Construction of SVS: Scale of Virtual Twin's Similarity to Physical Counterpart in Simple Environments

Xuesong Zhang
xuesong.zhang@kuleuven.be
KU Leuven
Leuven, Belgium

Adalberto L. Simeone
adalberto.simeone@kuleuven.be
KU Leuven
Leuven, Belgium

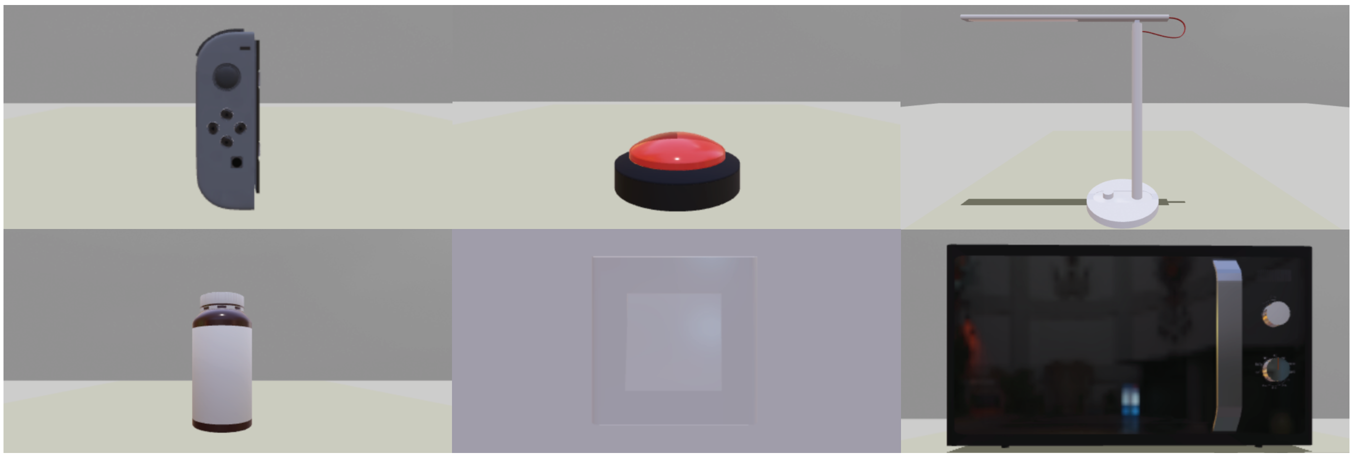

Figure 1: High-Fidelity Virtual Twins in study 1 (from left to right, top to bottom): Joy-Con, button, lamp, bottle, light switch, microwave.

ABSTRACT

Due to the lack of a universally accepted definition for the term "virtual twin", there are varying degrees of similarity between physical prototypes and their virtual counterparts across different research papers. This variability complicates the comparison of results from these papers. To bridge this gap, we introduce the Scale of Virtual Twin's Similarity (SVS), a questionnaire intended to quantify the similarity between a virtual twin and its physical counterpart in simple environments in terms of *visual fidelity*, *physical fidelity*, *environmental fidelity*, and *functional fidelity*. This paper describes the development process of the SVS questionnaire items and provides an initial evaluation through two between-subjects user studies to validate the items under the categories of visual and functional fidelity. Additionally, we discuss the way to apply it in research and development settings.

CCS CONCEPTS

• **Human-centered computing** → **User studies**; **Virtual reality**.

KEYWORDS

Virtual twin, Evaluation, Fidelity, Similarity, Questionnaire

SUI '24, October 7–8, 2024, Trier, Germany

ACM Reference Format:
Xuesong Zhang and Adalberto L. Simeone. 2024. Construction of SVS: Scale of Virtual Twin's Similarity to Physical Counterpart in Simple Environments. In *ACM Symposium on Spatial User Interaction (SUI '24), October 7–8, 2024, Trier, Germany*. ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3677386.3682100

1 INTRODUCTION

The use of virtual prototypes in Virtual Reality (VR) is gaining increased research attention, from their use in training to the evaluation of early designs ([2, 26, 32]), a practice also referred to as *Immersive Virtual Reality Evaluations* (IVREs) [37]. This type of virtual prototype is referred to as a "virtual twin" or "virtual replica" across research papers ([6, 21]). In this work, we define "virtual twin" as a digital replica intended to closely mirror its physical counterpart in terms of *three-dimensional appearance*, *interactive capabilities*, and *usability* within a virtual environment. This definition serves to distinguish it from the more widely recognised term "digital twin" [27], which emphasises real-time data exchange between the virtual copy and its physical entity, whereas a three-dimensional representation is not always necessary.

