Virtual Reality

https://doi.org/10.1007/s10055-026-01418-y

Article in Press

NumbersPlayRoom VR: a serious game for supporting numerical cognition in primary school children

Received: 28 November 2025

Accepted: 15 June 2026

Published online: 18 July 2026

Cite this article as: Fietta V., Gaggi O., Gatto S. *et al.* NumbersPlayRoom VR: a serious game for supporting numerical cognition in primary school children. *Virtual Reality* (2026). https://doi.org/10.1007/s10055-026-01418-y

Valentina Fietta, Ombretta Gaggi, Salvatore Gatto & Matteo Noro

NumbersPlayroom VR: a Serious Game for Supporting Numerical Cognition in Primary School Children

Valentina Fietta[2†], Ombretta Gaggi[1†], Salvatore Gatto[1†], Matteo Noro[1*†]

[1]Department of Mathematics "Tullio Levi-Civita", University of Padua, Via Trieste, 63, Padua, 35121, Veneto, Italy.
[2]Department of Psychology, University of Padua, Via Venezia, 8, Padua, 35121, Veneto, Italy.

*Corresponding author(s). E-mail(s): matteo.noro@math.unipd.it;
Contributing authors: valentina.fietta.1@phd.unipd.it; gaggi@math.unipd.it; salvatore.gatto@math.unipd.it;
[†]All authors contributed equally to this paper. M.N. designed and developed the system.

Abstract

This paper presents *NumbersPlayRoom VR*, an exploratory study on a virtual reality serious game designed to support the learning of fundamental numerical skills in primary school children, particularly those with numerical learning difficulties such as Developmental Dyscalculia. Grounded in numerical cognition and multimedia learning theories, the system contains three mini-games that target complementary components of mathematical reasoning: arithmetic fluency (*Mental Calculation*), procedural column-based calculation (*Column Counting*), and magnitude representation and mental number line (*Block Sorting*). The experience places the learner in a virtual classroom where they can interact with numerical content through embodied actions in an immersive, distraction-free environment. We report a two-phase empirical evaluation. In the first phase, adult users tested the application and provided usability and appreciation feedback, indicating high perceived usability and positive responses to the game design. In the second phase, the system was deployed in a primary school with children, including students with certified learning disorders. Quantitative data from an emoji-based survey and in-game performance metrics were complemented by qualitative observations from teachers and supervisors. Overall, the results

suggest that *NumbersPlayRoom VR* is usable, engaging and provides preliminary evidence of its potential as a complementary tool for supporting numerical cognition in primary school children.

Keywords: VR, Serious Games, Unity, Dyscalculia, Math

1 Introduction

The acquisition of numerical skills is a fundamental stage in child development. However, for children with *Developmental Dyscalculia (DD)*, who represent approximately 5–6% of the school population (Dowker (2024), Shalev and Gross-Tsur (2001)), this process is hindered by persistent deficits in numerical cognition, i.e., the set of cognitive processes involved in understanding and manipulating numerical quantities. Mammarella et al. (2010) have highlighted the complexity of this domain, emphasising that numerical cognition encompasses various representations, including symbolic (numbers), nonsymbolic (quantities), and verbal (number words) formats.

Cognitive models, such as Dehaene's Triple Code (Dehaene (1992), Dehaene and Cohen (1995)), describe numerical cognition as the interaction of three systems: the analog magnitude system, the visual-Arabic code, and the verbal code. Deficiencies in the analog code, specifically the visuospatial representation of number, i.e., the mental number line, are seen as a hallmark trait of *DD*. Evidence from phenomena, such as the distance and SNARC effects, supports the importance of spatial-numerical associations in early number understanding (Cipora et al. (2020), Bachot et al. (2005)).

In addition, children with *DD* often show deficits in visuospatial working memory and executive functions, which further hinder calculation skills and algebraic manipulation (Berch et al. 1999). Therefore, effective interventions should not be limited to practice arithmetic procedures, they should reinforce this fundamental link between number and space. Recent neuroscientific research suggests that training with active video games can improve brain plasticity and visual attention (key factors in learning disorders) by remodeling fronto-parietal attentional networks (Bertoni et al. 2024).

In this context, many serious games have been developed to support children with *DD*, but the majority of them relies on traditional 2D screens. These standard solutions often fail to fully leverage spatial interaction and embodied cognition, potentially limiting their engaging impact and leading to boredom or frustration over time (Ferreira-Brito et al. 2019). Furthermore, mathematics anxiety often hinders the performance of many students with dyscalculia (Broitman and Davis 2013) and the immersive nature of VR can create a safe space for exploration without the pressure of traditional school settings. This reduction in anxiety allows students to experiment and learn from their mistakes in a supportive environment.

VR environments offer immersive, distraction-controlled spaces (Su et al. 2022) where children can directly manipulate quantities (enabling the embodiment of mathematical concepts), receive an immediate multimodal feedback, and engage with numerical concepts through embodied interaction rather than abstract symbolic manipulation, enhancing retention. Literature on immersive learning highlights how

VR can increase engagement, sustain attention, and support understanding of abstract concepts (Belter and Lukosch 2022; Acevedo et al. 2024). These features align with *Cognitive Theory of Multimedia Learning (CTML)* as the integration of verbal, visual, and kinesthetic channels can reduce extraneous cognitive load and promote richer learning (Mayer 2005).

Despite existing studies highlighting VR's educational potential, its use in this domain remains limited. Current research is mostly confined to laboratory settings and lacks comparison with standard methods or other game-based solutions (Mikropoulos and Natsis 2011). Furthermore, many existing VR solutions lack specific design considerations for children with learning difficulties, e. g., simplified interaction and game design, and the presence of accessible, non-text-heavy instructions.

To address this gap, *NumbersPlayRoom VR* was developed as a virtual reality serious game based on cognitive theories of numerical processing and designed specifically for children with numerical learning difficulties, in particular *DD*. The project aims to offer an inclusive, playful, and immersive environment where mathematical concepts can be explored through spatial interaction. Our solution targets specific components of numerical cognition through three distinct mini-games: "*Mental Calculation*" for arithmetic fluency, "*Column Counting*" for procedural skills, and "*Block Sorting*" to reinforce the mental number line and magnitude representation. The system adopts an inclusive design approach that minimises cognitive load, provides multimodal feedback, simplified controls, narrated instructions and visual cues that support learners with *Specific Learning Disorders (SLD)*.

This study aims to intersect inclusivity and educational applications in *VR*, addressing the critical need for digital tools that are not only supportive but also engaging and built on specific needs.

To mitigate the potential isolation caused by Head Mounted Displays (HMDs), which can weaken engagement due to a lack of immediate feedback (Makransky et al. 2019), and in line with collaborative educational practices, the application is conceived as a teacher-supervised activity. This approach is critical for students with learning or self-regulation difficulties, who need constant guidance to avoid disengagement (Wang et al. 2024; Makransky et al. 2017). By leveraging real-time screen mirroring, offered by the most commonly used devices on the market, the experience becomes socially mediated: educators can monitor students' progress, providing them with verbal support, thus ensuring a safe and guided learning environment that bridges the gap between the physical and the virtual worlds.

Our goal is to show how theoretical knowledge about *DD* and visuospatial processing can be translated into *VR* game mechanics, creating a complementary didactic tool that supports traditional education. By addressing specific cognitive difficulties through targeted design choices, we aim to provide a safe, engaging, and inclusive environment where primary school students can reinforce their skills without the anxiety often associated with math. To this end, this study is exploratory in scope: it addresses questions of usability and engagement rather than learning efficacy, as it does not include pre/post cognitive assessments or a control group.

The following research questions guided our study:

RQ1 Does the immersive nature of the VR environment effectively sustain engagement and motivation in primary school children, making it a usable and engaging complementary resource in educational settings?

RQ2 Is it possible to design a VR serious game that is usable also by very young children, including primary school students with diverse learning needs?

The paper principal contributions are:

- The operationalization of numerical cognition models and *DD*-related difficulties into specific VR mechanics. This work demonstrates how abstract cognitive theories can be translated into a concrete, standalone application tailored for cognitive reinforcement;
- An exploratory two-phase empirical evaluation: a preliminary usability study with adults and a field test in a primary school involving 33 children. The results provide qualitative and quantitative insights into the potential of *VR* as a complementary inclusive tool in an educational setting.

A preliminary prototype of this system was presented in (Noro and Gaggi (2025)).

This paper is organised as follows: Section 2 reviews the related work in VR in education. Section 3 details the design and implementation of the *NumbersPlayRoom VR* system. Section 4 describes the evaluation methodology. Section 5 presents the results of the user studies, which are summarised in Section 6. Finally, we conclude and outline future work in Section 7. This game is not intended for diagnostic purposes.

2 Related Work

Inclusive educational design has been a central concern in learning sciences for several decades, with researchers exploring how to design environments that accommodate the full diversity of learners through engagement, motivation, and flexible representation of content. A key contribution in this direction is the Universal Design for Learning (UDL) framework, developed at CAST in the 1990s, which advocates for multiple means of representation, action, and engagement to support learners with diverse needs (Rose et al. 2013). Building on these principles, educational technology research has progressively explored how digital tools, and serious games in particular, can operationalize inclusive design by offering adaptive, engaging, and low-anxiety learning experiences for children with learning disabilities (Jadán-Guerrero et al. 2023).

More recently, the emergence of immersive virtual reality environments has opened new possibilities in this direction, offering unprecedented levels of embodied interaction and environmental control. Within this broader context of inclusive educational technologies, several approaches have been specifically developed to support children with mathematical learning difficulties, such as Developmental Dyscalculia (DD).

Numerous serious games have been proposed for this population, yet most rely on traditional 2D platforms. Virtual Reality, by comparison, offers unique advantages for mathematics learning. According to Belter and Lukosch (2022), the key benefits include enhanced engagement and motivation, a multisensory learning experience, simulation of real-world applications, and reduced anxiety in learning. VR creates an immersive learning environment that can motivate students to actively participate in their education and make learning more enjoyable. By engaging visual, auditory, and kinesthetic senses, VR can help solidify mathematical concepts in a way that is more memorable and easier to understand. Simulating real-world scenarios where math is applied can be particularly beneficial for students who find it challenging to connect math to everyday life.

Wang et al. (2024) showed that the use of VR technologies not only benefits learning but can lead to an improvement in self-regulated abilities regardless of the user's level of scholastic ability. Self-regulated learning (SRL) is defined as a learning process where learners control and manage their own learning, including setting goals, self-monitoring and self-assessment.

The rich and interactive learning experiences offered by VR are proven to enhance learners' cognitive abilities and engagement (Makransky et al. 2017). The same authors also showed that it is really important to provide feedback or guidance during the use of VR otherwise learners may feel isolated and this feeling can weaken their engagement and cognitive abilities. Without immediate and concrete feedback, students may face greater learning difficulties and potentially disengage from the learning process (Makransky et al. 2019).

Students with lower learning achievements benefit from hands-on, authentic learning materials that make concepts tangible and memorable (DiFrancesca et al. 2016). Their success in learning is closely linked to environments that promote active participation and engagement, which helps sustain their motivation to learn. Furthermore, since many of these students struggle with self-regulation, they need prompt and constructive guidance during their learning journey (Xu 2022).

Despite existing studies highlighting VR's educational potential ((Acevedo et al. 2024)), the use of VR in this domain remains limited. Current research is mostly confined to laboratory settings and lacks comparison with standard methods or other game-based solutions (Mikropoulos and Natsis 2011). Athynos (Avila-Pesantez et al. 2019) is an augmented (not virtual) reality serious game for children with Dyscalculia, tested with 40 children aged between 7–9 years with difficulties in learning mathematics. Results showed that children took significant advantage of the training regarding basic numeracy, sequential order and mathematical reasoning. Differently from our system, Athynos runs on the Windows platform using a Kinect sensor 2.0 and cards that integrate AR tags. Moreover, the game was not designed based on numeric cognition and visuospatial reasoning.

