strategies, as they defaulted to the trusted, familiar source of authority when navigating new and challenging learning environments. This hesitancy could prompt them to adopt a verification strategy, spending more time cross referencing the AI's feedback with the task at hand to ensure its correctness. While this behaviour reflects an adaptive approach to interacting with a novel

feedback source, it also highlights the importance of building trust in AI systems within educational contexts. Establishing such trust might involve explicitly demonstrating the AI's reliability and aligning its feedback with familiar pedagogical strategies. By fostering confidence in the AI's feedback, educators can help students leverage the full potential of GenAI as a supportive tool for learning.

The distinct attention patterns across conditions highlight the complementary strengths of human and AI feedback strengthening the concept of HI (Roschelle et al., 2020). Teachers excel in promoting critical reflection and error correction, while AI provides efficient and adaptive support for verification. Learning technology researchers should consider integrating these approaches, enabling teachers to oversee and customize AI feedback to align with their instructional goals. Teachers could adopt hybrid models where GenAI systems handle task-specific feedback, allowing instructors to focus on fostering metacognitive skills and critical thinking. This hybrid model could provide the benefits of both conditions, supporting iterative reasoning while reducing cognitive load. The integration of GenAI into educational settings represents a transformative opportunity to enhance traditional teaching methods. The study's findings suggest that leveraging the strengths of both GenAI and human instructors can create a more adaptive learning environment. Teachers can focus on fostering critical thinking and self-evaluation skills, while GenAI can offer immediate, tailored feedback that helps students navigate the tasks more efficiently. However, integrating AI into education requires frameworks that address trust, equity and accessibility. As researchers and practitioners, we should ensure that AI tools are transparently designed, explaining both how feedback is generated and why certain recommendations are made. Training programs for educators should also address how to effectively integrate AI tools into their pedagogy, emphasizing the complementary strengths of AI and human feedback. This study contributes to the growing discourse on the role of AI in education by demonstrating how GenAI can complement traditional teaching practices. The synergy between GenAI and teachers has the potential to create more engaging, personalized and impactful learning experiences for students, opening new opportunities for innovative educational practices that leverage the strengths of both approaches.

LIMITATIONS AND CONCLUSION

This study highlights the complementary strengths of human and AI-generated feedback in embodied mathematics education. Human teachers excel in fostering critical reflection and self-evaluation through rich, multimodal interactions, while GenAI provides efficient, task-specific support that reduces cognitive load and enhances focus. Together, these insights lay the groundwork for future hybrid models that leverage both approaches to create adaptive, personalized learning environments. The ethical and practical implications of AI-based learning require careful consideration. Students and educators must understand how AI feedback is generated and why certain recommendations are made. Transparent design principles, coupled with effective training programs for educators, will be essential for fostering trust and maximizing the impact of these technologies. Furthermore, ensuring equitable access to AI tools is paramount to avoid exacerbating existing disparities in education. However, it also has certain limitations. First, the sample size of 30 students is relatively small, which can provide useful insights, but also restrict the generalizability of the results to a broader population. Second, while this study provides valuable insights into the immediate effects of feedback modalities, longer term studies are necessary to understand the sustained impact of GenAI and human feedback on the learning outcomes and retention of skills, particularly in complex and evolving learning tasks. While this study provides evidence

