Evaluating Tutorial-Based Instructions for Controllers in Virtual Reality Games

DOMINIC KAO, Purdue University, USA
ALEJANDRA J. MAGANA, Purdue University, USA
CHRISTOS MOUSAS, Purdue University, USA

Virtual reality (VR) has disrupted the gaming market and is rapidly becoming ubiquitous. Yet differences between VR and traditional mediums, such as controllers that are visible in the virtual world, enable entirely new approaches to instruction. In this paper, we present four studies, each using a different VR game. Within each study, we compared three different modalities of tutorials: *Text* (text-only), *Text+Diagram* (text with controller diagrams), and *Text+Spatial* (text with controller tooltips appearing on top of the player's virtual controllers). Data from our studies show that the importance of tutorial modality depends greatly on game type. In a third-person shooter, *Text+Spatial* led to significantly higher controls learnability than *Text* and *Text+Diagram*, and also led to significantly higher performance, player experience, and intrinsic motivation than *Text*. In a puzzle game, *Text+Spatial* led to significantly higher controls learnability and performance than *Text*. Additionally, *Text+Diagram* led to significantly higher controls learnability than *Text*. However, in a wave shooter and a rhythm game, differences between conditions were negligible on all measures. Our studies show that game type is an important factor to consider when designing tutorial modality.

CCS Concepts: • Human-centered computing~Empirical studies in HCI

KEYWORDS: VR; games; tutorials; text; diagrams; tooltips; instruction; interactive cues

1 INTRODUCTION

The virtual reality (VR) and augmented reality (AR) market is expected to reach $72.8 billion USD globally by 2024 [1]. Today, 78% of Americans are familiar with VR—a number that was 45% in 2015 [2]. Despite rapid growth in virtually all industries (e.g., healthcare [3], manufacturing [4, 5], employee training [6]), research on how best to instruct users in VR is limited. Because tutorials are often a user's first encounter with a game, effective tutorials are crucial for retaining players [7]. As Andersen et al. have noted, designers of modern games (e.g., VR games) rely on intuition, personal experience, existing examples, and user testing to create tutorials [7]. Although these games implement a large variety of different tutorial styles, the relative effectiveness of those tutorial styles is not well understood [7]. However, understanding
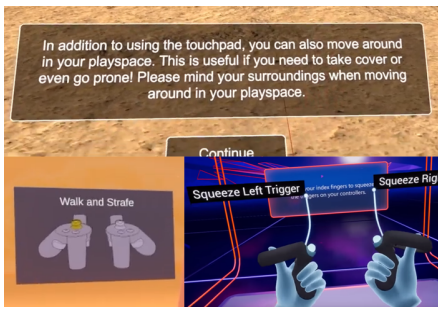
Figure 1. Tutorials from Onward (top)—text-only;
Windlands 2 (bottom-left)—text with diagrams;
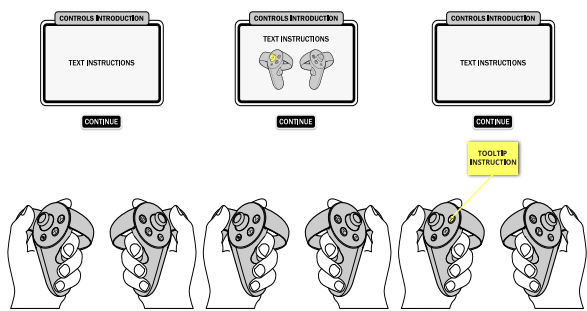Oculus First Steps (bottom-right)—text with tooltips.

Figure 2. The three conditions we studied (left-to-right):
Text; *Text+Diagram*; *Text+Spatial*.

how different tutorial implementations influence users would allow designers to save resources and help maximize learnability in VR games.

Tutorials can be crucial for increasing engagement in games. In one study, tutorials (compared to no tutorials) increased play time by 29% and in-game progress by 75%, but only in the most complex game [7]. In a VR wave shooter game, Frommel et al. found that tutorials presented just-in-time (i.e., at the same time that the mechanics can be practiced) can foster greater positive affect and motivation [8]. While it is clear that tutorials significantly affect players, VR games use a variety of different approaches in presenting their tutorials (see Figure 1 for examples), and there do not currently exist any studies on tutorial modality in VR games. Given that tutorials are a common staple of games [9], it is crucial to understand whether or not there are trade-offs in these approaches. Importantly, certain approaches (such as having controller tooltips that highlight relevant controller buttons on the player's virtual controllers) have only become more widespread within VR. We chose to focus our study on three specific modalities of tutorials: text (*Text*); text with controller diagrams (*Text+Diagram*); and text with controller tooltips (*Text+Spatial*). See Figure 2, which shows these modalities. We examined the effects of each of these modalities in four between-subjects experiments across four different VR games: Meme Dragons (a wave shooter), Beat Miner (a rhythm game), Hammer 2 (a third-person shooter), and Hack.VR (a puzzle game). We chose four different game genres because tutorial effectiveness may vary across game type. We measured three aspects of gameplay: *controls learnability* (how hard were the controls to learn?), *performance* (how well did players perform?), and *engagement-related outcomes* (how engaged were players?).

We present results from 433 participants demonstrating that tutorial modality is important in two of the four games studied. In a third-person shooter, *Text+Spatial* led to significantly higher controls learnability than *Text* and *Text+Diagram*, and also led to significantly higher performance, player experience, and intrinsic motivation than *Text*. In a puzzle game, *Text+Spatial* led to significantly higher controls learnability and performance than *Text*. *Text+Diagram* additionally led to significantly higher controls learnability than *Text*. However, we found that tutorial modality had a negligible effect in the wave shooter and the rhythm game. Our results suggest that tutorial modality is an important aspect of tutorial design for VR games which have higher controls complexity, and the tutorial's influence can extend to performance and engagement-related outcomes as a whole.

2 RELATED WORK

Virtual reality has long been touted as having potential to transform activities ranging from video games to educational applications [10, 11]. Through immersion, VR allows users to feel as if they are actually present in a computer-generated world [12]; this can make for highly engaging gameplay experiences [13]. Nevertheless, effective interactions in VR are the basis for engaging experiences [14]. In VR games, it is extremely common for the game to begin with a tutorial, teaching the player how to play [15]. However, very little work exists on how different modalities of these tutorials influence players' experiences. Here, we survey existing research on multimedia learning and tutorials in VR.

2.1 Multimedia Learning

For this work, we draw upon the cognitive theory of multimedia learning [16], which broadly describes the most effective principles in designing multimedia for learners. This theory is based on three core cognitive principles. The *dual channels* principle argues for the existence of separate channels for processing visual and audial information. The *limited capacity* principle argues that there is a limited capacity for each of these channels at any given time. The *active processing* principle argues that learning occurs through a process involving filtering, selecting, organizing, and integrating information based on learners' existing knowledge. In other words, learning is situated both in the learner's pre-existing context, and learners have a limited capacity for incoming information. This limited capacity is often referred to as cognitive load.

2.1.1 Cognitive Load. Cognitive load refers to the amount of working memory being used [17]. Multimedia learning theory argues that there are three important types of cognitive processing that affect cognitive load: 1) *extraneous processing*, which is the processing of information that is not directly related to the learning objective – e.g., pictures of lightning when learning about how lightning strikes; 2) *essential processing*, which refers to selecting information that is relevant to the learning objective; and 3) *generative processing*, which involves mentally organizing new information into a coherent structure and integrating that information with preexisting knowledge. All three types of processing increase cognitive load. Good multimedia learning design should aim to minimize extraneous processing, manage essential processing, and foster generative processing [18]. Based on the cognitive theory of multimedia learning, researchers have derived several key principles to follow when it comes to designing multimedia learning. We summarize several relevant principles below.

2.2.2 Use Words + Graphics Rather Than Words Alone. One key principle of multimedia learning is to combine graphics with words, rather than using words alone. However, there exist several types of graphics in multimedia learning [19]. *Decorative graphics* are decorative and do not improve the instructional message (e.g., a picture of a person riding a bicycle in a lesson on bicycle pumps). *Representational graphics* consist of a single graphic representing the appearance of an object (e.g., an image of a bicycle pump). *Relational graphics* show quantitative relationships (e.g., a line graph showing age on the x-axis and probability of a bicycle accident on the y-axis). *Organizational graphics* show qualitative relationships (e.g., a bicycle pump diagram with each part labeled). *Transformational graphics* show changes over time (e.g., a series of annotated frames depicting how a bicycle pump works). *Interpretive graphics* illustrate theories or principles through making invisible relationships visible (e.g., an animation showing the movement of air represented by small dots into the bicycle pump).

Multimedia learning theory recommends minimizing decorative graphics and representational graphics. Instead, multimedia learning theory recommends adopting images that help the learner understand material: organizational, transformational, and interpretive [18]. Organizational graphics, or diagrams, that annotate different controller buttons with their purpose are commonly found in VR tutorials. Empirical evidence for this multimedia learning principle is significant. Across 11 different studies, it has been found that test performance is better for students who learn from text material augmented with graphics [16]. Studies had varied learning material including lightning, vehicle brake systems, air pumps, and electrical generators. Words and graphics produced a 55% to 121% increase in the number of correct solutions on test knowledge. Other studies on learning material on the human heart [20], flight [21], and the life cycle of stars [22] have shown similar results. This principle is widely accepted and established. This principle also suggests that combining diagrams with text would provide more help in VR games compared to text alone.

2.1.3   Align Words to Graphics. Another important multimedia learning principle is to align words that describe graphics close to the graphics themselves. A number of studies support this principle – e.g., studies on learning about lightning formation and vehicle brake systems [23–25], technical training tasks [26], physical therapy training [27], and learning about the human heart [28]. In a systematic review of 50 studies, this principle was supported with an average effect size of 0.72 [29]. This principle suggests that any text describing graphics in VR tutorials should appear close to the graphics themselves. For example, instructional text should appear directly alongside diagrams. We might also consider the player's controllers themselves (in VR) as a type of "graphic," in which case the principle suggests that any tooltips describing controller buttons should appear close to the controllers. This also aligns with the Gestalt law of proximity, in which objects that are closer together will be seen as being grouped together [30].

2.1.4   Coherence Principle: Extra Material Hurts Learning. Another principle of multimedia learning theory is to avoid extra material that is unnecessary to the instructional goal (the coherence principle). This includes both extraneous words and extraneous graphics [18]. Extraneous graphics are also called *seductive details* in the literature, and researchers have posited that these images interfere with cognitive processing in three ways [31]: *distraction* (guiding the learner's attention towards unnecessary material), *disruption* (preventing the learner from building a mental model because of unnecessary material), and *seduction* (priming knowledge from an irrelevant domain). There are several studies that support the coherence principle [18, 20]. For example, Mayer et al. [32] found that inserting high-interest details about the instructional topic (e.g., "...people who make love once or twice a week are more immune to colds than folks who abstain from sex...") versus low-interest details (e.g., "A virus is about 10 times smaller than a bacterium, which is approximately 10 times smaller than a typical human cell...") decreased problem-solving transfer in two different experiments. There are several possible implications of this principle for VR. For example, because a visual representation of VR controllers already exists in the virtual world (in the player's hands), additional in-the-world diagrams could be viewed as unnecessary. Leveraging these virtual controllers could facilitate a reduction in cognitive processing since interpreting the diagrammatic information (such as mentally mapping that diagram to the player's own controllers) is no longer needed.

2.1.5   Signaling Principle: Using Visual Cues. The signaling principle, which is the last principle that we summarize, involves using visual cues to direct learners' attention (the signaling

principle). Visual cues have been shown to speed up the learning of information [33] and can be done in different ways, including using arrows, a larger text size, bolded text, and color. There is a preliminary body of literature that supports this principle [18]; however, the principle is less established in the literature than the prior principles reviewed. There is, however, evidence from HCI that such signaling techniques can prove useful. For example, Kelleher and Pausch used "Stencils-based" tutorials, in which they created translucent colored stencils containing a hole to direct the user's attention to a specific interface element. Additionally, they use an arrow that points to the interface element and a sticky note to describe its purpose. They find that this technique improves tutorial completion time, reduces errors, and reduces reliance on human assistance [34]. Conceptually, this approach is similar to the tooltips studied in this paper, which direct attention towards a specific controller button and explain its purpose. More recently, Dillman et al. [35] created a framework for interaction cues in virtual and augmented reality. Based on an analysis of 49 video games, Dillman et al. define an interaction cue as being a signal to tell users what is interactable, where to look, or where to go. Specifically, interaction cues can differ in their *purpose*, *markedness* (visual design), and *trigger* (context that leads to the cue appearing). In virtual reality training, studies have shown that cues that appear immediately as the user is expected to perform an action lead to faster learning and better retention of learned material as compared to cues that are time delayed [36]. Our tooltips in this paper can be considered as a type of interaction cue. They get the attention of the user immediately by having users look at the tooltip and understand which controller button is being referring to (purpose); the button is then emphasized through a yellow highlight (markedness); and finally, the tooltips appear when the appropriate tutorial phase is reached (trigger). We chose this design to mimic how other VR applications have taught users – applications such as the official Oculus tutorial [37]. The signaling principle, in theory, can help free up cognitive resources since players do not need to determine what to focus on in the game themselves [38].