Hornos-Arias et al. (2025) evaluated the relevance of digital serious games to detect and rehabilitate developmental dyscalculia in children aged 5 to 12 years. The study analysed 21 serious games, which focus more on remediation than on screening tools. The most used platform is mobile phones; no game implements VR.

Pizza Al Lancio is a mobile application that implements a serious game to help children understand fractions, in particular, equivalent and complementary ones. The game is based on a delivery boy who accidentally eats a slice of pizza while transporting it, and the player is asked to determine the fraction of the pizza eaten or left. The game has been tested with two groups of primary school pupils (Gaggi et al. (2018)).

To the best of our knowledge, the only VR solution for improving mathematical skill is *Math World VR*[1]. The students are placed around different areas in the world to solve simple arithmetic problems in a fun way. The game implements 12 very simple mini games like distinguishing between even/odd numbers with darts or selecting the result of a multiplication or division by shooting a bow and arrow at the right number. No studies are available to discuss the effectiveness of this game.

3 NumbersPlayRoom VR System

NumbersPlayRoom VR is a serious VR game designed to support the development of numerical skills in primary school children, with particular attention to those experiencing learning difficulties such as Developmental Dyscalculia (DD). The mini-games are grounded in principles of numerical cognition and multimedia learning theory, as introduced in the previous sections. The system leverages the immersive and interactive nature of VR to create an engaging and cognitively accessible environment, fostering a "*flow experience*", i.e., a state of focused concentration achieved by balancing task difficulty with the learner's abilities (Perttula et al. (2017)).

A key requirement was to minimise both physical and cognitive barriers. The system was developed using the Unity engine and the open-source OpenXR framework, enabling standalone execution across multiple VR devices while allowing full freedom of movement within the virtual space. From a usability perspective, interactions, navigation, and game mechanics were intentionally kept simple and predictable, allowing children to focus on mathematical tasks rather than on mastering complex controls.

3.1 Main environment

The main gaming environment is a virtual classroom (Figure 1). This setting was chosen to provide a familiar context that reduces anxiety linked to novel environments, while transforming the classroom into a more playful and stimulating space. For this goal, a low-poly graphic style has been adopted, characterized by simple shapes and bright colours. This style is also profitable in terms of performance and essential to guarantee a smooth experience.

Consistent with cognitive load reduction principles, specifically the Coherence Principles from CTML *Coherence Principle* (Mayer 2005; Wil 2016), the environment is designed to focus attention on the relevant information. Core user interface elements and the main components of each mini-game are centrally positioned on a virtual board (Figure 2), which serves as a natural and familiar focal point for the target user. Optional interactive elements (e.g., books, decorative objects, tools) are positioned towards the edges of the classroom. This layout encourages exploration and interaction while minimizing distraction from the primary mathematical tasks[2].

[1] https://www.meta.com/it-it/experiences/math-world-vr/4923914040997217/?srsltid=AfmBOoppa1hIywkrDWeANcLJRk8qjuzQTD8qSLbCWTCBqQasc-me0yg3

[2] A video of the games is available at https://drive.google.com/file/d/1i4oj5VfZpgjdmZTHQgekmwYMbcGHtu35/view?usp=sharing. The whole system was developed using the Italian language.

Fig. 1: The classroom environment.

Fig. 2: The blackboard.

3.2 Mini-games

The three mini-games developed so far target specific components of numerical cognition that are often impaired in DD. Each game includes progressive difficulty levels (easy, medium, hard) and focuses on distinct mathematical skills.

3.2.1 Mini-game 1: *Mental Calculation*

The *Mental Calculation* mini-game aims to strengthen mental calculation fluency and arithmetic fact retrieval, with a focus on verbal and symbolic components of mathematical reasoning. Players are presented with a simple arithmetic operation (i.e., addition, subtraction, or multiplication) that can be solved mentally, without external aids. They must then select the correct result from three possible options (Figure 3). To increase difficulty and inhibit impulsive heuristic responses (i.e., calculating only the unit), higher levels introduce distractors that share the same unit digit with the correct answer. Immediate multimodal feedback, including visual and auditory cues, reinforces learning, and incorrect responses can be retried until the correct solution is provided. At the end of each round, a score is assigned. The difficulty parameters for this mini-game were designed reflecting the transition from arithmetic fact retrieval to active working memory manipulation. In *easy mode* (i.e., numbers ranging from 0 to 10) with multiplication operations (restricted to times tables), users rely on the verbal-lexical system for direct memory retrieval (Daniela Lucangeli 2010). As the difficulty increases, the range of numbers will exceed this automatic threshold (0–20 in *medium* difficulty and 0–30 in *hard*), progressively requiring a switch to active working memory manipulation.

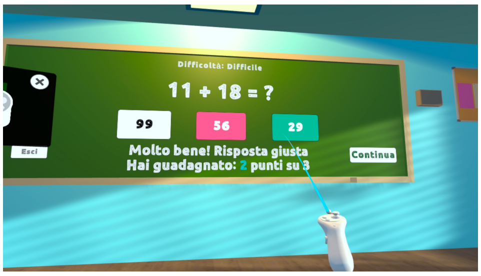


Fig. 3: Screenshot from *Mental Calculation* mini-game, where the user provides the correct answer after an incorrect attempt. Addition exercise, hard difficulty.

Fig. 4: Screenshot from *Block Sorting* mini-game, at the beginning of a round. The player must put the block in the right position over the yellow blocks. Medium difficulty.

3.2.2 Mini-game 2: *Block Sorting*

The *Block Sorting* mini-game focuses on numerical magnitude representation and spatial processing of numbers (i.e., mental number lines). Players must grab 3D blocks with random numbers and place them on pedestals in ascending or descending order (Figure 4). The physical act of manipulating, comparing, and placing blocks in virtual space aims to reinforce the association between numerical symbols, quantities, and spatial position (potentially supporting mechanisms such as the SNARC effect). The number of blocks increases with difficulty, requiring more planning and working memory. The number of items (3, 5, and 8 blocks) was set according to the known limits of the visuospatial working memory and magnitude representation. The *easy mode* (3 blocks) falls within the subitizing range, requiring minimal cognitive load (Kaufman et al. 1949). The progression to 5 and 8 blocks forces the learner to move beyond immediate perception, requiring active serial ordering and continuous magnitude comparison within the semantic system, areas where children with developmental dyscalculia typically show profound deficits. When requested by the child, the game provides colour feedback on the sorting correctness and also supports trial-and-error learning in a low-stress environment: in case of error, the student can freely rearrange the blocks by picking them up again from the pedestals.

3.2.3 Mini-game 3: *Column Counting*

The *Column Counting* mini-game has been designed to support the procedural skills and visual-spatial processing required for column calculation, an area in which children with DD and Nonverbal Learning Disorder (NLD) often struggle (Crollen and Noël 2015; Daniela Lucangeli 2010). Children must solve an arithmetic operation (i.e., addition, subtraction, or multiplication) by placing numbered cards (0–9) in the correct slots on a grid on the blackboard, following the column counting rules (Figure 5). The system validates the placement of each digit and provides granular feedback, helping to pinpoint specific procedural errors in the process. Difficulty levels systematically manipulate syntactic and procedural load. The *easy mode* limits operations to two digits (up to 99) without carry-over or borrowing, while *medium mode* opens the possibility to use these procedures within the same numerical range. Finally, *hard mode* expands the syntactic complexity to three digits (up to 999), requiring greater use of visuospatial working memory. When multiplication is selected, the game explicitly focuses on multiplication tables, and the number of rounds is adapted accordingly. As in the other mini-games, errors can be corrected freely: children may reposition cards or discard them by throwing them into a virtual bin. Completion time is recorded and used as the primary performance metric.

Fig. 5: Screenshot from *Column Counting* mini-game, where the user miscalculated the tens digit in the result. Addition exercise, Medium difficulty.

Fig. 6: Screenshot from the tutorial, many objects are placed on a table, the player is encouraged to interact with them.

3.3 Tutorial

VR may be an unfamiliar technology for many children, particularly for those who are very young. For this reason, an optional tutorial has been developed to facilitate the learning of basic interactions. The tutorial is structured in four progressive, non-rigid steps, each focusing on a specific in-game mechanic:

1. UI interaction: learning to point and click virtual UI buttons;
2. Movement: practicing with the teleportation system to explore the environment;
3. Grabbing: grab and manipulate virtual objects;
4. Socket interaction and Integration: use sockets (i.e., attachment point). This interaction combines all the skills learned previously into a small, gamified task that asks the player to find a hidden key and insert it into the door lock (Figure 6).

To maximize clarity and reduce the learning curve, each step uses voice instructions, concise text, and dynamic highlighting of relevant buttons on virtual controllers.

3.4 Multimodal Interaction and Feedback

The interaction system has been designed to be as intuitive as possible, even for novice VR users, and to minimize physical fatigue, such as the "gorilla arm syndrome" (Hansberger et al. 2017). The main interaction takes place via raycasting from the controllers, reducing the need for large and extended arm movements (e.g., to move the cards in the *Column Counting* game. Direct interaction (e.g., grabbing the blocks in the *Block Sorting* game) is also supported to enhance the sense of presence and physical manipulation. Spatial movement mainly uses teleportation, a snapping system that effectively prevents motion sickness (Golding 2016; Kim et al. 2018).

The experience uses a multimodal feedback system, which is particularly useful for children with DD who can benefit from clear and immediate reinforcement. In particular, we provide:

- Visual feedback: highlighting of targeted objects and useful control buttons, using colours to identify correct answers (green) or errors (red), which are always furnished with motivational and non-punitive text messages. This applies the CTML *Signal Principle* to guide attention;
- Haptic feedback: light vibrations on controllers confirm selections, object grabs, or correct/incorrect results;
- Auditory feedback: different sound effects support each interaction (click, grab, release, snap), transitions between game states (start/end of rounds), and provide positive reinforcement or a discreet indication of error.

The system also integrates synthesized voice instructions to briefly outline the objectives and necessary actions at the beginning of each activity in the mini-games and the tutorial. This choice aims to reduce the visual overload caused by reading text on screen, improve accessibility, and allow children to focus more on the interactive aspect of the task, as requested by the CTML principles of *Segmentation* and *Coherence*. The introduction of synthesized voice instructions is complementary to on-screen instructions that are always present in floating side canvases, which, however, individually suffer from the invisible gorilla effect or inattention blindness (Simons and Chabris 1999) as we discovered from the results obtained during the first testing phase with adult users, described in Section 5.

4 Evaluation Methodology

A two-phase evaluation strategy was adopted to assess the system usability, engagement, and educational potential. This approach combined:

1. a quantitative usability and appreciation assessment with impartial adult users (Phase 1, $N = 33$);
2. a quantitative and qualitative in-class study with the target user group, i.e., primary school children (Phase 2, $N = 33$).

Regarding the in-class study, it was conducted in accordance with ethical principles for research involving minors. Written informed consent was obtained from the parents or legal guardians of all participating children prior to the sessions. The school administration formally authorized both the use of Head-Mounted Displays (HMDs) and the collection of anonymized data. To minimize potential distress, supervisors were present throughout all sessions, safe movement zones were established before each session, and participants were free to interrupt their session at any time. Data were collected and stored anonymously. Given the age of the participants, particular attention was paid

to physical comfort: session duration was kept short (average 20 minutes), and children were asked about dizziness or discomfort immediately after headset removal.
This sequential methodology allowed us to validate and refine the application with adults before deploying it in the school setting with children.
All tests were conducted using a Meta Quest 2 device with the application installed stand-alone on the device.