However, the term "virtual replica" or "virtual twin" currently lacks an established definition, resulting in varying degrees of similarity between physical prototypes and their virtual counterparts across different research papers. This variation makes it challenging to compare the results of those papers. Factors such as display fidelity and interaction fidelity have been shown to influence user performance, presence, engagement, and usability [17]. Existing research indicates that perceived visual accuracy can affect user-identified usability problems [37]. Additionally, the use of supernatural interaction techniques in VR may lead to misunderstandings about real-world working mechanisms [38]. Consequently, researchers suggest that virtual prototypes should closely resemble their physical counterparts to achieve comparable usability results during IVREs [37]. Nevertheless, understanding what is the optimal balance between the effort necessary to create virtual twins that accurately simulate their physical counterparts not only visually but also interactively, and the impact of these fidelities on user performance remains an open question.

To address this gap, this work introduces a questionnaire designed to quantitatively assess the perceptual similarity of virtual twins compared to their counterparts within a simplified virtual environment, called the ***S**cale of **V**irtual Twin **S**imilarity* (SVS). This questionnaire aims to 1) assess virtual twins created during the design and implementation processes, thereby improving the reliability of the usability inspection outcomes in subsequent phases; and 2) provide researchers and practitioners with a tool to compare results from user studies or other forms of evaluations, in light of the level of fidelity of the virtual twins assessed through this questionnaire.

To construct the questionnaire, we used the methodological framework proposed by Boateng et al. [1], starting with domain identification and generating items through both deductive and inductive approaches. Subsequently, we conducted a preliminary validation of the questionnaire via two user studies. In these studies, participants engaged in observations and interactions with a variety of virtual twins, which featured a wide range of visual fidelities and interaction modalities. These objects were deliberately selected to showcase various grasp types and manipulation techniques. Subsequently, participants were asked to complete this new questionnaire.

Through a comprehensive analysis of the collected data and participant feedback, we aim to validate the items of the proposed questionnaire. This process involves iterative refinements, including clarifying ambiguous items, consolidating similar items, eliminating redundancies, and incorporating any missing elements. The results show the proposed questionnaire can help users assess the similarity of virtual twins in terms of visual and functional fidelity.

In this work, the contribution is two-fold: 1) we propose a novel questionnaire to quantify the perceptual similarity of a virtual twin to its physical counterpart in a simple virtual environment; 2) We introduce how to integrate this questionnaire into IVREs.

2 RELATED WORK

In the following, we take a closer look at the concept of virtual twin, relationship between fidelity and realism and its impact on user experience in VR.

2.1 Virtual Twin

As discussed earlier, a virtual twin is a digital replica of a physical counterpart that mirrors its appearance, interactions, and usability characteristics in virtual environment. Designed for use in IVREs, virtual twins allow users to interact with and evaluate these digital replicas in a controlled virtual setting, providing insights into their real-world counterparts. A virtual twin of a physical product or environment has been applied in many research fields.

One of the first examples of a one to one mapping between the physical prop and its virtual counterpart was demonstrated by Hoffman, who investigated the impact of passive haptics on the realism of the VR experience [12]. In a between-subjects study participants were assigned to either a "see only" or "see and touch" group. The task consisted in manipulating a virtual plate and the whole experience was set in a virtual kitchen. Participants in the "see only" group could grab the virtual plate by moving their "cyberhand" (a fingerless cycling glove tracked via a Polhemous sensor) inside the virtual object and pressing a button on a 3D-mouse. The physical prop used in the "see and touch" consisted of a real plate that was paired to the virtual plate and tracked through a Polhemus sensor placed on the bottom of the plate. They could pick up the object by grabbing the real plate, which would cause its movement to be mirrored in the VR system. A digitized texture of the real plate was applied to the virtual plate. Participants were asked to evaluate their experience by scoring five questions about the realism of the prop. Results indicated that participants of the "see and touch" group rated these questions with significantly more positive scores than the "see only" group.

In the industry product development process, researchers have proposed employing virtual twins in certain testing activities to obtain feedback from customers [8]. Similarly, in the field of architecture design, a virtual twin of a real building has been employed; researchers created a virtual twin which replicated interior details and ambience. They then compared the user behaviour in the real building and its virtual twin, finding consistent behaviour patterns [14]. Virtual twins can also serve as a communication cue. In remote collaboration, AR and VR users have leveraged (semi-transparent) virtual twins, especially for task guidance [6, 20]. Additionally, in product design process, virtual twin can be applied in usability evaluation [37], as well as privacy and security testing [15].