2.1.6 Applicability to VR. Research on multimedia learning theory has traditionally been done on flat 2D surfaces, such as computer displays. VR, however, creates new opportunities for instruction [39]. For example, through the head-mounted display (HMD) and handheld controllers common in many VR systems, it is possible to see virtual versions of the user's physical controllers. Studies have so far not investigated if these differences can be leveraged for more efficacious instruction. Our goal here is to investigate if novel instructional methods, such as using virtual tooltips on the user's virtual controllers, confers any benefits. Since prior studies have shown that tutorials can influence gameplay (e.g., playtime [40], emotion and motivation [8]), we are interested in how novel instructional methods can influence different gameplay factors. Moreover, multimedia learning theory principles may manifest differently in VR. For example, VR games can induce a higher level of immersion than their desktop game counterparts [41], which would suggest that players in VR have additional visual and auditory load. One study found that a more immersive educational game led to increased presence, which led to worse learning outcomes mediated through higher cognitive load [42]. Therefore, our work here is both of interest to how we create VR game tutorials and to multimedia learning theory applied to VR.

2.2 Tutorials in Games and VR Games

Research on tutorials in games has shown that the usefulness of tutorials depends greatly on the complexity of the game. Andersen et al. found that tutorials were justified in only the most

complex game, in which they increased play time by 29% [7]. Andersen et al. argue that tutorials may not be necessary for simpler games because their game mechanics can be discovered through experimentation. In a VR wave shooter game, Frommel, Fahlbusch, Brich, and Weber compared a tutorial that gave all of the instructions before practice to a tutorial in which the information was presented just-in-time as mechanics were being introduced. They found that the just-in-time tutorial led to higher positive emotions, lower negative motions, and higher motivation [8]. Kao compared different modalities of help facilities in game-making (text, interactive, intelligent agent, video), and found that both interactive help and video help had the highest positive impact on time spent, controls learnability, learning motivation, total editor activity, and game level quality [43]. In another study, Kao et al. found that teaching health-related knowledge through gameplay yielded similar outcomes to teaching the same knowledge through video, but that gameplay was more intrinsically motivating [44].

Hint systems have also been studied in games. O'Rourke et al. found that hint systems actually negatively impacted player performance overall compared to having no hints at all in an educational puzzle game [45]. Wauck and Fu performed a study in which they compared adaptive, automatic, and on-demand hint systems in a puzzle game. In addition to performance, they also studied player experience. They found that in some cases, there existed a type of placebo effect in which players (and especially novices) who won were more likely to attribute their good performance to hints regardless of whether the hints actually helped [46]. Similarly to Wauck and Fu, we decided to incorporate additional measures beyond performance in our study. Specifically, we look at both engagement-related measures and controls learnability.

User interface design in VR is also applicable to designing tutorials. Rzayev, Mayer, Krauter, and Henze studied the effect of notification placement location in VR [47]. They compared four conditions: notifications anchored to the headset, notifications anchored to the controller, notifications appearing floating in-the-world, and notifications appearing on a wall in the environment. Headset-anchored notifications were the most intrusive, but also least likely to be missed. Notifications appearing on a wall were easily missed. Notifications anchored to the controller or floating in-the-world resulted in the highest usability scores. Alexandrovsky et al. looked at several design choices for in-VR questionnaires [48]. They compared questionnaires in VR along two dimensions: anchoring (in the world vs. anchored to the body) and interaction method (laser-pointer control vs. VR controller trackpad). Completion time, usability scores, and interviews clearly showed that world anchoring with laser-pointer control was the optimal condition.

Almost all VR experiences are preceded by some form of instruction [49, 50]. These instructions can take on different modalities, such as text [51], text and pictures [52, 53], and text and animations [54–56]; additionally, some have additional verbal instructions [21, 28]. For example, in a VR game that teaches Japanese language and culture, Chen, Yang, and Andersen created a tutorial with minimal text that would prompt the player to bow. The game successfully taught players how and when to bow, and increased players' sense of involvement in Japanese culture [57]. Nevertheless, little is known about the relative efficacy of different tutorial modalities in VR. This is despite that most commercial VR games contain tutorials. Popular examples include Skyrim VR (Bethesda Softworks, 2017), Beat Saber (Beat Games, 2019), Robo Recall (Epic Games, 2017), and Windlands 2 (Psytec Games Ltd, 2018), which all have tutorial levels prior to gameplay in varying instructional modalities. While it is well accepted that tutorials impact game players in some contexts, there is a lack of research on the relative effectiveness of different tutorial modalities. Further research in this domain can help

develop guidelines that will help game designers and researchers make more informed design decisions.

3 EXPLORING EXISTING HELP MODALITIES IN GAMES

We first performed a directed content analysis of 50 existing VR games using the search phrase "popular VR games" in the Google search engine in mid-2019[1] [58]. Our goal was to determine common tutorial modalities used in popular VR games. The games involved a mix of shooting games (e.g., Fallout 4 VR), puzzle games (e.g., The Talos Principle), role-playing games (e.g., Skyrim VR), and other types (e.g., rhythm, racing, climbing). A little over half of the games surveyed had shooting as a core mechanic of the game (i.e., firing projectiles from a gun, bow, or some type of vehicle). For each of the games, two researchers studied the in-game tutorial either by playing the game itself, or by watching online gameplay videos. We played or watched each game until we felt we had a good sense of how gameplay worked and what controls players had to use in the game, taking a similar approach to Dillman et. al [35] (see section *Method and Analysis*). Researchers watched multiple different walkthrough gameplay videos for each game to ensure that we had significant coverage of all aspects of gameplay. We then used a deductive approach to derive codes based on the cognitive theory of multimedia theory learning through iterative discussions. From this deductive approach, we created the following core categories: *text* (text instructions), *verbal* (verbal instructions), and *diagram* (diagrams or images as part of the instruction). Based on the games themselves, we inductively [59] decided to add the following category as it did not fit into any of the other existing categories: *tooltip* (labeling the player's physical controller that appeared in VR). We used the signaling principle in multimedia theory learning as the main grounding theoretical principle for tooltip. Here, tooltip represents a unique aspect of VR in which the embodied interaction and virtual presence within the game world itself is leveraged, wherein the player's own virtual self (i.e., the virtual controllers) constitutes a form of visual scaffolding for learning the game's affordances. Finally, consistent with theory in the broader literature around diegetic help (i.e., learning the game being diegetically scattered through the core game play experience and not as a separate beginner level) versus a specific tutorial typically present at the beginning of the game or accessible through the menu [40], we added the additional categories of "Tutorial Help" and "Diegetic Help."

Next, anchor samples were created for each of the main categories, which three researchers used to code the games independently. A third researcher was added in this step to: 1) reduce the chances that a code was missed for any given game; and 2) ensure greater consistency. Coding was done by again playing or watching multiple walkthrough gameplay videos. After the first round of coding, researchers extensively discussed differences in coding, interpretation, and the use of categories. This round of coding produced refinement and greater specificity in how our categories were defined. All games were then iteratively recoded over 5 rounds of coding until all codes were mutually agreed upon and consensus was reached. We find that the majority of the games contain a tutorial level (90%). We formally defined a tutorial level as (1) a specific training session at the start of the game, (2) a first "level" in which the goal was to learn the game, or (3) a tutorial that was selectable from the main menu. 40% of the games studied

---

[1] As of April 28, 2021, 40% of the games on this 2019 list continue to appear on a Google search for "popular VR games." Of the remaining 60%, half of them are on Steam's top sellers for VR experiences [134]. Altogether, at minimum 70% of the games on this 2019 list are still widely popular as of April 28, 2021.

have some form of embedded help – e.g., loading screen hints, in-game hints posted on walls, just-in-time cues when the player was being introduced to a new mechanic. For coding modalities (Text, Verbal, Diagram, Tooltip), the modality could be present in either a tutorial or as diegetic help, so long as it was considered *instructional* in nature (i.e., it helped teach the player how to play the game). We wanted to understand broadly how help was being presented in VR games, regardless of whether it was in a tutorial. We found that a majority of games had some form of text help (88%), just over half had some form of diagrams or images (56%), and a small number leveraged the player's controllers that appeared in VR and used labels on those controllers to instruct the player (22%). Exactly half of the games studied used verbal instruction (50%). See Table 1.

Table 1: Fifty VR games and types of instruction contained in each game.

	Text	Verbal	Diagram	Tooltip	Tutorial Help	Diegetic Help		Text	Verbal	Diagram	Tooltip	Tutorial Help	Diegetic Help
Skyrim VR	✓	✗	✓	✓	✓	✓	Brass Tactics VR	✓	✓	✗	✗	✓	✗
Beat Saber	✓	✓	✗	✗	✓	✗	Red Matter	✓	✓	✓	✗	✓	✗
Astro Bot: Rescue Mission	✓	✗	✓	✗	✓	✓	Sairento VR	✓	✗	✓	✗	✓	✓
Onward	✓	✗	✗	✗	✓	✗	To The Top	✓	✓	✗	✗	✓	✓
Fallout 4 VR	✓	✗	✓	✓	✓	✓	Hellblade: Senua's Sacrifice	✓	✗	✗	✗	✗	✓
Superhot	✓	✗	✓	✗	✓	✓	Contractors VR	✗	✗	✗	✗	✓	✗
Lone Echo	✓	✓	✓	✓	✓	✓	In Death	✓	✗	✓	✗	✓	✗
Borderlands 2 VR	✓	✓	✓	✗	✓	✓	Arizona Sunshine	✓	✗	✗	✓	✓	✗
Wipeout Omega Collection VR	✓	✗	✗	✗	✗	✓	Alien Isolation VR	✓	✗	✗	✗	✓	✓
Rec Room	✓	✓	✓	✗	✓	✗	The Gallery II	✓	✗	✗	✗	✓	✗
Resident Evil 7 VR	✓	✗	✓	✗	✓	✗	Tetris Effect	✗	✗	✗	✗	✗	✗
Farpoint	✓	✓	✗	✓	✓	✗	London Heist VR	✗	✓	✗	✗	✓	✗
Robo Recall	✓	✓	✓	✓	✓	✓	Creed: Rise To Glory	✓	✗	✓	✗	✓	✗
Echo Arena	✓	✓	✓	✓	✓	✗	Ultrawings	✓	✓	✓	✗	✓	✓
Orbus VR	✓	✗	✓	✗	✓	✓	Blade & Sorcery	✓	✗	✗	✗	✓	✗
The Talos Principle	✓	✓	✓	✗	✓	✓	Windlands 2	✓	✓	✓	✗	✓	✗
GORN	✓	✓	✓	✗	✓	✗	Job Simulator	✓	✓	✓	✗	✓	✗
Pavlov VR	✓	✗	✗	✗	✓	✗	Payday 2 VR	✓	✓	✓	✗	✓	✓
Elite: Dangerous	✓	✓	✓	✗	✓	✓	Space Pirate Trainer	✓	✗	✗	✓	✓	✗
Subnautica	✓	✓	✓	✗	✓	✓	Vanishing Realms	✓	✗	✗	✓	✗	✓
Ace Combat 7	✓	✓	✗	✗	✗	✓	The Climb	✓	✓	✓	✗	✓	✗
Moss	✗	✗	✓	✗	✓	✗	The Persistence	✓	✓	✓	✗	✓	✗
Firewall: Zero Hour	✓	✓	✗	✗	✓	✗	BlazeRush	✓	✗	✗	✗	✓	✗
Rush of Blood	✗	✓	✓	✗	✓	✗	Serious Sam VR: First Encounter	✗	✗	✗	✓	✓	✗
From Other Suns	✓	✓	✓	✓	✓	✗	Star Trek: Bridge Crew	✓	✗	✓	✗	✓	✗

Figure 3. Screenshots of testbed games (clockwise from upper left): Meme Dragons; Beat Miner; Hack.VR; Hammer 2.
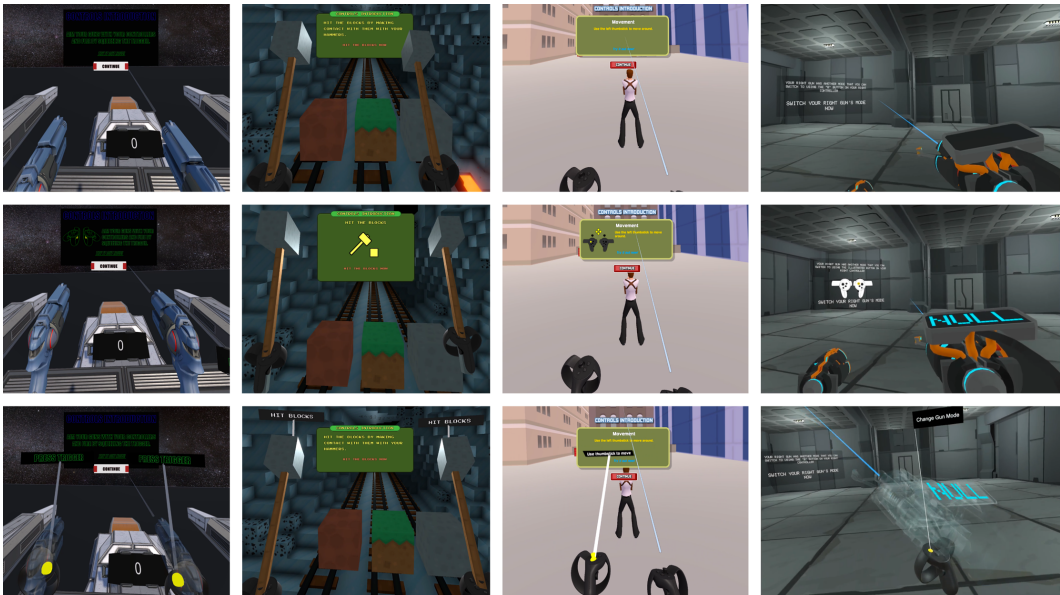

Figure 4. The three conditions *Text* (top row), *Text+Diagram* (middle row), and *Text+Spatial* (bottom row). Each action was taught, then practiced. In the pictured examples, practice involved (left-to-right): shooting 3 non-moving enemies, hitting 3 non-moving blocks, moving in each of 4 directions, and changing the gun's "mode." Practice was identical between conditions. High-resolution versions of each image can be found in the supplementary materials in section A.5.

