Are Game Design and User Research Guidelines Specific to Virtual Reality Effective in Creating a More Optimal Player Experience? Yes, VR PLAY

Heather Desurvire[1(✉)] and Max Kreminski[2(✉)]

[1] User Behavioristics + University of Southern California, Los Angeles, CA 90230, USA
heather@userbehavioristics.com
[2] University of California at Santa Cruz, Santa Cruz, CA 95064, USA
mkremins@ucsc.edu

Abstract. Virtual reality (VR) presents new usability, human-computer interaction, and playability challenges for developers, user-experience researchers, and designers. In addition to facing the traditional challenges, developers and researchers of VR games and VR experiences must contend with issues of physicality (including physical activity and physical discomfort), spatiality, and new or intensified physiological, psychological, and social considerations. However, many existing resources intended to help designers and game-user researchers work through usability and playability issues do not address these VR-specific challenges. This paper introduces the Virtual Reality PLAY (VR PLAY) guidelines, a set of guidelines intended to help developers, designers, and user researchers create more usable and playable VR games and experiences by optimizing the user and player experience for virtual reality.

Keywords: Player experience · Heuristics · Game principles · VR PLAY GAP · PLAY · Game design · Game-user research · Game usability Virtual reality

1 Introduction

The VR PLAY guidelines have been created to assist in evaluating and designing an improved virtual reality (VR) user experience. They are based on experience with previous VR development and design; multiple research sessions on consumer VR games and other playful or creative VR experiences; a review of existing research in the field of VR; and discussions with colleagues who study VR. Based on our success in developing and utilizing design principles for use by researchers and designers of AAA games to optimize the player experience, we developed an understanding of some of the unique challenges of VR not covered by previous work. This motivated the creation of the VR PLAY guidelines.

A. Marcus and W. Wang (Eds.): DUXU 2018, LNCS 10918, pp. 40–59, 2018.
https://doi.org/10.1007/978-3-319-91797-9_4

The VR PLAY guidelines are intended to assist designers and game-user researchers of VR games when working through usability and playability issues that are unique to, or intensified by, VR. The guidelines can be found on a public site, https://userbehavioristics.squarespace.com/vr-play. They consist of five categories: Usability, Playability, VR Immersion, Creative VR, and New Player Experience. Each guideline contains a short summary suggesting a principle that designers and researchers may use to gain insight into their own games; a longer explanation of the reasoning and principles behind the guideline; and examples of existing VR games and software that adhere to or violate the guideline, demonstrating how the guideline may be used. This paper introduces the guidelines, describes how they were developed, and presents a study that lends support to the validity of the guidelines as a design tool.

In our study, the guidelines were introduced to the designer of an existing VR experience developed for the Vive VR system. The experience was sponsored by and developed for Steelcase in conjunction with the Mobile & Environmental Media Lab at the University of Southern California. After being introduced to VR PLAY, the designer created a revised version of the experience. Each player in the study was exposed to two versions of the VR experience: Version A (designed without use of the guidelines) and Version B (designed with the guidelines in mind). We conducted interviews with the designer to evaluate his understanding of the VR PLAY guidelines, and we performed a side-by-side evaluation of player experiences in the two versions to determine if the guidelines had a significant positive impact on the designer's understanding of player-experience issues and the usability and playability of the revised design. These findings support the utility of the VR guidelines in improving the player experience.

The study identified many useful principles. These included UI design, spatiality, physicality, and some playability components. Most of these were specific to VR and not found in other guidelines. Two guidelines concerning the halo effect and VR ethics were found to be not directly actionable but nevertheless helpful to the designer in an advisory context. In addition, several guidelines—especially those concerning challenge, pace, and intermediate and long-term goals of the game—were not applicable in this study. These guidelines pertained to specific gameplay experiences not present in the current study's VR design. To evaluate these guidelines, the authors plan to conduct a study with a VR game that includes the relevant gameplay experiences.

Altogether, the study found the guidelines to be significantly useful when evaluating the usability and playability of existing games and when designing and optimizing the player experience in new VR games.

1.1 Motivation for VR PLAY

Virtual reality is an emerging medium with significant potential to enrich experiences in fields ranging from games and entertainment to health and creativity tools. As VR technology becomes increasingly accessible and affordable, the medium is undergoing a process of rapid expansion.

VR experiences, in comparison to traditional game console and computer player experiences, pose several unique usability and playability challenges. The increased immersion has the potential to accentuate the negative and positive aspects of an

experience, for instance by inducing physical discomfort or motion sickness [34] even as the higher fidelity of the experience instills in players a greater sense of satisfaction and delight [1]. In addition to considering all the usability and playability issues already known to exist in other digital games, designers of VR experiences must contend with issues of physicality (including physical activity and physical discomfort), spatiality, and new or intensified psychological and social phenomena [15]. Moreover, users and players of VR software and games are impacted by these new issues in a variety of ways. Some players may find they are especially sensitive to certain novel issues while for other players these very same issues may pass entirely unnoticed.

Usability and playability in video games are well-studied subject areas. There are resources [2, 6–8, 21] available to designers who want to improve the usability and playability of their game designs. Despite the growing popularity of VR, however, few of these resources have been designed with VR in mind. As a result, designers and evaluators of VR games and other VR play experiences have largely been left to their own devices in dealing with some of the novel challenges posed by VR.

To address such issues, we have created the VR PLAY guidelines. These guidelines for usability and playability can be used in two ways: as a set of heuristics against which existing VR games and play experiences can be evaluated, and as a design tool intended to inform the development of new VR games. To demonstrate the effectiveness of these guidelines as a design tool, we have employed them in a case study involving the revision of a work-in-progress playful VR experience.

1.2 Development of VR PLAY

In developing the VR PLAY guidelines, we drew on existing research in usability and playability for games and other software and on research in VR itself, including physical, psychological, social, and ethical perspectives on VR. We also conducted our own user research on VR games, have extensive experience with player research on VR games, and gleaned further information from discussions with other VR researchers. We divided the information we collected into several categories, each of which we concluded the guidelines should represent.

