Grasping Derivatives: Teaching Mathematics through Embodied Interactions using Tablets and Virtual Reality

Conference Paper

Author(s):
Chatain, Julia; Ramp, Virginia; Gashaj, Venera; Fayolle, Violaine; Kapur, Manu; Sumner, Bob; Magnenat, Stéphane

Publication date:
2022-06

Permanent link:
https://doi.org/10.3929/ethz-b-000543892

Originally published in:
https://doi.org/10.1145/3501712.3529748

Grasping Derivatives: Teaching Mathematics through Embodied Interactions using Tablets and Virtual Reality

Julia Chatain*†, Virginia Ramp*, Venera Gashaj†, Violaine Fayolle*
Manu Kapur†, Robert W. Sumner*, Stéphane Magnenat*
julia.chatain@inf.ethz.ch
ETH Zurich, *Dept. of Computer Science, †Dept. of Humanities, Social and Political Sciences
Switzerland

Figure 1: Three different ways of learning derivatives using embodied interaction: *direct-embodied* interaction on tablet, *direct-embodied* interaction in Virtual Reality, and *enacted* interaction in Virtual Reality

ABSTRACT

Grasping mathematics can be difficult. Often, students struggle to connect mathematical concepts with their own experiences and even believe that math has nothing to do with the real world. To create more concreteness in mathematics education, we focus on the role of the body in learning, and more specifically, embodied interactions for learning derivatives. In this project, we designed an embodied game to teach derivatives, and validated our design with a panel of experts. We then used this prototype to explore different embodied interactions in terms of usability, sense of embodiment, and learning outcomes. In particular, we evaluated different degrees of embodied interactions, and different types of embodied interactions in Virtual Reality. We conclude with insights and recommendations for mathematics education with embodied interactions.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**; **Empirical studies in HCI**; **Virtual reality**; *User studies*; *Gestural input*.

IDC '22, June 27–30, 2022, Braga, Portugal

ACM ISBN 978-1-4503-9197-9/22/06…$15.00
https://doi.org/10.1145/3501712.3529748

KEYWORDS

embodied interaction; embodied cognition; embodiment; mathematics education; virtual reality; problem solving followed by instruction

ACM Reference Format:

Julia Chatain, Virginia Ramp, Venera Gashaj, Violaine Fayolle, Manu Kapur, Robert W. Sumner, Stéphane Magnenat. 2022. Grasping Derivatives: Teaching Mathematics through Embodied Interactions using Tablets and Virtual Reality. In *Interaction Design and Children (IDC'22), June 27–30, 2022, Braga, Portugal.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3501712.3529748

1 INTRODUCTION

Because of its abstract nature, mathematics is a difficult topic to teach and learn. These difficulties can lead to students developing math anxiety: a feeling of panic arising when solving a mathematical problem or managing numbers [3]. Math anxiety affects children and increases until adulthood [11, 15, 61], with women reporting more math anxiety than men [4]. Another issue with the way we teach math is that it can lead students to hold wrong beliefs about mathematics that can impair their learning [52]. For example, students believe that mathematics has little to do with the real world, or that ordinary students cannot expect to understand mathematics and should memorize it instead [52]. The US National Research Council explained that mathematics education should focus on "seeking solutions, not just memorizing procedures", "exploring patterns, not just memorizing formulas", and "formulating conjectures, not just doing exercises" [12]. In turn, Papert explained that, as the best way to learn French is still to

Table 1: The four degrees of the Embodied Education Taxonomy [33]. The degrees explored in our work are highlighted.

Degree	1	2	2	2	3	3	3	4
Sensorimotor engagement	Low	High	Low	Low	Low	High	High	High
Gestural congruency	Low	Low	High	Low	High	Low	High	High
Immersion	Low	Low	Low	High	High	High	Low	High

spend a year in France, the best way to learn mathematics might be to spend a year in a so-called "MathLand" and interact with mathematical objects directly [49]. Moreover, modern approaches such as Realistic Mathematics Education invite designers to consider a bottom-up approach to instruction: starting from the "informal mathematical activities of the students" and identifying support symbolizations [23, 24].

Technological advances, such as Virtual Reality, enable us to build "MathLands" where students can manipulate and explore mathematical objects and concepts. The motivation of our work is to provide approaches to de-abstract mathematics and help address students' misconceptions. We ground our work in two frameworks. First, *embodied cognition* is the idea that cognition transcends the Cartesian divide between mind and body, and that pedagogical activities should acknowledge the primordial role of the students' bodies[1] in learning [2]. Second, *embodied interaction* is the idea that the users' bodies should be at the core of the interaction with digital content [14]. Despite growing evidence for the benefits of embodied learning and digital experiences, the influence of design choices is underexplored.

Applying embodied cognition and interaction, we design a game to teach derivatives to high-school students. With a panel of experts, we validate our prototype and identify the strengths of such an embodied activity. Through a quantitative study, we compare degrees and types of embodied interaction and their impact on usability, sense of embodiment, and learning. Through this mixed-methods approach [13], we contribute with design recommendations as well as quantitative evidence for the appropriate degree and type of embodied interaction.

2 RELATED WORK

Our work takes root in two major fields: embodied cognition and embodied interaction. In this section, we give an overview of the theory and impact of these paradigms, and detail the frameworks grounding our work.

2.1 Embodied Cognition

Embodied cognition rejects the Cartesian notion that the body and the mind are separate entities [2]. From this perspective, thinking is simply a form of truncated action, and considering thinking or learning without considering the bodies of the learners is a fallacy, or at least, incomplete [45]. Although mathematics is often thought as disembodied, experts argue that mathematics is grounded in "situated, spatial-dynamical, and somatic phenomenology" [1, 38].

The effectiveness of embodied approaches to learning resides in several aspects. Indeed, using their bodies to learn math can help students discover and develop intuitions about a topic without first having to familiarize themselves with abstract formalisms. It can also alleviate some of the cognitive load by anchoring part of the information in the spatial context, and connect the concepts to the tangible world to make them more concrete [36, 50, 57]. Moreover, studies have shown a developmental link between motor skills and mathematical abilities [19, 20], indicating a deeper importance of learners' bodies for mathematics.

The potential of embodied approaches for mathematics education has already been explored in various ways. Nathan et al. designed a game to teach students how to represent certain mathematical objects with gestures [47]. They found that teaching gestures to anchor their cognitive process helps students perform better mathematical proofs. Similarly, Howison et al. showed how an embodied activity can support sense-making and conceptual understanding when learning about proportions [29]. The use of concrete physical manipulatives has also shown to improve learning in terms of retention, problem solving, transfer, and justification [8]. More generally, embodied approaches have been used to teach integrals, exponential growth, slopes, and many other mathematical topics [57].

2.2 Embodied Interaction

When defining embodied interaction, Dourish emphasizes that the physical and social aspects of reality cannot be discarded when designing interaction [14]. Moreover, he explains that "embodied interaction is the creation, manipulation, and sharing of meaning through engaged interaction with artifacts". Dourish recommends engaged practice, as opposed to "disembodied rationality", and insists on meaning creation. Thus, embodied interaction encompasses physical as well as social aspects and focuses on engaged practice and making sense of purpose: all aspects that are missed in classical mathematics education.