Based on these findings, we decided to develop three experimental conditions that compare text, text with diagrams, and text with tooltips. It is important to note that many other permutations of these conditions are possible, e.g., diagrams and tooltips. However, these specific conditions were designed for ecological validity based on our content analysis (e.g., those seen in Figure 1), which would allow us to determine how typical implementations of tutorial modalities affect users. Additionally, multimedia learning theory would rule out certain conditions as increasing cognitive load due to redundancy, e.g., diagrams and tooltips which leverage the controllers as diagrams; although these hypotheses would still need to be confirmed in VR. These conditions are described in the next section.

4   EXPERIMENTAL TESTBED

Our experimental testbed consisted of 4 different VR games[2] (see Figure 3). Each of the 4 games comes from a distinct VR game genre, as classified by the developer (wave shooter, rhythm, third-person shooter, and puzzle). We selected these 4 VR games since they represented four distinct types of VR games. These genres were selected to represent popular VR game genres currently available [60][3]. Foxman, Leith, Beyea, and Ratan find that action (45%), shooter (30%), simulation (22%), adventure (13%), and puzzle (12%) are the most common VR game genres on Steam [60]. Our own selected games contain elements of action (3/4 games), shooter (2/4 games), adventure (2/4 games), and puzzle (1/4 games), while differing in their genres and complexity. Additionally, we selected these specific games because they all could be played in-place with arm movements only and could be played either standing or sitting. This was to deal with potential space constraints present in players' play spaces and to ensure greater conformity in experience [61, 62]. None of the 4 games contained existing tutorials prior to beginning work on this study. Working in coordination with the game developers of each game, we implemented three different types of tutorials at the start of each game (see Figure 4):

- **Text**: Text instructions only.
- **Text+Diagram**: Text instructions with diagrams depicting the actions required.
- **Text+Spatial**: Text instructions with a controller tooltip, and the appropriate button highlighted. To ensure the visibility of the highlighted button on the controller, any obscuring object (such as a gun) had its opacity set to 50%. Tooltip text always faces the player's line-of-sight.

We modified each of the 4 VR games to ensure a consistent experience across players. Gameplay options (e.g., difficulty level) were removed from the game menu; these options were left at default settings. We also enforced a linear play-style—all game levels were completed sequentially, and it was not possible to 'replay' sections of the game. Additionally, tutorials were created to match the existing aesthetic (color, font style, etc.) of the game as much as possible. For a complete list of modifications to each game from its base version, see supplementary materials section A.1. The four games are briefly described below:

---

[2] Three of the games are available on the Steam store:

https://store.steampowered.com/app/748600/Meme_Dragons/

https://store.steampowered.com/app/769540/Beat_Miner/

https://store.steampowered.com/app/730620/Hammer_2/

The fourth game was developed by one of the authors for a purpose unrelated to the current study [135]:

https://youtu.be/TGc8H3Nw-3M

[3] As can be seen in the supplementary materials section A.2, these games differ by genre *and* complexity of controls. Therefore, results will inherently be a reflection of both of these aspects of each game.

Meme Dragons is a wave shooter. Enemies appear at various spawn points and attack the player. The player can dodge incoming enemies and projectiles by moving their head. Players shoot by pointing the controller and pressing the trigger. Game performance is measured through a single playthrough high score. Higher scores are achieved through progression to higher levels and by shooting enemies.

Beat Miner is a rhythm game. Players mine blocks of ore using their controllers, which act as hammers as they speed along in a mine trolley. Players progress to each subsequent level by hitting blocks and avoiding bombs. Game performance is measured through the highest level attained. Gameplay is procedurally generated based on the loaded song's rhythm. For that reason, we use a hardcoded looping pop song so that all players have the same experience.

Hammer 2 is a third-person shooter. Players complete levels by shooting enemies, reaching specific locations, and other miscellaneous objectives. The game's controls include moving, jumping, using weapons, entering/exiting vehicles, and a special time-slowing mode called "hammer time." Game performance is measured by total cumulative score, which is increased by shooting enemies and other environmental features.

Hack.VR is a puzzle game. Players solve short puzzles using a VR flow-based programming language. Game controls include movement as well as multiple gun modes that allow manipulation (moving, rotating, and dropping objects), modification (changing program values), and program node connection/disconnection (rearranging programs' logical flow). Game performance is measured by the number of puzzles completed.

Each of these VR games is compatible with both the Oculus Rift and the HTC Vive. We decided to focus exclusively on these two devices, since at the time of the study they were among the most common VR devices owned by users [63]. Each tutorial condition, in each game, was built for both the Rift and the Vive. Therefore, in total, there were 6 tutorial versions (3 conditions and 2 device types). Inside of the game, the VR device is automatically detected, and the appropriate tutorial is loaded by the game.

4.1 Condition Validation

Although each condition in our tutorials was designed to be ecologically valid (and based on our content analysis), we wanted to confirm that our final implementations reflected that this was the case. We performed a validation study with 5 professional VR developers recruited from an online freelancing platform, whom were each paid $20. The recruitment job posting was to "Provide Feedback on VR Game Tutorials," and the job description stated that we were looking for VR developers who could critically evaluate a set of VR game tutorials. Developers had an average of 2.80 (SD=0.84) years of VR game development experience, and an average of 4.00 (SD=1.00) years of experience playing VR games. Developers all had work experience and portfolios which reflected recent game development experience (all within one year). Developers were instructed to give their honest opinions and were told their responses would be anonymous, and proceeded to our survey.

In the survey, we randomly selected 9 total games from Table 1. Three of these were randomly selected from games containing text instruction in the tutorial (*Text*), three were randomly selected from games containing diagrams in the tutorial (*Text+Diagram*), and three were randomly selected from games containing tooltips in the tutorial (*Text+Spatial*). Each game was displayed with the title of the game, then a link to a walkthrough video demonstrating that game's tutorial. Each set of three games was placed beside high-resolution images demonstrating the corresponding tutorial from our own set of 4 games for comparison, which

was then followed by a Likert-scale question. Developers were asked to review the tutorials for each game shown in the videos/images, then answered: "The tutorials in the images are implemented similarly to the tutorials in the videos," on a scale from 1:*Strongly Disagree* to 7:*Strongly Agree*. On average, scores were close to or above the mid-point between *Slightly Agree* and *Agree* for all three of *Text* (M=5.80, SD=1.30), *Text+Diagram* (M=5.40, SD=1.14), and *Text+Spatial* (M=6.20, SD=0.84). Therefore, our tutorials exhibited a degree of ecological validity.

5 HYPOTHESES

In this paper, we are interested in the effects of different VR tutorial modalities on: 1) controls learnability; 2) engagement-related outcomes (i.e., player experience and intrinsic motivation); and 3) performance. Controls learnability gauges the learnability of controls. For capturing engagement-related outcomes, player experience of need satisfaction and intrinsic motivation are both well-established frameworks with extensive empirical validation [13, 64, 65]. Both of these frameworks help us determine the influence of tutorial modality across a broad range of engagement-related outcomes. Finally, we measure the player's performance. Altogether, these measures are broadly important in gameplay experiences as they refer to the ease with which the controls can be learned (controls learnability), player experience (player experience of need satisfaction) and enjoyability (intrinsic motivation), and objective metrics (performance). These measures were selected since: 1) an increase in the effectiveness of tutorials will likely lead to an increase in controls learnability; 2) more learnable gameplay will likely lead to an increase in performance [40]; 3) performance is inherently linked to feelings of *competence* which is central to need satisfaction [13] and therefore higher performance should lead to an increase in player experience; and 4) higher levels of competence is associated with increased player motivation [13]. Therefore, we hypothesize that certain VR tutorial modalities will lead to increased controls learnability, performance, and engagement-related outcomes within the game[4].

The goal of the study is to understand the impacts of three different types of VR tutorial modalities. Our review of the cognitive theory of multimedia learning has shown that instruction using words and graphics outperform words alone. This suggests that using diagrams in addition to text will provide more efficacious tutorials than text alone. Additionally, the visual representations of the controllers themselves in VR can also be considered as an in-the-world 3D graphic, which suggests that using them as part of instruction will be more efficacious than text alone. We hypothesize the following:

H1: Adding diagrams to text will increase controls learnability, performance, and engagement-related outcomes.

H2: Adding controller tooltips to text will increase controls learnability, performance, and engagement-related outcomes.

Based on the coherence principle of multimedia learning (extra material hurts learning), drawn diagrams representing the player's controllers can be considered a type of "extra material" given that the player's controllers are already visible in VR (in the player's hands). Therefore, tooltips appearing on the virtual VR controllers themselves may facilitate less cognitive load along two dimensions: 1) reducing extra material by excluding the extraneous diagrams and 2) require less interpretative effort (i.e., generative processing) in understanding

---

[4] We choose not to include other measures such as presence since there is little evidence from our theoretical framing that tutorial modality alone will influence presence.

and internally mapping the diagram to the player's own controllers. Therefore, we predict that tooltips will outperform diagrams. We hypothesize the following:

H3: Adding controller tooltips to text will increase controls learnability, performance, and engagement-related outcomes compared to adding diagrams to text.

6 METHODS

6.1 Conditions

The conditions were divided by the tutorial types described earlier: *Text* (text instructions only), *Text+Diagram* (text instructions with diagrams), and *Text+Spatial* (text instructions with tooltips). Each controller instruction followed the same sequence. First, the instruction for a specific action was presented. Then, the user was asked to perform the action in a simulated tutorial setting. When the action being instructed involved interaction, dummy game objects (e.g., enemies, vehicles, props) were spawned. The player was required to interact with the objects in an appropriate way before continuing. For a complete description of specific actions in the tutorials (and corresponding controls), see supplementary materials section A.2. Across the conditions, only instruction for the actions differed (in modality). All other aspects— including practice of each individual action—were completely identical in all conditions.

6.2 Measures

Broadly, we are interested in determining the effects of tutorial modality on controls learnability (how easy did participants find the controls to learn?), performance (how well did participants perform in the game subsequent to the tutorial?), and engagement-related outcomes (how engaged were participants in the game subsequent to the tutorial?).

6.2.1 Controls Learnability. In order to assess the learnability controls, we use the *controls* subscale from the Player Experience of Need Satisfaction (PENS) scale [13], which has three questions assessing controls learnability: "Learning the game controls was easy," "The game controls were intuitive," and "When I wanted to do something in the game, it was easy to remember the corresponding control." These were all on a 7-point Likert scale (1:*Do Not Agree* to 7:*Strongly Agree*). Cronbach's alpha was 0.89.

Additionally, we measure cognitive load, with measures adapted from [66] and [67]. It consists of 8 items on a 6-point Likert scale (1:*Strongly Disagree* to 6:*Strongly Agree*). There are two subscales: mental load (e.g., "The game's controls were difficult to learn for me") and mental effort (e.g., "I needed to put lots of effort into learning the game's controls"). See supplementary materials section A.4 for the full set of cognitive load questions. Average Cronbach's alpha for each dimension of cognitive load was 0.90 and 0.88.