GAP and PLAY Principles. The Game Approachability Principles (GAP) [6, 14] and the Playability Principles (PLAY) [7] are two sets of heuristics or principles pertaining to general game usability and playability. The GAP heuristics deal primarily with issues of new-player experience and the processes by which players learn how to play a new game. The PLAY heuristics are broader in scope, encompassing a wide variety of issues that range from enduring play and player motivation to usability and user interface design. Both sets of principles have proven useful in our own research on over 60 games with AAA clients over the last several years, and countless other companies have used them successfully too. They have also been taught to over 150 students at the University of Southern California, one of the top universities that offer interactive media and games as a major. The students report using them in their own work now, including work for AAA publishers and indie studios.

VR PLAY Guidelines. Due to their generality, both the GAP and the PLAY heuristics are largely platform agnostic. Consequently, most of the advice they offer is applicable to VR. At the same time, however, these principles do not specifically address usability or playability issues that are unique to VR. Thus, adherence to these principles could be described as "necessary but not sufficient" for designers who wish to create positive player experiences in VR. In assembling the VR PLAY guidelines, we chose to include adapted and consolidated versions of the most important GAP and PLAY heuristics to highlight the general usability, playability, and new-player-experience issues that are likely to have the greatest impact on VR games.

Presence. One advantage of VR experiences over traditional video games and other software is the user's or player's increased sense of presence: the subjective perception that they are physically present in the virtual environment [5, 22]. Several forms of presence have been proposed, including social presence [10] and behavioral and cognitive presence [11, 24]. Presence can also be measured in several ways [3, 12, 17, 24, 27, 29, 34].

Presence is a key element of immersion, as more coherent experiences—in which multiple aspects of the experience, such as visuals and audio, work together to create the illusion of a "real" continuous world—are generally perceived by players as more immersive overall. Players who experience a heightened sense of presence feel that their experiences in the virtual environment are more "real" and thus more enjoyable when the experience itself is a positive one. They may also feel that their actions are more meaningful or consequential. As a result, creating and maintaining a sense of presence can be an important tool for improving the player experience in VR and non-VR games alike.

Embodiment. A factor closely related to presence is embodiment: the extent to which the player in a virtual environment identifies with the body of the player character and feels as though they are truly inhabiting it. Embodiment of the player character can be facilitated in a variety of ways, including by allowing the player to observe the way their actions directly map to equivalent actions by the player character (e.g., in a virtual mirror) [30]. Embodiment of certain characters may also pose ethical concerns, such as when a player inhabits the body of a character who commits violent actions [19].

Embodiment, like presence, is a key element of immersion. It can serve to further engage players in the experience by presenting them with the opportunity to embody characters they find particularly compelling. Many players find it enjoyable to engage in role-playing by temporarily taking up some of the attitudes and decision-making strategies of the character they are currently embodying, rather than maintaining their own real-life perspective during gameplay. In this sense, embodiment can be seen as a key contributor to player delight [2].

Ethics. Due to the heightened sense of presence and embodiment, VR environments often feel more similar to the real world than virtual environments in traditional video games. This greater sense of reality is accompanied by increased ethical issues [9, 19, 30]. For instance, research has shown that playing a first-person shooter game with a realistic gun-replica controller rather than a traditional mouse can lead to heightened levels of aggression in players [16]. Because players in VR experiences may experience

"bleed" between the real and virtual worlds to a greater degree than players of traditional video games, it is especially pressing that designers of VR experiences take ethical aspects into consideration [2].

An especially acute example of an ethical issue related to VR can be found in the interaction between VR and post-traumatic stress disorder (PTSD). Due to their perceived "reality," VR experiences may trigger PTSD flashbacks, causing players to suddenly and unexpectedly relive their memories of past traumatic events. Since a player who is experiencing a PTSD flashback may not have the presence of mind to remove the VR headset or otherwise disengage with the VR experience, it may be necessary to provide an in-game safe space where players can recover. However, for the same reasons that VR experiences may act as triggers, they may also be applied to the treatment of PTSD in the form of exposure therapy taking place in a VR environment [26].

Physicality. Unlike traditional video games, which generally do not require high levels of physical activity or exertion, VR experiences often involve a significant degree of physical activity. Designers of VR experiences must consider the players' physical comfort; their physical ability to perform the actions that the game asks of them; their need for breaks or cooldown time between periods of intense physical activity; and their differing physical traits and capabilities, including physical fitness, height, and physical disability [21, 33]. For instance, since VR games often require players to exert themselves physically, it is possible for a VR game to ask players to perform physical feats that are beyond their actual capabilities. If players are not conditioned for intense physical activity, but the game encourages them to keep playing anyway, players who are not aware of their own physical limits may find themselves taking on more than they can handle in terms of physical exertion. Simultaneously, physical activity can be used to increase player immersion by giving them the sensation that they are performing the same physical actions that the player character performs in the game world—for instance, physically wielding a motion controller as if it were a gun or a sword to control a weapon of this type in the game.

Organization into Categories. We organized the information gathered from these resources into five broad categories: Usability, Playability, VR Immersion, Creative VR, and New Player Experience. We synthesized the information in each category to create between 4 and 9 specific guidelines, for a total of 33 guidelines across all five categories. Each guideline includes an overall explanation of a principle for designers to follow, followed by examples of the principle being adhered to and violated. This way, it demonstrates the outcome both with and without the principle. The finalized VR PLAY guidelines were then published on a website to be easily accessible for designers.

2 Procedure

2.1 Study Design

We created a study with two phases to test the impact of the VR PLAY guidelines on the design of an example VR experience.

In Phase 1, a designer without prior knowledge of the guidelines created an initial version, Version A, of the VR experience. He was then introduced to the guidelines and given a period in which to revise the VR experience using the guidelines. This culminated in the design of Version B of the same experience.