While previous research has focused on the development of virtual twins using 3D Computer-Aided Design (CAD) models with the aim of achieving high fidelity, as evidenced in studies by Oda et al. [20], Wang et al. [33], and Zhang et al. [36], a notable gap exists in assessing users' perceptions of the similarity of these virtual twins.

2.2 Similarity, Fidelity and Realism

VR technology is capable of replicating diverse phenomena within virtual environments. Consequently, these virtual representations may contain elements that lack direct relevance in the physical actuality. An illustrative example is the conceptualisation of an imaginary spacecraft. While such entities are still difficult to manufacture in the tangible realm today, they can be seamlessly accommodated in the virtual environment.

In the field of Human-Computer Interaction (HCI), researchers have sought to evaluate the degree of similarity between virtual objects or worlds and their real-world counterparts. Within this context, the terms "realism" and "fidelity" are often used interchangeably to describe this attribute [25]. Especially, in the realm of digital games, researchers pointed out that realism involves narrative, perceptual, interactive, and psychological realism (fidelity) [25].

However, those two terminologies are not always equivalent. *fidelity* was considered as "the degree of accuracy to which a simulation, whether it is physical, mental, or both, represents a given frame of reality in terms of cues and stimuli, and permissible interactions" [31], where *realism* delves deeper into perception and resulting user experience [4]. A virtual representation could have high fidelity by accurately reproducing technical details, but it might lack realism if it does not evoke a natural or immersive user experience.

Due to the lack of an established standard for prototype fidelity, researches have brought forward different proposals to explain how it contributes to the overall realism of the simulation. In 2015, Ragan et al. categorised fidelity into three distinct types: interaction fidelity, display fidelity, and scenario fidelity [23]. While McMahan defines fidelity as comprising display fidelity, interaction fidelity, and simulation fidelity [17]. Later, in 2019, Within the domain of simulation-based serious games (SSGs), researchers have developed the General Conceptual Framework of Fidelity (GCFF), which categorises fidelity into objective and subjective dimensions. The objective dimensions include physical and functional fidelity, while the subjective dimensions encompass sensory, conceptual, and emotional fidelity [35]. Notably, this framework is not explicitly designed for virtual reality applications.

Researchers also took an in-depth look at each fidelity category. McMahan developed Framework of interaction fidelity analysis (FIFA), which suggests that interaction fidelity can be assessed through biomechanical symmetry, control symmetry, and system appropriateness in 2011 [17]. However, the literature currently lacks a standardised methodology for quantifying a technique's interaction fidelity, as indicated by recent studies [29]. Especially for haptic fidelity, Muender et al. assess the realism of haptic feedback with 14 factors with three categories: Sensing, Hardware and Software [18].

In the context of HCI prototyping, researchers also identified five dimensions to guide the planning and characterisation of prototypes. These are intended to enhance resource distribution efficiency in design and development: Level of Visual Refinement, Breadth of Functionality, Depth of Functionality, Richness of Interactivity, Richness of Data Model [16].

We proposed this questionnaire to not only quantify objective similarity and technical detail, but also to assess the extent to which the representation is perceived and experienced.

3 GENERATING QUESTIONNAIRE ITEMS

The development of the questionnaire in this study adhered to the methodology introduced by Boateng et al. [1]. It involves a nine-step process including item development, scale development and scale evaluation.

3.1 Step 1: Identification of the Domain(s) and Item Generation

The process of creating the questionnaire began with a deductive approach, focusing on literature review to identify relevant domains and generate questionnaire items [11, 24]. Our search spanned the ACM Digital Library and the IEEE Xplore Library, employing specific keywords: virtual reality AND fidelity AND simulation, virtual reality AND fidelity AND product design, virtual reality AND fidelity AND simulation, virtual reality AND realism AND simulation. Since the end of 2019, the original Oculus Quest has incorporated hand tracking feature, marking a further advance for standalone VR devices [22]. Given this milestone, the review covered publications from January 2020 to September 2023, during which we screened 726 papers for their relevance. Additionally, related references were also taken into investigation. Despite our thorough investigation, we found no validated questionnaires specifically designed to quantify the concept of similarity in this particular context. Other related frameworks, as listed in subsection 2.2, inspired our subsequent categorisation of the generator indicators.