of distinct cognitive and interaction patterns between the two feedback modalities, we acknowledge the limitations in fully interpreting these patterns without qualitative data. The lack of student or teacher feedback constrains our ability to validate some of the interpretative claims. These should be regarded as hypotheses for further investigation rather than definitive conclusions. Future research should integrate qualitative methodologies to explore how students and teachers perceive and interact with AI-generated feedback, providing a richer understanding of the observed patterns. Another limitation of this study is the potential challenge to ecological validity, as the controlled experimental set-up may not fully capture the complexity and variability of real-classroom environment. This research underscores the potential of GenAI to enhance educational feedback mechanisms, demonstrating that AI-generated feedback can be as effective as traditional human feedback in promoting student learning. The integration of GenAI into educational environments provides a promising opportunity to create more adaptive learning experiences. While this study did not implement a scenario where AI and human instructors worked collaboratively within the same condition, it was designed to explore their distinct strengths in isolation. The separation of conditions (human vs. AI feedback) allowed us to systematically compare their effects on learning outcomes and engagement. These insights provide a foundational understanding of how their complementary strengths could be integrated in future hybrid intelligence models. For example, a hybrid feedback model could involve AI providing task-specific feedback, such as hints or corrective prompts, while human instructors focus on fostering critical thinking and metacognitive skills. From a design perspective, this integration would align with hybrid intelligence principles by leveraging the adaptability and efficiency of AI with the nuanced, reflective support provided by human teachers (Luckin & Cukurova, 2019).

We acknowledge this as a limitation of the current study and future research should aim to operationalize hybrid intelligence delving deeper into the integration of these approaches, exploring various ways to combine the strengths of human and AI feedback. Additionally, investigations into how different demographic factors such as age, prior knowledge and learning preferences—affect the interaction between students and these feedback systems would provide valuable insights into optimizing their use in diverse educational contexts.

ACKNOWLEDGEMENTS

We would like to thank all the children and teachers who participated in this study. Without their enthusiasm and dedication, this work would not have been possible.

FUNDING INFORMATION

The collaboration between the University of California, Berkeley and the Norwegian University of Science and Technology is funded by the Peder Sather program.

CONFLICT OF INTEREST STATEMENT

There is no potential conflict of interest in this study.

DATA AVAILABILITY STATEMENT

As it is possible to identify participants from the data, ethical requirements do not permit us to share participant data from this study.

ETHICS STATEMENT

Participation was voluntarily, and all the data collected anonymously. Appropriate permissions and ethical approval were requested and approved.

ORCID

Giulia Cosentino  <https://orcid.org/0000-0001-7569-1019>

Kshitij Sharma  <https://orcid.org/0000-0003-3364-637X>
Dor Abrahamson  <https://orcid.org/0000-0003-2883-4229>

Endnote

¹This index is in contrast with information processing theory, which states that the human mind is comparable to a computer system that processes incoming information through a series of stages: sensory memory, short-term memory and long-term memory (Lachman et al., 2015). Rather, we seek to utilize the IPI to distinguish between global and local processes.