In Phase 2, a user study was conducted to identify which of the two versions adhered more closely to the VR PLAY guidelines, and which was closer to an optimal player experience overall. Both versions were evaluated on all five of the top-level criteria addressed by the guidelines. This user study followed a within-subject design and included eight players, all of whom played through both Version A and Version. The order in which each player went through these versions was counterbalanced to randomize any order effects.

2.2 Hypotheses

Within this study, we tested two hypotheses: Hypothesis A and Hypothesis B. Hypothesis A was that the designer's overall understanding of issues related to VR experience, user experience, and player experience would improve between Version A and Version B. This could be determined by examining the designer's impressions of user- and player-experience issues through the interviews and surveys, as well as by comparing the number of guidelines that were adhered to and violated in Version A and Version B. Note that the designer had to complete Version A as their best effort at creating an optimal player experience.

Hypothesis B was that Version B would consequently contain fewer violations of and more adherences to the VR PLAY guidelines than Version A across all five of the categories. Thus, we also predicted that Version B would be closer to an optimal player experience overall.

2.3 Phase 1. VR Experience: Designer Study

The existing VR experience that was revised within this case study is known as the "Tracked Chair" experience because it makes use of VR spatial tracking to create a virtual counterpart of a physical (real-world) office chair. The virtual chair is co-located with and follows the movements of the physical chair, meaning that a participant who is wearing a VR headset can see and sit down in the chair. This experience was constructed for the Vive platform and makes use of a Vive controller to track the location of the physical chair.

In the Tracked Chair experience, the participant must first "unveil" the chair by removing a virtual cloth covering. Then, they must sit in the chair and make use of several virtual interface elements: a set of virtual screens attached to the chair that can be displayed and hidden at will and a button that can change the color of the virtual chair.

The Tracked Chair experience was one entry in a larger series of VR experiences created within a University of Southern California research lab in cooperation with industry partners at the furniture company Steelcase. The experiences in this series adapt existing industrial design processes at Steelcase into a fictional future world in which virtual and augmented reality technologies have become deeply integral to the

Steelcase design process. The Tracked Chair scene portrays a final "unveiling" or "presentation" stage in the design process in which a finished furniture design is presented to a larger audience (outside the original design team) for the first time. Altogether, this series of experiences forms an "immersive design fiction" intended to inspire innovative thinking about the potential future of industrial design. This immersive design fiction uses VR to "realistically" portray a vision of the future that would currently be thought of as science fiction, allowing designers to directly interact with simulated mockups of futuristic design tools and social practices that could not otherwise be realized today [20].

Preliminary Interview with Designer. The designer of the existing VR experience had no prior exposure to usability or playability guidelines of any kind. We conducted a preliminary interview with the designer to discover what kinds of changes he already intended to make to the design, and whether he believed a set of VR-specific usability and playability guidelines would aid him in making these revisions.

Exposure to VR PLAY Guidelines. We then introduced the designer to the VR PLAY guidelines and provided him with a brief explanation of how the guidelines could be applied. To gauge any change in his perception of the guidelines' potential utility, we asked him several additional questions. Then, we instructed him to spend a week revising his existing design with the guidelines in mind.

After Exposure to VR PLAY Guidelines; Before Revisions. Immediately after being introduced to the guidelines, but before having a chance to apply them to the revision process, the designer stated that he could now think of many specific improvements he wanted to try. Many of these potential improvements made reference to ideas or language drawn from individual guidelines; in particular, while continuing to talk about "feedback" (using terminology introduced by guideline A1, Provide Feedback), he began to adopt terminology from other guidelines, framing several of the improvements using terms like "demonstration" (E4, Build Self-Efficacy through Demonstration & Practice); "knowledge transfer" (A5, Build on Real-World Knowledge + E1, Design for Knowledge Transfer); "goals" (B4, Provide Clear Goals); and "error prevention" (A2, Prevent Player Error) for the first time. He found these guidelines especially helpful because they gave him the vocabulary he needed to discuss issues he had noticed in some form already but did not previously know how to coherently express.

After Revisions. After a week-long revision process incorporating the guidelines, the designer rated Version B as 2.5 (where 1 = totally unlike the desired player experience and 5 = exactly like the desired player experience) in closeness to the desired player experience, noting that there had been substantial improvement over Version A, but that he would still not want to present Version B as a "finished" product. He rated the usefulness of the guidelines as 5 overall (where 1 = not useful and 5 = very useful), while rating their usefulness for the specific tasks (a) and (b) at 4.5 and 5, respectively. Overall, he felt that the guidelines had been especially useful in inspiring ideas for changes, but that the limited timeframe of the revision process meant that "if I went into

Unity [the game design tool] again for five minutes with the guidelines, I'm sure I'd come up with a whole bunch of other stuff to address." He also stated that he intends to use the guidelines from the very beginning of the design process when working on virtual reality projects in the future.

The designer's rating of his own understanding of how to use the guidelines dropped slightly, from 5 to 4.5 (where 1 = no understanding and 5 = complete understanding). He suggested, however, that this was because his initial assessment was over-optimistic, and that he still "definitely understood how to use them" overall. The guidelines, in his mind, discussed problems that "everyone [designing for virtual reality] deals with," and that—although it sometimes took a bit of searching—he was able to identify at least one applicable guideline for every design problem he encountered during the revision process.

Debriefing Interview with Designer. At the end of the revision process, we met with the designer to conduct a debriefing interview. The purpose of this interview was to develop an understanding of his design thinking and to gather evidence regarding whether and how his thinking had changed as a result of exposure to the VR PLAY guidelines. The designer was asked about any changes he made to the design, his use of the guidelines during the revision process, how useful he perceived the guidelines to be, and why he found them useful.