In our project, we involve the users' physical bodies in meaning-making interaction, in the playful context of a game. When designing interactions for human bodies, there are different approaches to consider. For example, somaesthetic appreciation design focuses on the role of bodies not only in the interaction, but also in the design process. It offers different concepts to consider, in particular to turn the users "inwards, towards their own body" [26, 27]. When designing embodied interaction for games there are two different perspectives: the flesh body (*Körper*) and the feeling body (*Leib*) [46]. Often, the users' bodies are only considered as physical entities utilized to push buttons, and not as feeling beings. Considering the latter changes the result of the design; designing a validation button for the *Körper* will result with a button located next to the resting position of the user's hand, avoiding large movements. However, designing the same button for the *Leib* might mean placing the button up high, inviting the user to adopt a winning pose [46].

The potential of embodied interaction has already been explored in several fields [43]. Chatain et al. brought the users' bodies to the

[1]In order to account for the diversity of bodies, we use plural [55].

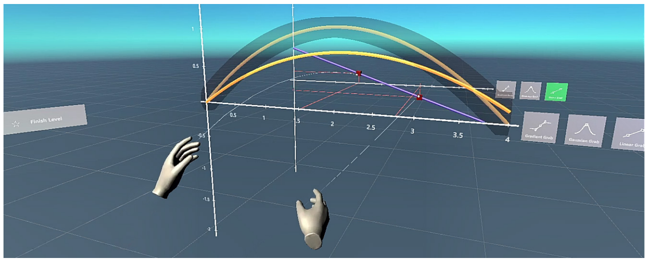

Figure 2: Example of a level from our first prototype, seen from a third-person perspective. The player manipulates the derivative curve (purple, in the back) in order to fit the function curve (yellow, in front) in the target area (grey).

core of the digital experience by using their hands as a controller and a display, both for games and playful educational activities [9]. Trajkova et al. used embodied interaction for data exploration in a museum context and found that the representation of the bodies impacted the gestures used to interact with digital content [56]. Focusing specifically on learning, Lindgren and al. demonstrated the benefits of embodied interaction for physics education in terms of learning, engagement, and positive attitude [39].

2.3 Frameworks

We use the term "embodiment" to describe embodied cognition supported by embodied interaction. We focus on congruent embodiment as defined by Johnson-Glenberg and Megowan-Romanowicz: "Embodied [...] means that the learner has initiated a physical gesture or movement that is well-mapped to the content to be learned" [32].

In this context, there are different ways of classifying embodied approaches:

Johnson-Glenberg and Megowan-Romanowicz defined four *degrees* of embodiment [33]. This taxonomy follows three axes: (1) amount of sensorimotor engagement, (2) congruency of the gestures to the content learned, (3) amount of immersion experienced by the user (Table 1). Importantly, different degrees of embodiment are often implemented with different technologies. To achieve higher immersion and sensorimotor engagement, Virtual Reality (VR) is a good solution, while lower degrees of embodiment are achieved with screens and tablets. Dourish explains that "Embodiment is not a property of systems, technologies, or artifacts; it is a property of interaction" [14]. However, as these interactions do happen within a technological context, it is important to understand how this impacts the meaning-making capabilities of the interaction. Indeed, there is a trade-off between implementing stronger embodiment with a more cumbersome technology, in particular in a classroom, and implementing weaker embodiment with a technology that is less space- and time-consuming, supports collaboration, and gives a better overview of the task.

In turn, Melcer and Ibister introduced a theoretical framework of embodied cognition [44, 48]. The "physicality" dimension of the framework describes five main *types* of embodiment: *direct-embodied* (body-centered), *enacted* (body-in-action), *manipulated* (object-centered), *surrogate* (object-in-action), and *augmented* (environment-centered). However, it is not clear how to select the proper embodied approach to teach a specific topic, in particular when several solutions are congruent with the topic at hand. For derivatives, several approaches involving bodily actions are relevant: We can select a *direct-embodied* approach to focus on derivatives as slopes, or an *enacted* approach to focus on derivatives as variations. Such a design decision might have a wider impact in terms of usability as an *enacted* approach is more indirect than a *direct-embodied* approach.

In our work, we present the design process of an embodied activity to teach derivatives to high-school students, and offer recommendations based on a quantitative study to advise educators and designers on how to address the trade-offs across degrees of embodiment and types of embodiment.

3 DESIGN

3.1 Initial design

We implemented a VR game supporting exploration and intuition-building of the derivative concept through embodied interaction. For this first prototype, we focused mainly on interaction and level design, from the perspective of embodied cognition [2]. As recommended in the *activities* design guidelines [1], we used no symbolic stimuli for this activity and focused on graphical representations. Each level of the game displays two curves (Figure 2): one curve represents the function (in front, in yellow), and one curve represents its derivative (in the back, in purple). The level also displays a target curve and a target area related to one of the curves (yellow/purple and grey).

In order to pass the level, the player needs to manipulate one curve to put the other curve in the target area. For example, on Figure 2, the player has to manipulate the derivative curve in order to give a bell shape to the function curve. Once the curve is in the target area, the "Finish Level" button turns green and the player can validate their solution. We provide percentage accuracy outcome-feedback computed according to the distance between the player's proposed curve and the target curve [31]. If they are perfectly aligned, this score is 100 %.

We explored the interaction space through three embodied interaction modes (Figure 3), focusing on congruent gestures [32]. With the *linear grab* mode, the player grabs the curve with one or two hands, and all the points of the curve are moved along the y-axis, by

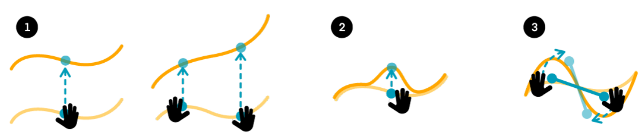

Figure 3: The three grab modes implemented in our prototype. (1) The *linear grab* mode applies a linear transformation corresponding to the hand movements. (2) The *Gaussian grab* mode adds a Gaussian shape to the selected node according to the hand movement. (3) The *gradient grab* mode modifies the local value of the derivative according to the slope between the two hands.

Table 2: Design Evaluation: Experts' profiles

ID	Age	Gender	Expertise	Math. training	Math. affect	Sports	VR
P1	25-29	F	Embodiment and haptic feedback	College (secondary)	High	Medium	Low
P2	25-29	M	Math learning (math graduate)	College (main)	High	High	Medium
P3	25-29	F	Math education (doctoral student)	College (main)	High	Medium	Low
P4	25-29	M	Art and game development	College (secondary)	Low	Low	High
P5	35-39	M	Embodied cognition (mathematics)	College (main)	High	Low	Low
P6	30-34	F	Cognitive sciences (children and adolescents)	High-school	High	High	Low

an offset linearly interpolated between the offsets of each hand. If only one hand is used, the curve is translated along the y-axis. With the *Gaussian grab* mode, the player moves a point of the curve up and down. The neighboring points are moved following a smooth Gaussian shape. Finally, with the *gradient grab* mode, the player manipulates the slope at a specific point by rotating their hands to the desired slope value.

Following the *materials* and *facilitation* guidelines [1], we start by immediate environmental outcome-feedback loops [31]. In the *normal levels*, the curve is updated in real time, as the player manipulates it. To help the students evaluate their understanding we also offer *delayed-feedback levels* where the curve is only updated once the player releases it. This way, the player can use the *normal levels* to explore the relationship between the curves and gain intuition, and is then invited to think deeper during the *delayed-feedback levels* as the interaction is less direct.

Finally, to help the player connect their interaction to numerical values, we added axes for each curve. When the player selects a point, the projected rays of the point towards the x and y axes are displayed in red (Figure 2).