6.2.2 Performance. Performance was measured using one performance metric per game that best captured the performance of the player in the game. These metrics are high score (Meme Dragons), highest level (Beat Miner), total score (Hammer 2), and puzzles completed (Hack.VR).

6.2.3 Engagement-Related Outcomes. In order to assess engagement-related outcomes, we use the *competence, relatedness, autonomy,* and *immersion* subscales of the PENS scale [13]. PENS is based on self-determination theory (SDT) [68]. PENS contends that the psychological "pull'" of games are largely due to their ability to engender three needs: *competence* (the need to develop mastery [69]), *relatedness* (the need to feel connected to others [70]), and *autonomy* (the need to

feel autonomous [71]) [13]. Additionally, the PENS measures *immersion* and *controls* in game contexts. All subscales were on a 7-point Likert scale (1:*Do Not Agree* to 7:*Strongly Agree*). PENS is used widely for assessing player experience [65]. Average Cronbach's alpha for each dimension of the PENS was 0.91, 0.94, 0.89, and 0.87.

As an additional measure of engagement-related outcomes, we use the Intrinsic Motivation Inventory (IMI), which assesses intrinsic motivation using four dimensions: 1) Interest/Enjoyment (e.g., I enjoyed doing this activity very much), 2) Effort/Importance (e.g., I put a lot of effort into this), 3) Pressure/Tension (e.g., I felt very tense while doing this activity), 4) Value/Usefulness (e.g., I believe this activity could be of some value to me) [72]. All subscales were on a 7-point Likert scale (1:*Not At All True* to 7:*Very True*). Average Cronbach's alpha for each dimension of the IMI was 0.96, 0.89, 0.78, and 0.97.

6.2.4   Pre-test Measures. We incorporated pre-test measures to ensure participants were comparable between conditions. We used the Simulator Sickness Questionnaire (SSQ) [73] to assess initial simulator sickness. The SSQ is rated on a 4-point Likert scale (0:*None* to 4: *Severe*). We used the Immersive Tendencies Questionnaire (ITQ) [74] to assess immersive tendencies (e.g., "Do you ever become so involved in a video game that it is as if you are inside the game rather than moving a joystick and watching the screen?"). The ITQ is rated on a 7-point Likert scale (0:*Never* to 7:*Often*). Finally, we ask the following questions to assess prior relevant experience: "I have a lot of previous experience playing *video games*," "I have a lot of previous experience playing *video games in VR*," and "I have a lot of previous experience *using VR*," (1: *Strongly Disagree* to 7:*Strongly Agree*).

6.3   Participants

After a screening process that disqualified participants with multiple surveys with zero variance, multiple surveys with ±3SD, or a failed attention check, a total of 433 Amazon Mechanical Turk participants were retained. See Table 2 for the number of participants in each experimental condition in each game.

Table 2: Participants across our four experiments.

	Meme Dragons	Beat Miner	Hammer 2	Hack.VR
Text	35	37	33	38
Text+Diagram	36	34	31	37
Text+Spatial	37	40	36	39
Total	108	111	100	114

The aggregate data set consisted of 370 self-reported male, and 63 self-reported female data points. Races/ethnicities were self-reported as white (363), Black or African American (10), Filipino (7), Vietnamese (5), American Indian (5), Chinese (3), Asian Indian (3), and other (37). Reported age of data points were between 18 and 47 (M = 28.6, SD = 7.6). On average, reported duration of having owned either an Oculus Rift or HTC Vive was 15.7 months (SD = 12.3), with a daily VR usage time of 1.66 hours (SD = 1.42). In total, 255 data points were collected while using the Oculus Rift, and 178 data points were collected while using the HTC Vive. Participants were reimbursed $5.00 to participate in the experiment.

6.4 Participant Sample Justification

Amazon Mechanical Turk (AMT) has been shown to be a reliable platform for providing: 1) data comparable in quality to participants recruited from college campuses [75], 2) a diverse population to sample from [76], and 3) a framework for conducting reliable experiments [75, 77]. Nevertheless, because we are recruiting a specific subset of participants from AMT (i.e., participants that own a specific HMD), there is a greater chance of bias. In a study that took place in the second half of 2020, Kelly et al. compared HMD owners versus non-owners on two online platforms (AMT and Prolific) and an undergraduate participant pool [78]. The study found that there is a larger proportion of men than women when looking at HMD owners. However, contrary to their hypotheses, HMD owners and non-owners did not have differing spatial task performance; in addition, HMD owners and non-owners did not differ on susceptibility to motion sickness. The only difference they find between HMD owners and non-owners is that female HMD owners report playing more video games than female non-owners. However, they find that this difference is reduced by the lack of a correlation between video game hours and spatial ability. The study concludes that few differences exist between HMD owners and non-owners, and Kelly et al. suggest that research on a specific population (e.g., HMD owners) may generalize well to the other population (e.g., HMD non-owners). The authors caution, however, that spatial ability in participants from online work sites (e.g., AMT and Prolific) may be lower than undergraduate participants, especially among men. They suggest one possible reason for this may have been a result of many of the undergraduates in their sample having come from a STEM major, in which success is generally correlated with spatial ability [79]. The study by Kelly et al. helps to alleviate the concern that HMD owners and HMD non-owners may be drastically different on important measures. That being said, we acknowledge that further work comparing these different populations needs to be performed to further validate Kelly et al.'s findings.

6.5 Design

Four experiments were conducted. These experiments were conducted identically and sequentially, differing only in the game being used. We ensure unique participants within each experiment by assigning a token to participants that completed the experiment. The order in which we conducted the experiments (i.e., games) was random. This was done by generating a random sequence of 4 numbers, with each number corresponding to one of the games. A between-subjects design was used: tutorial type was the between-subject factor. Within each experiment, participants were randomly assigned to a condition.

6.6 Procedure

For recruitment, we use the same procedure as outlined in Ma et al. [80]. First, each participant was asked to upload a photo. The photo had to contain their VR device (Oculus Rift or HTC Vive) *and* a piece of paper with the hand-written last four digits of their unique Amazon Worker ID. This method is easy to verify, but difficult to fake.[5] Participants that make it past this verification are then invited to participate in the actual study.

In the main study, participants first complete the SSQ, the ITQ, and prior VR/gaming experience questions. Participants then downloaded a version of the VR game corresponding to

---

[5] Even if a participant successfully fakes the photo, our games automatically detect if an appropriate device (Oculus Rift or HTC Vive) is connected—otherwise, the game will not boot.

their randomly assigned condition. At the end of each tutorial, and before the game began, participants were informed that they could exit the game *at any time* via the in-game menu. Participants then began playing the game. Participants were given no explicit goal, and were able to play the game for as little or as long as they liked. Data was automatically collected during gameplay and stored in a local binary file, which participants later uploaded. We take into account several precautions, including if the game process was terminated unexpectedly. For additional details, see supplementary materials section A.3.

After leaving the game, players complete the player experience of need satisfaction scale, the intrinsic motivation inventory, and the cognitive load questionnaire. Players answered several open-response questions to ensure that no technical problems interfered during the experiment, and they filled out demographics.

6.7 Analysis

Data was extracted and imported into Statistical Package for Social Science (SPSS) version 22 for data analysis using multivariate analysis of covariance (MANCOVA). Since age, sex, and device type have been shown to influence performance and engagement [13, 81–83], the analysis controls for these variables. Separate MANCOVAs are run for each of the following: *Controls Learnability (PENS Controls and Cognitive Load), Engagement-Related Outcomes (PENS Competence, PENS Relatedness, PENS Autonomy, PENS Immersion, IMI), Pre-Test Measures (SSQ, ITQ, Prior Experience Questions)* – with the independent variable being *tutorial condition*. We use ANCOVA for *Performance* since it consists of only one variable per game. The covariates are *age*, *sex*, and *device type*. All the dependent variables are continuous variables. The independent variable tutorial condition (i.e., 0 = *Text*, 1 = *Text+Diagram*, 2 = *Text+Spatial*) is a trichotomous variable. In terms of covariates, age is continuous, while sex and device type are dichotomous variables. Prior to running MANCOVAs, we check both assumption of homogeneity of variance and homogeneity of covariance by the test of Levene's Test of Equality of Error Variances and Box's Test of Equality of Covariance Matrices. When either of these assumptions were violated ($p < 0.05$ for Levene's or $p < 0.001$ for Box's), we use the more conservative Pillai's Trace [84]; otherwise, we use Wilk's Lambda. Post-hoc testing is performed using Dunn–Šidák, which controls for Type I errors (i.e., false positives) [85]. Effect sizes are reported using partial-eta squared (.01 for small, .06 for medium, and .14 for large effects [86, 87]).

7 RESULTS

7.1 Summary Descriptives

Based on a post-game question asking participants if they had played the game previously, across all 4 experiments none of the participants had previously played the same game[6]. Average time spent in each tutorial was 27.5s (Meme Dragons), 61.2s (Beat Miner), 132.3s (Hammer 2), 213.8s (Hack.VR). Average time actually spent playing (e.g., *excludes* time spent in the main menu, controls introduction, and paused) was 279.3s (Meme Dragons), 445.3s (Beat Miner), 738.1s (Hammer 2), and 661.1s (Hack.VR). Participants across all studies exhibited low pre-test SSQ scores. From 0:*None* to 3:*Severe*, the mean SSQ symptom score across participants

---

[6] This was asked post-game because we felt a user would be able to better recall if they had played the game after experiencing it.

was M=0.22 (SD=0.52). No participants were excluded from the analysis based on overall SSQ scores. For an overview of results, see Figures 5, 6, and 7.
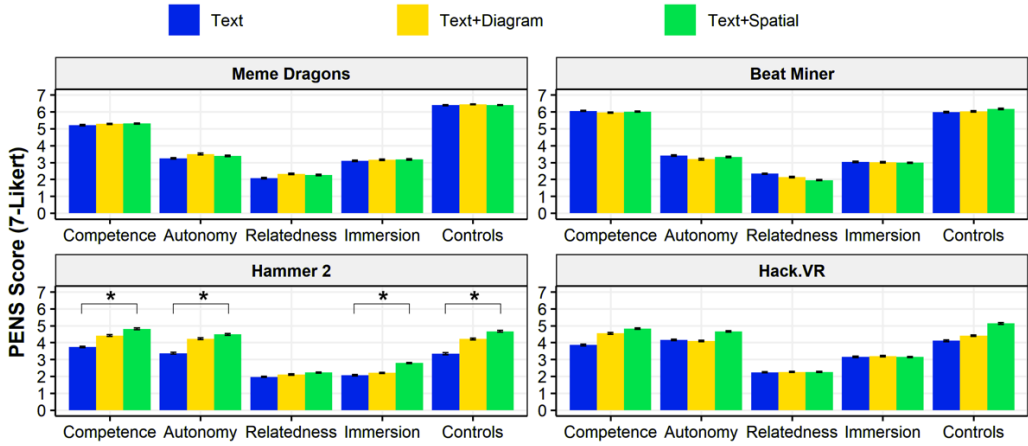

Figure 5. Player Experience of Need Satisfaction results. Asterisk (*) indicates a significant difference between tutorial condition types. Error bars show SEM.
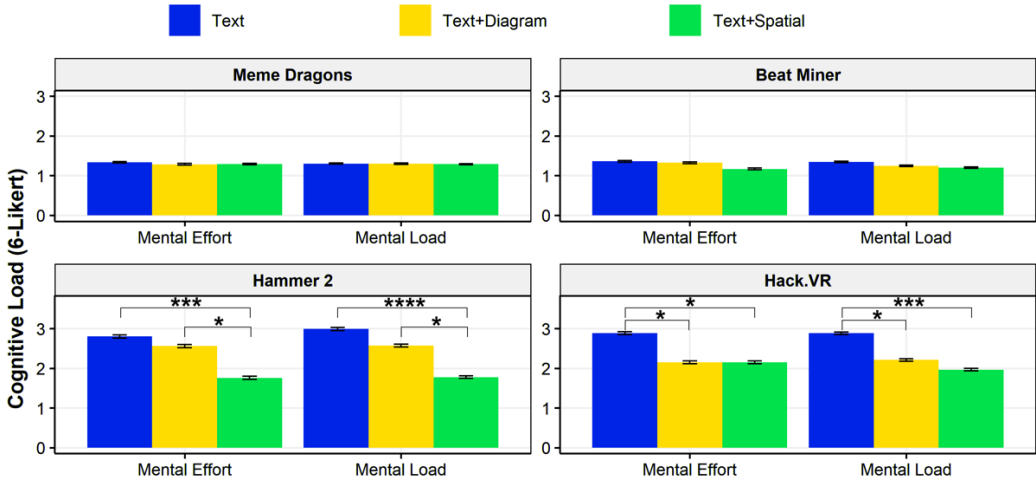

Figure 6. Cognitive Load results. Asterisk (*) indicates a significant difference between tutorial condition types. p < .05 (*); p < .01 (**); p < .005 (***); p < .001 (****). Error bars show SEM.