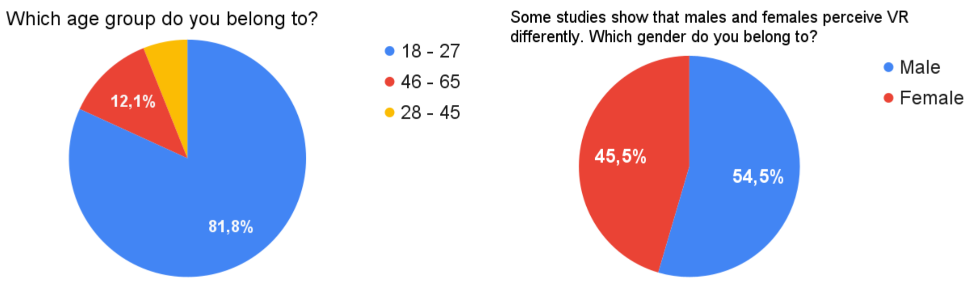
4.1 *Phase 1*: Preliminary evaluation with Adults

The first phase of the study consisted of a preliminary usability evaluation involving a generic adult sample. This session was designed to assess overall usability, instruction clarity, and system stability with a broader, heterogeneous group before involving the intended target users, i.e., children from primary schools, a vulnerable group. This preliminary step enabled the identification and resolution of any technical or design issues in a controlled setting, thereby reducing the risk of frustration or safety concerns during subsequent classroom testing.

All participants followed a structured protocol that included free exploration of the virtual environment, completion of the tutorial sequence, and interaction with the mini-games at predefined difficulty levels. After the session, they completed two questionnaires: the **System Usability Scale** (SUS, (Brooke 1995)), measuring perceived usability, and an **Appreciation Survey**, gathering additional feedback on enjoyment, instructional clarity, perceived discomfort, and overall user experience. The insights collected during this phase were used to identify critical usability issues and refine the application before the formal evaluation with children. It is important to clarify that this group was not intended to serve as a proxy for primary school children or to evaluate educational efficacy. Rather, the primary objective was to identify and resolve major usability flaws, interaction bottlenecks, and system bugs in a controlled setting before exposing a vulnerable target group (children, potentially with learning difficulties) to the technology.

4.1.1 Participants

Thirty-three adult volunteers, aged between 18 and 67 years, took part in the evaluation of our application. $81,8\%$ of users were in the 18–27 age range, with fewer in the 46-65 range $(12,1\%)$ and only a small number $(6,1\%)$ in the 28-45 range (Figure 7a). Participants were recruited through personal contacts, members of the university community, and a snowball sampling approach, where respondents were invited to recruit future subjects among their acquaintances (Parker et al. 2019). The sample aimed for a balanced gender distribution, with 18 males and 15 females (Figure 7b). Each participant completed the test individually to ensure independent and unbiased feedback. Detailed demographic distributions are reported in Subsection 5.1.



(a) Age ranges of participants in *Phase 1.* (b) Gender dist. of participants in *Phase 1.*

Fig. 7: Demographic distribution of adult participants

Prior knowledge of VR experience

Before using the application, participants were asked to rate their familiarity with *VR* headsets on a 1–5 Likert scale. The average score was **1.97**. Figure 8 shows that most participants had little or no prior experience with virtual reality. A Wilcoxon signed-rank test confirmed significantly low prior experience (V = 11, p ¡ .001, HL estimate = -1.50, r = -0.93), with most reporting little to no exposure.

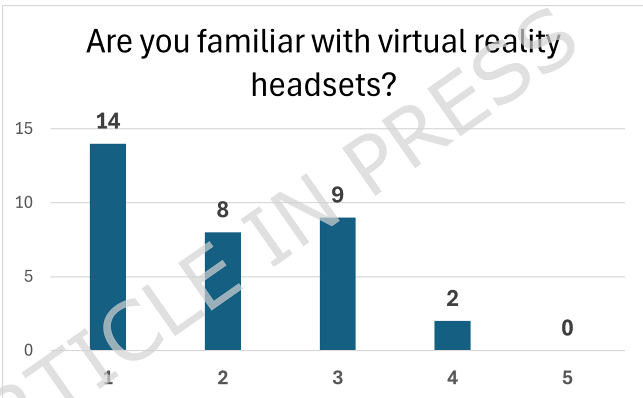


Fig. 8: Familiarity Chart, "Are you familiar with virtual reality headsets?"

4.1.2 Testing procedure

Each participant underwent an individual test session following a standardised protocol, which consisted of:

1. An onboarding phase, during which the supervisor introduced the VR controllers and adjusted the Head-Mounted Display (HMD) to ensure optimal fit and visual clarity;

2. Free exploration for a short period;
3. Completion of the in-game tutorial;
4. Playing mini-games at predefined level of difficulty (i.e., *Mental Calculation* - Hard; *Column Counting* - Medium; *Block Sorting* - Hard);
5. Complete the *SUS questionnaire*;
6. Complete the follow-up *Appreciation survey.*

Participants were encouraged to explore interactions and intentionally make errors to evaluate feedback systems. Sessions were monitored via screen mirroring for observation and assistance if needed.

4.1.3 Testing method

After completing the interactive session, participants filled in two online questionnaires built with Google Forms. The questionnaires were presented in separate sections and responses were collected anonymously.

1. **SUS Questionnaire**: consists of 10 statements that users rate on a 5-point Likert scale (1 = "strongly disagree" to 5 = "strongly agree") (Likert 1932). These items alternate between positive and negative phrasing to mitigate response bias (Brooke 1995). The proposed questionnaire, visible in Table 1, presents its standard formulation except for question Q1, where the participant is asked: "*I think I would have liked to use this application frequently* ***when I was a child***"; since we did not ask participants to impersonate a child.
2. **Appreciation Survey**: participants were also asked to rate various aspects of the game on a 5-point Likert scale, where higher scores indicated a more positive response. The survey, composed of 15 questions codified as AQn where n is the question number, presents a first section that focuses on the three mini-games: *Mental Calculation* (AQ1–AQ2), *Column Counting* (AQ3–AQ4), and *Block Sorting* (AQ5–AQ6). Each game is evaluated in terms of user enjoyment and clarity of instructions. The following set of questions examines the helpfulness of in-game instructions (AQ7), feedback mechanics (AQ8), and the tutorial's effectiveness in teaching gameplay mechanics (AQ9). The questionnaire also explores potential discomforts, such as muscle fatigue (AQ10) and motion sickness (AQ12), as well as the overall engagement and immersiveness of the game environment (AQ11). Additionally, it assesses whether the game helps users focus by reducing distractions (AQ13). Finally, demographic information is collected regarding the user's gender (AQ14) and age group (AQ15), with predefined age ranges from 18 to over 65. We collect data about gender to understand if, as stated by some research, there was evidence of a different biological perception of virtual reality between males and females (Chattha et al. 2020; Golding 2006; Klosterhalfen et al. 2005; Munafo J. 2017). The questionnaire is fully reported in Table A1 of Appendix A.

4.2 Phase 2: In-Class testing with target user

After validating the games with adults and refining the whole system based on their feedback, the system was deployed in a real classroom environment with the target user group. The main objective of this phase was to investigate whether the game could support the learning of mathematics in an engaging way. Due to time constraints, it was not possible to evaluate long-term learning gains or diagnostic validity; these aspects are left as directions for future work. This initial study aims at verifying the usability of the VR technology in this context and, in particular, of our system.

4.2.1 Participants

Testing was conducted in a primary school in Belluno, Italy, within the "*Istituto Comprensivo Statale Belluno 3*". This school was chosen due to facilitated contacts with school personnel who were willing to bring a new and challenging experience to their classrooms. Participants included 33 pupils:

- 14 children from the third-grade class, aged 8-9.
- 19 children from the fifth-grade class, aged 10-11.

Of this sample, five students were identified with certified special educational needs: two children in the third grade (presenting with ADHD, and autism) and three children in the fifth grade (diagnosed with opposition defiant disorder, ADHD, and dyscalculia).

4.2.2 Testing procedure

Sessions were conducted over three days in a designated classroom equipped with two separate testing stations. Supervisors monitored the children's gameplay via remote mirroring. A unified protocol has been established to conduct the tests uniformly to ensure proper data collection. Safe zones were previously set by supervisors, setting the limits as stationary to maintain an appropriately narrow boundary so the child encountered no risk moving too far outside the designated spaces. The specific selection of mini-games was agreed with teachers to ensure that the tasks were aligned with curricular objectives and appropriate for the classes. Once selected, the presentation order of the mini-games was chosen randomly. The testing procedure for this main session was structured as follows:

1. Onboarding phase: the supervisor introduced the VR controllers and adjusted the Head-Mounted Display (HMD) to ensure optimal fit and visual clarity;
2. Completion of the in-game tutorial;
3. Play a sequence of two mini-games. The first one was differentiated by grade level:
   - Third-grade class: *Mental Calculation* (multiplication, medium difficulty);
   - Fifth-grade class: *Column Counting* (addition, medium difficulty).

   The second mini-game was the same for all children: *Block Sorting* (medium difficulty).
4. Free-play time and exploration;
5. Removal of the HMD;
6. Completion of the *Emoji Survey* on a tablet (described in Section 4.2.3)

An extra testing session was conducted on the last day, in order to collect additional data on the involvement and perceived difficulty of the mini-games. For this session, a random sample of 6 children from the fifth grade was selected to test a set of mini-games at *Hard* difficulty. The specific protocol for this session was structured as follows:

1. Selection of participants using a random draw, ensuring gender balance;
2. Participants were asked to wear the headset independently and proceed directly to the games without repeating the tutorial;
3. Play a specific sequence of mini-games, all set to *Hard* difficulty:
   - *Column Counting*: first subtraction, then addition;
   - *Mental Calculation*: first addition, then subtraction;
   - *Block Sorting*;
4. Short free-play session;
5. Completion of the specific *Emoji Survey* for the extra session.

Fig. 9: A screenshot from *Emoji Survey* App. The question is "Do you like the game *Mental Calculation*" ?

4.2.3 Testing Method

Given the age group, the use of a *SUS questionnaire* was not considered appropriate. Consequently, we create a more child-friendly system that we called *Emoji Survey*.

We used three assessment tools:

1. **Emoji Survey**: a web application designed specifically for this study, featuring an intuitive, bias-free interface (Figure 9) where children rated their experience using a 5-point emoji scale (e.g., from very unhappy to very happy face, similar to the Smileyometer scale tool (Read and Horton 2025)). For each child, the same questions were asked[3]. For the main testing phase, the questions aimed to investigate the appreciation of individual mini-games (EQ1, EQ2 and EQ3), checking for any difficulties (EQ4, EQ5, and EQ6) and understanding the predisposition to use this tool for learning (EQ7), while also considering engagement (EQ8). At the end, it was asked if there was any physical discomfort or nausea (EQ9).

[3]We only changed the reference to games *Mental Calculation* or *Column Counting* according to class, since the former was only proposed to the third-grade class, and vice versa, the latter was only proposed to the fifth-grade class.

For the extra testing session, a specific 3-question survey was proposed to the selected fifth-grade sample. These questions focused on the enjoyment of replaying the game (SEQ1), whether the increased difficulty made the games more fun (SEQ2), and the perceived difficulty level compared to the first session (SEQ3). The full questionnaires are reported in Table A2 and A3 of Appendix A. Data were collected anonymously on tablets and stored locally.

2. **Qualitative Notes**: supervisors systematically recorded observations about children's interactions, strategies, difficulties, spontaneous comments, and overall engagement.
3. **In-game scores**: quantitative performance data for each game session (scores, completion times) were automatically logged locally by the application's leaderboard system for each mini-game.

4.3 Data Analysis

Quantitative data from *Phase 1* were analysed by computing individual *SUS* scores using the standard scoring procedure. Answers from the Appreciation Survey were summarized using descriptive statistics, such as means and standard deviations, to provide an overview of user enjoyment, instruction clarity, perceived discomfort, and overall user experience.