3.2 Design Improvement

We invited six experts to evaluate our prototype, individually (Table 2). Each expert tried the game for 15 minutes and was invited to speak about their experience while trying out the game (*think-aloud* comments). We then conducted a semi-structured interview with each expert where we asked general questions about the game and questions related to their specific expertise. We analysed the transcripts using an inductive thematic approach [6], to identify areas of improvement. From the experts' comments, we improved our design in several ways, detailed in the following sub-sections. We discuss our decisions in relation to the design principles for embodied interaction in VR [32].

3.2.1 Embodied Interaction. Overall, participants found the interaction with the curves intuitive: "It is very self-explanatory, it works, it is nice" (P2), "It was really easy to use, even for inexperienced people" (P1), "The actual interaction felt fine, there wasn't anything unnatural" (P5). However, sometimes is was not clear to them which curve they should interact with, nor which grab mode was activated: "I just was not sure which lines I was able to move, what tool do I have?" (P2). To address these issues, we focused on only one mode of interaction, the gradient grab, and only one curve to manipulate, the function curve.

Regarding the embodied part of the interaction, participants found it enjoyable: "It feels good when I change stuff" (P2), "I felt very comfortable" (P1). P5 found the *normal levels* more enjoyable than the *delayed-feedback levels*: "I just liked moving them and seeing a response. [...] When it was not being updated, it was not like I disliked it, but I did not get that sort of kinesthetic enjoyment out of it". P5 also enjoyed large movements more than restricted movements: "It was a more enjoyable thing to stretch out more". To include this in our design, we implemented the interaction techniques without any restriction on movement amplitude.

Moreover, embodied interaction creates a hands-on experience with mathematical concepts: "If you engage the whole body, you are automatically more engaged [...], just because you have this experience of being there with the curves. You are immediately closer to the topic" (P6), "Because I was really moving the [curve], I appropriated the curves to what I was doing and I learned that there are links" (P4). This way, the mathematical objects are manipulable and perceptible, and, therefore, concrete. However, participants would have preferred an even more direct interaction with the curves: 'I could not do something the way I wanted to, because I could not really 'grab' the curve" (P3), "But to really feel embodiment I would need to really move the things without any distance" (P4). We addressed this by using hand-tracking over controllers "for active, body-based learning" [32], without any physical distance between the user and the curves, and we included a skin color selection panel.

Finally, the interview with P5, an embodied cognition expert, revealed that an amount of desirable difficulty in the embodied interaction can actually benefit learning: "If you're trying to create a good user interface, then you want to make it seamless, but if you're trying to get people to learn, then it oftentimes helps to throw in some difficulty or something that makes them think". He also mentioned that the *delayed-feedback levels* play in that direction: "For example, the fact that the line was not updating is good for that. Even though, personally, it did not make me feel good, that is not a bad thing". He also suggested several ideas in that area, such as making only certain parts of the curve manipulable, or restricting the movements. Indeed, some difficulty such as lack of feedback can be beneficial to learning [5, 18]. Considering desirable difficulties in our prototype, and aligning with the "Use guided exploration" design principle [32], we reduced the interactability of the curve to a set of specific points, defined per level, and composed of the minimum amount of points necessary to define the curve. However, we decided to stay in alignment with our previous findings and not reduce the movements.

3.2.2 Mathematical Understanding. Overall, our panel of experts agreed that our activity is a novel and interesting approach to mathematics that helped them sharpen their intuition, through exploration: "You can, in a fun way, gain intuition and see" (P2), "It's a cool new dimension that I didn't know before, It was great to see this connection directly" (P3), "It makes it less like a recipe and more like the gradient actually has something to do with how the function looks" (P1), "I gained some sense of quantity of difference: 'If I do about this much to this line, the other line should move by about this much' " (P5). P3, mathematics education expert, mentioned that such an approach could be useful for students: "High school students would benefit from introducing the first derivation to sharpen their intuition, but also people in first years of college to get a different approach than only formulas and rules." We reinforced this exploratory approach by adding a short text at the beginning of the activity, inviting the students to explore the relationship between the two curves. We "minimized text reading" [32].

To solve the problems, participants used strategies focused on intuition or trial and error: "I could still call up certain intuitions" (P3), "I don't have any tactics. I just like how it feels" (P1), "If something did not work, I would try something else immediately" (P4). *Delayed-feedback levels* invited some participants towards deeper reasoning: "You do not just try things out but rather you have to think about it" (P3), "They were important, because they made me realize my difficulties" (P1), "I could have still done trial and error. It would just take longer and be less satisfactory" (P5), "I gained some intuition, which I then tried to apply on those *delayed-feedback levels*" (P2). We kept this mechanism in our new prototype, adding *delayed-feedback levels* at the end of each section, as a mean to "design in opportunities for reflection" [32], and align with the need for desirable difficulties previously identified.

Finally, we also identified the need to reconnect our activity with a more formal or traditional form of instruction: "It would be optimal if you connect it with the underlying theories" (P2), "Once the students sharpened their intuition, you can say 'Yes, but what does that mean now?' " (P3), "It should [...] have another kind of learning in the session. [...] You want people to learn the logic, and not guessing" (P4). Having a phase of exploration followed by instruction is a well-known pedagogical pattern, more generally called PS-I for "Problem Solving followed by Instruction", that has shown great potential for mathematics education [54]. This approach relies on three main mechanisms: activation of prior knowledge, awareness of knowledge gap, and recognition of deep features [40]. Our activity seems particularly suitable for this approach: participants can connect to the exercise, as well as identify knowledge gap in the *delayed-feedback levels*. In our final design, we integrated the activity in a PS-I pedagogical pattern by adding an instructional video after the activity.

3.2.3 Interface Design. We identified several issues related to interface design. Several participants mentioned that the grid was difficult to use, and, even difficult to see: "Having clear numbers there would be nice" (P2), "I did not even see the grid" (P6). We improved the readability of the grid: we made the unit graduations more visible, and highlighted the exact values corresponding to the selected points.

The participants also mentioned that the positioning of the curves creates occlusions: "Sometimes you cannot see the second line because of the front line" (P2), "It is a bit unfortunate sometimes that the two graphs were behind one another" (P2). We resolved this issue by placing both curves in the same plane, and adding a mini-display to provide an overview of the level.

Finally, P4, digital artist, mentioned that the visuals should be improved to be more appealing and attractive: "It is always a challenge to make math appealing. [...] Maybe visuals that could be a bit more enjoyable, you could have something more colourful." With this aim in mind, we need to also be mindful about our color choices, as the purple was difficult to see for P3: "I find the purple line at the back difficult to see". We improved our prototype by designing a colorful VR room with windows, plants, and we changed the colors of the curves to yellow and pink.

3.2.4 Virtual Reality. When designing embodied activities in VR, it is important to remember that VR is novel and requires an adjustment period: "Assume every learner is a VR newbie—start slow" [32]. Adding this on top of some potential math anxiety might also make the experience overwhelming for some users [41]. This was reflected in several comments: "I am struggling way more with the technology than with the task itself, I am just inexperienced" (P1). In particular, P6, who has very little experience with VR and game controllers, felt overwhelmed by the system: "I was too focused on everything that was so new to me. I was also focused on the... what are they called... the controllers. [...] Because [of that], I felt I could not do this".

To some extent, VR can even restrict the users' movements: "Because I do not have a lot of experience with VR, I am very careful when moving because I do not know if I am going to hit anything." (P1), "I just did not want to walk into something." (P5). Beyond the risk of colliding with the real world, the imprecision of the tracking can also impact the experience: "It is pretty difficult to aim easily" (P4), "With keyboard and mouse as input, we are much more precise. And precision is good in some aspect, you can reach 100 % at every level if you are exact" (P2).