Overview of Designer and Design Portion of the Study. Altogether, the designer's involvement in the study can be summarized as a sequence of several distinct steps:

- Initial interview and survey (before exposure to guidelines)
- Design of Version A
- Initial exposure to VR PLAY guidelines
- First post-exposure interview and-survey (immediately after exposure to guidelines)
- Design of Version B (with access to guidelines)
- Second post-exposure and -survey (after finishing the design of Version B)

2.4 Phase 2. User Study Prep

Following the revision process, two versions of the designer's VR experience existed. To determine what impact the VR PLAY guidelines had on the revisions that were made, we used the guidelines to conduct a comparative heuristic evaluation of the two versions [13, 23].

User Study of Tracked Chair Experience. To test whether the guidelines resulted in significant improvements to the overall usability and playability of the VR experience, we conducted a user study. We recruited eight participants with levels of VR experience ranging from none to experienced.

To acclimate participants to the experience of being in a VR environment, each participant was first given a few minutes in an empty VR world. This was intended to mitigate any first-time-experience issues, including difficulties with the Vive controllers. Participants were then instructed to play through Versions A and B of the VR experience. To minimize (counterbalance) order effects, odd-numbered participants played through Version A first while even-numbered participants played through

Version B first. After playing through both versions, each participant took part in a debriefing interview and filled out a survey about their experiences.

Hypotheses. Our first hypothesis (H:A) was that, after working with the guidelines for a week, the designer would have a greater overall understanding of user- and player-experience issues. This hypothesis was supported by the interviews with the designer and by the comparative heuristic evaluation of Versions A and B. In later interviews (both immediately after being introduced to the guidelines and after an entire week of working with the guidelines), the designer began to use terminology from the guidelines to articulate a more sophisticated understanding of the reasons for certain player-experience issues, and how these issues could potentially be solved. Additionally, Version B was found to contain substantially more instances of adherence to the guidelines and fewer violations of the guidelines than Version A. (Version A had 43 instances of adherence and 28 violations; Version B had 79 instances of adherence and 15 violations.) This suggests that the designer understood the guidelines, internalized them, and successfully applied them to the design revision process.

Our second hypothesis (H:B) was that Version B would contain fewer violations of and more adherences to the VR PLAY guidelines than Version A across all five of the issue categories (Usability, Playability, VR Immersion, Creative VR, and New Player Experience) that the guidelines address, and, thus, that Version B would be closer to an optimal player experience overall. This hypothesis was mostly supported by the comparative heuristic evaluation of the two versions, which found that Version B performed better (i.e., had more instances of adherence and fewer violations) than Version A in four of the five categories (Playability, VR Immersion, Creative VR, and New Player Experience). The one exception to this hypothesis was in the Usability category, which saw a decline in both the total number of adherences and the total number of violations from Version A to Version B; this is believed to be largely the result of a bug in Version B that was absent from Version A and had a negative impact on usability but was not part of the intended revisions to the design. (See Table 2 for the exact changes in total numbers of adherences and violations across each category from Version A to Version B.)

Analysis of User Study. Players went through the experience one at a time, observed by three coders. The coders took notes on what happened during the players' time in each version of the experience, which they then used to identify VR PLAY guidelines that had been adhered to and/or violated. To ensure inter-rater reliability, all coders were experienced in using the VR PLAY guidelines and were given the same guidelines on how to code the player experience.

Each coder submitted a single set of data, documenting how many players had experienced each guideline being adhered to and/or violated within each version of the VR experience. These sets of data were then aggregated by taking the mean, median, and mode of each value. Since there was little variance between these values, we report only the means (see Table 1).

Table 1. VR PLAY adherences and violations, Version A vs. Version B (Dark gray = clear improvement from A to B; light gray = all zeros/not applicable) N = 8 Players.

	VERSION A		VERSION B	
	Adherences	**Violations**	**Adherences**	**Violations**
A. Usability	***21 [A1, A5, A8]***	***8 [A1]***	***18 [A1, A5]***	***7 [A2, A5, A7]***
A1. Feedback	8	8	14	0
A2. Error prev	0	0	0	2
A3. Burden	0	0	0	0
A4. Status	0	0	0	0
A5. Real-World	5	0	4	2
A6. Review	0	0	0	0
A7. UI	0	0	0	3
A8. Navigation	8	0	0	0
B. Playability	***3 [B1]***	***3 [B4]***	***11 [B3, B6]***	***1 [B4]***
B1. Control	3	0	0	0
B2. Challenge	0	0	0	0
B3. Engage	0	0	8	0
B4. Goals	0	3	0	1
B5. Variety	0	0	0	0
B6. Social	0	0	3	0
C. VR Immersion	***11 [C5, C7]***	***4 [C3]***	***15 [C1, C7]***	***4 [C3]***
C1. Presence	0	0	7	0
C2. Embody	0	0	0	0
C3. Side Effects	0	4	0	4
C4. Comfort	0	0	0	0
C5. Safe Space	3	0	0	0
C6. Inclusion	0	0	0	0

C7. Obstruct	8	0	8	0
C8. Halo Effect	0	0	0	0
C9. Ethics	0	0	0	0
D. Creative VR	***8 [D1]***	***1 [D3]***	***8 [D1]***	***0 [N/A]***
D1. Empower	8	0	8	0
D2. Social	0	0	0	0
D3. Phase	0	1	0	0
D4. Inspiration	0	0	0	0
E. NPE	***0 [N/A]***	***12 [E5, E6]***	***27 [E1, E4, E5, E6]***	***3 [E1, E3]***
E1. Transfer	0	0	3	1
E2. Sandbox	0	0	0	0
E3. Scaffolding	0	0	0	2
E4. Practice	0	0	8	0
E5. Integrate	0	4	8	0
E6. Stepwise	0	8	8	0
Total	**43**	**28**	**79**	**15**

3 Analysis

3.1 Highly Impactful Guidelines

List of Impactful Guidelines. The data show an especially clear improvement from Version A to Version B in adherence to the following specific guidelines.