REFERENCES

- Abrahamson, D., & Bakker, A. (2016). Making sense of movement in embodied design for mathematics learning. *Cognitive Research: Principles and Implications*, 1, 1–13.
- Abrahamson, D., Shayan, S., Bakker, A., & Van Der Schaaf, M. (2016). Eye-tracking piaget: Capturing the emergence of attentional anchors in the coordination of proportional motor action. *Human Development*, 58(4–5), 218–244.
- Akata, Z., Balliet, D., De Rijke, M., Dignum, V., Dignum, F., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., Hung, H., Jonker, C., Monz, C., Neerincx, M., Oliehoek, F., Prakken, H., Schlobach, S., van der Gaag, L., van Harmelen, F., ... Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53(8), 18–28. <https://doi.org/10.1109/MC.2020.2996587>
- 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).
- Baker, R. S. (2016). Stupid tutoring systems, intelligent humans. *International Journal of Artificial Intelligence in Education*, 26, 600–614. <https://doi.org/10.1007/s40593-016-0105-0>
- Barrocas, R., Bahnmueller, J., Roesch, S., Lachmair, M., & Moeller, K. (2023). Design and empirical evaluation of a multitouch interaction game-like app for fostering early embodied math learning. *International Journal of Human-Computer Studies*, 175, 103030.
- Bofferding, L., & Hoffman, A. J. (2014). Learning negative integer concepts: Benefits of playing linear board games. In *International Group for the Psychology of Mathematics Education*. <https://api.semanticscholar.org/CorpusID:231261619>
- Chen, M. (2022). Small leans into big steps: A mixed-reality environment to support embodied, ensembled mathematics learning. In *Proceedings of the 21st Annual ACM Interaction Design and Children Conference* (pp. 564–568).
- Chi, M. T., & Wylie, R. (2014). The icap framework: Linking cognitive engagement to active learning outcomes. *Educational Psychologist*, 49(4), 219–243. <https://doi.org/10.1080/00461520.2014.965823>
- Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M., & Abrahamson, D. (2024). Hybrid teaching intelligence: Lessons learned from an embodied mathematics learning experience. *British Journal of Educational Technology*, 56(2), 621–649. <https://doi.org/10.1111/bjet.13525>
- Cosentino, G., Gelsomini, M., Sharma, K., & Giannakos, M. (2023). Interaction modalities and children's learning in multisensory environments: Challenges and trade-offs. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference* (pp. 397–410).
- Cosentino, G., & Giannakos, M. (2023). Multisensory interaction and analytics to enhance smart learning environments: A systematic literature review. *IEEE Transactions on Learning Technologies*, 16(3), 414–430. <https://doi.org/10.1109/TLT.2023.3243210>
- Csikszentmihalyi, M. (2020). *Finding flow: The psychology of engagement with everyday life*. <https://doi.org/10.5860/choice.35-1828>
- Cukurova, M., Giannakos, M., & Martinez-Maldonado, R. (2020). The promise and challenges of multimodal learning analytics. *British Journal of Educational Technology*, 51(5), 1441–1449. <https://doi.org/10.1111/bjet.13015>
- Czikszentmihalyi, M. (1990). *Flow: The psychology of optimal experience*. Harper & Row.
- Deng, R., & Gao, Y. (2023). A review of eye tracking research on video-based learning. *Education and Information Technologies*, 28(6), 7671–7702.
- Dourish, P. (2001). *Where the action is*. MIT Press.
- Druga, S., Vu, S. T., Likhith, E., & Qiu, T. (2019). Inclusive ai literacy for kids around the world. In *Proceedings of Fablearn 2019* (pp. 104–111).
- Duchowski, A. T., Krejtz, K., Krejtz, I., Biele, C., Niedzielska, A., Kiefer, P., & Giannopoulos, I. (2018). The index of pupillary activity: Measuring cognitive load vis-a-vis task difficulty with pupil oscillation. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13).