To mitigate these issues, we added a tutorial to our activity where participants can explore the VR space and grow trust for the digital boundary appearing as they reach the limit of the space. We also included an interaction tutorial where users can get familiar with hand tracking. Moreover, to improve precision, we selected a pinching gesture, over a grabbing gesture, for curve manipulation.

Finally, all experts agreed that the use of VR benefited their experience and connection with the mathematical content, in particular compared to a screen or a tablet activity: "I think that Virtual Reality is just more similar to reality than a tablet, or a computer screen and a mouse" (P6), "It was really cool to have the whole room and to see the curve in front of you, and not only in the screen. It felt like I was really there with the curve" (P6), "I would prefer [VR] over [a tablet] because there is some feeling here, which is not just like pointing, it is also grabbing" (P5).

3.3 Final design

In this final prototype, the game goes as follows: First, the participant explores the VR environment to understand the space and feel safe. Second, the height and the hand size of the participant are calibrated. Third, the participant picks their natural hand color across 12 different tones, and, finally, the activity starts (Figure 4).

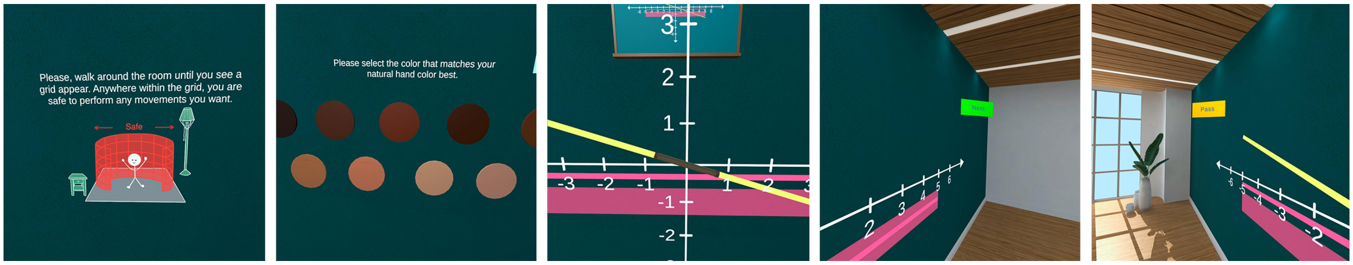

Figure 4: Different steps of the final activity in first person view: Familiarization with VR space, skin color selection, level solving, "Next" button to go to the next level, "Pass" button appears after one minute.

The activity contains a prompting text to explore the relationship between the yellow and pink curves, as well as a tutorial level with animated hands demonstrating the interaction technique, followed by 21 levels. In each level, the participant manipulates the function curve (yellow) to move the derivative curve (pink) into the target area (pink). The function can be manipulated at specific points (wooden handle). The resulting curve is approximated using constrained cubic splines [37], for smooth interpolation, and small movements having small effects on the interpolated curve.

After level completion, a "Next" button — positioned above the user, to provide an embodied interaction focused on the feeling body (*Leib*) [46], on the right side, to align with wide-spread interaction paradigms — can be pressed to proceed. One minute into a level, a "Pass" button appears. The levels progress in difficulty, focusing on different topics. Each topic contains several *normal levels* and finish with *delayed-feedback levels*. On a mini-display, the user can keep track of the level and score, which represents how close the manipulable points of derivative are to the target derivative.

4 QUANTITATIVE USER STUDY

After having shown the potential of our prototype for teaching derivatives with embodied interaction, we designed a study to quantitatively answer several research questions:

- How do different embodied interactions compare in terms of usability and resulting manipulations?
- Which embodied interaction brings the greatest sense of embodiment and sense of agency?
- How do different embodied interactions influence learning outcomes?

4.1 Embodied Interactions

Addressing our research questions, we consider the degree and type of embodiment. According to the embodiment matrix (Figure 5), for the type of embodiment, we compare the conditions direct embodiment on tablet (*TAB*, left on Figure 1) to direct embodiment in VR (*DIR*, center on Figure 1). To compare the type of embodiment, we compare the *DIR* condition to the *enacted* embodiment in VR condition (*ENA*, right on Figure 1).

The degree of embodiment. We compare low (degree 2/tablet) to high (degree 4/VR) embodiment (first row on Figure 5). We expect students in the lower embodiment condition to experience lower sensorimotor engagement due to gestures of a smaller amplitude (pointing versus grabbing), partial body engagement, as well as reduced immersion due to a limited coverage of the field of view.

The type of embodiment. Similarly, we compare bodily action: In the *direct-embodied* condition, the position of the user represents the derivative, while in the *enacted* condition, the movement of the user represents the derivative.

In the *direct-embodied* interaction, the user holds a proxy of the slope of the curve and manipulates it to influence the derivative. The slope between the user's hands represents the local slope of the curve, that is, the derivative. This approach emphasizes the derivative as a slope. For the *enacted interaction*, the user draws the desired slope by hand. The hand movement thus describes the derivative. This approach emphasizes the derivative as a variation. Both conditions are illustrated on the last column of Figure 5.

4.2 Demographics

We recruited 40 public high school students, from two different classes taught by the same teacher, who chose to be taught in English. No participant was repeating the class. Two students dropped out, resulting in a final sample of $n = 38$ students with a mean age of $M = 17.6$ ($SD = 0.61$) and 21 different mother tongues. 19 students identified as male, 18 as female, and 1 as other. The study took place a couple of weeks before the lesson on derivatives and the students knew how to read functions' graphs. The study was conducted in English and at the schools. Due to health regulations, face masks were mandatory at all times. The students received a financial compensation for their participation.

4.3 Protocol

We followed an in-between experimental design to avoid carry-over effects and fatigue effects. The study took place during class-time, at the school, in a room large enough to host the VR spaces (2.5 m * 1.7 m each). A preparatory intervention (20 min), and a PS-I intervention (1 h) were conducted on different days, with 1-7 days in between to avoid fatigue effects.

In the preparatory intervention, the students filled out several questionnaires followed by a 10 minutes VR game, "*Elixir*", heavily focused on hand tracking [42]. Through the questionnaires we obtained information about prerequisites (reading functions' graphs, identifying minimum and maximum on the graph of a function, and reading coordinates of vectors in the 2D plane), demographics, math anxiety [28], and body awareness [53].

After the preparatory intervention, we randomly assigned students to the conditions and balanced for prior knowledge, math grade, gender, VR experience, math anxiety, and body awareness. 13 participants were assigned to the *TAB* condition (7 male, 6 female), 12 to the *DIR* condition (6 male, 6 female), and 13 to the *ENA* condition (6 male, 6 female, 1 other).

During the PS-I intervention, the students filled in a Simulator Sickness Questionnaire (SSQ) [35]. Subsequently, they performed the derivative activity, either on tablet or in VR according to their condition. Then they filled in the same SSQ, and a System Usability Scale (SUS) questionnaire [7]. The participants in the *DIR* and *ENA* conditions filled in a Sense of Embodiment questionnaire, adapted to focus on the hands [51]. All participants filled in a questionnaire about felt agency on the mathematical objects adapted from an Avatar Embodiment questionnaire [22], and their usage of the tool. After this, the students watched an instruction video about derivatives, using the same color scheme as the exploratory activity, and recorded by an English native speaker. Then the students took a 5 minutes break where they could read comics, in order to avoid fatigue effects. Finally, the participants solved a post-test evaluating their understanding of derivatives, and a selection of questions on first derivatives from the Calculus Concept Inventory (CCI) [16]. The post-test focused on specific properties of the derivative and was presented in a visual style, while the CCI required to combine several properties and resembled classical math problems.

