Computers & Education: X Reality 7 (2025) 100103

Contents lists available at ScienceDirect

Computers & Education: X Reality

journal homepage: www.journals.elsevier.com/computers-and-education-x-reality

Pause for success: Harnessing interaction delay and target selection difficulty in VR hands-on learning environments

Sara Khorasani [a,*], Stephain Hsu [a], Rui Guan [b], Jorge Goncalves [c], Andrew Irlitti [a], Jarrod Knibbe [c], Eduardo Velloso [d]

[a] University of Melbourne, Australia
[b] Monash University, Australia
[c] The University of Queensland, Australia
[d] University of Sydney, Australia

A R T I C L E I N F O

Keywords:
Virtual reality embodied learning VR learning
VR education VR training
Interaction techniques target selection
difficulty interaction delay makerspace

A B S T R A C T

Human-computer interaction (HCI) theory suggests that we should minimize interaction delays and reduce target selection difficulty to optimise performance. However, in learning scenarios, delays have been shown to cause 'forced learning' and difficulty can be an intrinsic motivator. Any interplay between delays, forced learning, difficulty, and the embodied, immersive explo-ration style of virtual reality (VR) remains poorly understood. We study the impact of delay and target selection difficulty on learning outcomes in VR. Using a VR makerspace training module with a 2x2 factorial, mixed-methods approach, we analyze the learning data from 124 participants who interacted with either a 5-s or zero delay post target selection, and *Easy* versus *Hard* target selection difficulties. The findings reveal that incorporating a 5-s delay post-interaction led to superior learning outcomes, providing users with more time to process and rehearse information. In contrast, altering the target selection difficulty showed negligible effects on learning outcomes, with participants reporting a simultaneous increase in engagement and distraction from the learning content. This research challenges conventional HCI theories within a VR context, suggesting potential educational benefits from strategically incorporated interaction delays.

1. Introduction

There is much excitement surrounding the possibilities of VR-based maker space learning experiences – getting repeated, hands-on access to complex equipment and scenarios, all the while lowering barriers to access (Khorasani et al. (2023); Baumgartner et al. (2022)). Much of the work here, however, is on higher-order factors, such as learning content, pedagogical considerations, and interaction modalities (e.g., Caratachea and Monty Jones (2024); Radu et al. (2021); Radu and Schneider (2023); Petersen et al. (2021); Petersen et al. (2022); Schwarze et al. (2019); Garcia et al. (2019)). The specific implications of low-level considerations, however, are complex and remain poorly understood (Khorasani et al. (2023)) .

Human-computer interaction (HCI) theories, for example, would suggest we should minimize interaction delays (eg., Nielsen (1993); Donald A. Norman (2002)) and reduce target selection difficulty (eg., Clark et al. (2020); Henrikson et al. (2020); Dalsgaard et al. (2021); Triantafyllidis and Li (2021)) to optimise general performance. Typically, this may help us address the extraneous processing and increased cognitive load that the embodied, immersive nature of VR has been shown to introduce (Mayer et al. (2022)). Conversely, however, slower interaction can promote thoughtfulness (Chan et al. (2022)) and increasing selection difficulty may be beneficial for engagement (Oliveira dos Santos et al. (2018)). What impact, then, do our low-level interaction design decisions have on learning outcomes in VR? If we are designing to optimise learning in VR, should we also be optimizing for interaction performance, or not?

We select two low-level interaction design concepts: *interaction delay* and *target selection difficulty*, and study their impact on learning outcomes in VR.

Delays in user interface response times have long been known to cause confusion and generate user dissatisfaction, sometimes even leading to perceived system breakdowns (Nielsen (1993); Shneiderman (1984)). In contrast, other research suggests that intentionally

This article is part of a special issue entitled: Extended Reality in Maker Education published in Computers & Education: X Reality.

* Corresponding author.
E-mail addresses: sarakhorasani@gmail.com, skhorasani@student.unimelb.edu.au (S. Khorasani).

https://doi.org/10.1016/j.cexr.2025.100103
Received 1 October 2024; Received in revised form 4 June 2025; Accepted 5 June 2025
Available online 21 June 2025

incorporating such delay may spur a "forced learning" mechanism by controlling.

intrinsic cognitive load - a cognitive process that slows deliberately during moments of significance, potentially enhancing retention (Chan et al. (2022); Drey et al. (2022); Riding (2013)). Therefore, we question whether interaction delays impact users' abilities to learn. In our study, we compare a 5-s blinking delay post target selection with zero delay, to assess their impact on learning.

For *target selection difficulty*, motivated by Fitts' Law (Fitts (1954)), we question whether target selection difficulty, the time required to select a target based on the target's size, could have an impact on learning outcomes (Fitts (1954)). Interface size and design choices have been shown to impact user distraction and cognitive load (Grahn and Kujala (2020); Shi et al. (2021)). However, difficulty has also been shown to increase intrinsic motivation and engagement (Lomas et al. (2017); Abuhamdeh and Csikszentmihalyi (2012)). We implemented two levels of target selection difficulty (Hard vs Easy) by adjusting target sizes in a raycasting selection interaction, to understand its impact on learning.

To understand the effect of these interaction techniques on learning outcomes, we developed a VR training module teaching participants how to operate a table saw. We conduct a 2 x 2 factorial, between-subjects, mixed-methods study, of 124 participants. We analyze our quantitative learning data using a Bayesian statistical model and use inductive thematic analysis for our qualitative interview data.

Our results show that introducing a 5-s delay post-interaction led to the best learning outcomes. Varying the target selection difficulty was found to have a negligible impact on learning outcomes. Our participants highlight how the delay gave them more time to process and rehearse the information. While the target selection difficulty was reported to increase engagement but also distract from the learning content.

Our study offers detailed and practical insights into the impacts of interaction delay and target selection difficulty on learning effectiveness within a VR environment. We contribute to the body of knowledge in HCI by questioning and investigating the conventional understanding of interface design within a VR context, revealing how interaction delays may yield potential benefits for learning when strategically incorporated into educational content design.

2. Related work

Over the last decade, the application of VR for education and training has been hailed as a game changer and researchers have attempted to study it's benefits (Oberdörfer et al. (2019); Rogers et al. (2018); Murphy and Higgins (2019); Oagaz et al. (2021); Rettinger et al. (2021); Petersen et al. (2021)). It's ability to immerse users in a rich environment leading to fully embodied learning has been highlighted as a key strength (Petersen et al. (2022); Johnson-Glenberg (2018)). However, many studies have highlighted the limitations or issues caused by its richness and immersive nature, such as extraneous cognitive load that prevent it from being used to it's full potential for learning applications (Krauß et al. (2021); Sweetser and Rogalewicz (2020); Matviienko et al. (2022)).

In pursuit of understanding these issues, the VR learning community typically focuses on higher-order factors such as learning content, pedagogical considerations and interaction modalities (e.g., Caratachea and Monty Jones (2024); Radu et al. (2021); Radu and Schneider (2023); Petersen et al. (2021, 2022); Schwarze et al. (2019); Garcia et al. (2019)). However, in more established domains like desktop-based learning environments, studies explore the impact of colours, button sizes, loading delays among many other lower-order factors that can drastically impact learning outcomes. In VR however, these potentially crucial lower-level factors remain poorly understood (Khorasani et al. (2023)). In this study we attempt to bridge this gap by exploring lower-level VR interaction characteristics that may impact cognitive load in VR. More specifically we focused on specific challenges in VR environments when using raycasting as an interaction mode. Raycasting is a technique predominantly used in VR for interacting with 3D objects using a laser pointer. When using raycasting, interaction delays and target selection difficulties significantly influence cognitive processing, making this exploration critical for understanding VR-specific learning dynamics (Mayer et al. (2022)).

2.1. Cognitive load in VR

Our exploration into lower-level VR interaction characteristics was motivated by the compelling evidence suggesting that design elements in VR can significantly influence cognitive load, thereby impacting the learning process (Pan and Steed (2019); Mayer et al. (2022)). Cognitive load, the total mental effort utilized in working memory (Sweller. (1994)), comprises three main components: intrinsic load, related to the inherent complexity of the material; extraneous load, is attention given to how information is presented instead of the learning material; and germane load, the cognitive effort dedicated to the process of learning, facilitating the construction and automation of schemas (**?**). While extraneous cognitive load, can be detrimental by overburdening the learner's capacity (Ouellette et al. (2019)), the strategic manipulation of intrinsic and germane loads can enhance engagement and learning outcomes. This is supported by studies that indicate that appropriately challenging scenarios, which effectively manage these types of cognitive load, can foster increased engagement and therefore enhance learning (Paas, Tuovinen, et al. (2003)). Our study seeks to further unravel how VR's immersive and interactive design elements can be optimized to balance these cognitive loads, ultimately promoting a more effective and engaging learning environment. Cognitive load theory suggests that learning is hampered when cognitive processing related to non-relevant activities (extraneous processing) exceeds the learner's capacity (Ratcliffe and Tokarchuk (2022); Sweller (1994)). A nuanced understanding of these principles can substantially inform the design of interaction characteristics in VR experiences (Nukarinen et al. (2018); Biermann et al. (2022)). For example, the structure and difficulty of tasks contribute substantially to intrinsic cognitive load and, therefore, should be carefully calibrated to optimise learning (Sweller et al. (1998)). Consequently, our study has focused its lens on lower-level interaction characteristics, as their design can significantly influence extraneous cognitive load and, consequently, learning outcomes.

2.2. Target selection difficulty & interaction delay

An important interaction characteristic is "target selection difficulty", as posited by Fitts' law, a well-established model in the field of human-computer interaction and human factors. According to this law, the time required to move to a target area is a function of the ratio between the distance to the target and the width of the target (Fitts (1954); MacKenzie (1992)). In essence, larger targets or those closer to the user are easier to select, making them less cognitively demanding. As previously mentioned, it is well established that VR environments can cause extraneous cognitive load and negatively impact learning (Mayer et al. (2022)). We question whether the target selection difficulty (influenced by the target's size) is a factor that causes increased extraneous processing and thus influences learning outcomes (Fitts (1954)). Besides the ease of interaction, the size of interactive elements, such as buttons, could significantly affect visual saliency and, thereby, the allocation of attentional resources (Theeuwes (2010)).