To complement our deductive approach, we employed an inductive method by conducting a two-hour workshop with three VR experts. During the workshop, all experts analysed images and videos of VR applications extracted from 35 papers presented at the 2021 CHI Conference on Human Factors in Computing Systems (CHI). CHI is the leading international venue for research in the field of HCI and has the highest impact [10]. We selected 35 papers that included visual representations and videos of VR applications, excluding review papers. After examination, all experts are requested to reason about their personal perceptions of the realism and similarity of virtual objects. They should also identify the factors that influence their judgements of realism and similarity in comparison to real-world objects, drawing upon both their daily experiences and their expertise in the field. All comments were recorded and then grouped up after the workshop. After this phase, we constructed the initial iteration of the scale, encompassing four categories with 21 indicator items: **model fidelity** (seven items), **physical fidelity** (five items), **lighting fidelity** (five items), **functional fidelity** (five items).

3.2 Step 2: Content Validity

In this phase, we invited both experts and members of the target population to assess the initial questionnaire items.

Evaluation by Experts. We recruited seven experts to inspect the appropriateness of the content. These were experts from different fields: HCI, computer graphics, and gaming design. Each of them attended an individual interview with the first author. During the interview, each interviewer provided feedback pertaining to the utilisation of terminology, the expounded explanations, as well as their perspectives on the potential reduction or augmentation of additional items. After this round, we interactively restructured the questionnaire with the following 4 categories, comprising 24 total indicator items: **visual fidelity** (VF)(7 items), **physical fidelity** (PF) (5 items), **environment fidelity** (EF)(6 items), and **functional fidelity** (FF)(6 items).

Evaluation by Target Population. We additionally extended invitations to three participants who comprise the target demographic, in order to assess this questionnaire through cognitive interviews. Within these interviews, participants articulated their cognitive processes while perusing the questionnaire items, affording us insight into the appropriateness of the formulations for their comprehension.

(a) (b) (c) (d) (e)

Figure 2: A bottle (a) and its virtual twins (b, c, d, e).

After the iterative process, we arrived at the finalised version of the questionnaire in the initial round, see Table 1. Each item is rated on a seven-point scale, ranging from 1 ("Strongly Disagree") to 7 ("Strongly Agree"), including a "Not Applicable" (N/A) option.

4 STUDY 1

We conducted a between-subjects user study to assess the preliminary questionnaire, focusing on the visual fidelity part. Participants were tasked with examining six distinct virtual objects (see Figure 1) within an immersive virtual environment and comparing them to their physical counterparts, presented either as tangible items or reference photographs.

4.1 Model Fidelity

Each of these six objects displayed four unique variations in visual presentation. Given that each participant observes only one of four fidelity variations per object, we selected six objects to ensure each variant is presented and to gather more comprehensive data across all variations.

These variations were introduced during the design process, where modifications were made to both modelling and material assignment. Modelling fidelity was assessed based on workload: simpler models were categorised as low fidelity, whereas more labour-intensive ones were classified as high fidelity. For instance, a low-fidelity bottle was constructed using a basic frustum of a cone combined with a cylinder (see Figure 2b, Figure 2c), while its high-fidelity counterpart involved detailed vertex adjustments and beveling (see Figure 2d, Figure 2e). The material for the low-fidelity model was created by merely adjusting the base diffuse colour in the Principled BSDF shader (see Figure 2b, Figure 2d); while the high-fidelity material incorporated advanced settings, such as additional subsurface colouring and reflective attributes, along with roughness adjustments using RGB Curves (see Figure 2c, Figure 2e). These modelling and material strategies collectively resulted in four distinct variations for each object.

4.2 Apparatus

To optimise participant recruitment, we employed an online distribution approach using the Babylon.js-based virtual environment—a real-time 3D engine that utilises a JavaScript library for rendering 3D graphics in web browsers via HTML5. The virtual object was initially modelled in Blender, after which it was refined with materials compatible with Babylon.js and export as *.glb* file. Accompanied