The tablet intervention was conducted on Apple iPad 5th Gen 32GB 9.7", and the VR interventions on Oculus Quest 2. During the activity, we logged general information about the participants (height, hand size, skin color), time to level completion, level completion or skipping, and manipulations the mathematical objects.

The implementation and study design were validated through a pilot study with 19 high-school students.

4.4 Results and analysis

The degrees (*TAB* and *DIR* conditions) and types (*DIR* and *ENA* conditions) of embodiment were compared using independent Welch

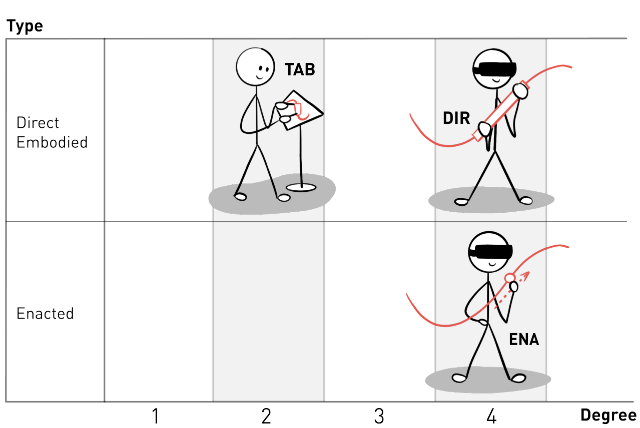

Figure 5: Embodiment matrix. The degree is compared along the horizontal axis, and the type is compared along the vertical axis.

t-test or Yuen test [60] following the results of the Shapiro-Wilk normality assumptions check (Table 3).

4.4.1 Usability and Resulting Manipulations. There was a significant difference in duration in degree of embodiment with a very large effect size. In average, it took $M = 9.27$ ($SD = 2.00$) minutes for the participants in the *TAB* condition to solve all the levels, compared to $M = 16.42$ ($SD = 3.45$) in the *DIR* condition. Similarly, a significant difference was found concerning duration in types of embodiment with a large effect size. In average, participants in the *ENA* condition took $M = 23.22$ ($SD = 3.45$) minutes to complete all the levels.

Regarding the number of manipulations with the curve, we found no significant difference across the degrees of embodiment with a medium effect size. However, we found a significant difference between the types of embodiment with a very large effect size. Participants in the *DIR* condition interacted $M = 133$ ($SD = 25$) times in average, while participants in the *ENA* condition interacted $M = 246$ ($SD = 83$) times in average. In the *direct-embodied* conditions, the participants usually grabbed the handle and adjusted until satisfaction. On the other hand, people in the *enacted* condition often released the knob and tried again. This difference also explains the duration difference.

Regarding usability, we computed the SUS scores for each condition: *TAB* scored 68 ($SD = 12$), *DIR* scored 69 ($SD = 12$) and *ENA* scored 62 ($SD = 16$). We noticed that the first question of the SUS "I think that I would like to use this system frequently" scored rather low ($M = 2.83$, $SD = 1.17$) because students did not necessarily want to study math frequently altogether. Therefore we should refrain from comparing these scores to general SUS scores. Comparing the degrees of embodiment, we expected the *DIR* condition to be less usable than the *TAB* condition because of the cumbersomeness of the VR hardware, and the limited accuracy of the hand tracking. However, the t-test was not significant, therefore we could not reject the null hypothesis of no effect. Moreover, the effect size was very small. When comparing the types of embodiment, we expected the *ENA* condition to be less usable than the *DIR* condition as the curve should be read left to right but enacting the slope in that direction with the left hand would cover the slope. Once again, the t-test was not significant, however the effect size was medium. This suggests that the *ENA* condition was slightly less usable than the *DIR* condition. Regarding the *ENA* condition in particular, we ran a Pearson correlation test and found no evidence of correlation between the percentage of left hand usage and the reported usability ($r = 0.15$, $p = 0.39$).

As the different degrees of embodiment use different technologies, we expected the *DIR* condition to create more simulator sickness than the *TAB* condition. We found no significant difference in delta SSQ scores between *TAB* and *DIR*, with a medium effect size. As expected, we found no significant difference between the *DIR* and *ENA* delta SSQ scores, with a small effect size.

In conclusion, regarding manipulations and usability, there is no counter indication against a higher degree of embodiment, even though it uses a more cumbersome technology. The only drawback is the significantly-longer duration of the activity. Regarding the type of embodiment, an *enacted* approach is more time-consuming,

Table 3: Inventory of the t-tests results. A result was considered significant (∗) when $p < 0.05$ and almost significant when $p < 0.10$ (·). Cohen's d and Cohen's U_3 effect sizes are reported [25, 59]. Cohen's U_3 represents distribution overlap and is the percentage of participants in the lower-mean condition scoring lower than the mean score of the participants in the higher-mean condition.

	Degree: *TAB* and *DIR*					**Type: *DIR* and *ENA***				
Dependent variable	df	t	p	d	U_3	df	t	p	d	U_3
Duration	17.38	−6.27	**<0.001**∗	2.6	100%	18.10	−3.19	**0.005**∗	1.25	100%
Number of manipulations	18.23	1.69	0.11	0.65	83%	14.38	−4.64	**<0.001**∗	1.79	100%
SUS score	22.95	−0.10	0.92	0.04	38%	20.43	1.23	0.23	0.50	67%
Delta SSQ score	14.86	1.23	0.24	0.47	77%	10.82	0.90	0.39	0.33	83%
Total hand movement						10.02	1.56	**0.001**∗	1.83	92%
Average hand movement						7.91	4.86	**0.001**∗	1.98	92%
Average amplitude						20.36	16.72	**<0.001**∗	6.57	100%
Sense of body ownership						20.87	−0.88	0.38	0.35	67%
Sense of body agency						22.87	0.24	0.81	0.1	46%
Sense of body change						22.02	−0.01	0.99	0.06	58%
Sense of curve agency	20.5	−0.08	0.94	0.03	46%	21.40	1.66	0.11	0.65	54%
Learning Post-test scores	9.52	1.45	0.18	0.58	50%	11.79	0.27	0.79	0.10	54%
CCI scores	22.70	0.54	0.59	0.22	75%	22.97	1.76	**0.09**·	0.70	69%

generates more superfluous manipulations, and might be less usable. Therefore, a *direct-embodied* approach should be preferred.

4.4.2 Sense of Embodiment and Curve Agency. First, we evaluated whether there are movement differences between the types of embodiment. We found that participants in the *DIR* condition moved their hands more than the participants in the *ENA* condition. We also found a significant difference in movement per manipulation, and in average amplitude.

With regards to the Sense of Embodiment across the types of embodiment, we found no significant difference for the sense of body ownership, the sense of body agency, and the sense of body change.

Concerning the sense of felt agency on the function curve, we found no significant differences across the different degrees of embodiment, nor across the different types of embodiment. However, we expected participants in the *ENA* condition to feel less agency on the curve, as the interaction is slightly more indirect, and the low p-value suggests that such effect might be identified with more participants.