A1. Provide Feedback. Four out of eight players experienced violation of this guideline in Version A, reporting that the buttons used to deploy screens around the chair were either too subtle to notice (a slight shift in color from gray to black) or occurred out of the player's line of sight (as players looked down at the buttons to use them, they missed the screens appearing and disappearing around the chair). These violations were absent from Version B, which used gestures to deploy and hide the screens. Some

supplementary feedback, such as bursts of confetti on successful completion of certain actions, was added to Version B, increasing player delight.

B3. Encourage Player Engagement. In Version B, all players (8/8) found that exploration was rewarded with new and engaging feedback that did not appear in Version A. Two interactions that players found especially engaging in Version B were swiping to remove the cloth covering from the chair and spinning the chair to "calibrate" it, neither of which were present in Version A.

C1. Create a Sense of Presence. Version B adds several features that highlight the physicality of the virtual environment, including the two interactions discussed under B3. In addition to encouraging player engagement, these improvements helped to create a sense of physical presence in the world.

E4. Build Self-efficacy through Demonstration and Practice. Version A contained no demonstration of any of the mechanics players needed to use to interact with the chair. Version B demonstrated how to use all these mechanics, showing a transparent "ghost" version of the Vive controller repeatedly acting out the motions the player would need to perform. This made players substantially more confident in the knowledge that they were performing the actions "correctly."

E5. Integrate Tutorial with Gameplay. Version A did not make any explicit attempt to teach players how to interact with the various mechanics. Version B added introductory text and demonstration for several of these mechanics and integrated them smoothly with gameplay, alternating sections of explanatory scaffolding with sections in which players could freely practice with the mechanics they had learned so far.

E6. Teach Mechanics One Thing at a Time (Stepwise Learning). Version A of the experience introduced several new interactions in quick succession (revealing the chair, interacting with it as a mixed physical/virtual object, sitting in it, and using the buttons to raise and lower screens) without pausing to explain any of them in further detail. Version B, on the other hand, clearly separated these steps by displaying introductory text each time a new interaction was introduced.

What Does This Mean? The clear improvements in adherence to these guidelines between Version A and Version B indicate that the designer both understood and internalized the principles of usability and playability that these guidelines introduce. He was successfully able to use the guidelines to revise Version A's design for improved player experience in Version B.

3.2 Guidelines Not Applicable

List of Non-applicable Guidelines. Several of the guidelines were not applicable to this design and, as such, were not observed to be violated or adhered to during the user study. These included the following:

A3. Avoid Burden on Memory. Neither Version A nor Version B required players to remember things they had learned for more than a few minutes at a time, and there were

relatively few things to remember in both versions. As such, players had no difficulty remembering the things they did need to recall, and no hints were needed to mitigate burden on memory in either version.

A4. Show Player Status and World State. In both versions, all elements of player status and world state were immediately visible in the virtual environment at all times. The VR experience did not keep track of any hidden, intangible, or abstract elements of status or state (such as a health bar, score, or inventory). As such, the visualization of these abstract elements did not pose any usability or playability concerns.

A6. Let Players Review What They've Learned. As with guideline A3, there was little for players to review in either version. Both versions were only a few minutes long, and players had no difficulty remembering how to use the few mechanics they needed to remember, even without any kind of review.

B2. Ramp Up Challenge Gradually. Unlike many games, neither Version A nor Version B featured any kind of scaling difficulty. There were no enemies for players to defeat, scores for players to earn, or time limits on players' exploration of the virtual environment.

B5. Support a Variety of Players. The brief duration of the experience and relatively simple mechanics present in both Version A and Version B ensured that most players would be able to play without difficulty.

C2. Facilitate Embodiment of Character. Neither Version A nor Version B attempted to show or tell players whom they were playing as. However, players did not seem to notice or comment on this, possibly suggesting that they did not notice, or mind the fact, that they did not know whom the player character was meant to be, or they might have assumed it was themselves.

C4. Keep the Player Comfortable. Player discomfort was not a significant issue in either version, due to the limited duration of the experience and the relative ease of performing the few physical gestures required. Neither version required any kind of strenuous physical exertion on the part of the players.

C6. Provide a Safe Space. Since neither version of the experience gave players the impression they were under any kind of threat, the entirety of both experiences could be seen as a safe space. Players could stop playing at any time without exiting the experience and would suffer no negative repercussions within the virtual world for doing so.

C8. Be Aware of the First-Time Halo Effect. Players will often be wowed by the VR experience, even if the interface and/or game is poorly designed, so their perception at first-time play is likely to be more positive than with more experience in VR. Therefore, a player's unusual enthusiasm is to be discounted as first-play enthusiasm. This guideline is primarily advisory in nature and does not make any specific suggestions about how the final experience should work. As such, instances in which it is adhered to or violated are not visible in the observation of player behavior. Instead, one must ask designers or observe the design process directly to determine whether this guideline was adhered to or violated.

C9. Design Ethically. To design ethically means, for example, to beware of triggering events in a person, such as triggering a PTSD event, without an escape area. Like guideline C8, this guideline is primarily advisory in nature. Instances of adherence to or violation of this guideline are not directly visible in the observation of player behavior.

D2. Facilitate Performance and Social Creativity. Neither version of the experience offered players the opportunity to perform for any kind of in-world audience. In addition, the experience was not primarily intended to facilitate creative expression, so player expression of creativity was sharply limited within both versions.

D4. Provide Creative Inspiration. As with guideline D2, neither Version A nor Version B was primarily intended to facilitate creative expression, and both versions sharply limited player expression of creativity. As such, it would not make sense for this experience to provide players with creative inspiration.

E2. Provide a Safe Place to Learn (Sandbox). As with guideline C6, neither Version A nor Version B ever gave the players the impression they were under any kind of threat. At no point were there significant negative consequences for any player action, meaning that players could stop to practice at any point without suffering any negative consequences.

What Does This Mean? Most of the guidelines that were found to be not applicable to this design are associated with game-like features, such as overarching goals, status, and scores. Due to this, these will likely be more applicable to games as opposed to the Tracked Chair experience. Their limited applicability to this design is primarily a consequence of the fact that this design lacks certain common game mechanics to which these guidelines apply.