Table 1: SVS questionnaire

Visual Fidelity *The degree to which the physical equivalent and the virtual object look similar in terms of visual display.*
1. I felt that the dimensions (size/thickness) of the virtual object were the same as the real-world equivalent. 2. I felt that the shape of the virtual object was the same as the real-world equivalent. 3. I felt that the virtual object was as detailed as the real-world equivalent (in terms of minor structure, decorations, markings, and other types of geometric detail that can be perceived visually). 4. I felt that the virtual object surface colour/pattern appeared the same as the real-world equivalent. 5. I felt that the virtual object was as opaque/ transparent/reflective as the real-world equivalent. 6. I felt that the virtual object looked physically the same as the real-world equivalent (in terms of perceived rigidness, ability to deform or bend itself, etc.) 7. I felt that the virtual object looked identical to the real-world equivalent.
Physical Fidelity *The degree to which the physical equivalent and the virtual sound similar, and feel similar in terms of controls, and audio; as well as the physics models driving each of these variables.*
8. I felt that the surface of the virtual object had the same smoothness/roughness as the real-world equivalent. 9. I felt that the virtual object had the same temperature as the real-world equivalent. 10. I felt that the virtual object had the same weight as the real-world equivalent. 11. I felt that when touching the virtual object, the virtual object material sounds the same as the real-world equivalent. 12. I felt that the virtual object physically behaved the same as the real-world equivalent.
Environmental Fidelity *The degree to which the physical environment looked and illuminated.*
13. I felt that the virtual object was placed at the same distance from me as the real-world equivalent. 14. I felt that the dimensions of the virtual environment were the same as in the real-world environment. 15. I felt that the virtual object was illuminated in the same way as the real-world equivalent (e.g., colour temperature, shadows, self-shadow* appeared in the same way) **self-shadow: a shadow cast by a partial object and still displayed on the object.* 16. I felt that the location of shadows cast by the virtual object in the virtual environment was the same as in the real world environment. 17. I felt the darkness and detail of the shadows cast by virtual objects in the virtual environment is the same as in the real-world environment. 18. I felt that the overall illumination throughout the whole environment was the same as in the real world environment.
Functional Fidelity *The degree to which the virtual simulation acts like the physical equivalent in reacting to the operation executed.*
19. I felt that the virtual object has the same working mechanism as the real-world equivalent. 20. I felt that the virtual object reacted to my interactions/operations in the same way as the real-world equivalent. 21. I felt that the virtual object was easier to use than the real-world equivalent. 22. I felt that the virtual object was harder to use than the real-world equivalent. 23. I felt that I was as effective when interacting with the virtual object as with the real-world equivalent. 24. I felt that I could interact with the virtual object in the same way as with the real world equivalent.

by HTML-based instructional content, participants accessed the virtual environment via the headset-integrated browser to complete the user study. Assessing the virtual content with conventional PC browsers does not allow for direct observation of the virtual objects correctly. The static website is hosted online.

4.3 Procedure

Participants are invited to engage in the user study by accessing the study website[1] through either a headset-integrated browser or a conventional PC browser. Initially, they are introduced to the study by perusing the participant information on the website, which includes the purpose of the user study, the procedure, and instructions on how to access and interact with the virtual content. A unique ID is generated for each participant. Subsequently, they are prompted to give their consent and complete a demographic questionnaire. Successively, the website forwards them to a page where they can enter the virtual environment and use their headsets to observe the virtual twin (see Figure 1). After a one-minute observation period, a button appears, directing participants to a new page. There, participants are presented with the physical counterpart of the virtual object, either as a tangible item or as a reference photograph. Participants then compared both virtual twins and physical counterparts while filling in the questionnaire. This observation and response cycle is repeated for the next five objects. The sequence in which the objects are presented is randomised, and the visual fidelity levels of each object are also chosen at random. Upon completion of the study, each of the four levels of visual fidelity, ranging from low to high fidelity, will have been presented to the participant at least once.

4.4 Demographics

In total, 24 participants (16 male, 8 female) attended this user study, recruited through mailing lists, word-of-mouth, and social media. They had an average age of 27.46 years ($SD = 3.05$). All participants had prior VR experience and played video games. The user study was conducted either in our VR lab or at the participants' own place. The study was approved by the university ethic board.

5 STUDY 2

To examine items under functional fidelity, we conducted a second between-subjects user study. Similar to the first user study, participants were asked to observe and interact with the various virtual twins and compare these experiences to their own interaction with the corresponding tangible physical objects. Afterwards, they filled out the questionnaire.

5.1 Interaction Technology

Two different interaction technologies were tested in this study: *controller-based* interaction and *hand-based* interaction (see Figure 3). All interactions techniques were implemented based on the Oculus Interaction SDK [3] with Unity (version 2021.3.16f1) [30]. Participants were allowed to manipulate the virtual twin by directly poking, pushing, or pulling it with their hand, or by using the controller, replicating the same interactions they could perform with the physical counterparts in real life. However, users could not deform or disassemble either the virtual or the physical objects.