- Gelsomini, M. (2023). *Sensei—Experience in all senses. Multisensory and multimodal research framework*. <https://sensei.space/>
- Giannakos, M., & Cukurova, M. (2023). The role of learning theory in multimodal learning analytics. *British Journal of Educational Technology*, 54(5), 1246–1267. <https://doi.org/10.1111/bjet.13320>
- Giannakos, M. N., Horn, M. S., Read, J. C., & Markopoulos, P. (2020). *Movement forward: The continued growth of child-computer interaction research*. Elsevier.
- Ginsburg, H. (1997). *Not a cookbook: Guidelines for conducting a clinical interview* (pp. 115–137). <https://doi.org/10.1017/CBO9780511527777.005>
- Holstein, K., Alevan, V., & Rummel, N. (2020). A conceptual framework for human-AI hybrid adaptivity in education. *Artificial Intelligence in Education*, 21, 240–254. <https://doi.org/10.1007/978-3-030-52237-7>
- Huang, Y., Cheng, Y., Chen, L., & Hsu, J. Y. (2019). Human-AI co-learning for datadriven AI. *CoRR*. <http://arxiv.org/abs/1910.12544>
- Järvelä, S., Malmberg, J., & Järvelä, H. (2022). Generation Z and beyond: Co-evolution of human capabilities and intelligent technologies in the twenty-first century. In *International Handbook on Education Development in Asia-Pacific* (pp. 1–13). Springer Nature Singapore. <https://doi.org/10.1007/978-981-19-6887-7>
- Järvelä, S., Nguyen, A., & Hadwin, A. (2023). Human and artificial intelligence collaboration for socially shared regulation in learning. *British Journal of Educational Technology*, 54, 1057–1076. <https://doi.org/10.1111/bjet.13325>
- Joseph, A. W., & Murugesh, R. (2020). Potential eye tracking metrics and indicators to measure cognitive load in human-computer interaction research. *Journal of Scientific Research*, 64(1), 168–175.
- Ke, F., Liu, R., Sokolijk, Z., Dahlstrom-Hakki, I., & Israel, M. (2024). Using eye-tracking in education: Review of empirical research and technology. *Educational Technology Research and Development*, 72, 1383–1418.
- Lachman, R., Lachman, J. L., & Butterfield, E. C. (2015). *Cognitive psychology and information processing: An introduction*. Psychology Press.
- Lee-Cultura, S., Sharma, K., & Giannakos, M. (2022). Children's play and problem-solving in motion-based learning technologies using a multi-modal mixed methods approach. *International Journal of Child-Computer Interaction*, 31, 100355.
- Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of ai: A learning sciences-driven approach. *British Journal of Educational Technology*, 50(6), 2824–2838. <https://doi.org/10.1111/bjet.12861>
- Moreno-Arjonilla, J., López-Ruiz, A., Jiménez-Pérez, J. R., Callejas-Aguilera, J. E., & Jurado, J. M. (2024). Eye-tracking on virtual reality: A survey. *Virtual Reality*, 28(1), 38.
- Moyer-Packenham, P. S., Shumway, J. F., Bullock, E., Tucker, S. I., Anderson-Pence, K. L., Westenskow, A., Boyer-Thurgood, J., Maahs-Fladung, C., Symanzik, J., Mahamane, S., & MacDonald, B. (2015). Young children's learning performance and efficiency when using virtual manipulative mathematics ipad apps. *Journal of Computers in Mathematics and Science Teaching*, 34(1), 41–69.
- Mutlu-Bayraktar, D., Ozel, P., Altindis, F., & Yilmaz, B. (2022). Split-attention effects in multimedia learning environments: Eye-tracking and eeg analysis. *Multimedia Tools and Applications*, 81(6), 8259–8282.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4.
- Paas, F. G., & Van Merriënboer, J. J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86(1), 122–133. <https://doi.org/10.1037//0022-0663.86.1.122>
- Palinko, O., Kun, A. L., Shyrokov, A., & Heeman, P. (2010). Estimating cognitive load using remote eye tracking in a driving simulator. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications* (pp. 141–144).
- Piaget, J. (1970). *Genetic epistemology*. Columbia University Press.
- Poole, A., & Ball, L. J. (2006). Eye tracking in hci and usability research. In *Encyclopedia of human computer interaction* (pp. 211–219). IGI Global.
- Riconscente, M. M. (2013). Results from a controlled study of the ipad fractions game motion math. *Games and Culture*, 8(4), 186–214.
- Roschelle, J., Lester, J., & Fusco, J. (2020). Ai and the future of learning: Expert panel report. *Digital Promise*. <https://doi.org/10.51388/20.500.12265/106>
- Shapiro, L. (2019). Embodied cognition. In *The oxford handbook of cognitive science* (pp. 118–136). Oxford University Press.
- Sharma, K., Lee-Cultura, S., & Giannakos, M. (2022). Keep calm and do not carry-forward: Toward sensor-data driven ai agent to enhance human learning. *Frontiers in Artificial Intelligence*, 4, 713176.
- Sheets-Johnstone, M. (2011). *The primacy of movement* (pp. 1–606). John Benjamins Publishing Company.
- Shute, V., Rahimi, S., Smith, G., Ke, F., Almond, R., Dai, C.-P., & Sun, C. (2021). Maximizing learning without sacrificing the fun: Stealth assessment, adaptivity and learning supports in educational games. *Journal of Computer Assisted Learning*, 37(1), 127–141. <https://doi.org/10.1111/jcal.12473>