Another characteristic in our study, interaction delay, can also impact cognitive load (Chan et al. (2022); Shneiderman (1984)). When a system fails to respond promptly, it can lead to confusion and a sense of system failure or breakdown (Nielsen (1993)). However, when employed strategically, interaction delays can serve as valuable pacing tools, potentially enhancing the learning experience (Chan et al. (2022); Drey et al. (2022); Riding (2013)). Pacing plays a vital role in learning, with slower-paced lessons often resulting in better comprehension and retention (Allen and Tanner (2005)). Therefore, interaction delays in VR could be an unconventional yet effective tool to control the learning pace, driving focus towards critical content and ensuring better assimilation of knowledge. Additionally, when designed with intent, delays can build anticipation towards significant learning moments in VR (Koob and Volkow (2016); Paas, Renkl, and Sweller (2003)). Similar dynamics are noted in game design, where unpredictability and difficulty spikes in games keep players engaged and motivated for more. This element of surprise and unpredictability has been shown to influence dopamine release, which in turn stimulates the pleasure and reward circuits in the brain (Schultz (1998)).

2.3. Interaction & learning design principles

Incorporating the above design principles into the design of interactive elements in a VR learning environment might enhance engagement, motivation, and learning outcomes. Linked to this idea, Zone of Proximal Development (ZPD), explores the optimal level of challenge for a learner (VYGOTSKY (1980)). The size of interactive buttons and the interaction delay may influence a learner's ZPD and may need to be accounted for to maximize learning. Furthermore, the concept of 'Flow'—a state of total immersion and heightened focus in an activity—could be instrumental in enhancing the VR learning experience (Csikszentmihalyi (1990); Oliveira dos Santos et al. (2018)). Crafting interaction characteristics like delay such that they can induce a state of flow may improve engagement and learning outcomes.

Beyond the aforementioned theories, there are numerous others that may explain the impact of target selection difficulty and selection delay on learning. One theory is the Levels of Processing framework, introduced by Craik and Lockhart (1972), which explains the significance of engaging deeply with learning materials for enhanced

Memory retention (Craik and Lockhart (1972)). This theory may suggest that the depth of cognitive engagement, may potentially be influenced by interaction delays. Furthermore, the concept of desirable difficulties, as proposed by Bjork (1994), highlights how incorporating challenges, such as interaction delays and target selection difficulties, can foster better learning outcomes by necessitating more active and engaged processing from learners (Bjork (1994)). This is particularly relevant in VR, where the immersive nature can both facilitate and hinder attention. Finally, the role of attention and attention guidance in learning, underscores the importance of designing VR experiences that minimize distractions (extraneous load) while maximizing engagement with the learning material (germane load) (Wickens et al. (2021); Hollender et al. (2010)). These perspectives collectively inform our exploration of how low-level VR interaction characteristics impact learning, providing a theoretical basis for our empirical findings.

2.4. Research gap & contributions

As we advance the frontiers of VR learning research, it becomes increasingly important to consider the lower- order factors that subtly yet profoundly influence learning experiences. These factors, such as interaction delays, target selection difficulty, and other fine-grained interaction design elements, often dictate how learners engage with and process information in virtual environments. In the specific context of VR maker spaces, where hands-on, active learning are central, these lower-level factors remain underexplored. Understanding how they affect cognitive load, engagement, and ultimately learning outcomes is critical.

With these multidimensional facets of interaction and cognitive load theory, our study aims to uncover insights that inform the design of effective maker space learning environments in VR. The challenge and opportunity lie in leveraging these principles to design a VR experience that reduces extraneous cognitive load and facilitates the creation of immersive experiences that are optimized for learning outcomes.

Fig. 1. 3D Model of a Table Saw showing the red lever used to adjust the fence, the wheels used to adjust the blade height & angle and the on and off switch. These parts of the machine had to be used to safely set up and cut the piece of wood.

Table 1
2x2 factorial experiment design.

	Target Selection Difficulty	
	Easy	Hard
Delay	C1 - Easy + Delay	C2 - Hard + Delay
No Delay	C3 - Easy + No Delay	C4 - Hard + No Delay

3. System design and implementation

To test how low-level interaction delay and target selection difficulty impact learning, we designed and built a VR application in Unity to teach participants how to safely operate a *SawStop* table saw. The exact saw is shown in Fig. 1.

We selected this activity due to its complex, procedural nature – the task includes physical movement and processes, with specific sub-tasks that must be completed at specific times. As such, the task can be rigorously assessed post-training. We identified a total of 20 sub-tasks (see Table 4) focused around 12 major steps that are necessary to operate the machine and incorporated them into our training protocol. This is based on our university's training procedure for this specific machine. The 12 major steps can be seen in the table below.

Training Tasks:

1.Removing all tripping hazards.
2.Putting safety glasses on.
3.Installing a riving knife to prevent kickback.
4.Unlocking and locking the fence.
5.Moving the fence in position against the wood being cut.
6.Adjusting the height of the blade to the safe cutting height.
7.Using a Mitre Gauge and push stick to push the material safely into the blade.
8.Not picking up cut parts before the machine is off. (Participant would cut their hand off if they grabbed the wood during the operation.)
9.Pushing the material fully past the blade to cut evenly.
10.Not allowing the push stick to touch the moving blade.
11.Turning the exhaust fan on before operation and turning it off after operation.
12.Turning the machine on and off at appropriate times.

3.1. Experimental design

We adopted a 2x2 factorial, between-subjects experimental design to investigate the effects of target selection difficulty and interaction delay

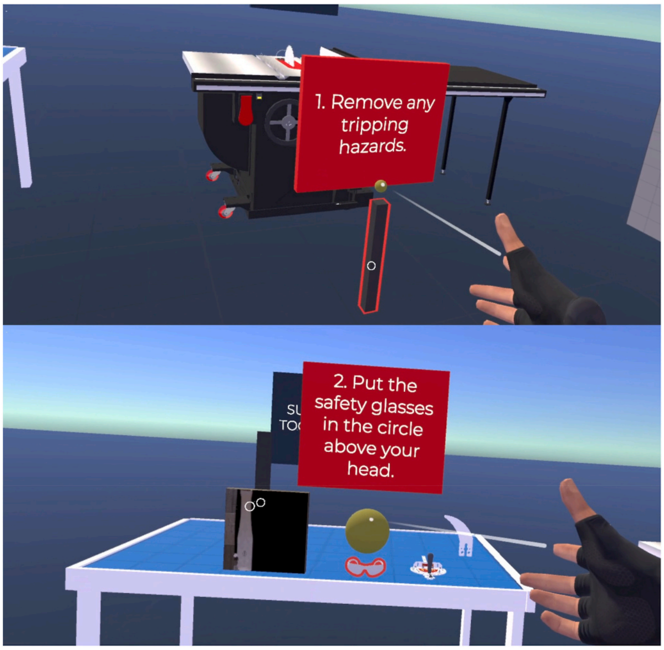


Fig. 2. Images showing the difference in the Hard (top) and Easy (bottom) TSD conditions as well as the laser pointer interaction where the colour of the sphere changes slightly to signal highlighting.

on learning outcomes in VR. We compare easy versus hard target selection and delay (5s) vs no delay (0s), see Table 1. We operationalise easy target selection as 17 cm spheres,[1] 2m from participants (yielding a viewing angle of $4^\circ\ 52'\ 0.03''$). Hard targets are 5 cm spheres,[2] the same distance from participants (yielding a viewing angle of $1^\circ\ 25'\ 0.94''$). Interaction delays occur immediately after target selection, accompanied by the yellow sphere blinking for feedback (in the 5s delay condition). During the delay condition participants saw the blinking sphere during the pause, with the sphere blinking once a second for 5 s. After which the associated animation showing them the correct procedure would begin playing. The delay duration was based on pilot testing among the authors and was determined to be sufficiently long to have an impact without being overwhelming. This technique is aligned with theoretical frameworks in Cognitive Load Theory, temporal spacing, Forced Reflection, and Pause Procedure which support the use of brief, strategic pauses to enhance processing and retention during learning tasks (Thaman (2014)). This design allows us to determine how distinct combinations of target selection difficulties and interaction delays impact participant scores, providing an understanding of their influence on learning within VR environments.

In the VR space, participants first complete a guided training exercise, before subsequently completing an un-guided practical test. During the practical test, no additional instructions were provided beyond “Repeat the steps you learnt during the training to cut this piece of wood. The steps should be repeated in the correct order. However, this time the wood must be cut at an angle”. This required them to do an additional step that wasn’t explicitly shown in the training but referred to.

3.2. The VR application

All conditions in our study were deployed in the same virtual environment. During the guided training, participants stood at a fixed location. They could observe a 360-degree view by moving their heads and crouching down if desired. This setting ensured the target selection difficulty of the yellow spheres remained the same (as in VR environments’

targets become larger as the user gets closer to them). Interaction with objects was indirect during the training, with participants targeting and selecting the spheres (using a laser pointer). The spheres acted as play buttons which then initiated animation demonstrations of the required task to be completed at that step. At each step, the participants received visual text instructions on how to operate the machine. These prompts were identical in content and order of appearance across all conditions (see Fig. 3). For example, one text prompt read, “always put your safety glasses on before cutting”. Finally, the object that the text prompts referred to were outlined in red in all conditions because we did not assume that participants knew the names of the objects and tools. For example, when the text prompt said, “install the riving knife behind the blade”, the system outlined the riving knife in red (as shown in Fig. 2 bottom).

During the assessment, participants were free to walk around the space and interact with objects directly.

3.3. The physical space & apparatus

The physical space was a large 5 × 10m room, and we mapped the VR space to the room’s exact dimensions. This mapping allowed participants to freely move around the space during the assessment task[3] without the need to teleport in VR to get to the desired location. We did not implement a teleport feature. Participants used a Meta Quest 2 with a 5m long link cable. The research team ensured participants could freely and safely walk around the space while tethered to the PC.

Participants were equipped with Meta Quest 2 controllers as their primary interaction devices. During training, the controller was used to mimic a virtual laser pointer, with the trigger affording target selection. During assessment, the grip button was deployed for grabbing and manipulating objects within the virtual space. The saw, levers, buttons, and all objects were fully interactive.