In *Phase 2*, children's *Emoji Survey* responses were analysed descriptively to quantify enjoyment, perceived difficulty, and engagement for each mini-game. In-game performance data, i.e., scores and completion times, were logged automatically by the application and used to contextualize behavioural patterns observed during the sessions. Qualitative notes from the supervisor were thematically synthesized to provide context and deeper insights into the user experience. This analysis focused particularly on usability issues, engagement factors, and the specific perceptions of children with learning difficulties.

From the statistical point of view, data were analyzed in JASP (2025 version), (Han and Dawson 2020), using non-parametric tests due to non-normal distribution and a low number of Likert-scale responses. The Wilcoxon signed-rank tests were used to assess if medians significantly exceeded neutrality (value 3) in the utilized Likert scales with adults and children (Field et al. 2012). The between-group comparisons (third vs. fifth graders, n = 13 and 19) employed Mann-Whitney U tests, to test differences across enjoyment and ease-of-use items for the Block Sorting exercise (Field et al. 2012). Furthermore, correlation analyses between in-game metrics and subjective ratings were conducted using Kendall's τ_b (Field et al. 2012).

5 Results

This section presents the results of the two evaluation phases designed to assess the usability, engagement, and educational potential of our *VR* application.

5.1 Adult User testing results

5.1.1 System Usability Scale Results

All 33 participants completed the *SUS questionnaire* after using the application. The mean *SUS* score was **84.17**, well above the threshold of 68, placing the application in the "excellent" usability category, confirmed also by the Wilcoxon signed-rank tests with positive responses across all 10 items (all $0 < V < 561$, all p ¡ .001). The ten questions, their average scores and the Wilcoxon analysis' results are reported in Table 1.

.

Table 1: Descriptive statistics of the *SUS questionnaire* responses (Mean= M, Standard Deviation= SD and the Wilcoxon signed-rank test (V) metrics (p= p value, HL= Hodges-Lehmann Estimate, r= Rank-Biserial Correlation))

Qn	Item	M	SD	V	p	HL	r
Q1	I think I would have liked to use this application often when I was a child	4.24	0.75	378	¡ .001	1.50	1.00
Q2	I found the application complex without needing it	1.58	0.61	0	¡ .001	-1.50	-1.00
Q3	I found the application very easy to use	4.21	0.86	448	¡ .001	1.50	0.93
Q4	I think I would need the support of someone who is already able to use the application	2.00	0.97	24	¡ .001	-1.5	-0.88
Q5	I found the various features of the application well integrated	4.61	0.50	561	¡ .001	1.50	1.00
Q6	I found inconsistencies among the various features of the application	1.33	0.48	0	¡ .001	-1.50	-1.00
Q7	I think most people could learn to use the application easily	4.45	0.62	496	¡ .001	1.50	1.00
Q8	I found the application very complex to use	1.55	0.62	0	¡ .001	-1.50	-1.00
Q9	I became very familiar with the application during its use	4.06	0.83	343	¡ .001	1.50	0.95
Q10	I needed to learn many processes before being able to use the application effectively	1.45	0.75	5	¡ .001	-2.00	-0.98

Analysis of Results

The overall results, i.e., the high mean scores and low standard deviations across most items, show that users strongly agree that the system is intuitive and coherent. Items related to integration of features (Q5, M = 4.61) and ease of learning (Q7, M = 4.45) receive particularly high and consistent ratings.

Slightly higher variability appeared in question Q4 (Need for support, SD = 0.97), Q3 (Easy of use, SD=0.86), Q9 (Confidence, SD=0.83) suggesting that while most users felt comfortable using the system autonomously, a small group required additional guidance.

This finding indicates that virtual reality may not be intuitive for everyone, specially for those who never had previous experience with *VR*-systems. Negatively phrased items (the even ones) scored low, confirming the system's simplicity and internal consistency. In summary, the application achieved excellent usability regardless of prior experience with VR systems.

5.1.2 *Appreciation Survey* Results

After completing the *SUS questionnaire*, participants rated their enjoyment and comprehension of each mini-game (see Table 2 and Figure 10). All three mini-games were highly appreciated: mean values for enjoyment are always greater than 4.6 and they are stastically different, accordingly to the Wilcoxon test, from the neutral value of 3 (See Table 2).

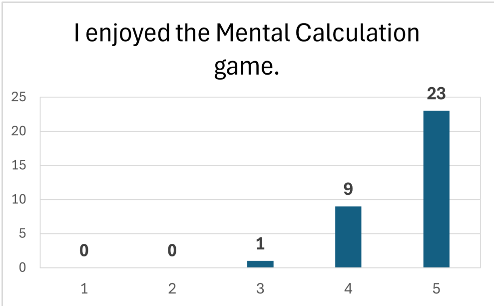
Table 2: Mini-Game Appreciation and Comprehension (N=33) (Mean= M, Standard Deviation= SD, and the Wilcoxon signed-rank test (V) metrics: p= p value, HL= Hodges-Lehmann Estimate, r= Rank-Biserial Correlation).

Mini-game	**Enjoyment**					**Confusion**				
	M ± SD	**V**	**p**	**HL**	**r**	**M ± SD**	**V**	**p**	**HL**	**r**
Mental Calculation	4.67± 0.53	528	¡ .001	2.00	1.00	1.39 ± 1.15	52.50	¡ .001	-2.00	-0.81
Column Counting	4.64 ± 0.56	561	¡ .001	1.50	1.00	1.76± 1.35	86.00	¡ .001	-1.50	-0.69
Block Sorting	4.76 ± 0.49	528	¡ .001	2.00	1.00	1.24 ± 0.78	17.50	¡ .001	-2.00	-0.93

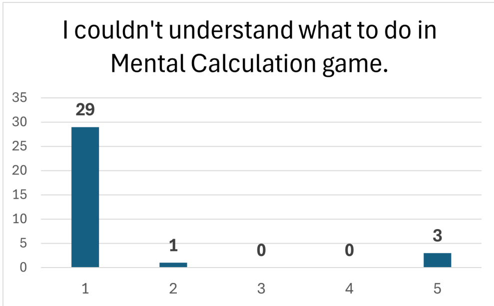
Block Sorting game achieved the highest appreciation scores and the lowest confusion, suggesting excellent clarity of mechanics. These results make this specific mini-game the most successful in terms of both engagement and clarity of mechanics. *Mental Calculation* received high appreciation, though some participants found it slightly more challenging. *Column Counting* maintained strong enjoyment but showed greater variability in comprehension, indicating a need for clearer in-game instructions.

These results show that all participants clearly enjoyed all three activities. The distinction lay not in enjoyment, but in comprehension: *Block Sorting* and *Mental*

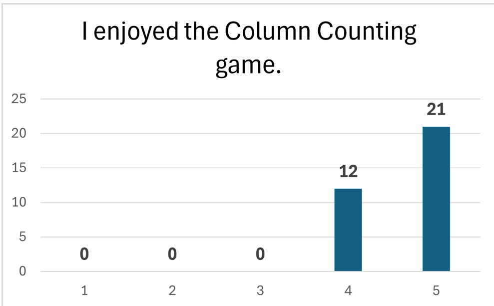
Calculation felt naturally intuitive to most users, while *Column Counting* is considered more challenging by a subset of participants.



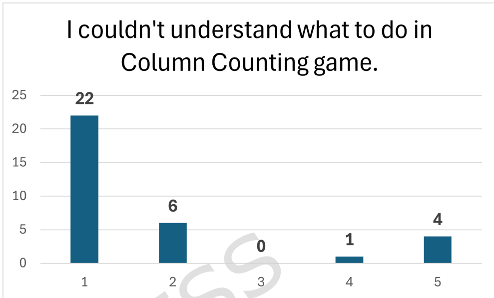
(a) *Mental Calculation* enjoyment (AQ1)



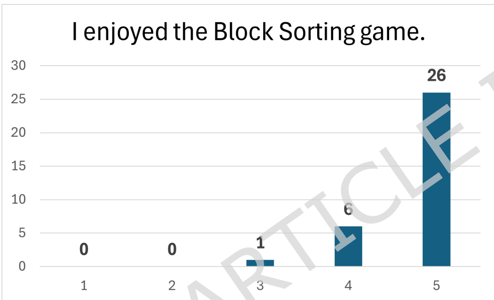
(b) *Mental Calculation* comprehension (AQ2)



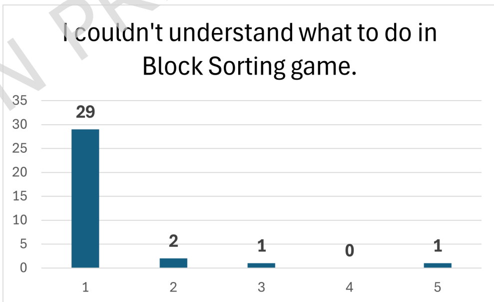
(c) *Column Counting* enjoyment (AQ3)



(d) *Column Counting* comprehension (AQ4)



(e) *Block Sorting* enjoyment (AQ5)

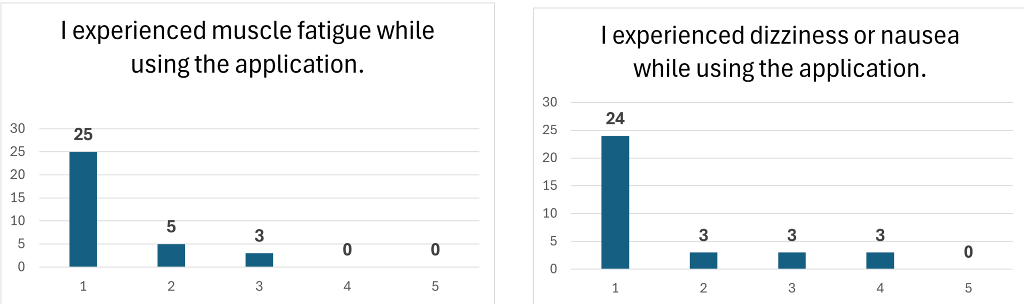


(f) *Block Sorting* comprehension (AQ6)

Fig. 10: Mini-game appreciation and comprehension ratings

Physical Comfort and Cybersickness

Physical comfort ratings were highly positive. Most participants reported minimal muscle fatigue (M = 1.33, SD = 0.64; V = 0, p ¡ .001, HL = -2.00, r = -1.00) and low levels of nausea or dizziness (M = 1.55, SD = 0.99; V = 10.50, p ¡ .001, HL = -2.00, r = -0.96), (Figure 11). All three players reporting nausea or dizziness are males. These results indicate that neither the headset's weight nor the movement required for the interaction caused significant physical strain during the testing sessions (Figure 11a). Discomfort, when reported, appeared linked to individual sensitivity rather than demographic factors. Notably, it often occurred when participants used the optional smooth locomotion mode instead of the default teleportation system. This confirms that teleportation is an effective and comfortable navigation method for reducing motion sickness risk.



(a) Muscle fatigue (AQ10)

(b) Nausea and dizziness (AQ12)

Fig. 11: Physical comfort assessment

5.1.3 Critical Issues Identified

The first phase of the tests reported some critical issues, i.e., the participants did not see the lateral instructions, or the conflicts with the snap turning control. Therefore, voice guidance was added to the games, and interaction mechanics were improved before classroom deployment.

Several participants reported difficulty noticing or following the on-screen written instructions, deliberately located laterally in the virtual space, particularly in the *Column Counting* game (see Figure 12). This issue is related to the phenomenon of **inattentional blindness**, also called **gorilla effect**. This phenomenon is the tendency to overlook peripheral information when concentrating on a task.

We note that participants often need the collaboration of the supervisor to understand the game mechanics. To mitigate this need, short voice prompts were added at the start of each mini-game and during the tutorial. This multimodal approach reduced cognitive load and supported different learning styles. Following this improvement, participants were able to understand the games immediately without further clarification from the instructor.