We also looked into the impact of body awareness on sense of embodiment and sense of curve agency. Using Pearson's correlation factor, we did not find evidence for correlation between body awareness and sense of body agency ($r = -0.28$, $p = 0.18$), sense of body change ($r = -0.13$, $p = 0.54$), and sense of curve agency ($r = -0.03$, $p = 0.88$). However, we found an almost significant correlation between body awareness and sense of body ownership ($r = -0.38$, $p = 0.06$), meaning that participants with higher body awareness felt less body ownership in the VR conditions.

In conclusion, the *direct-embodied* interaction generated more movement and more amplitude per manipulation than the *enacted* approach. However, this did not translate into a higher sense of embodiment. We also did not find any differences in curve agency across degrees of embodiment, nor types of embodiment. We would therefore recommend favoring a *direct-embodied* approach if an emphasis on movement is desired. We would also advise to be particularly careful on avatar personalisation in VR as higher body awareness led to less body ownership.

4.4.3 Learning and Concept Inventory. The prerequisites scores were very high ($M = 85\%$, $SD = 11.0$), especially for reading the graph of a function ($M = 95\%$, $SD = 16.4$), reading the sign of a function graph ($M = 92\%$, $SD = 17.2$), and reading vector coordinates ($M = 89\%$, $SD = 22.0$). Students scored lower on the questions about reading local maximum and minimum ($M = 64\%$, $SD = 22.3$) but this is less primordial in our activity. There was no significant difference in prerequisite scores between the degrees of embodiment ($t(15.68) = 1.26$, $p = 0.23$), nor between the types of embodiment ($t(12.65) = 0.20$, $p = 0.85$).

Regarding learning, we found no significant difference in post-test scores across degrees of embodiment and types of embodiment. The effect size of the type of embodiment is very small, suggesting that we would not observe an effect with more participants. We then compared the CCI results, and found no significant differences across degrees of embodiment and an almost significant difference with medium to large effect size across types of embodiment in favor of the *DIR* condition. We also ran a Pearson correlation test using data from the *DIR* and *ENA* conditions, and found no correlation between the average amplitude of movement and the post-test scores ($r = 0.17$, $p = 0.44$), nor the CCI scores ($r = 0.08$, $p = 0.72$).

Concerning the number of successfully completed levels, a Fisher exact test yielded no significant difference between the *TAB* and *DIR* condition ($p = 0.46$), but an almost significant difference between the *DIR* and *ENA* condition ($p = 0.08$, $M_{DIR} = 21.58$, $M_{ENA} = 19.15$), meaning the *ENA* participants skipped more levels than the *DIR* participants.

Finally, we carried out a multiple regression to investigate the role of math grade, prerequisites, math anxiety, and body awareness in the final post-test and CCI scores. For the experimental condition, we used a contrast comparing the degree of embodiment (*TAB* and *DIR*), as well as the type of embodiment (*DIR* and *ENA*). For the post-test score, the results of the regression indicated that the model explained 34% of variance and significantly reflected the underlying data ($F(6, 30) = 2.57$, $p = 0.04$). Math grade was the only predictor contributing significantly to the model ($B = 10.00$, $p = 0.035$). In particular, the degree of embodiment ($B = 5.13$, $p = 0.18$) and the type of embodiment ($B = 4.18$, $p = 0.27$) did not contribute significantly. Looking at CCI, the model explained 48% of the variance and significantly reflected the underlying data ($F(6, 30) = 4.49$, $p = 0.002$). Again, math grade was the only significant predictor ($B = 0.79$, $p = 0.040$), while the degree and the type of embodiment were not (resp. $B = 0.24$, $p = 0.43$ and $B = 0.41$, $p = 0.19$).

In conclusion, we found no differences in learning across different degrees of embodiment. This might mean either one of two things: there is no effect of the degree of embodiment on learning, or this effect is counterbalanced by the cumbersomeness of VR. Regarding the type of embodiment, the *enacted* approach resulted in worse learning, and a higher quitting rate. Therefore, we would recommend against an *enacted* approach except if the topic at hand requires it. It is also important to note that the math grade was the only significant predictor of the post test scores.

5 DISCUSSION

Although mathematics is considered inherently abstract, mathematics learning can benefit from initial concrete examples and representations [8, 17, 58]. Moreover, students learn mathematics for different reasons. While some students might decide to dedicate their career to the topic, others will only use their mathematics skills as a tool in other contexts: focusing the lesson solely on abstract symbols and formalism does not reflect such individual differences. Similarly, students suffering from math anxiety can benefit from hands-on experiences [10].

With our work, we offer an embodied activity to discover and explore concrete derivatives, while gaining intuition. From our design process and empirical results, we present several aspects to consider when designing embodied interaction for learning mathematics. Indeed, although VR is promising when it comes to highly embodied interaction, such technology is also time consuming and spatially cumbersome. As designers, we ought to make the experience worth the logistics, and go beyond the increased motivation tied to the technology [34].

First, although VR can indeed increase the sense of embodiment and movement amplitude, this is not always automatic. For example, participants with less VR experience might feel afraid and reduce their movements. Moreover, selecting a more indirect form of interaction might result in reduced sense of agency, as well as less movements and smaller amplitude. We recommend preceding the VR activity by an exploration phase where students discover the virtual space and its limits, as well as the interaction possibilities. Moreover, we recommend favoring more direct forms of interaction, and, if using hand tracking, being mindful of expectations students bring from the real world [9].

Second, we ought to consider the role of precision in our activity. For example, in an activity with percentage accuracy outcome-feedback such as a score [31], accuracy is of importance, and picking a less precise gesture, such as a grabbing, over a more precise gesture, such as pinching, will increase unnecessary frustration. This aspect goes even further: While high-achieving students will rather focus on the general shape of a graph, other students put a strong emphasis on accuracy when gesturing function representations [21]. We recommend designing interaction matching the precision required by the activity, but also by the target audience.

Third, when designing interaction for learning, we recommend acknowledging the discrepancy between a good interaction from a usability perspective, and a good interaction from a learning perspective, and, in particular, thinking in terms of desirable difficulties [5]. As we saw, delaying visual feedback can create opportunities for reflection, and knowledge gap awareness [31, 32]. Moreover, focusing the interaction on specific areas of the problem at hand can help the student focus on the critical aspects.

Fourth, we highlighted different types of embodiment, in particular *direct-embodied* and *enacted* [44, 48]. While *direct-embodied* interaction focuses on the body as "the primary constituent of cognition", an *enacted* approach emphasizes "learning by physically doing". However, in the case of derivatives, the implications go further: a *direct-embodied* interaction focuses on derivatives as a quantity or a slope, while an *enacted* interaction focuses on derivatives as variation rates. We recommend aligning the interaction design with the activity design when it comes to the type of embodiment. For example, in our activity design, we focused on derivatives as slope, and our empirical results showed that, in that context, the *enacted* interaction resulted in less learning and less persistence.

Finally, we want to emphasize the importance of designing embodied interaction not for physical bodies, but for feeling bodies [46]. Although we did not focus our study on this aspect, we did notice that these design choices were particularly enjoyed by the students. For example, following the recommendation of Mueller et al., we placed the button to finish a level on the top right of the user [46]. As a result, students soon turned this interaction into a "high-five" motion, and, we believed, appreciated their achievement at an embodied level.

5.1 Limitations and Future Work

The main limitation of our work is the sample size of the quantitative study. Although this is not an issue for the first two research questions as the expected effect sizes are rather large, it can be an issue for the question on learning outcomes as expected effect sizes are smaller. In particular, the effect sizes regarding the effect of the degree of embodiment on learning outcomes are small to medium, in favor of the weaker embodiment, suggesting that further investigation is necessary. We want to pursue this question with a large-scale study based on our design. Moreover, the learning assessment happened directly after the study: assessments over an extended period of time should be used to address medium and long term effects.