Two of these guidelines, however, represent a special case: guidelines C8, Be Aware of the First-Time Halo Effect, and C9, Design Ethically. Both guidelines are advisory in nature. Rather than making concrete, directly actionable statements about the qualities of a successful game, they simply recommend that designers educate themselves about and remain aware of certain potential pitfalls during the design process. Thus, adherence to or violation of these two guidelines cannot readily be determined through observation of player behavior. In the future, it may be desirable to rewrite these guidelines using more action-oriented language or move them into a supplemental resource separate from (but attached to) the main guidelines list.

3.3 Player Perception of Changes Between Versions A and B

Usability. Due to the short time (one week) allotted for revisions, Version B contained several bugs that impeded the functionality of key features. None of these bugs were present in Version A. When players were asked to compare the usability of Versions A and B, they made frequent references to these bugs in their descriptions of the usability issues they encountered. As such, it is not possible to make a reliable direct comparison between player usability ratings for the two versions. However, observation of their experiences with the two versions demonstrated that, excepting the issues introduced by the bugs, players generally found Version B easier to use. In particular, the players who were least impacted by the bugs seemed to find the gestural controls in Version B

intuitive, and all players (even those impacted significantly by the bugs) were immediately able to intuit that the gestures in Version B could be reversed to perform inverse actions.

Delight. In addition, players clearly felt that Version B was more delightful than Version A overall, with 6/8 players stating that B was more delightful than A. Players commonly cited the increased feedback, the physicality of the cloth-removal gesture, and the overall sense that completing objectives felt more "rewarding" in Version B as reasons for this preference.

Comfort. According to their responses, most players (5/8) felt that both versions were equally comfortable to use, with no significant differences between Versions A and B in terms of physical comfort. However, there was a slight preference among some players for Version B; 2/8 players said they felt B was more comfortable (as the surrounding environment appeared to be physically open rather than enclosed), while only one player felt A was more comfortable (as she felt the gestural interactions with the chair in B could become physically tiring after a while). This player's preference in particular underscores the importance of principle C5, Provide a Safe Space: VR experiences that ask players to perform physically demanding tasks as a regular part of play should provide a safe in-world place where players may physically rest.

3.4 Categories of Each Guideline Adhered to and Violated

The guidelines are divided into five categories: Usability, Playability, VR Immersion, Creative VR, and New Player Experience (see Table 2). We hypothesized that, for each of these categories, Version B would violate fewer of the guidelines and/or adhere to more than Version A. This was found to be true for four categories (Playability, VR Immersion, Creative VR, New Player Experience and Usability). (See Table 2 for the exact changes in total adherences and violations in each category from Version A to Version B.)

Table 2. Changes in VR PLAY adherences and violations by category from Version A to Version B

	Δ # Adherences	Δ # Violations
Usability	−3	−1
Playability	+8	−2
VR immersion	+4	±0
Creative VR	±0	−1
NPE	+27	−9

By far the most significant improvements were seen in the New Player Experience category, with 27 additional instances of adherence to the guidelines and 9 fewer instances of violation. This can largely due to the substantial improvements that were made to the new-player experience in Version B, which added a variety of features intended to teach players how to use the mechanics at their disposal. No such features

were present in Version A, which left players almost entirely to their own devices in attempting to learn how to play.

The Usability category, meanwhile, was the only category to see an overall decline from Version A to Version B: although it had 1 fewer violation in Version B, it also had 3 fewer adherences. This can largely be explained by the impact of a bug introduced in Version B—due to the brief period (one week) allotted for the second iteration—that made certain gesture-based interactions in Version B function unreliably. For the players impacted by the bug, this had a substantial negative effect on the usability of Version B in comparison to Version A. Since the introduction of the bug in Version B was not deliberate but did have a definite impact on usability, it is not possible to judge whether the guidelines would have been shown to have a net positive effect on usability from Version A to Version B if the bug had not occurred.

4 Conclusion

4.1 Impact of the VR PLAY Guidelines

Between Versions A and B of the VR experience considered in this case study, a comparative heuristic evaluation found that Version B—the version created after the designer was introduced to the VR PLAY guidelines—violated fewer of these guidelines (28 versus 15 total violations for Versions A and B, respectively) and adhered to more (43 versus 79 total adherences). This indicates that the designer was successful in applying the guidelines to the revision process and confirms both Hypothesis A (that the designer's understanding of usability and playability principles increased with exposure to the guidelines) and Hypothesis B (that Version B would contain fewer violations of and more adherences to the guidelines than Version A).

Impact on the Designer. The results of the side-by-side heuristic evaluations of Version A and Version B suggest that the designer both understood and internalized the guidelines and was able to successfully leverage his understanding to improve the usability and playability in the revised design. In addition, his statements in interviews suggest that he acquired an expanded vocabulary for discussing usability and playability issues from his time spent with the guidelines as he continued to use terminology he learned from the guidelines in his interview responses, even after the end of the revision process.

In his own words, the designer felt that the guidelines were especially helpful as a tool to help him generate solutions to problems he was already aware of but did not yet know how to address. He also expressed a particular desire to use the guidelines again on VR projects in the future—this time ideally "from the very beginning" of the design process, and he rated the guidelines as 5 on a five-point scale (where 1 = not useful and 5 = very useful).

Impact on Version B as Experienced by Players. When surveyed, 6/8 players found Version B more delightful than Version A. This suggests that the changes made by the designer when working with the guidelines resulted in significant improvements to the overall enjoy-ability of the VR experience.

Due to the short playtime of the VR experience and the limited amount of physical activity required, players generally did not feel that physical comfort was an issue for them in either version. Consequently, 5/8 players expressed that the versions were approximately equal in terms of physical comfort. However, 2/8 players did express a slight preference for Version B over A while only 1/8 players expressed a slight preference for Version A in terms of comfort.