- Sinclair, N., & Heyd-Metzuyanim, E. (2014). Learning number with touchcounts: The role of emotions and the body in mathematical communication. *Technology, Knowledge and Learning*, 19, 81–99.
- Suraworachet, W., Seon, J., & Cukurova, M. (2024). Predicting challenge moments from students' discourse: A comparison of GPT-4 to two traditional natural language processing approaches. Proceedings of the 14th Learning Analytics and Knowledge Conference, 473–485. <https://doi.org/10.1145/3636555.3636905>
- Thompson, P. W. (2013). In the absence of meaning.... In *Vital directions for mathematics education research* (pp. 57–93). Springer.
- Ting, C.-C., & Gluth, S. (2024). Unraveling information processes of decision-making with eye-tracking data. *Frontiers in Behavioral Economics*, 3, 1384713. <https://doi.org/10.3389/frbhe.2024.1384713>
- Van Duijvenvoorde, A. C., Zanolie, K., Rombouts, S. A., Raijmakers, M. E., & Crone, E. A. (2008). Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *Journal of Neuroscience*, 28(38), 9495–9503. <https://doi.org/10.1523/JNEUROSCI.1485-08.2008>
- Van Gog, T., & Jarodzka, H. (2013). Eye tracking as a tool to study and enhance cognitive and metacognitive processes in computer-based learning environments. In *International handbook of metacognition and learning technologies* (pp. 143–156). Springer.
- Varma, S., & Schwartz, D. L. (2011). The mental representation of integers: An abstract-to-concrete shift in the understanding of mathematical concepts. *Cognition*, 121(3), 363–385. <https://doi.org/10.1016/j.cognition.2011.08.005>
- Vesin, B., Mangaroska, K., & Giannakos, M. (2018). Learning in smart environments: User-centered design and analytics of an adaptive learning system. *Smart Learning Environments*, 5(1), 1–21. <https://doi.org/10.1186/s40561-018-0071-0>
- Wojcik, E. H., Prasad, A., Hutchinson, S. P., & Shen, K. (2022). Children prefer to learn from smart devices, but do not trust them more than humans. *International Journal of Child-Computer Interaction*, 32, 100406.
- Xue, Y., Zhu, F., & Li, J. (2024). Research on a quantification model of online learning cognitive load based on eye-tracking technology. *Multimedia Tools and Applications*, 1–15. <https://doi.org/10.1007/s11042-024-19814-4>
- Yang, C., Jen, C.-H., Chang, C.-Y., & Yeh, T.-K. (2018). Comparison of animation and static-picture based instruction: Effects on performance and cognitive load for learning genetics. *Journal of Educational Technology & Society*, 21(4), 1–11.
- Yang, W. (2022). Artificial intelligence education for young children: Why, what, and how in curriculum design and implementation. *Computers and Education: Artificial Intelligence*, 3, 100061. <https://doi.org/10.1016/j.caeai.2022.100061>
- Yang, W., Hu, X., Yeter, I. H., Su, J., Yang, Y., & Lee, J. C.-K. (2024). Artificial intelligence education for young children: A case study of technology-enhanced embodied learning. *Journal of Computer Assisted Learning*, 40(2), 465–477.
- Zhai, C., Wibowo, S., & Li, L. D. (2024). The effects of over-reliance on ai dialogue systems on students' cognitive abilities: A systematic review. *Smart Learning Environments*, 11(1), 28.

How to cite this article: Cosentino, G., Anton, J., Sharma, K., Gelsomini, M., Giannakos, M., & Abrahamson, D. (2025). Generative AI and multimodal data for educational feedback: Insights from embodied math learning. *British Journal of Educational Technology*, 56, 1686–1709. <https://doi.org/10.1111/bjet.13587>