Fig. 3d, shows the 3D-rendered hand interacting with a lever. All participants practiced with these controls in tutorial scenarios before they moved to the training and assessment phases as shown in Fig. 4.

4. Methods

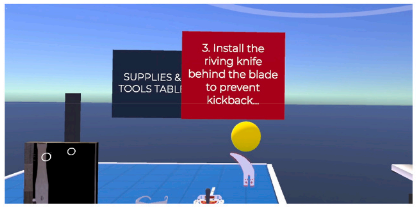
4.1. Participants

We recruited 125 participants (43 males and 82 females), see Table 2. One participant was eliminated as they misunderstood the study and could not perform the test. This resulted in exactly 31 participants per condition. Participants were recruited from digital notice boards and channels on The University’s website. Each participant was informed about the study and was offered a $15 gift voucher as compensation for their time and effort.

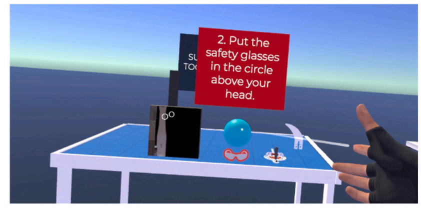
As part of the recruitment process, participants were asked about their experience with machine tools, such as a table saw. To ensure the validity of our study on learning, we excluded individuals who reported any prior experience, whether extensive or limited. The rationale was to avoid any bias that could be introduced by previous experience with such tools, maintaining the integrity of our learning experiment.

However, participants with prior VR experience were not excluded, as our primary focus was on the learning of the procedural tasks rather than VR navigation. Of the total participants, 121 (96.8 %) were novice VR users, and 4 (3.2 %) were experienced. Due to the small number of experienced users, their impact on learning outcomes was not analyzed. Future work could explore how prior VR exposure interacts with task difficulty and learning outcomes.

[1] the approximate size of a watermelon.
[2] the approximate size of a lime.
[3] participants were stationary during the training phase.

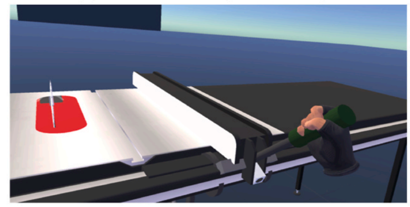


(a) Spheres appear bright yellow before hovered/clicked on.



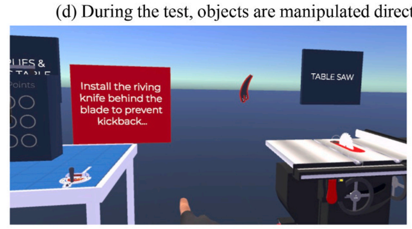
(b) Once successfully clicked, they turn blue.

(c) Objects are outlined in red when referred to in all conditions.

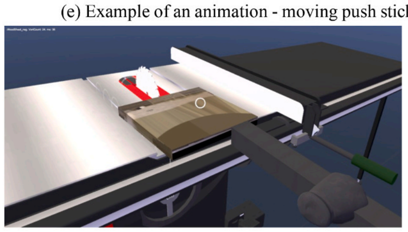


(d) During the test, objects are manipulated directly.

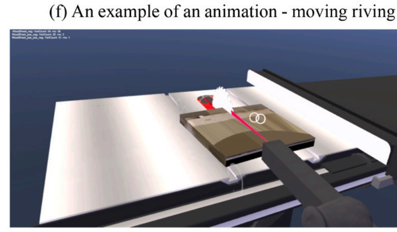
(e) Example of an animation - moving push stick.



(f) An example of an animation - moving riving knife.



(g) The test: participants use a push stick to cut the wood.



(h) The wood being cut.

Fig. 3. Images from the VR environment for all conditions and the test. The images show examples of animations, laser pointer interactions and the interactions required in the test. It is important to note all interactions in the training use the laser pointer and spheres as "play buttons" to initiate animations. In the test, interactions mirroring the real world were required.

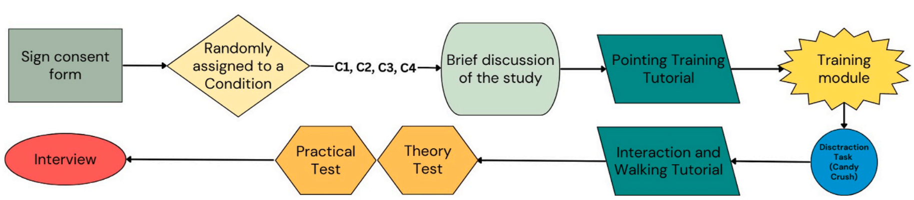


Fig. 4. Study workflow for all participants.

Table 2
Demographic characteristics of research participants.

	N	%
Gender		
Male	43	34.4
Female	82	65.6
Age Group		
18–24	72	57.6
25–34	45	36
35–44	8	6.4
VR Experience		
Novice	121	96.8
Experienced	4	3.2

4.2. Procedure

The study took between 40 min and 1 h, averaging 50 min per participant. Participants were first shown the environment and interaction controls. Next, they completed the table saw training phase in one of the four conditions, followed by a brief distraction task, before completing a practical and verbal theory test. At the end, participants were interviewed about their experience. We break down the steps in more detail below.

Upon arrival, participants were first required to read a plain language statement and sign the informed consent form. They were then randomly assigned to one of the 4 conditions.

The study started with a brief overview of the procedure. We familiarized participants with the 4 buttons they would use on the controllers. These included buttons A and B to cycle through the instructions, the trigger button to click, and the grip button to grip/hold objects. To ensure that participants had mastered the controls, they went through a 2–3 min tutorial. We only used this task for participants to gain practice with the controls, and we did not collect any data from this stage. A researcher was always in the room with participants to ensure they were safe and that the wire attached to the headset was not a tripping hazard.

Upon completing the control tutorials, participants went through a training module corresponding to their assigned condition. Participants were only given instructions for the controls needed in the training phase. For example, participants in all conditions learned how to use the laser pointer to click on the yellow spheres. We also provided another tutorial before the practical test to ensure the participants were comfortable with walking around the space.

During the training module, all participants were allowed to ask for help if they did not understand a step, and a scripted instruction was read out to them. If participants skipped a step or an animation, they were told they could not go back. The training took between 10 and 15 min. After completing the training, participants were asked to play Candy Crush for 5 min on an iPad. This distraction task was intended to clear their working memory to prevent rote learning before the test. It was also intended to give them a break from wearing the headset.

Participants then completed another interaction tutorial prior to the assessment phase. This covered how to interact with objects by picking them up and holding them. Participants were allowed to repeat this tutorial, prior to the assessment, if they wished. The assessment environment did not differ from the training environment, however in the assessment participants were able to fully interact with the VR environment naturally. Whereas in the training they walked around and watched pre-recorded animations using only a laser pointer to press “start” and “next”. Thus the additional tutorial ensured all participants were able to use the new natural interaction techniques.

During the assessment, if participants faced any difficulty with the controls, they were offered support, but no help or clues were given as to how to complete the task. The tests consisted of all the same steps as the training plus an additional step that was not included in the training. This additional step was to adjust the angle of the blade on the table saw by using a rotary wheel. At the start of the assessment scenario, participants were asked to repeat all of the same steps in the same order with the addition of this extra step. Tasks were only considered incorrect if they were not attempted at all or attempted in the wrong order. For example, if the safety glasses were put on before the saw was turned on that would be considered correct. It did not matter if the safety glasses were put on in the 1st, 2nd or 3rd order, if they were put on before the saw was turned on. As it should be in a real workshop. The assessment took 5–10 min.

Participants then took part in a structured interview about their experience. Interviews lasted on average 10 min.

The questions used can be seen below.

1. “How difficult was pointing and clicking the yellow spheres? From 1 to 10, where 10 is the most difficult”.
2. “How did the size of the yellow sphere influence your ability to learn?”
3. “How did the blinking delay affect your learning process?”
4. “What caused the most significant effect in remembering or forgetting content?”
5. “Are there any interesting aspects of your learning experience you would like to share?”

4.3. Measures

We recorded participants’ performance during the test. We operationalised performance by giving 1 point per correct implementation of the 20 identified tasks. Tasks were only completed correctly if they occurred in the right order. We recorded the number of interactions and the order in which they conducted those tasks. For example, locking the fence after the wood was cut was recorded as an interaction but considered incorrect. In this case, though the participant performed the required tasks, they did not complete them in the correct order, which reflects their lack of understanding of the correct procedure.

4.4. Analysis

To analyze the interviews, we used inductive thematic analysis (Nowell et al. (2017)). We transcribed and coded relevant quotes from the interviews that we believed helped shed light on the research questions and the themes we identified in the process.

For the quantitative test results, we modeled the data using an Item Response Theory approach to Bayesian modeling (Bürkner (2019); Hernando (2011); Khajah et al. (2014)). The rethinking Bayesian package was used (McElreath (2020)). We described the model and its implementation procedures below. We used the 4 tasks with the highest discrimination index in our model.

We employ Bayesian statistical methods to analyze our results, opting for this approach due to its added flexibility, capability to quantify uncertainty, better handling of small samples, and greater potential for future extensibility. For a comprehensive rationale advocating the use of Bayesian methods over traditional frequentist statistics in the field of Human-Computer Interaction (HCI), see Kay et al. (2016). Readers who may not be familiar with these methods can find an introduction in McElreath McElreath (2020) and can see examples of their practical application in.

HCI in Schmettow (2021). In this manner, we shift the focus away from p-values and dichotomous significance testing, directing our discussion towards causal modelling and parameter estimation.

4.4.1. Discrimination index

In our study, we applied the Discrimination Index (DI) method to evaluate the effectiveness of our tasks in discriminating between high and low performers in our sample. This measure, proposed by Thorndike (1918), is a robust tool for testing the discriminative power of individual items within a test.

We divided our sample into high scorers (top 27 %) and low scorers (bottom 27 %). The DI was then calculated using the formula $DI = P(H) - P(L)$, where $P(H)$ represents the proportion of high-scoring students who got a question correct and $P(L)$ is the proportion of low-scoring students who also got the question correct.