To inform our design, we invited several experts. P3, in particular, has teaching experience in mathematics, but is not an experienced teacher. Inviting a mathematics teacher as well as high-school students from the first step of the design might have revealed interesting findings. The latter was not possible at the time of the design, due to corona-related restrictions.

Another concern is the fatigue effect of the quantitative study as the participants had to fill in several questionnaires. However, as we included a break before the learning outcome questionnaires, we believe that this effect is mitigated.

Similarly, as VR is still a novel technology for most people, some of them might feel anxious when participating in the activity, as was indeed the case in the qualitative study. Conversely, VR might feel exciting for some participants and generate a positive novelty effect [30]. However, we mitigated these effects for the quantitative study by including a preliminary activity where the participants could discover the technology. This was particularly useful as only 3 participants already had VR experience.

Finally, all participants wore a face mask due to the local health regulations, increasing the discomfort of the VR condition. However, they might have already been used to wearing a mask.

Regarding future work, there are several main directions left to explore: First, evaluate whether our recommendations generalize to other topics; Second, evaluate whether they generalize to embodiment of different natures, for example temporal instead of spatial; Third, evaluate in more details how the degrees and types of embodiment influence conceptual understanding, and, in particular, the gestures used to communicate understanding; Finally, evaluate the relation between the design of the embodied interaction and the learning strategies of the users, accounting for individual preferences for gestured graphs [21].

6 CONCLUSION

In our work, we implemented an activity to help high school students build intuition about derivatives. First, we validated our prototype with a panel of experts, and drew conclusions about embodied interaction design; First, although VR is good for embodiment, it requires an adjustment period and can restrict the user's movements. Second, embodied interaction with curves is intuitive and enjoyable, and creates a hands-on experience. Moreover, direct interaction is favored over indirect interaction, although desirable difficulty in interaction can actually benefit learning. Finally, embodied activities offer a novel approach to mathematics and helps building mathematical intuition. However, such activity should be reconnected to formal instruction in order to be truly beneficial to the students.

We then used our validated prototype to compare different degrees of embodiment (weak embodiment on tablet and strong embodiment in VR), and different types of embodiment (*direct-embodied* and *enacted*). Our results show that even though VR technology is more cumbersome and more time consuming, it does not significantly reduce the usability of the prototype nor increases simulator sickness. Moreover, we show that participants using a more indirect interaction, the *enacted* interaction, tend to give up more often and learn less than participants using a more direct interaction. Finally, we did not find differences in learning outcomes across different degrees of embodiment, suggesting that Virtual Reality is not necessary for a successful embodied design.

7 SELECTION AND PARTICIPATION OF CHILDREN

The pilot and quantitative studies were approved by the ETH Zurich Ethics Commission as proposal 2021-N-169. The teenagers were recruited through different mailing lists of math teachers in Switzerland. The teachers hosted the study in their classrooms. The teenagers and their legal guardians received an information sheet and a consent form weeks before the study, and were free to chose to participate. All obtained information was stored on a shared folder hosted by the university and only available to the project members conducting the analysis. This anonymized information was indexed by participant id.

ACKNOWLEDGMENTS

The authors would like to thank Keny Chatain, Vera Baumgartner, Nino Antulov-Fantulin, Simona Bochsler, Dragan Trninić, Dražen Popović, Anna Ettlin, Emma Eglinton, GTC, and LSE for the help and feedback, the colleagues who helped conducting the study, the Future Learning Initiative of ETH Zurich for the support, the participants of the user studies and their teachers for making our project possible, and the reviewers for their insightful comments.

REFERENCES

[1] Dor Abrahamson and Robb Lindgren. 2014. *Embodiment and embodied design.*
[2] Dor Abrahamson, Mitchell J Nathan, Caro Williams-Pierce, Candace Walkington, Erin R Ottmar, Hortensia Soto, and Martha W Alibali. 2020. The future of embodied design for mathematics teaching and learning. In *Frontiers in Education*, Vol. 5. Frontiers, 147.
[3] Mark H Ashcraft, Elizabeth P Kirk, and Derek Hopko. 1998. On the cognitive consequences of mathematics anxiety. (1998).
[4] Connie Barroso, Colleen M Ganley, Amanda L McGraw, Elyssa A Geer, Sara A Hart, and Mia C Daucourt. 2021. A meta-analysis of the relation between math anxiety and math achievement. *Psychological Bulletin* 147, 2 (2021), 134.
[5] Elizabeth L Bjork, Robert A Bjork, et al. 2011. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society* 2, 59-68 (2011).
[6] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.
[7] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.
[8] Kira J Carbonneau, Scott C Marley, and James P Selig. 2013. A meta-analysis of the efficacy of teaching mathematics with concrete manipulatives. *Journal of Educational Psychology* 105, 2 (2013), 380.
[9] Julia Chatain, Danielle M Sisserman, Lea Reichardt, Violaine Fayolle, Manu Kapur, Robert W Sumner, Fabio Zünd, and Amit H Bermano. 2020. DigiGlo: Exploring the Palm as an Input and Display Mechanism through Digital Gloves. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 374–385.
[10] Yu-ching Chen. 2019. Effect of mobile augmented reality on learning performance, motivation, and math anxiety in a math course. *Journal of Educational Computing Research* 57, 7 (2019), 1695–1722.
[11] Krzysztof Cipora, Monika Szczygieł, Klaus Willmes, and Hans-Christoph Nuerk. 2015. Math anxiety assessment with the Abbreviated Math Anxiety Scale: Applicability and usefulness: Insights from the Polish adaptation. *Frontiers in Psychology* 6 (2015), 1833.
[12] National Research Council et al. 1989. *Everybody counts: A report to the nation on the future of mathematics education.* National Academies Press.
[13] Filitsa Dingyloudi and Jan-Willem Strijbos. 2018. Mixed methods research as a pragmatic toolkit: Understanding versus fixing complexity in the learning sciences. *F. Fischer, CE Hmelo-Silver, SR Goldman &P. Reimann (eds.). The International Handbook of the Learning Sciences* (2018), 444–454.
[14] Paul Dourish. 2004. *Where the action is: the foundations of embodied interaction.* MIT press.
[15] Ann Dowker, Amar Sarkar, and Chung Yen Looi. 2016. Mathematics anxiety: What have we learned in 60 years? *Frontiers in psychology* 7 (2016), 508.
[16] Jerome Epstein. 2007. Development and validation of the Calculus Concept Inventory. In *Proceedings of the ninth international conference on mathematics education in a global community*, Vol. 9. Charlotte, NC, 165–170.

[17] Emily R Fyfe, Nicole M McNeil, Ji Y Son, and Robert L Goldstone. 2014. Concreteness fading in mathematics and science instruction: A systematic review. *Educational psychology review* 26, 1 (2014), 9–25.

[18] Emily R Fyfe and Bethany Rittle-Johnson. 2017. Mathematics practice without feedback: A desirable difficulty in a classroom setting. *Instructional Science* 45, 2 (2017), 177–194.

[19] Venera Gashaj, Nicole Oberer, Fred W Mast, and Claudia M Roebers. 2019. The relation between executive functions, fine motor skills, and basic numerical skills and their relevance for later mathematics achievement. *Early education and development* 30, 7 (2019), 913–926.

[20] Venera Gashaj and Dragan Trninic. 2022. Adding up fine motor skills: developmental relations between manual dexterity and numerical abilities. *PsyArXiv. January* 2 (2022).