Fig. 12: Example of a floating instruction canva

Other two issues were solved before testing with target users: (1) repositioning of blocks in the *Block Sorting* to prevent unintentional user movements that frequently resulted in blocks falling out of the world, and (2) removal of the snap-turning control, which occasionally conflicted with left-stick movement and object manipulation.

In summary, the adult testing phase revealed minor usability issues mainly related to instruction visibility and control smoothness. Implementing voice guidance and refining interaction mechanics significantly enhanced clarity and comfort, preparing the system for classroom testing with the target group.

5.2 In-class Testing with children

The second evaluation phase examined how primary school children interacted with *NumbersPlayRoom VR* in a real classroom environment. These tests aimed to observe the children to assess their interaction with *VR* environment, their enjoyment, and the potential use of our system as an educational tool for learning mathematics. Quantitative data were collected using the *Emoji Survey*, while qualitative data were derived from supervisor observations and in-game performance logs.

The user engagement during all sessions was remarkably high. Children approached activities with enthusiasm and curiosity, and once immersed in the virtual environment, they demonstrated focused behaviour and purposeful interaction. Teachers reported that the level of sustained attention often exceeded what is typically observed during traditional math exercises. Common engagement indicators included:

- virtual objects were examined in detail by moving closer to them;
- the process of reasoning was expressed verbally in an impulsive manner, e.g., “I need one more block” or “This column is six”;
- the desire to either try again or to make the game more challenging;
- recurrent comments about their wish to have their own *VR* headset and play again.

5.2.1 Quantitative Results from the *Emoji Survey*

Children rated each mini-game and their overall experiences on a 5-point Likert-like emoji scale, where higher scores correspond to more positive evaluations. The results of the survey (see Figure 13) are shown in Table 3.

Table 3: Enjoyment and perceived ease of use of each mini-game by class (Mean= M, Standard Deviation= SD, and the Wilcoxon signed-rank test (V) metrics: p= p value, HL= Hodges-Lehmann Estimate, r= Rank-Biserial Correlation).

Class	Mini-game	Enjoyment					Perceived ease of use				
		M ± SD	**V**	**p**	**HL**	**r**	**M ± SD**	**V**	**p**	**HL**	**r**
Third	*Block Sorting*	4.85± 0.38	91	¡.001	2.00	1.00	4.85± 0.38	91	¡.001	2.00	1.00
Fifth	*Block Sorting*	4.68± 0.75	187.50	¡.001	1.50	0.97	4.47± 0.96	183	¡.001	1.50	0.93
Fifth	*Column Counting*	4.47± 0.61	171	¡.001	1.50	1.00	4.42± 0.84	167	¡.001	1.50	0.95
Third	*Mental Calculation*	4.92± 0.28	91	¡.001	2.00	1.00	4.77± 0.60	78	¡.001	2.00	1.00

Block Sorting - Third grade

Block Sorting received a very positive reaction from third-grade students. Enjoyment scores were extremely high (M = 4.85, SD = 0.38), indicating that the tasks were enjoyable and appealing to the target users. Consistent with these scores, the perceived ease of use scores were the same (M = 4.85, SD = 0.38), thereby mirroring the ease of the task and its approachability. Wilcoxon signed-rank tests against the neutral value of 3 confirmed highly positive perceptions of *Block Sorting*, (see Table 3). These results suggest that this mini-game could be used as an introductory or confidence-building activity within a set of mathematical tasks in our *VR* system.

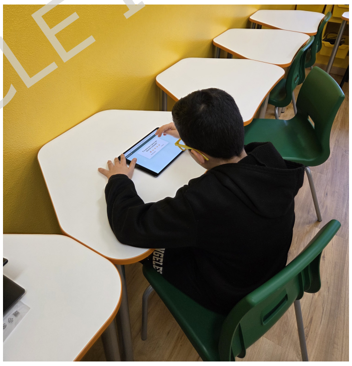


Fig. 13: Child completing the *Emoji Survey*.

Block Sorting - Fifth grade

For fifth-grade pupils, *Block Sorting* remained the most appreciated activity. Enjoyment registered across children was still high (M=4.68, SD=0.75), and although the variability grew compared to third grade, the overall experience remained strongly positive. The perceived ease of use of the game was slightly lower than for third-grade children (M = 4.47, SD = 0.96), with a more varied range of answers. Even though, Wilcoxon signed-rank tests for the fifth-grade class (n = 19) against the neutral value of 3 confirmed highly positive perceptions of this game (see Table 3). This pattern suggests that older children may approach the task using different strategies or have different expectations. Some may find it trivially easy, while others may perceive it as mildly cognitively demanding. The positive reaction from both groups of children confirms that this mini-game is widely usable and appealing to different users. However, it also highlights that older participants may require more challenging content to maintain engagement.

Moreover, Mann-Whitney U tests revealed no significant differences between third- and fifth-grade students across items about *Block Sorting*'s Enjoyment (U = 115.50, p = 0.67, r = 0.07) and Perceived ease of use (U = 101.50, p = 0.28, r = 0.18).

Column Counting - Fifth grade

The Wilcoxon signed-rank tests against the neutral value of 3 confirmed, also in this case, positive perceptions of *Column Counting*, (see Table 3). Although enjoyment remained high (M = 4.47, SD = 0.61), the higher variability suggests that the activity had a different impact on fifth-grade children. Perceived ease of use followed the same trend (M=4.42, SD=0.84), indicating that some users experienced this game as more challenging than others. The game's arithmetic complexity and its need for continuous visual tracking may partly explain these results. While the game is still enjoyable, the data suggest that it would benefit from greater precision when placing tiles or from more difficulty tiers to address the varying skill levels within this age group.

Mental Calculation - Third grade

The *Mental Calculation* mini-game proved highly successful with third-grade pupils (see Table 3 for the statistically significant Wilcoxon test's results). Enjoyment levels were very high (M = 4.92, SD = 0.28), indicating the overall appreciation and exceptionally low variability. Not only did the children like the activity, but they also did so consistently, with no notable outliers. The perceived ease of use also showed excellent results (M=4.77, SD=0.60): pupils rated the game easy to use and most of them experienced little or no challenge in completing the tasks. This aligns with the game's simple mechanics and short action–feedback loop. Overall, the findings suggest that, at its current level of difficulty, *Mental Calculation* is highly usable and enjoyable for this age group. However, it may be too easy to stimulate deeper engagement or cognitive challenges.

5.2.2 Qualitative Observations

A few usability considerations emerged across the sessions. Some participants often expressed the desire to throw virtual objects presented in the classroom freely.

Although this interaction might be a bit distracting, it shows that the system can effectively immerse players in the *VR* environment. Adding short, controlled moments to the game could enhance its engagement. During *Block Sorting* task, many children asked whether the blocks should be arranged from left to right or in the opposite direction. This revealed a degree of ambiguity in the task, as well as some confusion between increasing and decreasing order, especially among fifth-grade pupils.

A subset of children found the teleportation-based movement system unintuitive or too static, preferring the continuous movement via the *VR* joystick. In one case, this preference led to mild dizziness, highlighting the importance of providing a hybrid movement option. Additional issues included difficulties with raycast pointing among shorter or more timid children, who tended to hold the controllers close to their chest (see Figure 14). In this case, the children required more time to familiarize themselves with the environment. It was important to reassure pupils who initially expressed fear of being judged on their performance; once reassured, they typically completed the tasks successfully and with visible satisfaction.

Fig. 14: A child playing *Column Counting* mini-game.

5.2.3 Extra Testing Session

Extra testing sessions were conducted after all the children had fulfilled the experiment. Six children from the fifth-grade class were randomly selected to play some of the mini-games again, with increased difficulty. Then, each child in this sample answered a second version of the *Emoji Survey* to gather information on enjoyment and perceived difficulty with the more challenging tasks compared to the first use of the system. The results of the additional testing sessions are reported in the Table 4 with the Wilcoxon signed-rank results for the extra session. They confirmed significantly positive perceptions of the sessions.

As expected, the results of this additional test provide valuable insights into how older pupils respond to increased difficulty. The first question (SEQ1) revealed a high level of enthusiasm (M=4.83, SD=0.41), suggesting that participants were willing to play again despite encountering more challenging tasks. This suggests strong interest retention and a positive attitude towards repeated gameplay. The second question

Table 4: Extra testing session: enjoyment, fun, and perceived ease of use at higher difficulty levels (Mean= M, Standard Deviation= SD, and the Wilcoxon signed-rank test (V) metrics: p= p value, HL= Hodges-Lehmann Estimate, r= Rank-Biserial Correlation).

Questions	**M ± SD**	**V**	**p**	**HL**	**r**	**Interpretation**
SEQ1 – Enjoyment (playing again)	4.83 ± 0.41	21	¡ .05	2.00	1.00	Very high enjoyment
SEQ2 – Increased fun	4.33 ± 0.52	21	¡ .05	1.50	1.00	Moderately higher fun
SEQ3 – Perceived ease of use	4.00 ± 0.63	15	¡ .05	1.00	1.00	Slightly harder tasks

(SEQ2) showed that children generally found the harder version of the games more enjoyable (M = 4.33, SD = 0.52), though there was slightly more variability. The final question (SEQ3) recorded the lowest mean score (M = 4.00, SD = 0.63), confirming that the advanced tasks were perceived as noticeably more difficult. However, it is important to note that difficulty did not reduce engagement; instead, it appeared to stimulate curiosity and motivation. These findings suggest that older pupils may benefit from higher complexity levels, as more challenging tasks can sustain interest, promote cognitive engagement and enhance the educational value of the mini-games.

5.2.4 In-Game Performance Metrics

In-game performance was logged automatically during *Phase 2* and provides an objective complement to the *Emoji Survey* and the supervisors' observations. Since one participant with specific needs only explored the environment without completing the protocol, a total of 32 valid gameplay sessions were analysed. Scores were interpreted relative to the maximum achievable points for each mini-game and difficulty level; for *Column Counting*, completion time was used as the primary performance indicator.

Block Sorting.

Block Sorting was completed by nearly all children across both grades. Twenty-nine participants played at medium difficulty (maximum score = 15), obtaining near-ceiling performance with an average score of **14.69/15**. The registered scores were clustered around the maximum value, with only minor variations and one lower outlier (12/15, see Figure 15). This indicates that the mechanics were understood immediately and that the task was appropriately calibrated for both age groups. Three additional children preferred the easy difficulty (maximum score = 9) due to low confidence in mathematics; nevertheless, their performance remained high (two perfect 9/9 scores and one 8/9). These data confirm *Block Sorting* as the most usable activity, coherently with its high enjoyment and ease ratings in the *Emoji Survey*.

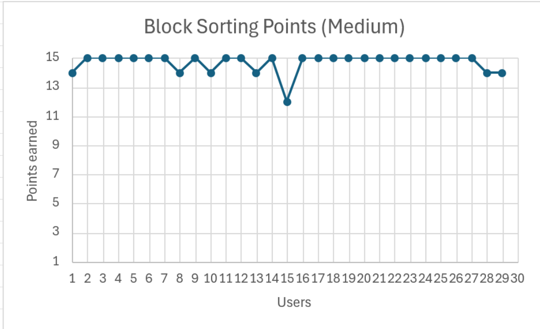


Fig. 15: Scores obtained in *Block Sorting* at medium difficulty.

Mental Calculation.

Performance in *Mental Calculation* was also consistently high among third-grade pupils. For addition and subtraction at hard difficulty (maximum score = 9), all fourteen participants completed the task with an average score of **8.86/9**, showing strong accuracy despite the higher cognitive demand. Multiplication at medium difficulty (maximum score = 15) was completed by thirteen children, with one missing log due to an incomplete session closure. The average score was **14.38/15**, and only a few attempts showed moderate drops (12–13 points). Overall, these results suggest that the mechanics were quickly learned and confirm that the mini-game was well-suited to the third-grade group (see Figure 16).