4.2 Which Guidelines Are Most Impactful?

Between Version A and Version B, we found a clear and significant improvement in adherence to several specific VR PLAY guidelines, including A1, B3, C1, E4, E5, and E6.

Notably, three of these highly impactful guidelines are from Category E (New Player Experience). This speaks to the potential significance of improvements in new-player experience relative to improvements in other areas: because Version A of the VR experience contained little to no guidance for new players, the addition of even relatively minimal guidance in Version B was a basic issue, yielding substantial improvements in overall player experience for relatively little investment in time and effort on the designer's part.

The absence of any guidelines from category D (Creative VR) from this list may be accounted for by the sharply limited nature of the creative expression possible within this VR experience. As the VR experience was not intended as a creative tool, players had few opportunities to exercise their creativity, limiting the applicability of creativity-specific guidelines. This leads us to conclude that some guidelines have more impact for some applications than for others.

Given the positive effects of the guidelines, we have evidence they are helpful in increasing the user experience.

4.3 Guidelines Not Applicable

Several guidelines, meanwhile, were found to be not applicable to the VR experience evaluated in this study. These guidelines included A3, A4, A6, B2, B5, C2, C4, C6, C8, C9, and D2.

The majority of these guidelines were not applicable because the VR experience being evaluated did not contain certain common game mechanics to which these guidelines apply. For instance, the experience did not include any form of ramping up the challenge or difficulty, did not track any kind of score, and took place over a very short span of time, limiting the applicability of guidelines pertaining to certain kinds of experiences and games. Almost all these guidelines would thus be applicable to games of different types.

Two of the guidelines that were not applicable stand apart from the others that were attributed to more game-like features not present in the current study. The guidelines C8, Be Aware of the First-Time Halo Effect, and C9, Design Ethically, are advisory in nature and do not make use of directly actionable language, instead merely recommending that designers keep certain potential pitfalls in mind during the design process. Adherence to and violation of these advisory guidelines thus cannot be

determined directly from observation of player behavior, making it difficult to apply these guidelines in an evaluative capacity. This might be addressed in the future, see below.

4.4 Next Steps

In order to evaluate the utility of the guidelines that were found to be not applicable in this VR experience study, we intend to perform follow-up studies that apply the VR PLAY guidelines to other types of VR games and experiences where they might be applicable. Candidates include strategy games, physical games (including exercise games), action games, first-person shooter games, and horror games. These game varieties include mechanics that were not present in the VR experience evaluated by this study, increasing the likelihood that the specific guidelines will be applicable to these other games.

The "advisory" guidelines C8 and C9 are potential candidates for future revision and reevaluation. They might be rewritten to use more actionable language, making it possible to evaluate whether they have been adhered to or violated directly through observation of player behavior. Alternatively, they could be moved to a supplementary appendix attached to the main guidelines list but not presented as equivalent in form and purpose to the other guidelines.

Acknowledgements. We want to gratefully thank all the brilliant and hardworking interns of User Behavioristics, including Jordan Klein, Laura Doumad, Clerisse Cornejo, Shing-Hoi Lau, and Eva Wierzbicki, and colleagues and reviewers such as Dennis Wixon, Charlotte Wiberg, and Katherine Isbister, who provided helpful comments on previous versions of this paper. The authors also gratefully acknowledge the support from Scott Fisher and Joshua McVeigh-Schultz of the Mobile & Environmental Media Lab at the University of Southern California and Donna Flynn, Mark Baloga, and Paul Noll from Steelcase Corporation, who partnered with us on their Tracked Chair VR experience.

References

1. Basdogan, C., Ho, C.H., Srinivasan, M.A., Slater, M.: An experimental study on the role of touch in shared virtual environments. ACM Trans. Comput.-Hum. Interact. **7**(4), 443–460 (2000)
2. Bowman, S.L.: Bleed: The Spillover Between Player and Character. Nordiclarp.org (2015). https://nordiclarp.org/2015/03/02/bleed-the-spillover-between-player-and-character/
3. Chastine, J.W., Nagel, K., Zhu, Y., Yearsovich, L.: Understanding the design space of referencing in collaborative augmented reality environments. In: Proceedings of Graphics Interface 2007, (GI 2007), pp. 207–214. ACM, New York (2007)
4. Chung, J., Gardner, H.J.: Measuring temporal variation in presence during game playing. In: Spencer, S.N. (ed.) Proceedings of the 8th International Conference on Virtual Reality Continuum and its Applications in Industry, (VRCAI 2009), pp. 163–168. ACM, New York (2009)
5. Crawley, D.: Introduction to Bayesian Statistics, April 2015. https://venturebeat.com/2015/04/18/were-not-talking-about-what-vr-is-doing-to-our-eyes-and-our-brains/. Accessed 18 Aug 2015