This calculation resulted in a value ranging between -1 and 1. A high positive DI (close to 1) indicates an effective question that differentiates between high and low performers well. Conversely, a DI close to 0 signifies that the question was ineffective in distinguishing performance levels. A negative DI could hint at a confusing or misleading question, as it was answered correctly by more low performers than high performers.

This analysis allowed us to select the most valuable test questions for further analysis. A task with a DI of 0.7 or greater is able to distinguish between students that have mastery and true comprehension of that task, and thus why we used this cutoff for our Bayesian analysis. This is a common practice in the educational data mining literature, to ensure that a test is designed as intended and is able to effectively discriminate between low and high performing students (Chiavaroli and Familari (2011)).

4.4.2. Bayesian model

In our data analysis process, we opted to utilize a Bayesian model, more specifically, a Bayesian Item Response Theory approach, as it lends itself particularly well to our research context given that tasks differ in their level of difficulty (Bürkner (2019)). We used a two-parameter item response theory (2 PL IRT) model with interaction delay and target selection difficulty (TSD) as fixed effects using the Bernoulli family and logit link function. The model also includes random intercepts for both task and the participant. The Bayesian model was implemented in the

Table 3
Test score mean for each of the conditions.

	C1 - Easy + Delay	C2 - Hard + Delay	C3 - Easy + No Delay	C4 - Hard No Delay
Practical Mean	15.0	14.9	14.5	13.4
Theory Mean	2.0	2.4	2.1	2.3
Total Mean	17	17.3	16.5	15.6

Table 4
Discrimination index for test tasks.

Test Tasks	Discrimination Index
Hazard Removal	0.41
Exhaust On	0.09
Exhaust Off	0.14
Glasses On	0.31
Riving Knife	0.73
Wood Placement	0.19
Fence Unlock	0.36
Fence Position	0.82
Fence Lock	0.73
Blade Height	0.63
Blade Angle	0.70
Saw On	0.04
Saw Off	0.09
Push Stick	0.29
Mitre Gauge	0.29
Push Both	0.36
Push Through	0.51
Cut Hands	0.53
Cut Stick	0.48
Grab Before Off	0.63

brms package in R with a total of 6 chains and 2000 iterations each. We analyzed our data using using the rethinking package in R (McElreath (2020)). More specifically, we modeled it such that we could understand the influence of different predictors on task performance (C), specifically: the task (t[T]), the participant (part[P]), delay (d[D]), and the target selection difficulty (tsd[TSD]).

The outcome variable C signifies task performance and is modeled as a binomial distribution reflecting partici-pants' correct or incorrect task responses, with the probability p. The logit function is used to model the log-odds of p.

The priors for the predictors — delay (d[D]) and target selection difficulty (tsd[TSD]) — are set as normal distributions centred at zero, which represents our neutral expectation before viewing the data. For the predictors t[T] and part[P], we employed adaptive priors— these are normal distributions with means (t_bar and p_bar respectively) and standard deviations (t_sigma and p_sigma respectively). This allows the model to learn the central tendency and variability of these predictors directly from the data. The model is further tuned by including hyperpriors. By allowing for uncertainty in the prior distribution, hyperpriors can help prevent overfitting and improve the robustness of the model's predictions. The hyperpriors for these adaptive priors are specified as normal distributions for the means and exponential distributions for the standard deviations, reinforcing our uncertainty around these hyperparameters.

In our model, we estimated two fixed effects, delay (d[D]) and target selection difficulty (tsd[TSD]). The estimated effect of the delay represents the difference in task performance between tasks with and without delay, on average, across all individuals and tasks. Similarly, the estimated effect of the target selection difficulty represents the difference in task performance as the level of difficulty changes, again averaged across all individuals and tasks.

The model includes random effects for the task (t[T]) and participant (part[P]). These effects represent the variation in task performance attributed to differences between the tasks and between the participants, respectively, beyond the fixed effects of delay and target selection difficulty.

In the context of our experiment, a significant random effect of task would suggest that the performance on different tasks varies even after accounting for the fixed effects. Similarly, a significant random effect for participants would suggest that individual differences in performance are not fully explained by the delay condition or the target selection difficulty. These random effects help us to better understand the sources of variability in task performance and the extent to which our fixed effects generalise across different tasks and participants.

5. Results

To understand the impact of interaction delay and target selection difficulty on learning in VR, we used a Bayesian model to analyze the test scores and a general inductive thematic analysis of the interview results. Our initial analysis of our data sets involved a brief overview of test score means across all 4 conditions as shown in Table 3. The results indicate that groups in the delay conditions had the highest scores.

Fig. 5 illustrates violin plots representing the distribution of raw test scores for different conditions. Each plot provides a view of the distribution shape and range of scores obtained in each condition, combining

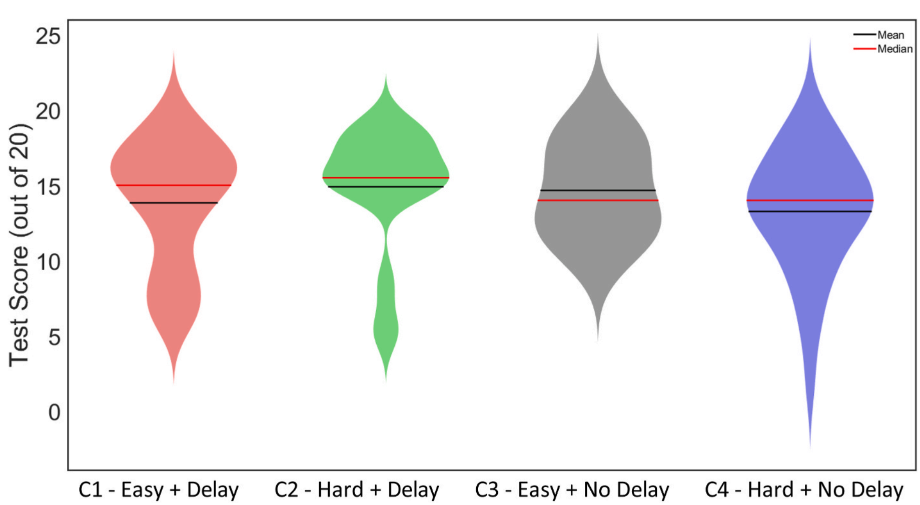


Fig. 5. Violin plots of the raw test scores (out of 20). Notice "C2 - Hard + Delay" is performing the best, with "C4 - Hard + No Delay" performing the worst. Additionally, the potential bimodal nature of "C1 - Easy + Delay".

Table 5
Summary of the bayesian model - with only the 4 tasks with the highest discrimination index.

	Mean	SE_Mean	SD	2.5 %	97.5 %
No Delay	−0.43	0.01	0.64	−1.68	0.83
Delay	0.41	0.01	0.63	−0.85	1.66
Easy TSD	0.13	0.02	0.64	−1.11	1.41
Hard TSD	−0.11	0.02	0.64	−1.33	1.15
Task 1 - Blade Angle	−0.72	0.02	0.63	−2.04	0.50
Task 2 - Fence Lock	−0.39	0.02	0.63	−1.68	0.84
Task 3 - Fence Position	0.55	0.02	0.63	−0.68	1.81
Task 4 - Riving Knife	0.71	0.02	0.63	−0.52	1.96

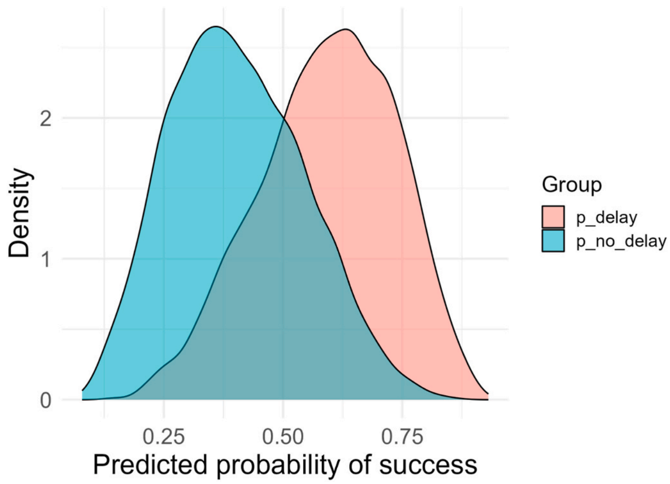


Fig. 6. Probability density plots for the delay and no delay conditions. The delay conditions have a higher probability of success (task performance), however it is important to note there is considerable overlaps of the two curves. This level of overlap is expected in a learning setting with participants having a wide range of abilities and it does not take away from the large effect that delay has on learning.

elements of a box plot and a kernel density plot.

In the figure, four conditions are represented, labelled as “C1 - Easy + Delay”, “C2 - Hard + Delay”, “C3 - Easy + No Delay”, and “C4 - Hard + No Delay”. The vertical axis represents the test score out of 20, including practical tasks (20 points) but not the theory questions (5 points). A higher score, therefore, indicates a better performance.

We can observe that Condition 2 (“C2 - Hard + Delay”) yielded the best performance, with most test scores concentrated towards the upper end of the scale. This suggests that most participants could successfully complete the majority of the practical tasks under this condition.

On the other hand, Condition 4 (“C4 - Hard + No Delay”) led to the poorest performance. The scores under this condition are generally distributed towards the lower end of the scale, indicating difficulties participants faced when performing tasks.

The violin plot for Condition 1 (C1 - Easy + Delay) reveals a potentially bimodal distribution, suggesting two distinct subgroups within this condition. This could be due to various factors; refer to the Discussion section for a thematic analysis of the interview results from participants that explain how the delay did or did not help them.