[21] Susan Gerofsky. 2011. Seeing the graph vs. being the graph. *Integrating gestures* (2011).

[22] Mar Gonzalez-Franco and Tabitha C. Peck. 2018. Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI* 5 (2018), 74. https://doi.org/10.3389/frobt.2018.00074

[23] Koeno Gravemeijer and Michelle Stephan. 2002. Emergent models as an instructional design heuristic. In *Symbolizing, modeling and tool use in mathematics education*. Springer, 145–169.

[24] Koeno P Gravemeijer, Richard Lehrer, HJ van Oers, and Lieven Verschaffel. 2013. *Symbolizing, modeling and tool use in mathematics education*. Vol. 30. Springer Science & Business Media.

[25] Paul HP Hanel and David MA Mehler. 2019. Beyond reporting statistical significance: Identifying informative effect sizes to improve scientific communication. *Public understanding of science* 28, 4 (2019), 468–485.

[26] Kristina Hook. 2018. *Designing with the body: Somaesthetic interaction design*. MIT Press.

[27] Kristina Höök, Martin P Jonsson, Anna Ståhl, and Johanna Mercurio. 2016. Somaesthetic appreciation design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 3131–3142.

[28] Derek R Hopko, Rajan Mahadevan, Robert L Bare, and Melissa K Hunt. 2003. The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment* 10, 2 (2003), 178–182.

[29] Mark Howison, Dragan Trninic, Daniel Reinholz, and Dor Abrahamson. 2011. The Mathematical Imagery Trainer: from embodied interaction to conceptual learning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1989–1998.

[30] Wen Huang. 2020. *Investigating the novelty effect in virtual reality on stem learning*. Ph.D. Dissertation. Arizona State University.

[31] Cheryl I Johnson, Shannon KT Bailey, and Wendi L Van Buskirk. 2017. Designing effective feedback messages in serious games and simulations: A research review. *Instructional techniques to facilitate learning and motivation of serious games* (2017), 119–140.

[32] Mina C Johnson-Glenberg. 2019. The necessary nine: Design principles for embodied VR and active stem education. In *Learning in a Digital World*. Springer, 83–112.

[33] Mina C Johnson-Glenberg and Colleen Megowan-Romanowicz. 2017. Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive research: principles and implications* 2, 1 (2017), 1–28.

[34] Sam Kavanagh, Andrew Luxton-Reilly, Burkhard Wuensche, and Beryl Plimmer. 2017. A systematic review of virtual reality in education. *Themes in Science and Technology Education* 10, 2 (2017), 85–119.

[35] Robert S Kennedy, Norman E Lane, Kevin S Berbaum, and Michael G Lilienthal. 1993. Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *The international journal of aviation psychology* 3, 3 (1993), 203–220.

[36] David Kirsh. 2013. Embodied cognition and the magical future of interaction design. *ACM Transactions on Computer-Human Interaction (TOCHI)* 20, 1 (2013), 1–30.

[37] CJC Kruger. 2003. Constrained cubic spline interpolation. *Chemical Engineering Applications* (2003).

[38] George Lakoff and Rafael Núñez. 2000. *Where mathematics comes from*. Vol. 6. New York: Basic Books.

[39] Robb Lindgren, Michael Tscholl, Shuai Wang, and Emily Johnson. 2016. Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education* 95 (2016), 174–187.

[40] Katharina Loibl, Ido Roll, and Nikol Rummel. 2017. Towards a theory of when and how problem solving followed by instruction supports learning. *Educational Psychology Review* 29, 4 (2017), 693–715.

[41] Janelle K Lorenzen. 2017. *The effect of instructional strategies on math anxiety and achievement: A mixed methods study of preservice elementary teachers*. Ph.D. Dissertation. The University of Southern Mississippi.

[42] Magnopus. 2020. *Elixir*. Oculus Quest 2. Last played December 2021.

[43] Laura Malinverni and Narcis Pares. 2014. Learning of abstract concepts through full-body interaction: A systematic review. *Journal of Educational Technology & Society* 17, 4 (2014), 100–116.

[44] Edward F. Melcer and Katherine Isbister. 2016. Bridging the Physical Divide: A Design Framework for Embodied Learning Games and Simulations. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI EA '16)*. Association for Computing Machinery, New York, NY, USA, 2225–2233. https://doi.org/10.1145/2851581.2892455

[45] Derek Melser. 2004. *The act of thinking*. MIT Press.

[46] Florian'Floyd' Mueller, Richard Byrne, Josh Andres, and Rakesh Patibanda. 2018. Experiencing the body as play. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[47] Mitchell J Nathan and Candace Walkington. 2017. Grounded and embodied mathematical cognition: Promoting mathematical insight and proof using action and language. *Cognitive research: principles and implications* 2, 1 (2017), 1–20.

[48] Erin R Ottmar, Candace Walkington, Dor Abrahamson, Mitchell J Nathan, Avery Harrison, and Carmen Smith. 2019. Embodied Mathematical Imagination and Cognition (EMIC) Working Group. *North American Chapter of the International Group for the Psychology of Mathematics Education* (2019).

[49] Seymour A Papert. 2020. *Mindstorms: Children, computers, and powerful ideas*. Basic books.

[50] Wim TJL Pouw, Tamara Van Gog, and Fred Paas. 2014. An embedded and embodied cognition review of instructional manipulatives. *Educational Psychology Review* 26, 1 (2014), 51–72.

[51] Daniel Roth and Marc Erich Latoschik. 2019. Construction of a validated virtual embodiment questionnaire. *arXiv preprint arXiv:1911.10176* (2019).

[52] Alan H Schoenfeld. 2016. Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics (Reprint). *Journal of Education* 196, 2 (2016), 1–38.

[53] Stephanie A Shields, Mary E Mallory, and Angela Simon. 1989. The body awareness questionnaire: reliability and validity. *Journal of personality Assessment* 53, 4 (1989), 802–815.

[54] Tanmay Sinha and Manu Kapur. 2021. When Problem Solving Followed by Instruction Works: Evidence for Productive Failure. *Review of Educational Research* (2021), 00346543211019105.

[55] Katta Spiel. 2021. The Bodies of TEI–Investigating Norms and Assumptions in the Design of Embodied Interaction. (2021).

[56] Milka Trajkova, A'aeshah Alhakamy, Francesco Cafaro, Rashmi Mallappa, and Sreekanth R Kankara. 2020. Move your body: Engaging museum visitors with human-data interaction. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–13.

[57] Cathy Tran, Brandon Smith, and Martin Buschkuehl. 2017. Support of mathematical thinking through embodied cognition: Nondigital and digital approaches. *Cognitive Research: Principles and Implications* 2, 1 (2017), 1–18.

[58] Dragan Trninic, Manu Kapur, and Tanmay Sinha. 2020. The Disappearing "Advantage of Abstract Examples in Learning Math". *Cognitive Science* 44, 7 (2020), e12851.

[59] Jeffrey C Valentine and Harris Cooper. 2003. Effect size substantive interpretation guidelines: Issues in the interpretation of effect sizes. *Washington, DC: What Works Clearinghouse* (2003), 1–7.

[60] Karen K Yuen. 1974. The two-sample trimmed t for unequal population variances. *Biometrika* 61, 1 (1974), 165–170.

[61] Jing Zhang, Nan Zhao, and Qi Ping Kong. 2019. The relationship between math anxiety and math performance: A meta-analytic investigation. *Frontiers in psychology* 10 (2019), 1613.