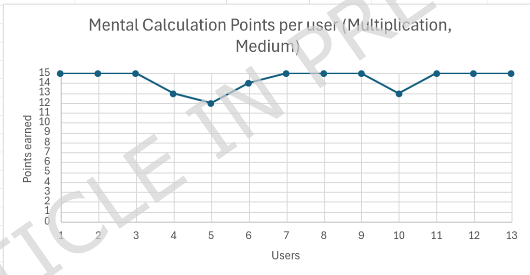


Fig. 16: Scores obtained in *Mental Calculation* tasks.

Column Counting.

Eighteen fifth-grade children completed the *Column Counting* task at medium difficulty; one pupil experienced difficulty with the structured task and was allowed to interact freely with the environment; for this reason, no valid performance record was collected. Task performance differences for this task emerged mainly in completion time rather than accuracy. Completion times ranged from **02:28** to **06:23**, with a mean of **04:14** (SD $\approx$ **01:05**). This variability, depicted in Figure 17, suggests a higher and less uniform cognitive load compared to the other tasks, likely due to the need to align digits, manage carry operations, and sustain visual attention across multiple elements. The dispersion in timing closely matches the *Emoji Survey* results, where *Column Counting* received the widest spread in perceived ease of use.

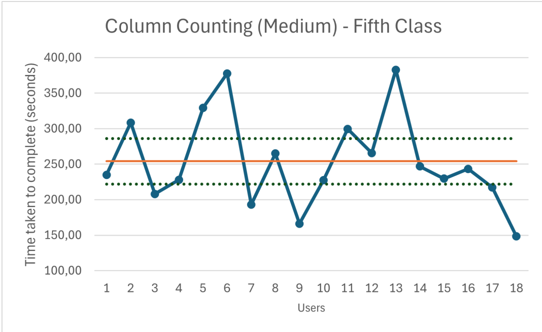


Fig. 17: Completion times (in seconds) for *Column Counting* at medium difficulty (fifth-grade); with mean line and confidence interval

Together, these quantitative performance metrics confirm that the children were able to engage effectively with the mathematical tasks. *Block Sorting* and *Mental Calculation* produced near-ceiling accuracy and rapid comprehension, while *Column Counting* displayed a more demanding profile reflected in longer and more variable completion times. These outcomes align with the subjective ratings and qualitative notes, reinforcing the suitability of the system for short classroom learning sessions.

Correlation Analysis between In-game Metrics and the Emoji Survey

To investigate the relationship between performance metrics and subjective experience, a correlation analysis was performed. Given the small sample size depending on the task and the non-normal distribution of the data (as confirmed by the Shapiro-Wilk test, p ¡ .001), non-parametric tests were employed. Overall, the correlations between the *Emoji Survey* ratings and objective performance metrics (scores and completion times) did not reach statistical significance in most tasks (see Table 5). This result suggests that students' subjective enjoyment or perceived difficulty are only in some cases directly related to their actual performance. However, some specific significant correlations emerged from the data. About the *Mental Calculation*, a significant positive correlation was found between "Enjoyment" and "Perceived ease of use" (τ_b= 0.60, p ¡ .05). This suggests that, for the third class, a higher perception of usability was associated with higher enjoyment with this minigame. Second, in the *Block Sorting* a significant positive correlation was observed between "Enjoyment" and "Perceived ease of use" (τ_b= 0.34, p ¡ .05). This suggests that also for the total sample of children from third and fifth classes, a higher perception of usability of the *Block Sorting* minigame was associated with higher enjoyment.

Table 5: Comprehensive Correlation Analysis between In-game Metrics and Subjective Ratings in the *Emoji Survey* (τ_b = Kendall's τ_b, p= p value, CI= Confidence Intervals).

Task	**Correlated variables**	τ_b	**p**	**95% CI** [*Lower*, *Upper*]
Mental Calculation	Enjoyment – Perceived Ease	0.60	¡ .05	[0.32, 0.88]
	Enjoyment – Points	-0.18	0.51	[-0.42, 0.06]
	Perceived Ease – Points	0.29	0.28	[-0.06, 0.64]
Column Counting	Enjoyment – Perceived Ease	0.37	0.09	[0.08, 0.67]
	Enjoyment – Completion Time	-0.11	0.61	[-0.48, 0.27]
	Perceived Ease – Completion Time	0.02	0.92	[-0.37, 0.41]
Block Sorting	Enjoyment – Perceived Ease	0.34	¡ .05	[0.16, 0.53]
	Enjoyment – Points	0.05	0.79	[-0.11, 0.20]
	Perceived Ease – Points	-0.05	0.77	[-0.21, 0.11]

5.2.5 Children with learning disorders

The following observations concern a small sample of children with certified learning disorders and should therefore be interpreted as exploratory qualitative insights rather than as systematic evidence of effectiveness. These cases are reported to highlight potential directions for future research. Despite the limited sample size, some recurring behavioral patterns were observed during the sessions.

In the third-grade class, two children presented certified conditions: one with ADHD and one with non-verbal autism. The child with ADHD demonstrated a good level of autonomy and engagement, completing the assigned tasks with only minor lapses of attention to the game instructions. The child with non-verbal autism, although not formally assessed within the study, explored the virtual environment with evident curiosity and engagement. This participant expressed a preference for continuous locomotion using the joystick, suggesting that the default teleportation system may not be optimal for all users.

In the fifth-grade class, three cases were observed. The child with oppositional defiant disorder experienced difficulties with the teleportation system, which led to an interruption of the session, highlighting the need for more flexible navigation options. The child with dyscalculia was able to complete the assigned calculations independently, maintaining focus throughout the activity, albeit at a slower pace. The second child with ADHD appeared more distracted during the initial interaction but showed increased focus in subsequent tasks, particularly at higher difficulty levels, and expressed interest in further interaction with the virtual environment after completing the exercises.

Teachers reported that some pupils who typically struggle with attention or arithmetic in traditional classroom settings appeared more focused and motivated during the VR activities.

Even if these observations are encouraging, they are based on a limited number of cases and do not allow for generalizable conclusions regarding effectiveness. Future studies with larger and more targeted samples, as well as controlled experimental designs, are needed to systematically evaluate the impact of VR on learners with specific educational needs.

6 Discussion

The present study aligns promisingly with recent benefits reported in the literature for the use of VR in school-based learning, including for children with learning difficulties such as DD (Jadán-Guerrero et al. 2023; Wang et al. 2024). More broadly, this study offers several contributions to the current state of the art. The majority of serious games reviewed by Hornos-Arias et al. (2025) focus on mobile platforms rather than VR environments and are oriented toward remediation without a systematic grounding in numerical cognition models. Athynos (Avila-Pesantez et al. 2019) and Math World VR are some examples. *NumbersPlayRoom VR* instead explicitly operates on cognitive models and develops distinctive game mechanics that focus on specific cognitive components known to be deficient in DD, going beyond the laboratory-limited research reported by Mikropoulos and Natsis (2011). Our primary aim was to explore the usability, engagement, and educational potential of our VR game in an ecological school environment. To achieve this, we designed a two-phase experiment.

A first testing phase was conducted with adults (N=33), confirming the system's high usability (SUS score = 84.17) and robustness, validating the inclusive design choices, and allowing for the correction of identified usability issues. The importance of this finding is also supported by prior work showing that early-stage testing with end users and other expert stakeholders is essential for designing VR tools that are functional and minimally confusing for children (Fowler et al. 2024).

During Phase 1, all mini-games scored high enjoyment (¿4.6), significantly above neutral and confirming educational appeal despite low VR experience in adults. *Block Sorting* excelled in clarity, followed by *Mental Calculation*, while *Column Counting* showed higher variability in user confusion, suggesting instructional refinements. These patterns align with VR mini-game studies where clear mechanics boost comprehension and motivation even across experience barriers (Agbo et al. 2023).

Moreover, the physical comfort and qualitative findings from Phase 1 highlight the VR application's success in minimizing cybersickness and addressing early usability issues through iterative design, supporting its readiness for educational deployment. This phase, therefore, bridged adult proxy testing and the target population, reinforcing the importance of design and co-design in the development of educational tools (Tobar-Muñoz et al. 2023).

After incorporating usability improvements identified in Phase 1, Phase 2 was conducted with children from third and fifth grade.

The in-class testing reveals *NumbersPlayRoom VR's* strong suitability as an engaging, accessible math tool for primary school children, with near-ceiling performance and enthusiasm across grades confirming its educational potential. In particular, Emoji Survey means (¿4.4) significantly outperformed the neutral value of 3, aligning with

VR literature showing immersive environments boost motivation and focus in 8-11-year-olds via embodied interaction (Kurniawan et al. 2019; Belter et al. 2023). *Block Sorting* and *Mental Calculation* excelled (third grade: M=4.92 Enjoyment, SD=0.28), while *Column Counting's variability* (fifth grade: SD=0.84 Ease of use) reflects age-appropriate challenge without deterring appeal. In addition, no significant differences emerged between third and fifth graders on *Block Sorting*, indicating broad usability of our serious game. Furthermore, the extra sessions demonstrated scalability: challenge increased perceived difficulty but sustained fun, suggesting adaptive difficulty for sustained engagement, as we will discuss in the following paragraph as a possible future line of research. The classroom testing with the target audience (N=33) provided preliminary evidence that the VR environment can support engagement and motivation (answering to **RQ1**). These results carry broader design implications for immersive educational tools.

An interesting point is the lack of correlation between enjoyment/ease of use and game metrics. This may indicate that children derive enjoyment from the intrinsic engagement of VR experience rather than from performance outcomes, supporting the tool's motivational value independently of skill level. Harris (2005), for example, found that high motivational scores in VR play among children were independent of performance, driven by intrinsic factors such as challenge and variability (not outcomes) (Harris and Reid 2005). This point remains interesting to explore in future studies, despite the numerical limitations of our data—which may have acted as a limiting effect (discussed in Section 6.1).

However, a contrast emerged between Phase 1 and Phase 2 (with children): voice guidance should be treated as a baseline design requirement rather than an optional enhancement, particularly for young or cognitively diverse learners. In line with Mayer's CTML, delivering instructions through the auditory channel reduces extraneous cognitive load and frees attentional resources for the mathematical task (Mayer 2005). Complementarily, our tests suggest that on-screen text instructions should be placed near the relevant visual cue or in a peripheral area within the virtual environment, ensuring they are available when needed without being intrusive (Su et al. 2022).

Moreover, the differential success of the three mini-games suggests that the closer a VR mechanic is to direct embodied manipulation of numerical concepts in space, as in *Block Sorting*, the more naturally it supports engagement and comprehension, consistent with theoretical accounts of spatial-numerical association (Belter and Lukosch 2022; Acevedo et al. 2024). This suggests that VR offers a uniquely suitable medium for grounding abstract numerical concepts in spatial, embodied experience, something that 2D screen-based tools structurally cannot replicate (Hornos-Arias et al. 2025). Conversely, tasks with higher procedural complexity, such as *Column Counting*, demand more careful interface design to prevent extraneous cognitive load from interfering with mathematical reasoning. Regarding navigation, our observations suggest that while teleportation effectively prevents cybersickness, often children (especially those with ADHD) showed a clear preference for continuous movement, perceiving it as more intuitive and less disruptive to their engagement. This points to the need for hybrid navigation solutions that preserve the comfort benefits of teleportation while offering the fluidity of continuous movement: for instance, through joystick-driven locomotion combined with discrete screen transitions (e.g., spatial "jumps" or brief fades to black) that prevent the visual-vestibular conflict responsible for motion sickness.

Qualitative observations revealed that children spontaneously treated free interaction with the virtual environment as an informal reward after task completion, expressing a desire to explore and manipulate objects beyond the structured exercises. Deliberately integrating such moments of free exploration and object manipulation after exercise completion could therefore serve as an intrinsic motivator, supporting engagement without compromising educational goals.