We then focus on Discrimination Index (DI) (Ustun et al. (2022)). This index shows us the most discriminating tasks, in other words, the

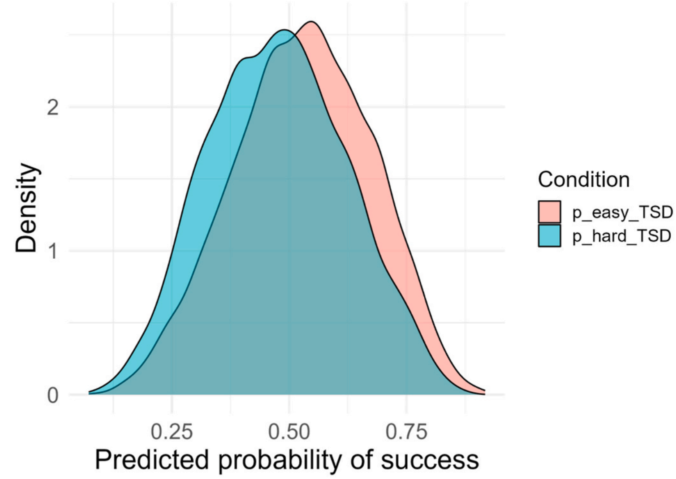


Fig. 7. Probability density plots for the Easy versus Hard Target Selection Difficulty conditions. There is only a marginal difference in task performance between easy and hard TSD, given the large overlap in the two curves this difference may be considered insignificant.

tasks which were most likely to be correct by high-performing participants and incorrect by low-performing ones - therefore, using them, we can more accurately decipher the impact of delay and target selection difficulty on learning outcomes. Our calculation resulted in 4 tasks above 0.7, shown below in Table 4. We will focus on these main tasks in this study and, consequently, in our Bayesian models. These 4 tasks were the most discriminating in our previous study which used a similar setup, thus validating our decision to use them exclusively. Using less discriminating tasks would impact our ability to reliably distinguish between the 4 conditions (Khorasani et al. (2023)). We can only speculate as to why these 4 tasks had high DIs. The blade angle has a high DI due to the fact that participants were not trained on how to adjust it in the training. The other 3 tasks had high DIs due to being tricky and sensitive tasks, but perhaps not appearing important to the task at hand (cutting).

5.1. Bayesian model

Our Bayesian model, estimated the impacts of interaction delay and target selection difficulty on task performance as shown in Table 5. Specifically, the model was used to estimate the probabilities of a correct task response under the conditions of delay and no delay as shown in Fig. 6 and Easy and Hard TSD in Fig. 7.

‘No Delay’, has a negative mean (−0.43), indicating a negative outcome in task performance when delay is absent. Conversely, ‘Delay’, has a positive mean (0.41), suggesting that the presence of delay affects task performance positively.

The positive mean (0.13) for ‘Easy TSD’, implies a better performance in tasks with a lower target selection difficulty. On the contrary, ‘Hard TSD’, shows a negative mean (−0.11), suggesting difficulties in the completion of these tasks.

The Rhat values, all in the proximity of 1.00, provide evidence of a good convergence in our model, hence offering confidence in the reliability of the estimates. Rhat is a convergence diagnostic that compares the variance within multiple chains of a parameter’s samples to the variance between the chains; values close to 1 indicate convergence. The

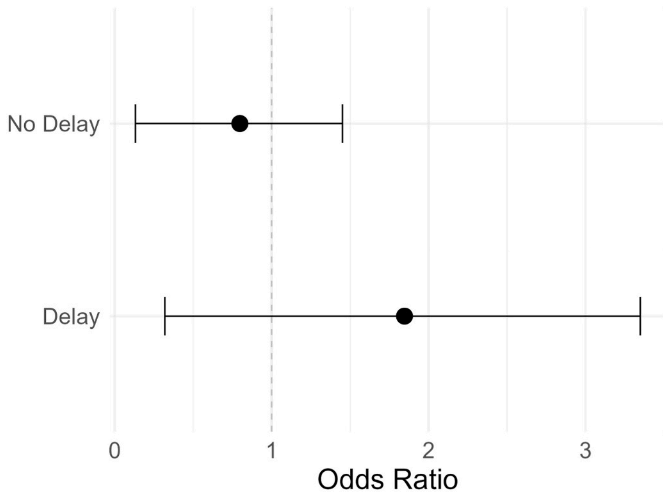


Fig. 8. Odds Ratio plots comparing delay with no delay from the Bayesian model. Odds ratio is a way of measuring how much the chance of an outcome (task performance) changes given a condition (delay). Comparing delay to no delay, we can see that the odds of delay positive changing task performance is significantly greater.

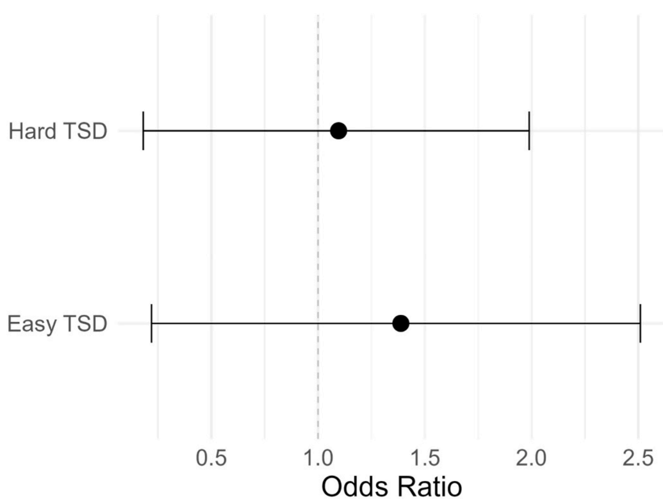


Fig. 9. Odds Ratio plots comparing Easy versus Hard Target Selection Difficulty from the Bayesian model. Odds ratio is a way of measuring how much the chance of an outcome (task performance) changes given a condition (Hard vs Easy TSD). Easy TSD is only marginally better than Hard TSD, and with overlapping credible intervals we may conclude no significant difference between Easy and Hard conditions.

n_eff values, although varying across predictors and tasks, ranging from 1736 to 1939, indicate a good working model. n_eff (Effective Sample Size) measures the number of independent samples equivalent to the correlated samples obtained, helping to assess the efficiency of the sampling process; higher values indicate more reliable estimates.

These observations collectively provide quantitative insights into the impact of interaction delay and target selection difficulty on task performance.

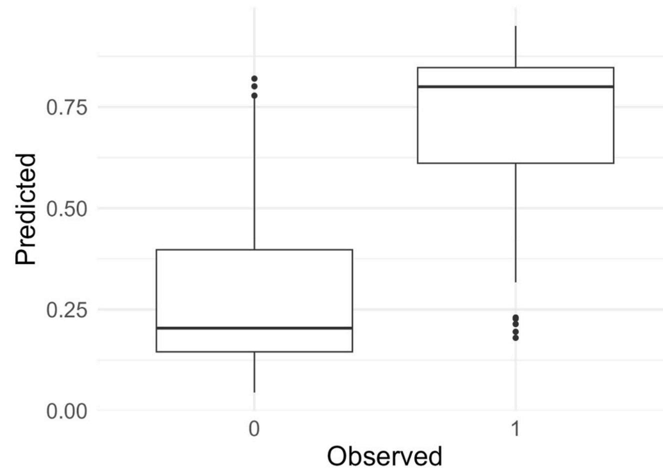


Fig. 10. Posterior predictive check of the Bayesian model showing that the model is doing well at predicting correct (1) and incorrect (0) task performance. The x-axis are actual observed data (correct versus incorrect), while the y-axis shows the predicted values by the model. We can see that the model tends to predict correctly.

5.1.1. Comparison of delay vs No delay

The estimated effects from the Bayesian model can be interpreted in terms of log-odds, odds ratios, and probabilities, each offering unique insights (Barreda and Silbert (2023)).

5.1.2. Log-odds and odds ratios

The model estimates the log-odds of success for the No Delay group as −0.43 and for the Delay group as 0.41. The difference in these log-odds is approximately 0.84. To interpret these log-odds in a more understandable way, we can convert them into odds ratios. The odds of success for the Delay group are $\exp(0.84) = 2.31$ times the odds of success for the No Delay group. This indicates a significant positive relationship between delay and success rate, with the Delay conditions having more than twice the odds of success than the No Delay conditions.

Additionally, as shown in Table 5 the Highest Posterior Density (HPD) or credible interval is the uncertainty around the odds ratios. Large intervals are expected in an educational setting with large variations in participant abilities, however they still highlight the better performance with delay.

Typically, if the 95 % credible interval for the odds ratio includes 1 (see Fig. 8), it's interpreted as the data being consistent with the null hypothesis, in other words Delay having no effect (Szumilas (2010)). However, Bayesian credible intervals have a different interpretation than frequentist confidence intervals. In Bayesian analysis, a 95 % credible interval can be interpreted as there being a 95 % probability that the true odds ratio is within the interval (see Fig. 9).

Therefore, the fact that the interval includes 1 doesn't mean that the null hypothesis is necessarily true; rather, it means that based on the data we have, we cannot rule out the possibility that the delay has no effect. It is most probable that the value isn't 1 (given that the center of the credible interval isn't 1).

Moreover, note that the width of the credible interval also provides information: A narrower interval indicates more precision, whereas a wider interval indicates more uncertainty. So our case, where the credible interval includes 1 but also the interval is very wide, it does not provide strong evidence in favor of the null hypothesis, because it also includes many values far from 1.

5.1.3. Probabilities

Though log-odds and odds ratios are typically used for interpretation, we can also convert these values to probabilities for a more intuitive understanding.

The probability of success for the No Delay group is approximately 0.40 (or 40 %), and for the Delay group, it is approximately 0.60 (or 60 %). Thus, the Delay group shows a 20 percentage point increase in the probability of success compared to the No Delay group.

While this does not capture the multiplicative relationship of the odds ratio, it does offer an easy-to-understand measure of the difference in success rates between the two groups.

However, it's important to consider the uncertainty around these estimates. The 95 % Bayesian credible intervals for the probabilities were calculated. This interval ranged from approximately 16 %–70 % for the no delay condition. For the delay condition, the interval ranged from approximately 30 %–84 %. This means that while the estimated probability of a correct task response was higher with delay than without, there is overlap in the credible intervals. As previously mentioned, this is expected in a learning setting given the variability in participant abilities - it does not take away from the significant positive impact of delay on learning outcomes.

5.2. Comparison of hard vs easy TSD

With the log odds of 0.13 for Easy TSD and −0.11 for Hard TSD translated to probabilities (Easy TSD - 53 % & Hard TSD - 47 %), and the considerable overlap in their credible intervals as shown in Fig. 7 we believe there is no considerable difference between Hard and Easy TSD. The difference translated into odds ratios is 1.37. While that may be considered a significant difference, the fact that the credible intervals overlap 100 %, we cannot conclude an effect.