7.2 Experiment 1: Wave Shooter (Meme Dragons)

Pre-test: There was no statistically significant difference in the pre-test measures of SSQ, ITQ, prior game, VR, and VR game experience based on tutorial condition, $F(16, 190) = 1.35$, $p = 0.17$; Wilk's Lambda = 0.806, partial $\eta^2 = 0.102$.

Controls Learnability: There was no statistically significant difference in controls learnability measures based on tutorial condition, $F(6, 200) = 0.04$, $p = 1.00$; Wilk's Lambda = 0.998, partial $\eta^2 = 0.001$.

Performance: There was no significant difference in performance based on tutorial condition, $F(2, 102)=0.015$, $p = .99$, $\eta^2_p = .000$.
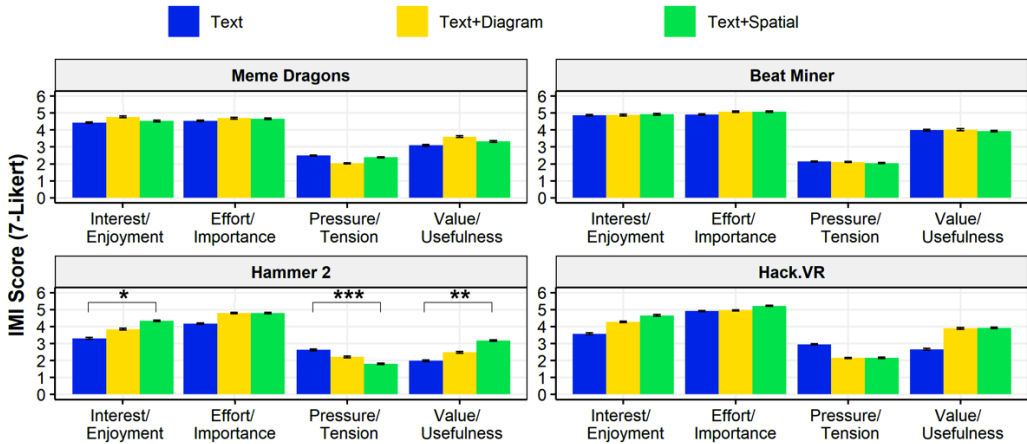
Figure 7. Intrinsic Motivation Inventory results. Asterisk (*) indicates a significant difference between tutorial condition types. p < .05 (*); p < .01 (**); p < .005 (***). Error bars show SEM.

Engagement-Related Outcomes: There was no statistically significant difference in engagement-related measures based on tutorial condition, $F(16, 190) = 0.44$, $p = 0.97$; Wilk's Lambda = 0.930, partial $\eta^2 = 0.036$.

Meme Dragons Summary: In Meme Dragons, there were no differences observed across any measures.

7.3 Experiment 2: Rhythm Game (Beat Miner)

Pre-test: There was no statistically significant difference in the pre-test measures of SSQ, ITQ, prior game, VR, and VR game experience based on tutorial condition, $F(16, 198) = 0.69$, $p = 0.80$; Pillai's Trace = 0.106, partial $\eta^2 = 0.053$.

Controls Learnability: There was no statistically significant difference in controls learnability measures based on tutorial condition, $F(6, 206) = 0.68$, $p = 0.67$; Wilk's Lambda = 0.962, partial $\eta^2 = 0.019$.

Performance: There was no significant difference in performance based on tutorial condition, $F(2, 105)=1.338$, $p = .27$, $\eta^2_p = .025$.

Engagement-Related Outcomes: There was no statistically significant difference in engagement-related measures based on tutorial condition, $F(16, 196) = 0.22$, $p = 1.00$; Wilk's Lambda = 0.966, partial $\eta^2 = 0.017$.

Beat Miner Summary: In Beat Miner, there were no differences observed across any measures.

7.4 Experiment 3: 3rd Person Shooter (Hammer 2)

Pre-test: There was no statistically significant difference in the pre-test measures of SSQ, ITQ, prior game, VR, and VR game experience based on tutorial condition, $F(16, 176) = 1.38$, $p = 0.16$; Pillai's Trace = 0.223, partial $\eta^2 = 0.111$.

Controls Learnability: There was a statistically significant difference in controls learnability measures based on tutorial condition, $F(6, 186) = 3.52$, $p < 0.005$; Pillai's Trace = 0.204, partial $\eta^2 = 0.102$. Univariate testing found that there was a significant difference in PENS Controls, $F(2, 94) = 4.41$, $p < 0.05$, partial $\eta^2 = 0.086$; Mental Load, $F(2, 94) = 7.81$, $p < 0.001$, partial

$\eta^2 = 0.143$; and Mental Effort, $F(2, 94) = 6.01$, $p < 0.005$, partial $\eta^2 = 0.113$. A Dunn–Šidák post-hoc test found that for PENS Controls, *Text+Spatial* was significantly higher than *Text* (p < 0.05). A Dunn–Šidák post-hoc test found that for Mental Load, *Text+Spatial* was significantly lower than *Text+Diagram* (p < 0.05) and *Text* (p < 0.001). A Dunn–Šidák post-hoc test found that for Mental Effort, *Text+Spatial* was significantly lower than *Text+Diagram* (p < 0.05) and *Text* (p < 0.005).

Performance: There was a significant difference in performance based on tutorial condition, $F(2, 94) = 4.229$, p = .02, η^2_p = .083. A Dunn–Šidák post-hoc test found that *Text+Spatial* was significantly higher than *Text* (p < 0 .05).

Engagement-Related Outcomes: There was a statistically significant difference in engagement-related measures based on tutorial condition, $F(16, 176) = 1.72$, $p < 0.05$; Pillai's Trace = 0.270, partial $\eta^2 = 0.135$. Univariate testing found that there was a significant difference in PENS Competence, $F(2, 94) = 3.65$, $p < 0.05$, partial $\eta^2 = 0.072$; PENS Autonomy, $F(2, 94) = 3.97$, $p < 0.05$, partial $\eta^2 = 0.078$; PENS Immersion, $F(2, 94) = 4.02$, $p < 0.05$, partial $\eta^2 = 0.079$; IMI Interest/Enjoyment, $F(2, 94) = 3.39$, $p < 0.05$, partial $\eta^2 = 0.067$; IMI Pressure/Tension, $F(2, 94) = 5.75$, $p < 0.005$, partial $\eta^2 = 0.109$; IMI Value/Usefulness, $F(2, 94) = 5.07$, $p < 0.01$, partial $\eta^2 = 0.097$. A Dunn–Šidák post-hoc test found that *Text+Spatial* scored significantly higher than *Text* in PENS Competence (p < .05); for PENS Autonomy (p < .05); and for PENS Immersion (p < .05). A Dunn–Šidák post-hoc test found that *Text+Spatial* scored significantly higher than *Text* in IMI Interest/Enjoyment (p < .05). *Text+Spatial* scored significantly lower than *Text* in IMI Pressure/Tension (p < .005). *Text+Spatial* scored significantly higher than *Text* in IMI Value/Usefulness (p < .01).

Hammer 2 Summary: There is no support for H1. H2 is supported. H3 is supported for controls learnability only.

7.5 Experiment 4: Puzzle (Hack.VR)

Pre-test: There was no statistically significant difference in the pre-test measures of SSQ, ITQ, prior game, VR, and VR game experience based on tutorial condition, $F(16, 204) = 0.85$, $p = 0.63$; Pillai's Trace = 0.124, partial $\eta^2 = 0.062$.

Controls Learnability: There was a statistically significant difference in controls learnability measures based on tutorial condition, $F(6, 214) = 2.81$, $p < 0.05$; Pillai's Trace = 0.146, partial $\eta^2 = 0.073$. Univariate testing found that there was a significant difference in PENS Controls, $F(2, 108) = 3.37$, $p < 0.05$, partial $\eta^2 = 0.059$; Mental Load, $F(2, 108) = 6.23$, $p < 0.005$, partial $\eta^2 = 0.103$; and Mental Effort, $F(2, 108) = 3.95$, $p < 0.05$, partial $\eta^2 = 0.068$. A Dunn–Šidák post-hoc test found that for PENS Controls, *Text+Spatial* was significantly higher than *Text* (p < 0.05). A Dunn–Šidák post-hoc test found that for Mental Load, *Text+Spatial* was significantly lower than *Text* (p < 0.005). A Dunn–Šidák post-hoc test found that for Mental Load, *Text+Diagram* was significantly lower than *Text* (p < 0.05). A Dunn–Šidák post-hoc test found that for Mental Effort, *Text+Spatial* was significantly lower than *Text* (p < 0.05). A Dunn–Šidák post-hoc test found that for Mental Effort, *Text+Diagram* was significantly lower than *Text* (p < 0.05).

Performance: There was a significant difference in performance based on tutorial condition, $F(2, 108) = 4.295$, p = .02, η^2_p = .074. A Dunn–Šidák post-hoc test found that *Text+Spatial* was significantly higher than *Text* (p < 0.05).

Engagement-Related Outcomes: There was no statistically significant difference in engagement-related measures based on tutorial condition, $F(16, 204) = 1.50$, $p = 0.10$; Pillai's Trace = 0.210, partial $\eta^2 = 0.105$.

Hack.VR Summary: In Hack.VR, H1 is supported for controls learnability only. H2 is supported for controls learnability and performance. H3 is not supported.

7.6 Summary

We summarize our results according to our hypotheses. See Table 3.

H1: Adding diagrams to text increases controls learnability, performance, and engagement-related outcomes. Supported only in the puzzle game for controls learnability.

H2: Adding controller tooltips to text increases controls learnability, performance, and engagement-related outcomes. Supported in the third-person shooter game for controls learnability, performance, and engagement-related outcomes. Supported in the puzzle game for controls learnability and performance.

H3: Controller tooltips and text will lead to increased controls learnability, performance, and engagement-related outcomes relative to diagrams and text. Supported only in the third-person shooter game for controls learnability.

Table 3: Hypothesis summary.

	W	R	T	P
H1.1 Adding diagrams to text increases controls learnability.	N	N	N	Y
H1.2 Adding diagrams to text increases performance.	N	N	N	N
H1.3 Adding diagrams to text increases engagement-related outcomes.	N	N	N	N
H2.1 Adding controller tooltips to text increases controls learnability.	N	N	Y	Y
H2.2 Adding controller tooltips to text increases performance.	N	N	Y	Y
H2.3 Adding controller tooltips to text increases engagement-related outcomes.	N	N	Y	N
H3.1 Adding controller tooltips to text increases controls learnability compared to adding diagrams to text.	N	N	Y	N
H3.2 Adding controller tooltips to text increases performance compared to adding diagrams to text.	N	N	N	N
H3.3 Adding controller tooltips to text increases engagement-related outcomes compared to adding diagrams to text.	N	N	N	N

W: Wave Shooter; R: Rhythm Game; T: Third-Person Shooter; P: Puzzle Game

8 DISCUSSION

Overall, the results support that *Text+Spatial* outperformed *Text* in controls learnability and performance in the third-person shooter and puzzle games. Additionally, *Text+Spatial* outperformed *Text* in player experience and intrinsic motivation in the third-person shooter game. Results also show that *Text+Diagram* outperforms *Text* in controls learnability in the puzzle game. Finally, the results support that *Text+Spatial* outperforms *Text+Diagram* in controls learnability in the third-person shooter game. Importantly, we found significant results only in the third-person shooter and puzzle games, while there were no significant differences found in the wave shooter and rhythm games. These results collectively show that game type is a significant factor in determining the influence of modality of tutorial instruction. Participants took on average at least twice as long in the tutorials for the third-person shooter and puzzle games compared to the wave shooter and rhythm games. Moreover, the intuitiveness of controls was rated as lower in the third-person shooter and puzzle games (PENS Controls scores: M=4.08 and M=4.55) compared to the wave shooter and rhythm games (M=6.41 and M=6.06). This suggests that the controls were more difficult and took longer to learn in the third-person shooter and puzzle games, potentially amplifying differences between tutorial modalities. This parallels Andersen et al.'s finding that tutorials appeared justified only in complex games [7].