6. Desurvire, H., Wiberg, C.: Master of the game: assessing approachability in future game design. In: CHI 2008 Extended Abstracts on Human Factors in Computing Systems, (CHI EA 2008), pp. 3177–3182. ACM, New York (2008)
7. Desurvire, H., Wixon, D.: Game principles: choice, change & creativity: making better games. In: CHI 2013 Extended Abstracts on Human Factors in Computing Systems, (CHI EA 2013), pp. 1065–1070. ACM, New York (2013)
8. Dimopoulos, Y., Moraitis, P., Amgoud, L.: Characterizing the outcomes of argumentation-based integrative negotiation. In: Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, (WI-IAT 2008), vol. 2, pp. 456–460. IEEE Computer Society, Washington, DC (2008)
9. Fink, P.W., Foo, P.S., Warren, W.H.: Obstacle avoidance during walking in real and virtual environments. ACM Trans. Appl. Percept. **4**(1), Article no. 2 (2007)
10. Greef, P.D., Ijsselstijn, W.A.: Social presence in a home tele-application. Cyberpsychol. Behav. **4**(2), 307–315 (2004)
11. Hayhoe, M.M., Ballard, D.H., Triesch, J., Shinoda, H., Aivar, P., Sullivan, B.: Vision in natural and virtual environments. In: Proceedings of the 2002 Symposium on Eye Tracking Research & Applications, (ETRA 2002), pp. 7–13. ACM, New York (2002)
12. Hung, Y.H., Parsons, P.: Assessing user engagement in information visualization. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, (CHI EA 2017), pp. 1708–1717. ACM, New York (2017)
13. Jeffries, R., Desurvire, H.: Usability testing vs. heuristic evaluation: was there a contest? SIGCHI Bull. **24**(4), 39–41 (1992)
14. Kajiyama, T.: Botanical data retrieval system supporting discovery learning. In: Proceedings of the 1st ACM International Conference on Multimedia Retrieval, (ICMR 2011). ACM, New York (2011). Article 36
15. Kerse, D., Regenbrecht, H., Purvis, M.: Telepresence and user-initiated control. In: Proceedings of the 2005 International Conference on Augmented Tele-Existence, (ICAT 2005), pp. 239–240. ACM, New York (2005)
16. Kim, K.J., Biocca, F., Jeong, E.J.: The effects of realistic controller and real-life exposure to gun on psychology of violent video game players. In: Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication, (ICUIMC 2011). ACM, New York (2011). Article 49
17. Lehtinen, V., Hänninen, R., Toivonen, V., Oulasvirta, A., Kuuva, S., Saariluoma, P.: Developing a rapid questionnaire-based metric for experience of form. In: Norros, L., Koskinen, H., Salo, L., Savioja, P. (eds.) European Conference on Cognitive Ergonomics: Designing beyond the Product—Understanding Activity and User Experience in Ubiquitous Environments, (ECCE 2009). VTT Technical Research Centre of Finland, VTT, Finland (2009). Article 4
18. Lobel, A., Granic, I., Engels, R.: Stressful gaming, interoceptive awareness, and emotion regulation tendencies: a novel approach. Cyberpsychol. Behav. Soc. Netw. **17**(4), 222–227 (2014)
19. Madary, M., Metzinger, T.K.: Real virtuality: a code of ethical conduct Recommendations for good scientific practice and the consumers of VR-technology. Front. Robot. AI **3**, 3 (2013)
20. McVeigh-Schultz, J., Kreminski, M., Prasad, K., Hoberman, P., Fisher, S.S.: Immersive Design Fiction: Using VR to Prototype Speculative Interfaces and Interaction Rituals within a Virtual Storyworld (in process)
21. Mueller, F., Isbister, K.: Movement-based game guidelines. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (CHI 2014), pp. 2191–2200. ACM, New York (2014)

22. Nesbitt, K.V.: Modelling human perception to leverage the reuse of concepts across the multi-sensory design space. In: Stumptner, M., Hartmann, S., Kiyoki, Y. (eds.) Proceedings of the 3rd Asia-Pacific Conference on Conceptual Modelling, (APCCM 2006), vol. 53, pp. 65–74. Australian Computer Society, Inc., Darlinghurst (2006)
23. Nielsen, J.: Enhancing the explanatory power of usability heuristics. In: Adelson, B., Dumais, S., Olson, J. (eds.) Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 1994), pp. 152–158. ACM, New York (1994)
24. Nunez, D., Blake, E.: Cognitive presence as a unified concept of virtual reality effectiveness. In: Proceedings of the 1st International Conference on Computer Graphics, Virtual Reality and Visualisation, (AFRIGRAPH 2001), pp. 115–118. ACM, New York (2001)
25. Pätsch, G., Mandl, T., Hacker, C.W.: Using sensor graphs to stimulate recall in retrospective think-aloud protocols. In: Proceedings of the 5th Information Interaction in Context Symposium, (IIiX 2014), pp. 303–307. ACM, New York (2014)
26. Rothbaum, B.O., et al.: Virtual reality exposure therapy for PTSD Vietnam veterans: a case study. J. Trauma. Stress **12**(2), 263–271 (1999)
27. Schloerb, D.W.: A quantitative measure of telepresence. Presence: Teleoper. Virtual Environ. **4**(1), 64–80 (1995)
28. Shin, J., An, G., Lee, K.: Integration of a precise indoor position tracking algorithm with an HMD-based virtual reality system. In: Proceedings of the 2nd ACM International Workshop on Immersive Media Experiences, (ImmersiveMe 2014), pp. 23–26. ACM, New York (2014)
29. Slater, M., Steed, A.: A virtual presence counter. Presence: Teleoper. Virtual Environ. **9**(5), 413–434 (2000)
30. Souza, C.S., Leitão, C.F.: Semiotic engineering methods for scientific research. HCI Synth. Lect. Hum.-Cent. Inform. **2**(1), 1–222 (2009)
31. Turpin, A., Scholer, F., Mizzaro, S., Maddalena, E.: The benefits of magnitude estimation relevance assessments for information retrieval evaluation. In: Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval, (SIGIR 2015), pp. 565–574. ACM, New York (2015)
32. Yee, N., Bailenson, J.N., Rickertsen, K.: A meta-analysis of the impact of the inclusion and realism of human-like faces on user experiences in interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, (CHI 2007), pp. 1–10. ACM, New York (2007)
33. Yoo, S., Parker, C., Kay, J.: Designing a personalized VR exergame. In: Tkalcic, M., Thakker, D., Germanakos, P., Yacef, K., Paris, C., Santos, O. (eds.) Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization (UMAP 2017), pp. 431–435. ACM, New York (2017)
34. Von Mammen, S., Knote, A., Edenhofer, S.: Cyber sick but still having fun. In: Proceedings VRST 2016 of 22nd ACM Conference on Virtual Reality, pp. 325–326. ACM, New York (2016)
35. Witmer, B.G., Singer, M.J.: Measuring presence in virtual environments: a presence questionnaire. Presence: Teleoper. Virtual Environ. **7**(3), 225–240 (1998)