5.3. Model validation

The model's ability to accurately generate data similar to the observed dataset during posterior predictive checks (as shown in Fig. 10) indirectly supports our claims regarding the effect of the delay. A successful predictive check indicates that the model, has managed to capture key patterns in the data. It is important to note that this does not directly confirm the impact of delay; however, it is suggestive of the model's overall appropriateness in representing the underlying processes that generated our data, including the role of delay. Consequently, the robust performance of the model in these checks enhances the credibility of our claims regarding the impact of delay on task performance.

In conducting our research, we utilized a Bayesian model to robustly analyze the data and to determine the effect of delay on learning outcomes. The results highlight that delay positively influences learning. This conclusion was substantiated through the use of predictive performance checks as well as PSIS-LOO and WAIC (discussed in detail in the appendix). These checks provided further confidence in the accuracy and reliability of our models' results.

5.4. Qualitative data

Through the analysis of our interview data, we reveal insights from our participants alongside our quantitative findings. Our analysis provides both expected and unexpected perspectives on the roles of target selection difficulty and interaction delay on learning.

5.4.1. Effects of interaction delay on learning outcomes

Our study produced compelling findings regarding the effect of delay on learning outcomes in a VR environment.

- we observed that including a 5-s delay led to improved learning outcomes overall (i.e., “C1 - Easy + Delay” and “C2 - Hard + Delay”). Participants reported that this delay gave them a chance to process the information: *“The delay made me feel more relaxed and decrease[d] pressure, it help[ed] me process the information, join all the small piece[s] together”* (P9,C1 - Score 17/20). Another participant also stated, *“Delay helped giving extra time to process and contextualize information, 5 s was just right to retain info”* (P74, C1 - Score 18/20). Further elucidating this idea, participants from “C2 - Hard + Delay” also indicated: *“… people will use the delay to help them repeat information in their minds, so the delay forced my learning”* (P22C2 - Score 18/20) As well as, *“The Delay helped me review the content, I will read at least 2–3 times and helped me slow down”* (P48C2 - Score 19/20).

These findings may be explained through the lens of the cognitive theory of Desirable Difficulties (Bjork (1994)). This theory posits that introducing difficulties or challenges into learning can enhance long-term memory retention and knowledge transfer. According to this perspective, the delay in our VR environment could have served as a “desirable difficulty,” prompting learners to engage more deeply with the educational content, leading to improved learning outcomes.

A small portion of participants reported lower learning outcomes, having less positive experiences with the delay. For example, P39 stated: *“My brain stops working for a few seconds when there is blinking and waiting because I am not sure when it will end”* (P39, C1: Score 7/20). Another participant felt similarly, saying *“Delay is a big distraction.*

I was thinking about something else and not sure when the blinking will finish” (P92, C1: Score 6/20). This theme resurfaced in the “C2 - Hard + Delay” group, with participants expressing the following: *“Once I finally get [clicked] the ball, the delay made it even more annoying. It was so distracting”* (P17, C2: Score 5/20). As well as, *“As I am not sure if loading is finished, it affect[ed] my learning. My mind was blank and broke me out of the flow and out of the learning zone”* (P134, C2: Score 9/20).

These participant responses shed light on the bimodal behaviour observed in our results and suggest that while the delay generally aids learning, it can also cause some individuals a momentary lapse in cognitive processing.

5.4.2. Effects of task selection dificulty on learning outcomes

Drawing from our Bayesian statistical analysis, we found that the different TSD levels did not significantly affect the learning outcomes. We can see the neutral reaction from participants across Easy and Hard TSDs. *“I learn [the pointing] also, [I] get distracted also but [am] still able to do the job”* (P29, C3: Score 12/20). The Hard TSD group also reported, *“I will follow this requirement by system. I didn't think about it that much in training and I don't think it affected me that much”*(P36, C4: 12/20). With a participant saying, *“I quickly adjusted to [the] new environment, started to focus more during steps 4–5 as I got better at using the system”* (P117, C2: 17/20).

5.4.3. Engagement and distraction effects

Participants expressed that the target selection process enhanced their engagement and provided a fun learning experience. This was especially true for novice VR users. For example, P31 stated: *"If first time in VR, the balls will help users feel more engaged and teaches users how to use the pointer. But if you are experienced, pressing A (to go next) might be the superior way to quickly get to the information you need and learn quicker"* (P31, C3: Score 13/20). P69, however, highlighted a potential risk here: *"As I get better at shooting the balls, I got more excited about the next one because I gain a sense of success. Can mistake the main goal of the training is shooting the ball …"* (P69, C4: Score 14/20).

On the other hand, Despite the perceived advantages, some participants found that the target selection process was a distraction from the learning content: *"[My] focus was more on shooting the balls, I could have use the time to memorize information on the boards"* (P36, C4: Score 11/20).

6. Discussion

We explored the impact of interaction delay and target selection difficulty on learning in VR. Our results show that delay had a significant impact (2.31x) on task performance while target selection difficulty did not (1.37x). In both instances, but more so for TSD, we see over lapping HPDs or credible intervals. This is expected in a learning setting given the variability in participant abilities, and does not negate our findings. In a learning setting, a 2.31x difference between two conditions is highly significant. Highlighting the importance of factoring in even lower-level characteristics in system design. Beyond pedagogy and higher-level factors like level of immersion, designers need to be aware of the impacts of these lower-level characteristics on learning outcomes.

The observed benefits of interaction delays in VR learning may be partially explained by several cognitive theories. The 5-s delay could encourage deeper engagement with content, aligning with the Levels of Processing theory (Craik and Lockhart (1972)), and introduce 'desirable difficulties' (Bjork (1994)), enhancing learning efficiency. Additionally, these delays might focus learners' attention, reducing extraneous load and optimizing learning conditions (Hollender et al. (2010)). These theories suggest that strategic design elements in VR could effectively balance cognitive challenges and support, potentially enhancing learning outcomes. Further exploration into how these theoretical frameworks apply to VR learning environments is warranted.

6.1. What do our findings mean for hands on learning in VR?

Our findings predominantly apply to environments where raycasting is used for object selection and interaction with UIs. But VR allows for embodied and fully interactive learning scenarios and the goal is often to create the most interactive and hands on training possible. So how would we apply these findings to a hands on environment where the user is interacting with objects beyond raycasting? Delay can be applied in a hands on scenario by pacing the learning content - eg. making the user pause before allowing them to interact with the next object, either by using instructional signs, or loading scenes. VR scenarios often naturally contain instances when a new scene has to load, and this is often a challenge where designers use different tactics to minimize the loading time to prevent frustration. Our findings however, are informing designers that minimizing delay is not always necessary, instead strategically placed loading screens will help users to pause and reflect.

Extensive testing is necessary to find the optimal timing of delays, especially if the training is long and has multiple steps. Too many instances of delays may become distracting and bothersome - instead we recommend thinking about the most important and crucial learning steps in the training scene. Similar to our discrimination index, the most important learning moments may be the only ones that require delay and time for reflection. However, it is important for the user to know the delay is intentional and not a bug as otherwise it could introduce confusion and lead to negative learning outcomes.

Beyond delay and target selection difficulty, other low-level characteristics should be considered when designing learning experiences. For example, how does walking versus being stationary impact learning? How does hands-on interaction versus pointing with a raycast influence learning? Are these time consuming and expensive development characteristics necessary for optimal learning?

The themes extracted from the interview results shed further light on our findings.

6.2. Observed themes

The interaction delay, a factor often perceived negatively in traditional system designs, surprisingly fostered enhanced learning outcomes. We identified two major themes in our analysis that may explain the impact delay had on users:

Delay Enhances Rehearsal Time: Participants reported that the interaction delay granted them additional time to process and rehearse the information. This is consistent with the cognitive theory positing that rehearsal within the working memory can bolster the migration of information to long-term memory, thereby enhancing learning (Ruchkin et al. (2003)). The delay thus seems to facilitate cognitive engagement, turning a seemingly disadvantageous feature into a valuable learning aid.

Distraction Due to Undefined Delay Duration: However, a small group of participants also noted the lack of a visible end-point to the blinking delay could be distracting, causing some to lose focus. This is an important insight, suggesting that interaction delay, while beneficial, should be thoughtfully implemented, with clear duration indicators to reduce cognitive load and avoid periods of mental blankness (Sweller (1994)).

The target selection difficulty surfaced another set of themes, emphasizing the dual role it played in the learning experience.

Interactivity Over Learning Content: Participants noticed that their attention was on the challenge of interacting with the yellow spheres, overshadowing the actual learning content. This suggests that while increased difficulty may enhance intrinsic motivation and engagement, it may inadvertently divert attention away from the primary learning goal (Lomas et al. (2017); Csikszentmihalyi (1990)).

Engaging and Fun Learning Experience: Despite the potential distraction, participants enjoyed the interactivity offered by the higher difficulty TSD. It contributed to a more engaging learning experience.

Distracting from Learning Content: However, this advantage is somewhat offset by the third theme, suggesting that the interactive and difficulty TSD feature could also distract from the main learning content.

Overall, our findings suggest a nuanced relationship between interaction delay, target selection difficulty, and learning outcomes. They underline the effects of these features on learning, highlighting the need to consider low-level interaction characteristics in the design of educational experiences to avoid cognitive overload and distraction from the learning objectives.

6.3. Design implications and recommendations

Our research into interaction delay and target selection difficulty (TSD) in a Virtual Reality (VR) learning environment provides valuable insights for designers and educators alike. Here, we distil key design implications from our study.

6.3.1. Strategically implement interaction delays

Contrary to prevailing beliefs, our research shows that interaction delays in VR can enhance learning outcomes, provided they are well-implemented. Our results demonstrate that a modest 5-s delay gives learners additional time to process and rehearse information, reinforcing their understanding. While it might appear counter-intuitive, this strategic 'slowing down' can promote 'forced learning,' which bolsters retention. Designers should thus consider incorporating intentional delays into VR environments when crafting educational content. Nevertheless, it's important to acknowledge that a small subset of our participants found the delay distracting, hinting at the need for personalized learning pathways where such factors could be adjustable according to the learner's preferences or performance.