Promising results were observed with children with specific learning disorders. Teachers reported that the results observed in VR appeared to exceed the typical classroom performance of these students, though these observations remain qualitative and should be interpreted with caution. They reported sustained attention and engagement rarely seen with traditional teaching tools. Essentially, these students were able to complete complex tasks (such as *Column Counting*) with a surprisingly low error rate and a degree of autonomy that they often struggle to achieve in standard lessons. This strong interest from teachers in supporting children suggests that external feedback, as well as feedback provided during gameplay, can act as guiding factors that help children using VR headsets to complete tasks and remain engaged (Makransky et al. 2019).

These observations, combined with data collected from questionnaires, suggest that the VR environment may help reduce external distraction, allowing them to focus entirely on mathematical logic in a safe and playful space, also according to the recent literature studies (Su et al. 2022). This is particularly relevant in the context of mathematics-related anxiety: the immersive, distraction-controlled nature of the VR environment appears to create a low-pressure space where students feel free to experiment and make mistakes without the social exposure typical of traditional classroom settings (Belter and Lukosch 2022), a condition that may be especially beneficial for learners with specific educational needs (Broitman and Davis 2013).

These results highlight that virtual reality may represent a promising complementary tool in education and is usable by very young children, even with diverse learning needs (**RQ2**). By bridging the gap between cognitive theory and teaching practice, *NumbersPlayRoom VR* can promote an engaging, targeted and inclusive learning environment in which every student, particularly those with greater difficulties, can find support and motivation. As an exploratory investigation, this study provides the basis for a tool that is technically feasible, instructionally useful, and enjoyable for children.

6.1 Limitations

Despite the promising results, several limitations must be acknowledged. Firstly, the sample size is relatively small ($N = 33$) and was drawn from a single school using convenience sampling. This introduces potential statistical power limitations and selection bias that limit the generalizability of the findings beyond the specific educational and cultural context in which the study was conducted.

Moreover, this is an exploratory study and did not include a possible control group as a waiting list or other active groups involved in traditional activities or 2D games, nor did it include standardised pre- or post-assessments of numerical competence. Consequently, it is not possible to isolate the specific contribution of virtual reality to learning progress or to attribute any observed improvements in performance or engagement directly to the VR system itself.

The outcome measures regarding cognitive improvement were mainly observational and qualitative, based on teacher feedback, emoji-based self-reports, and in-game performance metrics. While these indicators provide useful insights into usability, engagement, and feasibility, they represent indirect proxies of learning and do not allow for a rigorous assessment of cognitive gains.

A possible methodological weakness is the fact that we employed the ad hoc Emoji Survey. This choice was motivated by the potential superior comprehension of visual/emotional cues for children with numerical difficulties, where traditional numeric Likert scales might introduce comprehension barriers. While ad hoc tools carry validation risks compared to established literature instruments (e.g., Read (2008)), future studies could employ different approaches in parallel (emojis or, for example, using a slider between two extremes) to directly compare and validate our emoji-based methodology against standardized measures, strengthening its generalizability. It is worth noting also that using emojis as a response scale, even if supported in literature (Massey 2022), can introduce response biases. Children may choose the "friendliest" or most positive-looking icon rather than the one that best matches their true answer, which is similar to an acquiescence effect (Kooijmans et al. 2022). And a second concern is interpretive difficulty. Emojis are not always understood in a consistent way, and emotionally loaded symbols may also add an extra layer of interpretation that could be harder to process for children with learning disabilities (Lesniak and Grzybowski 2021). In this sense, the qualitative findings from interviews and even more the in-game objective metrics are crucial for providing interpretive coherence and contextual consistency to our quantitative self-report results.

Although the system is designed to support children with numerical learning difficulties, the empirical sample included only a very limited number of students with certified learning disorders (see 5.2.5). This mismatch between the target population and the evaluated sample further constrains the ability to draw conclusions regarding the system's applicability to learners with specific educational needs.

These limitations reflect practical constraints, including restricted access to schools and limited time for experimentation, which prevented the implementation of a longitudinal design or controlled comparisons with alternative learning methods. Moreover, while no significant correlations were observed between enjoyment/ease of use and game performance metrics, this null finding may be attributable to two main limitations: a ceiling effect in task performance, where most participants achieved near-maximal scores, and the limited sample size, which reduced statistical power to detect potentially meaningful associations. Future studies should address these limitations by expanding the sample size and implementing Dynamic Difficulty Adjustment (DDA) to create more challenging, personalized learning experiences (Lucia et al. 2026; Lopes and Lopes 2022). DDA would enable adaptive learning paths that dynamically scale task difficulty according to individual performance, preventing ceiling effects and better discriminating between participants' abilities.

We also cannot exclude the possible role of the novelty effect, although the sustained engagement observed in the additional testing session provides preliminary evidence that motivation is not attributable solely to the initial exposure to the system.

Teacher training and device management represent practical barriers for full-sized classroom adoption of *NumbersPlayroom VR*. Scalability can therefore be considered a key limitation, as is common with emerging educational technologies (Soltani and Rostami 2025): devices may not be available in all schools or in sufficient numbers to cover entire classes, requiring time-based sharing of resources. In addition, teachers would need adequate training to effectively integrate such tools into their practice. In this regard, a future co-design study involving teachers could represent a valuable step, both to support their inclusion in the development process and to provide an initial form of training (Han et al. 2025). However, the use of a standalone platform such as the Meta Quest 2, which requires no external PC and represents a relatively low-cost entry point compared to more resource-intensive VR solutions, partially mitigates hardware cost concerns.

For these reasons, the findings of this study should be interpreted as exploratory and usability-oriented, rather than as evidence of learning efficacy through our VR mini-games. A further consideration concerns the use of HMD displays with school-age children, a topic of ongoing debate in the research community (Bexson et al. 2024; Kaimara et al. 2022). While our data indicate minimal physical discomfort, the long-term effects with young children remain insufficiently studied. The sessions were deliberately kept short and conducted under continuous supervision to mitigate potential risks. Future longitudinal studies should carefully monitor exposure duration and physical wellbeing alongside cognitive outcomes.

7 Conclusions and Future Works

This paper presents the design, development, and evaluation of *NumbersPlayRoom VR*, a serious game tailored to enhance numerical skills in primary school children, with a specific focus on inclusivity for learners with Developmental Dyscalculia (DD).

Future work will be dedicated to address some usability issues based on user feedback. Specifically, we will implement targeted HCI improvements to reduce extraneous cognitive load observed in the *Column Counting* task. These include refining the snap-to-grid mechanics, introducing subtle visual indicators to guide card movement and release, and adding dedicated visual slots for carry-over and borrowing operations.

We also plan to investigate a "hybrid" movement system to solve problems encountered with the teleportation system, which causes difficulties, especially to children with ADHD. One possible solution is a continuous movement system that uses transitions to black to hide movements to avoid motion sickness. Furthermore, children's desire to interact freely with objects suggests that incorporating more physically active mechanics could further increase motivation and engagement without compromising educational goals. To further support gamification, a reward system is also essential, and we plan to implement new mini-games that cover broader mathematical aspects and a more granular, adaptive difficulty system (useful for maintaining flow state).

Moreover, the limited number of children and the absence of a control group limit this study to an exploratory investigation, which provides the basis for a tool that is technically feasible, instructionally useful, and enjoyable for children. A longitudinal study with a larger number of children and a control group will allow us to obtain a clearer, quantitative picture of cognitive improvement and skill retention by comparing the current VR experience with traditional exercises.

Future studies could replicate this research while accounting for gender differences. We believe that implementing serious games for learning represents one of the first opportunities to bridge the gender gap in mathematics, while also serving as an initial approach to spark curiosity among young girls toward computer science.

Finally, to facilitate adoption in schools and collaboration between students and teachers, future studies will investigate the use of a companion app that will enable the exercises customisation and evaluation, in addition to real-time monitoring.

Acknowledgements. The authors would like to express their gratitude to the students and PhD candidates who collaborated on the project for their valuable assistance. Special thanks are extended to the Belluno primary school for their warm welcome and crucial cooperation, which made this study possible.

Declarations

- **Funding**: The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.
- **Competing interests**: The authors have no relevant financial or non-financial interests to disclose.
- **Ethics approval and consent to participate**: The study protocol and the testing sessions with the HMDs were authorized by the "Istituto Comprensivo Statale Belluno 3" (Belluno, Italy). Informed consent was obtained from the parents or legal guardians of all children involved in the testing sessions.
- **Consent for publication**: The "Istituto Comprensivo Statale Belluno 3" (Belluno, Italy) provided informed consent for the publication of the anonymized data and images included in this article.
- **Data availability**: The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.
- **Code availability**: The source code of the developed application is available from the corresponding author on reasonable request.
- **Author contribution**: All authors contributed equally to this paper.

Appendix A Questionnaires

Table A1: Appreciation Questionnaire Items.

ID	Item	Type/Notes
AQ1	I enjoyed the *Mental Calculation* game.	Likert Scale
AQ2	I couldn't clearly understand what to do in the *Mental Calculation* game.	Likert Scale
AQ3	I enjoyed the *Column Counting* game.	Likert Scale
AQ4	I couldn't clearly understand what to do in the *Column Counting* game.	Likert Scale
AQ5	I enjoyed the *Block Sorting* game.	Likert Scale
AQ6	I couldn't clearly understand what to do in the *Block Sorting* game.	Likert Scale
AQ7	I found the instructions within the games useful and well-integrated.	Likert Scale
AQ8	I found that the hints and feedback (e.g., in case of error) within the games were unclear.	Likert Scale
AQ9	I found the tutorial useful for learning the game mechanics.	Likert Scale
AQ10	I experienced a feeling of muscle fatigue while using the application.	Likert Scale
AQ11	I found the game environment fun and engaging.	Likert Scale
AQ12	I experienced dizziness or nausea while using the application.	Likert Scale
AQ13	I found that the game helped me remove distractions and focus on the goal.	Likert Scale
AQ14	Some studies shows a different biological perception of virtual reality between males and females. To what gender do you belong?	Selection (Male/Female)
AQ15	To which age group do you belong?	Drop Selection*

*AQ15 Possible Answers: 6-14, 15-27, 28-45, 46-65, Over 65

Table A2: Emoji Survey Items.

ID	Question	Polar Labels
EQ1	Did you like the *Mental Calculation* game?	No at all / Yes very much
EQ2	Was the *Mental Calculation* game difficult?	Very difficult / Very easy
EQ3	Did you like the *Column Counting* game?	No at all / Yes very much
EQ4	Was the *Column Counting* game difficult?	Very difficult / Very easy
EQ5	Did you like the numbered blocks game?	No at all / Yes very much
EQ6	Was the numbered blocks game difficult?	Very difficult / Very easy
EQ7	Would you like to use these games for math?	No at all / Yes very much
EQ8	Were the games fun?	No at all / Yes very much
EQ9	Are you feeling dizzy (head spinning)?	Yes very much / No at all

Emoji Scale: very angry face, angry face, neutral face, slightly smiling face, smiling face

Table A3: Emoji Survey Items - Extra Session.