Our results are supported by prior work on multimedia learning theory. The principle of combining graphics with words (instead of using words alone) [19] would support the use of either pictorial diagrams or controller tooltips using the player's controller as an in-the-world graphic. For example, text and diagrams together have been shown to better support mental model development when compared to only text [20]. This can be understood through the dual channels principle of multimedia learning (based on Paivio's dual-coding theory), in which humans have two distinct channels for processing information: verbal and non-verbal [88]. Text is processed through the verbal channel, while diagrams and virtual controllers are processed through the non-verbal channel. Because the verbal and non-verbal channels do not overlap, this may be one reason for *Text+Spatial* and *Text+Diagram* being more effective than *Text* alone in some circumstances. Both the principle of avoiding extra material and using visual cues are also relevant. In light of the player's controller that already exists in VR, the pictorial diagrams may be viewed as a type of extra material that is unnecessary, and creates additional interpretation (i.e., greater generative processing). In terms of using visual cues, work has demonstrated that they can improve learning efficacy [33], reduce cognitive load [89, 90], and improve the speed and accuracy of completing tasks [34]. For example, we can see from Figure 6 that *Text+Spatial* has the lowest cognitive load scores as compared to the other two conditions. Visual cues in VR may be providing benefits both through spatial relevance (e.g., cues directly attached to the controller) and by reducing the need for interpreting other materials, such as text and images. Tooltips which are connected to and highlight relevant buttons on VR controllers may therefore help users more easily learn controls.

Controller tooltips, however, are not typically leveraged on traditional platforms (e.g., desktop displays), and therefore, further research could study different implementations of these tooltips. Our four experiments contribute to this area of research by showing that tooltips on the player's physical controllers, under certain game types, can be effective at increasing controls learnability, performance, and engagement-related outcomes in VR games. However, further research is needed to understand whether these results can generalize to additional game types; the interaction effects between game type and controls complexity; and additional varied implementations of text, diagram, and tooltip instructions. For example, prior work has shown that in a VR game (as opposed to a monitor-display game) players may focus more on the sensory experience [91]. Moreover, VR games can induce a higher level of immersion and flow compared to monitor-display games [41]. This suggests that players have a higher likelihood of being immersed in VR (i.e., both the visual and audial channels are working at higher load than for non-HMD applications), which means that cognitive load is especially important to manage. For example, Schrader and Bastiaens showed that a more immersive educational game led to increased virtual presence, which in turn led to a decrease in learning, which was mediated through an increase in cognitive load [42]. In other words, extraneous processing occurring with images and text that are not relevant to the instructional objective may especially be burdensome in an already demanding immersive VR environment. Our studies suggest tutorial modality can, for some game types, play an important role in moderating effects on cognitive load.

However, it is important to discuss that many of the results we expected to occur did not. H1 and H3 were for the most part not supported. In looking at Figures 5, 6, and 7, however, we do see that directionally the conditions do match our hypotheses for the third-person shooter and puzzle games, although these were not statistically significant. For example, *Text+Diagram* does score (non-significantly) higher than *Text* on most measures (H1). Similarly, *Text+Spatial* does

score (non-significantly) higher than *Text+Diagram* on most measures (H3). Although concrete claims cannot be made based on non-statistical significance, because the descriptive trends align with what we would expect based on learning theories, these results do call for additional studies on this topic to more concretely determine influence of tutorial modality in VR.

This work extends prior work on game tutorials. In Andersen et al. [7], the authors found that only complex games benefit from tutorials; in this paper, we find a similar result in that tutorial modality is important, but only in games with certain game types (which do inherently vary in complexity[7]). Therefore, for some game types it may not be worthwhile to invest considerable resources in developing more complex modality tutorials. These game types may be sufficiently intuitive to learn, or they are common enough that the majority of users already have some level of experience with them and do not benefit from multimedia learning principles embedded in tutorials. In Frommel et al. [8], the authors found that tutorials given just-in-time as mechanics were being practiced heightened positive affect and motivation of VR game players. In the present study, we show that how a tutorial is shown (i.e., the modality of instruction) can have a significant role in affecting players in complex VR games. Our results suggest that tutorial modality makes little difference in simpler VR games. However, in complex VR games, our results suggest that tutorial modality can influence controls learnability, engagement-related outcomes, and performance. For complex VR games, VR developers should consider using tooltips which highlight and label each controller button (while simultaneously showing the tutorial text) during tutorials. As one of the first VR studies to contrast tutorial modalities from the perspective of multimedia learning theory, this study is a first step towards understanding the effects of modality in VR games. Our results were tested on four different VR games (three of them commercially available) to increase generalizability.

9   LIMITATIONS AND FUTURE WORK

One key difference in our study and other VR studies is our recruitment method. Because participants in our study already owned a VR headset, our participants were much less naïve to VR than typical recruitment methods, such as through university mailing lists. However, participants who already use VR would be the most likely to benefit from these studies. Moreover, participants are playing the games in their own playing area, and this has the potential to increase external validity. That being said, the consideration of place (i.e., the physical space where the participant experience the VR game) can be an important factor in driving the player experience [61, 92] and can be a consideration of future studies.

The study is limited to tutorials that take place at the start of a game. While it is common in VR games to have a tutorial at the beginning, games sometimes opt for diegetic help. One future direction is to study tutorial modalities in the context of diegetic help, different ways of triggering the help (e.g., an "I don't know" shoulder shrug gesture, a 3D help object, through the pause menu, etc.), and role of users' cognitive styles [93]. Increasingly, VR games look to provide help when it is automatically detected that users need help, proposed by Drey, Jansen, Fischbach, Frommel, and Rukzio [94]; this is a highly promising direction for future work since VR users will vary in their prior experience, mirroring a trend in educational technology

---

[7] Although we operationalize complexity here as subjective measures such as intuitiveness of controls and cognitive load, we could imagine that controls complexity could refer to a variety of aspects, e.g., the complexity of the controller instructions, the number of controls, or the difficulty of using the controls. Future work could distinguish between these individual factors to study these measures as a function of complexity in more detail.

towards personalized [95] and adaptive [96] help. Incorporating finer-grained analysis of user actions [97] would be helpful in determining specific in-game actions which were well-learned or not-well-learned based on tutorial modality conditions. Lastly, although this was a controlled online study, looking at the long-term consequences of VR tutorials would also be important – e.g., user retention (or lack thereof) as a result of better or worse initial learning outcomes.

Many other game genres exist with different mechanics and level progressions. As such, more work needs to be done to understand the specific contexts in which these results may occur or fail to occur. Many factors vary between games including genre [98–100], complexity [40], length [101], immersiveness [101, 102], haptic feedback [103], and physical controllers [104, 105]. Other factors include varying levels of embodiment [106–110], collaboration [111–116], and demographic preferences [117–119]. Implementation-specific details of tutorials such as immediate error correction [120], the colors of tutorial elements [121], type of feedback [122–124], and other qualitative aspects [125] may influence their effectiveness. Additionally, the games we selected in this paper cannot be said to be representative of *all* games in their respective genres. For example, even within specific genres, such as puzzle games, there exists significant variance in controls/gameplay complexity (e.g., [126–131]). Therefore, more work is needed before it is possible to make broader generalizations across genre.

One other promising future direction would be the addition of biometrics, such as eye tracking [132], to better understand the cognitive processes of players during tutorials. Finally, we acknowledge a gender imbalance with the smaller number of female participants overall (63), which we posit is likely due to the male skew in terms of VR device ownership [133]. An additional limitation is that the present study did not allow participants to choose a non-binary gender when reporting demographics.

10 CONCLUSION

This research paper demonstrates that tutorial modality influences some VR game types' controls learnability, engagement-related outcomes, and performance. Many VR games feature tutorials, and these results can directly inform the design of VR games. Our content analysis found that the vast majority of popular VR games have a tutorial level and some form of text help, about half have some form of instructional diagrams/images, and about one quarter used instructional tooltips highlighting relevant controller buttons.

We then compared three tutorial modalities: *Text*, *Text+Diagram*, and *Text+Spatial*. In the third-person shooter and puzzle games, our results show that *Text+Spatial* leads to the highest controls learnability, player experience, intrinsic motivation, and performance on average. *Text+Diagram* was slightly worse, and *Text* was the least beneficial of the three conditions studied. Although we focused our studies on VR games, these results (supported by multimedia learning theory) *may* also be applicable to VR training environments more generally. Our results suggest that tutorial modality makes little difference in simpler VR games. However, in complex VR games, our results suggest that tutorial modality can influence controls learnability, engagement-related outcomes, and performance. Therefore, developers should consider utilizing more advanced tutorial modalities when instructing players in complex VR games, such as controller tooltips. This work is a first step towards investigating tutorial presentation modalities in VR.

REFERENCES

[1]   Forecast augmented (AR) and virtual reality (VR) market size worldwide from 2016 to 2024: 2020. *https://www.statista.com/statistics/591181/global-augmented-virtual-reality-market-size/*.
[2]   New Consumer Data Finds VR Headset Usage Expected To Increase In 2019, According To Greenlight Insights: 2018. *https://greenlightinsights.com/new-consumer-data-finds-vr-headset-usage-expected-increase-2019-according-greenlight-insights/*.
[3]   Rose, T., Nam, C.S. and Chen, K.B. 2018. Immersion of virtual reality for rehabilitation - Review. *Applied Ergonomics*.
[4]   Lawson, G., Salanitri, D. and Waterfield, B. 2016. Future directions for the development of virtual reality within an automotive manufacturer. *Applied Ergonomics*. 53, (2016), 323–330. DOI: https://doi.org/10.1016/j.apergo.2015.06.024.
[5]   Jou, M. and Wang, J. 2013. Investigation of effects of virtual reality environments on learning performance of technical skills. *Computers in Human Behavior*. (2013). DOI:https://doi.org/10.1016/j.chb.2012.04.020.
[6]   Bertram, J., Moskaliuk, J. and Cress, U. 2015. Virtual training: Making reality work? *Computers in Human Behavior*. 43, (2015), 284–292. DOI:https://doi.org/10.1016/j.chb.2014.10.032.
[7]   Andersen, E., O'Rourke, E., Liu, Y.-E., Snider, R., Lowdermilk, J., Truong, D., Cooper, S. and Popovic, Z. 2012. The impact of tutorials on games of varying complexity. *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI '12*. Pajitnov 1984 (2012), 59–68. DOI:https://doi.org/10.1145/2207676.2207687.
[8]   Frommel, J., Fahlbusch, K., Brich, J. and Weber, M. 2017. The Effects of Context-Sensitive Tutorials in Virtual Reality Games. (2017), 367–375. DOI:https://doi.org/10.1145/3116595.3116610.
[9]   Adams, E. 2014. *Fundamentals of game design*. Pearson Education.
[10] Burdea, G.C. and Coiffet, P. 2003. *Virtual reality technology*. John Wiley & Sons.
[11] Wann, J. and Mon-Williams, M. 1996. What does virtual reality NEED?: Human factors issues in the design of three-dimensional computer environments. *International Journal of Human Computer Studies*. 44, 6 (1996), 829–847. DOI:https://doi.org/10.1006/ijhc.1996.0035.
[12] Slater, M. and Sanchez-Vives, M. V. 2016. Enhancing our lives with immersive virtual reality. *Frontiers Robotics AI*.
[13] Ryan, R.M., Rigby, C.S. and Przybylski, A. 2006. The Motivational Pull of Video Games: A Self-Determination Theory Approach. *Motivation and Emotion*. 30, 4 (2006), 344–360. DOI:https://doi.org/10.1007/s11031-006-9051-8.
[14] Frommel, J., Sonntag, S. and Weber, M. 2017. Effects of controller-based locomotion on player experience in a virtual reality exploration game. *ACM International Conference Proceeding Series*. Part F1301, August (2017). DOI:https://doi.org/10.1145/3102071.3102082.
[15] Ho, J.C.F. 2017. Practice in Reality for Virtual Reality Games: Making Players Familiar and Confident with a Game. *IFIP Conference on Human-Computer Interaction*. (2017), 147–162. DOI:https://doi.org/10.1007/978-0-387-35175-9.
[16] Mayer, R.E. 2014. Cognitive theory of multimedia learning. *The Cambridge Handbook of Multimedia Learning, Second Edition*.
[17] Sweller, J. 1988. Cognitive load during problem solving: Effects on learning. *Cognitive Science*. (1988). DOI:https://doi.org/10.1016/0364-0213(88)90023-7.
[18] Clark, R.C. and Mayer, R.E. 2016. E-learning and the Science of Instruction important: Fourth Edition. *Published by John Wiley & Sons, Inc., Hoboken, New Jersey*. (2016).
[19] Clark, R.C. and Mayer, R.E. 2011. E-learning and the Science of Instruction : Proven Guidelines for Consumers and Designers of Multimedia Learning. (2011).
[20] Butcher, K.R. 2006. Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*. 98, 1 (2006), 182–197. DOI:https://doi.org/10.1037/0022-0663.98.1.182.
[21] Cuevas, H.M., Fiore, S.M. and Oser, R.L. 2002. Scaffolding cognitive and metacognitive processes in low verbal ability learners: Use of diagrams in computer-based training environments. *Instructional Science*. (2002). DOI:https://doi.org/10.1023/A:1020516301541.
[22] Yue, C.L., Bjork, E.L. and Bjork, R.A. 2013. Reducing verbal redundancy in multimedia learning: An undesired desirable difficulty? *Journal of Educational Psychology*. (2013). DOI:https://doi.org/10.1037/a0031971.
[23] Mayer, R.E. 1989. Models for Understanding. *Review of Educational Research*. (1989). DOI:https://doi.org/10.3102/00346543059001043.
[24] Mayer, R.E., Steinhoff, K., Bower, G. and Mars, R. 1995. A generative theory of textbook design: Using annotated illustrations to foster meaningful learning of science text. *Educational Technology Research and Development*. (1995). DOI:https://doi.org/10.1007/BF02300480.
[25] Moreno, R. and Mayer, R.E. 1999. Multimedia-supported metaphors for meaning making in mathematics. *Cognition and Instruction*. (1999). DOI:https://doi.org/10.1207/S1532690XCI1703_1.
[26] Chandler, P. and Sweller, J. 1991. Cognitive Load Theory and the Format of Instruction. *Cognition and Instruction*. (1991). DOI:https://doi.org/10.1207/s1532690xci0804_2.
[27] Pociask, F.D. and Morrison, G.R. 2008. Controlling split attention and redundancy in physical therapy instruction. *Educational Technology Research and Development*. (2008). DOI:https://doi.org/10.1007/s11423-007-9062-5.
[28] Erhel, S. and Jamet, E. 2006. Using pop-up windows to improve multimedia learning. *Journal of Computer Assisted Learning*. (2006). DOI:https://doi.org/10.1111/j.1365-2729.2006.00165.x.
[29] Ginns, P. 2006. Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*. (2006). DOI:https://doi.org/10.1016/j.learninstruc.2006.10.001.
[30] Wertheimer, M. 1938. Gestalt theory. (1938).