6.3.2. Rethink the role of target selection dificulty

Our findings challenge the conventional understanding of the Target Selection Difficulty (TSD) within a VR context. Despite variations in perceived difficulty between the "hard" and "easy" conditions, we observed no significant impact of TSD on learning outcomes. Interestingly, participants reported the "hard" TSD condition to be more engaging. This suggests that designers should view the TSD not just as a parameter to be minimized but as a tool for shaping learner engagement. It is necessary to conduct further investigations to better understand the nuanced role of TSD in VR learning environments.

6.3.3. Consider cognitive load theory in VR design

Our results have implications for how we understand and apply Cognitive Load Theory (CLT) in the context of VR learning. Specifically, our study suggests that increasing extraneous cognitive load, often viewed negatively, may be carefully managed through manipulating lower-level interaction characteristics of the experience. Similarly intrinsic and germane cognitive loads, may also be influenced similarly. Designers should, therefore, adopt a more nuanced approach when dealing with cognitive load in VR learning, balancing the need for engagement and interactivity with effective cognitive load management.

7. Limitations & future work

Reflecting upon the results of our study and feedback from participants, specific limitations and opportunities for future research emerge.

A key limitation in our experiment lies in the efficacy of our operationalization of the Target Selection Difficulty (TSD). While we aimed to create a significantly different experience in terms of cognitive demand between the Easy and Hard TSD conditions, participant feedback indicated a notable disparity between our expectations and their experiences. Participants rated the difficulty of the Easy TSD as 2 out of 10, which aligns with our expectations. However, the Hard TSD was rated only 5 out of 10, falling short of our goal of creating a high-difficulty experience at 9 or above. This discrepancy suggests that our design may have not sufficiently increased cognitive demand or offered an adequately challenging scenario for the Hard TSD condition.

In future research, this limitation opens up an important opportunity: to adjust and calibrate the Hard TSD condition to provide a more challenging interaction experience. Future studies could experiment with smaller interactive buttons or more complex navigation, potentially leading to a better understanding of the effects of a genuinely high-difficulty VR interaction scenario on cognitive load and learning outcomes.

A second noteworthy limitation of our study was the number of discriminating tasks presented to participants. Future studies should attempt to ensure a greater selection of tasks of varying levels of difficulty to increase the number of sufficiently discriminating tasks for modelling.

8. Conclusion

We explored the impact of lower-level interaction characteristics on learning in VR. We specifically evaluated the impact of interaction delay and target selection difficulty on learning outcomes in a VR table saw training module. We found that a short 5-s interaction delay improved learning outcomes. The qualitative results show that the delay gave participants more time to pause, think and reflect. Finally, the target selection difficulty had no impact on learning outcomes in this context. It was evident in our interviews that many participants found the hard TSD condition more enjoyable and engaging. We conclude by noting that it is essential for VR learning designers to take into account how seemingly minor and potentially counter-intuitive characteristics can drastically impact learning outcomes. Instead these findings can be used strategically to enhance learning in VR. Further work must be done to fully understand concepts such as slow cognition and forced learning to truly grasp the impact of slowing down learning to minimize extraneous processes in VR environments.

CRediT authorship contribution statement

Sara Khorasani: Writing – review & editing, Writing – original draft, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Stephain Hsu:** Writing – original draft, Data curation. **Rui Guan:** Formal analysis. **Jorge Goncalves:** Writing – review & editing, Supervision, Conceptualization. **Andrew Irlitti:** Supervision. **Jarrod Knibbe:** Writing – review & editing, Supervision, Conceptualization. **Eduardo Velloso:** Writing – review & editing, Supervision, Conceptualization.

Table 6
Leave-One-Out Cross-Validation (LOO CV) results for the bayesian model.

Metrics	Model Estimate	Model SE
Expected Log Pointwise Predictive Density (elpd_loo)	241.9	6.5
Effective Number of Parameters (p_loo)	13.9	0.3
LOO Information Criterion (looic)	−483.8	12.9

Table 7
Watanabe-Akaike information criterion (WAIC) results for the bayesian model.

Metrics	Model Estimate	Model SE
Expected Log Pointwise Predictive Density (elpd_waic)	241.9	6.5
Effective Number of Parameters (p_waic)	13.8	0.3
Watanabe-Akaike Information Criterion (waic)	−483.8	12.9

Declaration of competing interest

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

A. Appendix.

A.1. Bayesian Model Code

Test Tasks	Discrimination Index
Hazard Removal	0.41
Exhaust On	0.09
Exhaust Off	0.14
Glasses On	0.31
Riving Knife	0.73
Wood Placement	0.19
Fence Unlock	0.36
Fence Position	0.82
Fence Lock	0.73
Blade Height	0.63
Blade Angle	0.70
Saw On	0.04
Saw Off	0.09
Push Stick	0.29
Mitre Gauge	0.29
Push Both	0.36
Push Through	0.51
Cut Hands	0.53
Cut Stick	0.48
Grab Before Off	0.63

A.2. Model Validation Details

We evaluated the effectiveness of our model using two widely accepted criteria in Bayesian statistics, namely the Leave-One-Out Cross-Validation (LOO CV) and the Watanabe-Akaike Information Criterion (WAIC).

Table 6 presents the LOO CV results. Our model demonstrated a high Expected Log Pointwise Predictive Density (elpd_loo = 241.9) and a low LOO Information Criterion (looic = −483.8). Moreover, the Effective Number of Parameters (p_loo = 13.9) shows the model's minimal complexity, thus hinting at a balance between model accuracy and complexity.

Consistently, the WAIC results in Table 7 mirror our LOO CV findings. Our model showcased a high Expected Log Pointwise Predictive Density (elpd_waic = 241.9) and a low Watanabe-Akaike Information Criterion (waic = −483.8). The Effective Number of Parameters (p_waic = 13.8) reaffirms the model's simplicity. Thus, our model exhibits robust performance and reasonable complexity, making it a fitting choice according to both LOO CV and WAIC.

References

Abuhamdeh, S., & Csikszentmihalyi, M. (2012). The importance of challenge for the enjoyment of intrinsically motivated. *Goal-Directed Activities Personality and Social Psychology Bulletin, 38*, 317–330. https://doi.org/10.1177/0146167211427147

Allen, D., & Tanner, K. (2005). Infusing active learning into the large-enrollment biology class: Seven strategies, from the simple to complex. *Cell Biology Education, 4*, 262–268. https://doi.org/10.1187/cbe.05-08-0113

Barreda, S., & Silbert, N. (2023). *Bayesian multilevel models for repeated measures data*. London: Routledge. https://doi.org/10.4324/9781003285878

Baumgartner, E., Ferdig, R. E., & Gandolfi, E. (2022). Exploring the impact of extended reality (XR) on spatial reasoning of elementary students. *TechTrends, 66*, 825–836. https://doi.org/10.1007/s11528-022-00753-6

Biermann, O. C., Ajisafe, D., & Yoon, D. (2022). Interaction design for VR applications: Understanding needs for university curricula. In *CHI conference on human factors in computing systems extended abstracts* (pp. 1–7). New York, NY, USA: ACM. https://doi.org/10.1145/3491101.3519859.

Bjork, R. A. (1994). *Memory and metamemory considerations in the instruction of human beings*. Cambridge: MA MIT Press.

Bürkner, P. C. (2019). *Bayesian item response modeling in R with brms and stan*.

Caratachea, M., & Monty Jones, W. (2024). Making in virtual reality environments: A case study of K-12 teachers' perceptions on the educational affordances of virtual reality for maker-centered learning. *Educational Technology Research & Development, 72*, 155–180. https://doi.org/10.1007/s11423-023-10290-5

Chan, J. Y. C., Ottmar, E. R., & Lee, J. E. (2022). Slow Down to speed up: Longer pause time before solving problems relates to higher strategy efficiency. *Learning and Individual Differences, 93*, Article 102109. https://doi.org/10.1016/j.lindif.2021.102109

Chiavaroli, N., & Familari, M. (2011). When majority doesn't rule: The use of discrimination indices to improve the quality of MCQs. *Bioscience Education, 17*, 1–7. https://doi.org/10.3108/beej.17.8

Clark, L. D., Bhagat, A. B., & Riggs, S. L. (2020). Extending fitts' law in three-dimensional virtual environments with current low-cost virtual reality technology. *International Journal of Human-Computer Studies, 139*, Article 102413. https://doi.org/10.1016/j.ijhcs.2020.102413

Craik, F. I., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience. *Journal of Leisure Research, 24*, 93–94.

Dalsgaard, T. S., Knibbe, J., & Bergström, J. (2021). Modeling pointing for 3D target selection in VR. In *Proceedings of the 27th ACM symposium on virtual reality software and technology* (pp. 1–10). New York, NY, USA: ACM. https://doi.org/10.1145/3489849.3489853.

Drey, T., Albus, P., der Kinderen, S., Milo, M., Segschneider, T., Chanzab, L., Rietzler, M., Seufert, T., & Rukzio, E. (2022). Towards collaborative learning in virtual reality: A comparison of Co-Located symmetric and asymmetric pair-learning. In *Proceedings of the 2022 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3491102.3517641. doi:10.1145/3491102.3517641.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381–391. https://doi.org/10.1037/h0055392

Garcia, S., Kauer, R., Laesker, D., Nguyen, J., & Andujar, M. (2019). A virtual reality experience for learning languages. In *Extended abstracts of the 2019 CHI conference on human factors in computing systems*. New York, NY, USA: ACM. https://doi.org/10.1145/3290607.3313253.

Grahn, H., & Kujala, T. (2020). Impacts of touch screen size, user interface design, and subtask boundaries on In-Car task's visual demand and driver distraction. *International Journal of Human-Computer Studies, 142*, Article 102467. https://doi.org/10.1016/j.ijhcs.2020.102467

Henrikson, R., Grossman, T., Trowbridge, S., Wigdor, D., & Benko, H. (2020). Head-coupled kinematic template matching: A prediction model for ray pointing in VR. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–14). New York, NY, USA: ACM. https://doi.org/10.1145/3313831.3376489.

Hernando, M. (2011). Student procedural knowledge inference through item response theory. In *UMAP'11: Proceedings of the 19th international conference on user modeling, adaption, and personalization* (pp. 426–429).

Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction. *Computers in Human Behavior, 26*, 1278–1288. https://doi.org/10.1016/j.chb.2010.05.031

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI, 5*. https://doi.org/10.3389/frobt.2018.00081

Kay, M., Nelson, G. L., & Hekler, E. B. (2016). Researcher-centered design of statistics. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 4521–4532). New York, NY, USA: ACM. https://doi.org/10.1145/2858036.2858465.

Khajah, M., Huang, Y., Gonzalez-Brenes, J., Mozer, M., & Brusilovsky, P. (2014). Integrating knowledge tracing and item response theory: A tale of two frameworks. In *CEUR workshop proceedings*.

Khorasani, S., Victor Syiem, B., Nawaz, S., Knibbe, J., & Velloso, E. (2023). Hands-on or hands-off: Deciphering the impact of interactivity on embodied learning in VR. *Comp. Educ. Real., 3*. https://doi.org/10.1016/j.cexr.2023.100037. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85206149447&doi=10.1016%2fj.cexr.2023.100037&partnerID=40&md5=2089364e15c0bf655a0c5c3afad0dadf.

Koob, G. F., & Volkow, N. D. (2016). Neurobiology of addiction: A neurocircuitry analysis. *The Lancet Psychiatry, 3*, 760–773. https://doi.org/10.1016/S2215-0366(16)00104-8

Krauß, V., Boden, A., Oppermann, L., & Reiners, R. (2021). Current practices, challenges, and design implications for collaborative AR/VR application development. In *Proceedings of the 2021 CHI conference on human factors in computing systems* (pp. 1–15). New York, NY, USA: ACM. https://doi.org/10.1145/3411764.3445335.

Lomas, J. D., Koedinger, K., Patel, N., Shodhan, S., Poonwala, N., & Forlizzi, J. L. (2017). Is difficulty overrated?. In *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 1028–1039). New York, NY, USA: ACM. https://doi.org/10.1145/3025453.3025638.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction, 7*, 91–139. https://doi.org/10.1207/s15327051hci0701{_}3

Matviienko, A., Müller, F., Zickler, M., Gasche, L. A., Abels, J., Steinert, T., & Mühlhäuser, M. (2022). Reducing virtual reality sickness for cyclists in VR bicycle simulators. In *CHI conference on human factors in computing systems* (pp. 1–14). New York, NY, USA: ACM. https://doi.org/10.1145/3491102.3501959.

Mayer, R. E., Makransky, G., & Parong, J. (2022). The promise and pitfalls of learning in immersive virtual reality. *International Journal of Human-Computer Interaction*, 1–10doi. https://doi.org/10.1080/10447318.2022.2108563

McElreath, R. (2020). *Statistical rethinking*. Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

Murphy, D., & Higgins, C. (2019). *Secondary inputs for measuring user engagement in immersive VR education environments*. ArXiv abs/1910.01586.

Nielsen, J. (1993). *Usability engineering*. Boston: Academic Press.

Norman, D. A. (2002). *The design of everyday things*. New York, NY: Basic Books, Inc. Division of HarperCollins.

Nowell, L. S., Norris, J. M., White, D. E., & Moules, N. J. (2017). Thematic analysis. *International Journal of Qualitative Methods, 16*, Article 160940691773384. https://doi.org/10.1177/1609406917733847

Nukarinen, T., Kangas, J., Rantala, J., Koskinen, O., & Raisamo, R. (2018). Evaluating ray casting and two gaze-based pointing techniques for object selection in virtual reality. In *Proceedings of the 24th ACM symposium on virtual reality software and technology* (pp. 1–2). New York, NY, USA: ACM. https://doi.org/10.1145/3281505.3283382.

Oagaz, H., Schoun, B., & Choi, M. H. (2021). Performance improvement and skill transfer in table tennis through training in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*. https://doi.org/10.1109/TVCG.2021.3086403

Oberdörfer, S., Heidrich, D., & Latoschik, M. E. (2019). Usability of gamified knowledge learning in VR and Desktop-3D. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13). New York, NY, USA: ACM. https://doi.org/10.1145/3290605.3300405.

Oliveira dos Santos, W., Bittencourt, I. I., Isotani, S., Dermeval, D., Brandão Marques, L., & Frango Silveira, I. (2018). Flow theory to promote learning in educational systems: Is it really relevant? *Revista Brasileira de Informática na Educação, 26*, 29. https://doi.org/10.5753/rbie.2018.26.02.29

Ouellette, M., Breeding, L., & Clark, C. (2019). Using applied cognitive load theory and difficulty analysis for educational game design for understanding and transference of literacy skills in adults. In *Proceedings of the 14th international conference on the foundations of digital games* (pp. 1–11). New York, NY, USA: ACM. https://doi.org/10.1145/3337722.3337725.

Paas, F., Renkl, A., & Sweller, J. (2003a). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist, 38*, 1–4. https://doi.org/10.1207/S15326985EP3801{_}1

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003b). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist, 38*, 63–71. https://doi.org/10.1207/S15326985EP3801{_}8

Pan, Y., & Steed, A. (2019). Avatar type affects performance of cognitive tasks in virtual reality. In *Proceedings of the 25th ACM symposium on virtual reality software and technology*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3359996.3364270. doi:10.1145/3359996.3364270.

Petersen, G. B., Mottelson, A., & Makransky, G. (2021). Pedagogical agents in educational VR: An in the wild study. In *Proceedings of the 2021 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411764.3445760. doi:10.1145/3411764.3445760.

Petersen, G. B., Petkakis, G., & Makransky, G. (2022). A study of how immersion and interactivity drive VR learning. *Computer Education, 179*. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-85122255182&doi=10.1016%2fj.compedu.2021.104429&partnerID=40&md5=ff75c7d975ebd94582feb0bf8659ebc0. doi: 10.1016/j.compedu.2021.104429.

Radu, I., Joy, T., & Schneider, B. (2021). Virtual makerspaces: Merging AR/VR/MR to enable remote collaborations in physical maker activities. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1–5). New York, NY, USA: ACM. https://doi.org/10.1145/3411763.3451561.

Radu, I., & Schneider, B. (2023). Designing augmented reality for makerspaces: Guidelines, lessons and mitigation strategies from 5+ years of AR educational projects. *Computers & Education: X Reality, 2*, Article 100026. https://doi.org/10.1016/j.cexr.2023.100026

Ratcliffe, J., & Tokarchuk, L. (2022). Rich virtual feedback from sensorimotor interaction may harm, not help, learning in immersive virtual reality. In *Proceedings of the 28th ACM symposium on virtual reality software and technology*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3562939.3565633. doi:10.1145/3562939.3565633.

Rettinger, M., Müller, N., Holzmann-Littig, C., Wijnen-Meijer, M., Rigoll, G., & Schmaderer, C. (2021). VR-Based equipment training for health professionals. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3411763.3451766. doi:10.1145/3411763.3451766.

Riding, R. (2013). School learning and cognitive styles. *David Fulton Publishers*. https://doi.org/10.4324/9780203821985

Rogers, C., El-Mounaryi, H., Wasfy, T., & Satterwhite, J. (2018). Assessment of STEM e-Learning in an immersive virtual reality (VR) environment. *The ASEE Computers in Education (CoED) Journal*.

Ruchkin, D. S., Grafman, J., Cameron, K., & Berndt, R. S. (2003). Working memory retention systems: A state of activated long-term memory. *Behavioral and Brain Sciences, 26*, 709–728. https://doi.org/10.1017/S0140525X03000165

Schmettow, M. (2021). In *New statistics for design researchers*. Springer.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology, 80*, 1–27. https://doi.org/10.1152/jn.1998.80.1.1

Schwarze, A., Kampling, H., Heger, O., & Niehaves, B. (2019). Is virtual reality the future of learning? A critical reflection. In *Proceedings of the 52nd Hawaii international conference on system sciences*. https://doi.org/10.24251/hicss.2019.214

Shi, A., Huo, F., & Han, D. (2021). Role of interface design: A comparison of different online learning system designs. *Frontiers in Psychology, 12*. https://doi.org/10.3389/fpsyg.2021.681756

Shneiderman, B. (1984). Response time and display rate in human performance with computers. *ACM Computing Surveys, 16*, 265–285. https://doi.org/10.1145/2514.2517

Sweetser, P., & Rogalewicz, Z. (2020). Affording enjoyment in VR games: Possibilities, pitfalls, and perfection. In *32nd Australian conference on human-computer interaction* (pp. 55–64). New York, NY, USA: ACM. https://doi.org/10.1145/3441000.3441050.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*, 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

Sweller, J., van Merrienboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review, 10*, 251–296. https://doi.org/10.1023/A:1022193728205

Szumilas, M. (2010). Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry = Journal de l'Academie canadienne de psychiatrie de l'enfant et de l'adolescent, 19*, 227–229.

Thaman, R. G. (2014). Effective use of pause procedure to enhance student engagement and learning. *Journal of Clinical and Diagnostic Research*. https://doi.org/10.7860/JCDR/2014/8260.4691

Theeuwes, J. (2010). Top–down and Bottom–Up control of visual selection. *Acta Psychologica, 135*, 77–99. https://doi.org/10.1016/j.actpsy.2010.02.006

Thorndike, E. L. (1918). The nature, purposes, and general methods of measurements of educational products. *Teachers College Record: The Voice of Scholarship in Education, 19*, 16–24. https://doi.org/10.1177/016146811801900702

Triantafyllidis, E., & Li, Z. (2021). The challenges in modeling human performance in 3D space with fitts' law. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems* (pp. 1–9). New York, NY, USA: ACM. https://doi.org/10.1145/3411763.3443442.

Ustun, A. B., Zhang, K., Karaoğlan-Yilmaz, F. G., & Yilmaz, R. (2022). Learning analytics based feedback and recommendations in flipped classrooms: An experimental study in higher education. *Journal of Research on Technology in Education*, 1–17doi. https://doi.org/10.1080/15391523.2022.2040401

Vygotsky, L. S. (1980). *Mind in society*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9vz4

Wickens, C. D., Helton, W. S., Hollands, J. G., & Banbury, S. (2021). *Engineering psychology and human performance*. New York: Routledge. https://doi.org/10.4324/9781003177616