ID	Question	Polar Labels
SEQ1	Did you like playing today as well?	No at all / Yes very much
SEQ2	Did the games seem more fun to you?	No at all / Yes very much
SEQ3	How did the games seem to you today?	Very difficult / Very easy

References

(2016) Applying the Coherence Principle: Adding Extra Material Can Hurt Learning, John Wiley & Sons, Ltd, chap 8, pp 151–178. https://doi.org/https://doi.org/10.1002/9781119239086.ch8

Acevedo P, Magana AJ, Benes B, et al (2024) A systematic review of immersive virtual reality in stem education: Advantages and disadvantages on learning and user experience. IEEE Access 12:189359–189386. https://doi.org/10.1109/ACCESS.2024.3489233

Agbo FJ, Olaleye SA, Bower M, et al (2023) Examining the relationships between students' perceptions of technology, pedagogy, and cognition: the case of immersive virtual reality mini games to foster computational thinking in higher education. Smart Learning Environments 10(1):16

Avila-Pesantez DF, Vaca-Cardenas LA, Delgadillo Avila R, et al (2019) Design of an augmented reality serious game for children with dyscalculia: A case study. In: Botto-Tobar M, Pizarro G, Zúñiga-Prieto M, et al (eds) Technology Trends. Springer International Publishing, Cham, pp 165–175, https://doi.org/https://doi.org/10.1007/978-3-030-05532-5_12

Bachot J, Gevers W, Fias W, et al (2005) Number sense in children with visuospatial disabilities: Orientation of the mental number line. Psychology Science 47:172–183

Belter M, Lukosch H (2022) Towards a virtual reality math game for learning in schools - a user study. In: 2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), pp 808–809, https://doi.org/10.1109/VRW55335.2022.00255

Belter M, Lukosch H, Lindeman RW, et al (2023) Exploring virtual reality (vr) to foster attention in math practice–comparing a vr to a non-vr game. In: International Conference on Human-Computer Interaction, Springer, pp 3–16

Berch DB, Foley EJ, Hill RJ, et al (1999) Extracting parity and magnitude from arabic numerals: Developmental changes in number processing and mental representation. Journal of Experimental Child Psychology 74:286–308. https://doi.org/https://doi.org/10.1006/jecp.1999.2518

Bertoni S, Franceschini S, Mancarella M, et al (2024) Action video games and posterior parietal cortex neuromodulation enhance both attention and reading in adults with developmental dyslexia. Cerebral Cortex 34. https://doi.org/10.1093/cercor/bhae152

Bexson C, Oldham G, Wray J (2024) Safety of virtual reality use in children: a systematic review. European journal of Pediatrics 183(5):2071–2090

Broitman J, Davis JM (2013) Overview of nvld. URL https://api.semanticscholar.org/CorpusID:146628163

Brooke J (1995) Sus: A quick and dirty usability scale. Usability Eval Ind 189

Chattha UA, Janjua UI, Anwar F, et al (2020) Motion sickness in virtual reality: An empirical evaluation. IEEE Access 8:130486–130499. https://doi.org/10.1109/ACCESS.2020.3007076

Cipora K, He Y, Nuerk HC (2020) The spatial–numerical association of response codes effect and math skills: why related? Annals of the New York Academy of Sciences 1477(1):5–19. https://doi.org/https://doi.org/10.1111/nyas.14355

Crollen V, Noël MP (2015) Spatial and numerical processing in children with high and low visuospatial abilities. Journal of Experimental Child Psychology 132:84–98. https://doi.org/https://doi.org/10.1016/j.jecp.2014.12.006

Daniela Lucangeli ICM (2010) Psicologia della cognizione numerica. FrancoAngeli, URL https://francoangeli.it/Libro/Psicologia-della-cognizione-numerica?Id=18406

Dehaene S (1992) Varieties of numerical abilities. Cognition 44(1):1–42. https://doi.org/https://doi.org/10.1016/0010-0277(92)90049-N

Dehaene S, Cohen L (1995) Towards an anatomical and functional model of number processing. Mathematical Cognition 1:83–120

DiFrancesca D, Nietfeld JL, Cao L (2016) A comparison of high and low achieving students on self-regulated learning variables. Learning and Individual Differences 45:228–236. https://doi.org/https://doi.org/10.1016/j.lindif.2015.11.010

Dowker A (2024) Developmental dyscalculia in relation to individual differences in mathematical abilities. Children 11(6). https://doi.org/10.3390/children11060623

Ferreira-Brito F, Fialho M, Virgolino A, et al (2019) Game-based interventions for neuropsychological assessment, training and rehabilitation: Which game-elements to use? a systematic review. Journal of Biomedical Informatics 98:103287. https://doi.org/https://doi.org/10.1016/j.jbi.2019.103287

Field A, Field Z, Miles J (2012) Discovering statistics using R. sage

Fowler LA, Vázquez MM, DePietro B, et al (2024) Development, usability, and preliminary efficacy of a virtual reality experience to promote healthy lifestyle behaviors in children: pilot randomized controlled trial. Mhealth 10:29

Gaggi O, Ciraulo F, Casagrande M (2018) Eating pizza to learn fractions. In: Proceedings of the 4th EAI International Conference on Smart Objects and Technologies for Social Good. Association for Computing Machinery, New York, NY, USA, Goodtechs '18, p 220–225, https://doi.org/10.1145/3284869.3284921

Golding J (2016) Chapter 27 - motion sickness. In: Furman JM, Lempert T (eds) Neuro-Otology, Handbook of Clinical Neurology, vol 137. Elsevier, p 371–390, https://doi.org/https://doi.org/10.1016/B978-0-444-63437-5.00027-3

Golding JF (2006) Motion sickness susceptibility. Autonomic Neuroscience 129(1):67–76. https://doi.org/https://doi.org/10.1016/j.autneu.2006.07.019, nausea and Vomiting: An Interdisciplinary Approach

Han H, Dawson KJ (2020) JASP (software)

Han X, Luo H, Wang Z, et al (2025) Using virtual reality for teacher education: a systematic review and meta-analysis of literature from 2014 to 2024. Frontiers in Virtual Reality 6:1620905

Hansberger JT, Peng C, Mathis SL, et al (2017) Dispelling the gorilla arm syndrome: The viability of prolonged gesture interactions. In: Lackey S, Chen J (eds) Virtual, Augmented and Mixed Reality. Springer International Publishing, Cham, pp 505–520

Harris K, Reid D (2005) The influence of virtual reality play on children's motivation. Canadian journal of occupational therapy 72(1):21–29

Hornos-Arias J, Grau S, Serra-Grabulosa JM (2025) Early detection and intervention of developmental dyscalculia using serious game-based digital tools: A systematic review. Information 16(9). https://doi.org/10.3390/info16090787

Jadán-Guerrero J, Avilés-Castillo F, Buele J, et al (2023) Gamification in Inclusive Education for Children with Disabilities: Global Trends and Approaches - A Bibliometric Review, pp 461–477. https://doi.org/10.1007/978-3-031-37105-9_31

Kaimara P, Oikonomou A, Deliyannis I (2022) Could virtual reality applications pose real risks to children and adolescents? a systematic review of ethical issues and concerns. Virtual reality 26(2):697–735

Kaufman EL, Lord MW, Reese TW, et al (1949) The discrimination of visual number. The American Journal of Psychology 62(4):498–525. URL http://www.jstor.org/stable/1418556

Kim HK, Park J, Choi Y, et al (2018) Virtual reality sickness questionnaire (vrsq): Motion sickness measurement index in a virtual reality environment. Applied Ergonomics 69:66–73. https://doi.org/https://doi.org/10.1016/j.apergo.2017.12.016

Klosterhalfen S, Kellermann S, Pan F, et al (2005) Effects of ethnicity and gender on motion sickness susceptibility. Aviation, space, and environmental medicine 76 11:1051–7. URL https://api.semanticscholar.org/CorpusID:7005762

Kooijmans R, Langdon PE, Moonen X (2022) Assisting children and youth with completing self-report instruments introduces bias: A mixed-method study that includes children and young people's views. Methods in Psychology 7:100102

Kurniawan C, Rosmansyah Y, Dabarsyah B (2019) A systematic literature review on virtual reality for learning. In: 2019 IEEE 5th international conference on wireless and telematics (ICWT), IEEE, pp 1–4

Leśniak E, Grzybowski SJ (2021) The role of emoticons in the comprehension of emotional and non-emotional messages in dyslexic youth–a preliminary study. Frontiers in Psychology 12:693287

Likert R (1932) A technique for the measurement of attitude. Archives of Psychology 140:1–55

Lopes JC, Lopes RP (2022) A review of dynamic difficulty adjustment methods for serious games. In: International Conference on Optimization, Learning Algorithms and Applications, Springer, pp 144–159

Lucia V, Eichhorn C, Johanna P, et al (2026) Dynamic difficulty adjustment in serious games: A literature review. Information 17(1):96

Makransky G, Lilleholt L, Aaby A (2017) Development and validation of the multimodal presence scale for virtual reality environments: A confirmatory factor analysis and item response theory approach. Computers in Human Behavior 72:276–285. https://doi.org/https://doi.org/10.1016/j.chb.2017.02.066

Makransky G, Terkildsen TS, Mayer RE (2019) Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learning and Instruction 60:225–236. https://doi.org/https://doi.org/10.1016/j.learninstruc.2017.12.007

Mammarella IC, Lucangeli D, Cornoldi C (2010) Spatial working memory and arithmetic deficits in children with nonverbal learning difficulties. journal of learning disabilities. Journal of Learning Disabilities

Massey S (2022) Using emojis and drawings in surveys to measure children's attitudes to mathematics. International Journal of Social Research Methodology 25(6):877–889

Mayer RE (2005) The Cambridge Handbook of Multimedia Learning. Cambridge University Press

Mikropoulos TA, Natsis A (2011) Educational virtual environments: A ten-year review of empirical research (1999–2009). Computers & Education 56(3):769–780. https://doi.org/https://doi.org/10.1016/j.compedu.2010.10.020

Munafo J. STDiedrick M. (2017) The virtual reality head-mounted display oculus rift induces motion sickness and is sexist in its effects. Exp Brain Res 235:889–901. https://doi.org/https://doi.org/10.1007/s00221-016-4846-7

Noro M, Gaggi O (2025) Virtual reality to enhance mathematical skills in children with learning difficulties: design and evaluation of numbersplayroom vr. In: Proceedings of SENSE-XR 2025: The Feeling of Virtual: XR, Haptics, and 3D Graphics, at the CHItaly conference, p 220–225

Parker C, Scott S, Geddes A (2019) Snowball sampling. SAGE Research Methods Foundations https://doi.org/10.4135/

Perttula A, Kiili K, Lindstedt A, et al (2017) Flow experience in game based learning – a systematic literature review. International Journal of Serious Games 4(1):57–72. https://doi.org/10.17083/ijsg.v4i1.151

Read JC (2008) Validating the fun toolkit: an instrument for measuring children's opinions of technology. Cognition, Technology & Work 10(2):119–128

Read JC, Horton M (2025) Using the smileyometer to measure ux with children. Interacting with Computers p iwaf016

Rose D, Meyer A, Gordon D (2013) Universal Design for Learning: Theory and practice.

Shalev RS, Gross-Tsur V (2001) Developmental dyscalculia. Pediatric Neurology 24(5):337–342. https://doi.org/https://doi.org/10.1016/S0887-8994(00)00258-7

Simons DJ, Chabris CF (1999) Gorillas in our midst: Sustained inattentional blindness for dynamic events. perception 28(9):1059–1074

Soltani N, Rostami A (2025) Immersive learning at scale: Exploring the feasibility of vr in education. In: Proceedings of the 2025 ACM International Conference on Interactive Media Experiences, pp 49–59

Su YS, Cheng HW, Lai CF (2022) Study of virtual reality immersive technology enhanced mathematics geometry learning. Frontiers in Psychology Volume 13 - 2022. https://doi.org/10.3389/fpsyg.2022.760418

Tobar-Muñoz H, Baldiris S, Fabregat R (2023) Co-design of augmented reality games for learning with teachers: A methodological approach. Technology, Knowledge and Learning 28(2):901–923

Wang WS, Lee H, Lin C, et al (2024) Enhancing students' learning outcomes in self-regulated virtual reality learning environment with learning aid mechanisms. British Journal of Educational Technology URL https://api.semanticscholar.org/CorpusID:271630249

Xu (2022) The impact of self-regulated learning strategies on academic performance for online learning during covid-19. Frontiers