[31] Harp, S.F. and Mayer, R.E. 1998. How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology.* 90, 3 (1998), 414–434. DOI:https://doi.org/10.1037/0022-0663.90.3.414.

[32] Mayer, R.E., Griffith, E., Jurkowitz, I.T.N. and Rothman, D. 2008. Increased Interestingness of Extraneous Details in a Multimedia Science Presentation Leads to Decreased Learning. *Journal of Experimental Psychology: Applied.* (2008). DOI:https://doi.org/10.1037/a0013835.

[33] Lin, L. and Atkinson, R.K. 2011. Using animations and visual cueing to support learning of scientific concepts and processes. *Computers and Education.* 56, 3 (2011), 650–658. DOI:https://doi.org/10.1016/j.compedu.2010.10.007.

[34] Kelleher, C. and Pausch, R. 2005. Stencils-Based Tutorials: Design and Evaluation. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '05.* (2005), 541. DOI:https://doi.org/10.1145/1054972.1055047.

[35] Dillman, K.R., Mok, T.T.H., Tang, A., Oehlberg, L. and Mitchell, A. 2018. A visual interaction cue framework from video game environments for augmented reality. *Conference on Human Factors in Computing Systems - Proceedings.* 2018-April, April (2018). DOI:https://doi.org/10.1145/3173574.3173714.

[36] Hu, X., Moore, A., Eubanks, J.C., Aiyaz, A. and Mcmahan, R.P. 2020. Evaluating Interaction Cue Purpose and Timing for Learning and Retaining Virtual Reality Training. *Proceedings - SUI 2020: ACM Symposium on Spatial User Interaction.* (2020). DOI:https://doi.org/10.1145/3385959.3418448.

[37] Oculus 2019. First Steps (https://www.oculus.com/experiences/quest/1863547050392688/). (2019).

[38] Klepsch, M., Schmitz, F. and Seufert, T. 2017. Development and validation of two instruments measuring intrinsic, extraneous, and germane cognitive load. *Frontiers in psychology.* 8, (2017), 1997.

[39] Lee, E.A.L. and Wong, K.W. 2008. A review of using virtual reality for learning. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).* 5080 LNCS, (2008), 231–241. DOI:https://doi.org/10.1007/978-3-540-69744-2_18.

[40] Andersen, E., O'Rourke, E., Liu, Y.-E., Snider, R., Lowdermilk, J., Truong, D., Cooper, S. and Popovic, Z. 2012. The impact of tutorials on games of varying complexity. *CHI* (2012).

[41] Pallavicini, F. and Pepe, A. 2019. Comparing player experience in video games played in virtual reality or on desktop displays: Immersion, flow, and positive emotions. *CHI PLAY 2019 - Extended Abstracts of the Annual Symposium on Computer-Human Interaction in Play.* (2019), 195–210. DOI:https://doi.org/10.1145/3341215.3355736.

[42] Schrader, C. and Bastiaens, T.J. 2012. The influence of virtual presence: Effects on experienced cognitive load and learning outcomes in educational computer games. *Computers in Human Behavior.* (2012). DOI:https://doi.org/10.1016/j.chb.2011.11.011.

[43] Kao, D. 2020. Exploring Help Facilities in Game-Making Software. *ACM Foundations of Digital Games* (2020).

[44] Kao, D., Joshi, A., Mousas, C., Peddireddy, A., Gopi, A.K., Li, J., Springer, J., McGowan, B.S. and Reed, J.B. 2021. Fighting COVID-19 at Purdue University: Design and Evaluation of a Game for Teaching COVID-19 Hygienic Best Practices. *ACM Foundations of Digital Games* (2021).

[45] O'Rourke, E., Ballweber, C. and Popović, Z. 2014. Hint systems may negatively impact performance in educational games. *L@S 2014 - Proceedings of the 1st ACM Conference on Learning at Scale.* (2014), 51–60. DOI:https://doi.org/10.1145/2556325.2566248.

[46] Wauck, H. and Fu, W.-T. 2017. A Data-Driven, Multidimensional Approach to Hint Design in Video Games. (2017), 137–147. DOI:https://doi.org/10.1145/3025171.3025224.

[47] Rzayev, R., Mayer, S., Krauter, C. and Henze, N. 2019. Notification in VR: The effect of notification placement, task, and environment. *CHI PLAY 2019 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play.* (2019), 199–211. DOI:https://doi.org/10.1145/3311350.3347190.

[48] Alexandrovsky, D., Putze, S., Bonfert, M., Höffner, S., Michelmann, P., Wenig, D., Malaka, R. and Smeddinck, J.D. 2020. Examining Design Choices of Questionnaires in VR User Studies. *Conference on Human Factors in Computing Systems - Proceedings.* (2020), 1–21. DOI:https://doi.org/10.1145/3313831.3376260.

[49] Kumaravel, B.T., Nguyen, C., DiVerdi, S. and Hartmann, B. 2019. TutoriVR: A video-based tutorial system for design applications in virtual reality. *Conference on Human Factors in Computing Systems - Proceedings* (2019).

[50] Yang, U. and Kim, G.J. 2002. Implementation and evaluation of "just follow me": An immersive, VR-based, motion-training system. *Presence: Teleoperators and Virtual Environments.* (2002). DOI:https://doi.org/10.1162/105474602317473240.

[51] Oliveira, D.M., Cao, S.C., Hermida, X.F. and Rodríguez, F.M. 2007. Virtual reality system for industrial training. *IEEE International Symposium on Industrial Electronics* (2007), 1715–1720.

[52] Chittaro, L. and Buttussi, F. 2015. Assessing knowledge retention of an immersive serious game vs. A traditional education method in aviation safety. *IEEE Transactions on Visualization and Computer Graphics.* 21, 4 (2015), 529–538. DOI:https://doi.org/10.1109/TVCG.2015.2391853.

[53] Bowman, D.A., Hodges, L.F., Allison, D. and Wineman, J. 1999. The educational value of an information-rich virtual environment. *Presence: Teleoperators and Virtual Environments.* 8, 3 (Jun. 1999), 317–331. DOI:https://doi.org/10.1162/105474699566251.

[54] Galvan-Bobadilla, I., Ayala-Garcia, A., Rodriguez-Gallegos, E. and Arroyo-Figueroa, G. 2013. Virtual reality training system for the maintenance of underground lines in power distribution system. *2013 3rd International Conference on Innovative Computing Technology, INTECH 2013* (2013), 199–204.

[55] Corato, F., Frucci, M. and Di Baja, G.S. 2012. Virtual training of surgery staff for hand washing procedure. *Proceedings of the Workshop on Advanced Visual Interfaces AVI* (2012), 274–277.

[56] Carlson, P., Peters, A., Gilbert, S.B., Vance, J.M. and Luse, A. 2015. Virtual training: Learning transfer of assembly tasks. *IEEE Transactions on Visualization and Computer Graphics.* 21, 6 (2015), 770–782.

DOI:https://doi.org/10.1109/TVCG.2015.2393871.

[57] Cheng, A., Yang, L. and Andersen, E. 2017. Teaching Language and Culture with a Virtual Reality Game. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* (2017), 541–549. DOI:https://doi.org/10.1145/3025453.3025857.

[58] Assarroudi, A., Heshmati Nabavi, F., Armat, M.R., Ebadi, A. and Vaismoradi, M. 2018. Directed qualitative content analysis: the description and elaboration of its underpinning methods and data analysis process. *Journal of Research in Nursing.* 23, 1 (2018), 42–55. DOI:https://doi.org/10.1177/1744987117741667.

[59] Elo, S. and Kyngäs, H. 2008. The qualitative content analysis process. *Journal of Advanced Nursing.* (2008). DOI:https://doi.org/10.1111/j.1365-2648.2007.04569.x.

[60] Foxman, M., Leith, A.P., Beyea, D., Klebig, B., Chen, V.H.H. and Ratan, R. 2020. Virtual Reality Genres: Comparing Preferences in Immersive Experiences and Games. *Extended Abstracts of the 2020 Annual Symposium on Computer-Human Interaction in Play* (2020), 237–241.

[61] Sharma, H.N., Alharthi, S.A., Dolgov, I. and Toups, Z.O. 2017. A framework supporting selecting space to make place in spatial mixed reality play. *CHI PLAY 2017 - Proceedings of the Annual Symposium on Computer-Human Interaction in Play.* (2017), 83–100. DOI:https://doi.org/10.1145/3116595.3116612.

[62] Mousas, C., Kao, D., Koilias, A. and Rekabdar, B. 2020. Real and Virtual Environment Mismatching Induces Arousal and Alters Movement Behavior. *Proceedings - 2020 IEEE Conference on Virtual Reality and 3D User Interfaces, VR 2020* (2020).

[63] Head-Mounted Display Market by Technology (AR, VR), Application (Consumer, Commercial, Enterprise & Industry, Engineering & Design, Military, Medical), Product (Head Mounted, Eyewear), Component, Connectivity, and Geography.: 2019. *https://www.marketsandmarkets.com/Market-Reports/head-mounted-display-hmd-market-729.html.*

[64] Deci, E.L. and Ryan, R.M. 2012. Self-determination theory. (2012).

[65] Deterding, S. 2015. The Lens of Intrinsic Skill Atoms: A Method for Gameful Design. *Human-Computer Interaction.* 30, 3–4 (2015), 294–335. DOI:https://doi.org/10.1080/07370024.2014.993471.

[66] Paas, F.G. 1992. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of educational psychology.* 84, 4 (1992), 429.

[67] Sweller, J., Van Merrienboer, J.J.G. and Paas, F.G.W.C. 1998. Cognitive architecture and instructional design. *Educational psychology review.* 10, 3 (1998), 251–296.

[68] Deci, E.L. and Ryan, R.M. 2012. *Motivation, personality, and development within embedded social contexts: An overview of self-determination theory.* Oxford University Press.

[69] White, R.W. 1959. Motivation reconsidered: The concept of competence. *Psychological Review.* 66, 5 (1959), 297.

[70] Baumeister, R.F. and Leary, M.R. 1995. The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin.* 117, 3 (1995), 497.

[71] Chirkov, V., Ryan, R.M., Kim, Y. and Kaplan, U. 2003. Differentiating autonomy from individualism and independence: A self-determination theory perspective on internalization of cultural orientations and well-being. *Journal of Personality and Social Psychology.* 84, 1 (2003), 97.

[72] McAuley, E., Duncan, T. and Tammen, V. V 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport.* 60, 1 (1989), 48–58.

[73] Kennedy, R.S., Lane, N.E., Berbaum, K.S. and Lilienthal, M.G. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology.* (1993). DOI:https://doi.org/10.1207/s15327108ijap0303_3.

[74] Witmer, B.G. and Singer, M.J. 1998. Measuring presence in virtual environments: A presence questionnaire. *Presence: Teleoperators and Virtual Environments.* (1998). DOI:https://doi.org/10.1162/105474698565686.

[75] Buhrmester, M., Kwang, T. and Gosling, S.D. 2011. Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on psychological science.* 6, 1 (2011), 3–5.

[76] Chandler, J. and Shapiro, D. 2016. Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology.* 12, (2016).

[77] Mason, W. and Suri, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods.* 44, 1 (2012), 1–23. DOI:https://doi.org/10.3758/s13428-011-0124-6.

[78] Kelly, J.W., Cherep, L.A., Lim, A.F., Doty, T. and Gilbert, S.B. 2021. Who Are Virtual Reality Headset Owners? A Survey and Comparison of Headset Owners and Non-Owners. (2021).

[79] Wai, J., Lubinski, D. and Benbow, C.P. 2009. Spatial Ability for STEM Domains: Aligning Over 50 Years of Cumulative Psychological Knowledge Solidifies Its Importance. *Journal of Educational Psychology.* (2009). DOI:https://doi.org/10.1037/a0016127.

[80] Ma, X., Cackett, M., Park, L., Chien, E. and Naaman, M. 2018. Web-Based VR Experiments Powered by the Crowd. (2018). DOI:https://doi.org/10.1145/3178876.3186034.

[81] Rice, M., Koh, R., Lui, Q., He, Q., Wan, M., Yeo, V., Ng, J. and Tan, W.P. 2013. Comparing avatar game representation preferences across three age groups. *CHI '13 Extended Abstracts on Human Factors in Computing Systems on - CHI EA '13.* (2013), 1161. DOI:https://doi.org/10.1145/2468356.2468564.

[82] Suznjevic, M., Mandurov, M. and Matijasevic, M. 2017. Performance and QoE assessment of HTC Vive and Oculus Rift for pick-and-place tasks in VR. *2017 9th International Conference on Quality of Multimedia Experience, QoMEX 2017* (2017).

[83]   Lim, S. and Reeves, B. 2009. Being in the Game: Effects of Avatar Choice and Point of View on Psychophysiological Responses During Play. *Media Psychology*. 12, 4 (2009), 348–370. DOI:https://doi.org/10.1080/15213260903287242.

[84]   Tabachnick, B.G. and Fidell, L.S. 2007. *Using multivariate statistics*. Allyn & Bacon/Pearson Education.

[85]   Tukey, J.W. 1949. Comparing individual means in the analysis of variance. *Biometrics*. 5, 2 (1949), 99–114.

[86]   Cohen, J. 2013. *Statistical power analysis for the behavioral sciences*. Routledge.

[87]   Miles, J. and Shevlin, M. 2001. *Applying regression and correlation: A guide for students and researchers*. Sage.

[88]   Paivio, A. 1991. Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology/Revue canadienne de psychologie*. 45, 3 (1991), 255.

[89]   Mayer, R.E. and Moreno, R. 2003. Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*. 38, 1 (Mar. 2003), 43–52. DOI:https://doi.org/10.1207/S15326985EP3801_6.

[90]   Wouters, P., Paas, F. and van Merriënboer, J.J.G. 2008. How to optimize learning from animated models: A review of guidelines based on cognitive load. *Review of Educational Research*.

[91]   Rogers, K., Ribeiro, G., Wehbe, R.R., Weber, M. and Nacke, L.E. 2018. Vanishing Importance. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*. (2018), 1–13. DOI:https://doi.org/10.1145/3173574.3173902.

[92]   Van Delden, R., Moreno, A., Poppe, R., Reidsma, D. and Heylen, D. 2017. A thing of beauty: Steering behavior in an interactive playground. *Conference on Human Factors in Computing Systems - Proceedings*. 2017-May, (2017), 2462–2472. DOI:https://doi.org/10.1145/3025453.3025816.

[93]   Alharthi, S.A., Raptis, G.E., Katsini, C., Dolgov, I., Nacke, L.E. and Toups, Z.O. 2018. Toward understanding the effects of cognitive styles on collaboration in multiplayer games. *Proceedings of the ACM Conference on Computer Supported Cooperative Work, CSCW*. (2018), 169–172. DOI:https://doi.org/10.1145/3272973.3274047.

[94]   Drey, T., Jansen, P., Fischbach, F., Frommel, J. and Rukzio, E. 2020. Towards progress assessment for adaptive hints in educational virtual reality games. *Conference on Human Factors in Computing Systems - Proceedings*. (2020). DOI:https://doi.org/10.1145/3334480.3382789.

[95]   Glassman, E.L., Lin, A., Cai, C.J. and Miller, R.C. 2016. Learnersourcing Personalized Hints. *CSCW*. (2016).

[96]   Odili Uchidiuno, J., Hammer, J., Koedinger, K. and Ogan, A. 2021. Fostering Equitable Help-Seeking for K-3 Students in Low Income and Rural Contexts. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), 1–14.

[97]   Harpstead, E., Myers, B.A. and Aleven, V. 2013. In search of learning: Facilitating data analysis in educational games. *Conference on Human Factors in Computing Systems - Proceedings* (2013).

[98]   Johnson, D., Nacke, L.E. and Wyeth, P. 2015. All about that Base: Differing Player Experiences in Video Game Genres and the Unique Case of MOBA Games. (2015).

[99]   Schrier, K. 2014. *Learning, education and games. Volume one: Curricular and design considerations*. Carnegie Mellon University.

[100] Klopfer, E., Scheintaub, H., Huang, W., Wendel, D. and Roque, R. 2009. The simulation cycle: Combining games, simulations, engineering and science using StarLogo TNG. *E-Learning and Digital Media*. 6, 1 (2009), 71–96.

[101] Tyack, A., Wyeth, P. and Klarkowski, M. 2018. Video game selection procedures for experimental research. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (2018).

[102] Cho, J.Y., Andersen, E. and Kizilcec, R.F. 2021. Delivery Ghost: Effects of Language Immersion and Interactivity in a Language Learning Game. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), 1–7.

[103] de Oliveira, A., Khamis, M. and Esteves, A. 2020. Using a VR field study to assess the effects of visual and haptic cues in "in-the-wild" locomotion. *CEUR Workshop Proceedings* (2020).

[104] Gerling, K.M., Klauser, M. and Niesenhaus, J. 2011. Measuring the impact of game controllers on player experience in FPS games. *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments, MindTrek 2011*. (2011), 83–86. DOI:https://doi.org/10.1145/2181037.2181052.

[105] Birk, M. and Mandryk, R.L. 2013. Control your game-self. January 2014 (2013), 685. DOI:https://doi.org/10.1145/2470654.2470752.

[106] Márquez Segura, E., Waern, A., Moen, J. and Johansson, C. 2013. The design space of body games: technological, physical, and social design. *Proceedings of the SIGCHI conference on Human Factors in computing systems* (2013), 3365–3374.

[107] Xu, Y., Barba, E., Radu, I., Gandy, M., Shemaka, R., Schrank, B., MacIntyre, B. and Tseng, T. 2011. Pre-patterns for designing embodied interactions in handheld augmented reality games. *2011 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities* (2011), 19–28.

[108] Pillat, R., Nagendran, A. and Lindgren, R. 2012. Design requirements for using embodied learning and whole-body metaphors in a mixed reality simulation game. *2012 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities (ISMAR-AMH)* (2012), 105–106.

[109] Kao, D. 2019. The Effects of Anthropomorphic Avatars vs. Non-Anthropomorphic Avatars in a Jumping Game. *The Fourteenth International Conference on the Foundations of Digital Games* (2019).

[110] Kao, D. and Harrell, D.F. 2015. Mazzy: A STEM Learning Game. *Foundations of Digital Games*. (2015).

[111] Bonsignore, E.M., Hansen, D.L., Toups, Z.O., Nacke, L.E., Salter, A. and Lutters, W. 2012. Mixed reality games. *Proceedings of the ACM 2012 conference on computer supported cooperative work companion* (2012), 7–8.

[112] Sobel, K., Bhattacharya, A., Hiniker, A., Lee, J.H., Kientz, J.A. and Yip, J.C. 2017. It wasn't really about the PokéMon: parents' perspectives on a location-based mobile game. *Proceedings of the 2017 CHI Conference on Human Factors in*

Computing Systems (2017), 1483–1496.

[113] Aleven, V., Dow, S., Christel, M., Stevens, S., Rosé, C., Koedinger, K., Myers, B., Flynn, J.B., Hintzman, Z., Harpstead, E. and others 2013. Supporting social-emotional development in collaborative inquiry games for K-3 science learning. *Proceedings of the 9th Games+ Learning+ Society Conference-GLS* (2013).

[114] Davis, K., Boss, J.A. and Meas, P. 2018. Playing in the virtual sandbox: Students' collaborative practices in Minecraft. *International Journal of Game-Based Learning (IJGBL)*. 8, 3 (2018), 56–76.

[115] Buffum, P.S., Frankosky, M., Boyer, K.E., Wiebe, E.N., Mott, B.W. and Lester, J.C. 2016. Collaboration and gender equity in game-based learning for middle school computer science. *Computing in Science & Engineering*. 18, 2 (2016), 18–28.

[116] Yuan, Y., Cao, J., Wang, R. and Yarosh, S. 2021. Tabletop Games in the Age of Remote Collaboration: Design Opportunities for a Socially Connected Game Experience. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021), 1–14.

[117] DiSalvo, B.J., Crowley, K. and Norwood, R. 2008. Learning in Context: Digital Games and Young Black Men. *Games and Culture*. 3, 2 (Feb. 2008), 131–141. DOI:https://doi.org/10.1177/1555412008314130.

[118] Richard, G.T. 2013. Gender and gameplay: Research and future directions. *Playing with virtuality: Theories and methods of computer game studies*. (2013), 269–284.

[119] Kumar, P., Vitak, J., Chetty, M., Clegg, T.L., Yang, J., McNally, B. and Bonsignore, E. 2018. Co-designing online privacy-related games and stories with children. *Proceedings of the 17th ACM Conference on Interaction Design and Children* (2018), 67–79.

[120] Easterday, M.W., Aleven, V., Scheines, R. and Carver, S.M. 2011. Using tutors to improve educational games. *International Conference on Artificial Intelligence in Education* (2011), 63–71.

[121] Kao, D. and Harrell, D.F. 2016. Exploring the Impact of Avatar Color on Game Experience in Educational Games. *Proceedings of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI 2016)*. (2016).

[122] Kao, D. and Harrell, D.F. 2016. Exploring the Effects of Encouragement in Educational Games. *Proceedings of the 34th Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI 2016)*. (2016).

[123] Kao, D. 2020. The effects of juiciness in an action RPG. *Entertainment Computing*. (2020). DOI:https://doi.org/10.1016/j.entcom.2020.100359.

[124] Denny, P., Mcdonald, F., Empson, R., Kelly, P. and Petersen, A. 2018. Empirical Support for a Causal Relationship Between Gamification and Learning Outcomes. (2018), 1–13. DOI:https://doi.org/10.1145/3173574.3173885.

[125] Hamilton, M., DiSalvo, B. and Fullerton, T. 2021. Mindful Gaming: User Experiences with Headspace and Walden, a Game. *International Conference on Human-Computer Interaction* (2021), 3–19.

[126] Kao, D. 2019. Exploring How Preference and Perceived Performance Vary in Different Game Genres Across Time of Day. *Proceedings of The Fourteenth International Conference on the Foundations of Digital Games* (2019).

[127] Almeda, M.V., Kleinman, E., Jemmali, C., Ithier, C., Rowe, E. and El-Nasr, M.S. 2020. Labeling debugging in may's journey gameplay. *Proceedings of the 51st ACM Technical Symposium on Computer Science Education* (2020).

[128] Chaffin, A., Doran, K., Hicks, D. and Barnes, T. 2009. Experimental evaluation of teaching recursion in a video game. *Proceedings of the 2009 ACM SIGGRAPH Symposium on Video Games - Sandbox '09*. 1, 212 (2009), 79. DOI:https://doi.org/10.1145/1581073.1581086.

[129] Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popović, Z. and Players, F. 2010. Predicting protein structures with a multiplayer online game. *Nature*. (2010). DOI:https://doi.org/10.1038/nature09304.

[130] Weintrop, D., Holbert, N., Horn, M.S. and Wilensky, U. 2016. Computational thinking in constructionist video games. *International Journal of Game-Based Learning (IJGBL)*. 6, 1 (2016), 1–17.

[131] Lee, M.J. and Ko, A.J. 2014. A demonstration of gidget, a debugging game for computing education. *2014 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)* (2014), 211–212.

[132] Esteves, A., Shin, Y. and Oakley, I. 2020. Comparing selection mechanisms for gaze input techniques in head-mounted displays. *International Journal of Human Computer Studies*. (2020). DOI:https://doi.org/10.1016/j.ijhcs.2020.102414.

[133] Nielsen 2017. Games 360 U.S. Report.

[134] Steam 2021. Steam Top Sellers (VR Supported) (https://store.steampowered.com/search/?filter=topsellers&vrsupport=402). (2021).

[135] Kao, D., Mousas, C., Magana, A.J., Harrell, D.F., Ratan, R., Melcer, E.F., Sherrick, B., Parsons, P. and Gusev, D. 2020. Hack.VR: A Programming Game in Virtual Reality. *ACM Foundations of Digital Games* (2020).