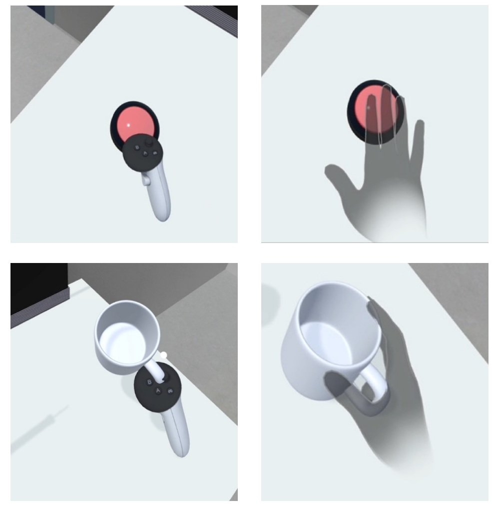

Figure 3: Interaction Technologies: controller-based interaction (left) and hand-based interaction (right).

5.2 Procedure

At the beginning, the participants received an introduction similar to that of the first study. Subsequently, participants are introduced to the interaction technology they will use in the following phases, with each participant experiencing only one interaction modality throughout the study. The user study comprised two phases. In the initial phrase, participants are asked to wear the headset and immerse themselves in the virtual environment, a virtual office (see Figure 4). In this setting, participants are assigned a series of predefined tasks (see subsection 5.3) to complete. Afterwards, they fill out the SVS questionnaire regarding the overall virtual environment (which will not be included in statistical analysis) and the System Usability Scale (SUS) focusing on the interaction technologies involved in the task. In the subsequent phase, participants wore the headset again and interacted with a selected virtual object from the virtual office. They then completed the SVS and SUS questionnaire regarding this chosen object. During the study, participants are encouraged to comment at any time. The entire study session was recorded.

5.3 Virtual Objects and Tasks

The virtual environment consisted of an office room containing the objects listed in Table 2. The object selection process was based on various grasp types employed during object interaction, as defined by the GRASP Taxonomy [7]. Additionally, we incorporated non-grasping actions, such as pushing or poking a button.

Participants appeared outside the office and had to first open the door, then activate the ceiling light using the toggle switch on the wall. Next, they are instructed to press a red button on the table to show the next task on the TV screen. Tasks include: observing and interacting with a mug, bottle, small office cabinet,

[1] Study website: https://hcisong.github.io/svs/

Figure 4: Virtual office setting

and large office cabinet, with the specific objective of finding a number to encourage detailed exploration and careful observation of each object; turning on the lamp; using the black rotary button to turn on the cube light, and adjusting its brightness and color. If participants are tasked with finding a number, they are required to verbally state the number once they locate it. When participants finish one task, they are asked to press the button again to reveal the next task.

Table 2: Objects in the Virtual Office

Object	Interaction Capability
Door with Handle	Rotatable, pushable, pullable
Table	No interaction abilities
Toggle switch	Toggleable
ceiling Light	Controlled through toggle switch
Red Button	Pressable
TV	Controlled through the red button
Mug	Movable
Bottle	Movable; lid is openable
Small Office Cabinet	Pushable and pullable drawers
Big Office Cabinet	Pushable and pullable doors

5.4 Demographics

We recruited 38 participants (20 male, 18 female) with an average age of 28.55 years ($SD = 4.64$). Partial participants from the first study also attended the second study after four months. They were recruited in the same way as in the first user study. Seven participants had no prior VR experience, and four stated they do not play video games in their daily life. This study received ethics approval, took place in our VR Lab, with each session lasting around 20 minutes.

6 RESULT

In Study 1, one participant quit after one observation, and another after two, resulting in 135 valid questionnaires. From Study 2, we collected 38 valid questionnaires. In total, 173 valid questionnaires were collected across both studies.

We examined the questionnaire with confirmatory factor analysis (CFA) along with Robust maximum likelihood estimation and Satorra-Bentler (SB) for the categories VF and FF.

Model Fit. The chi-square test indicated a significant difference between the model and observed data ($\chi^2 = 149.014$, df = 64, $p < 0.001$). Based on the modification indices, we identified covariations in the error terms for term 2 (*I felt that the shape of the virtual object was the same as the real-world equivalent*), term 19 (*I felt that the virtual object has the same working mechanism as the real-world equivalent*), term 22 *(I felt that the virtual object was harder to use than the real-world equivalent)*, hence, those terms were dropped. The revised questionnaire was as shown in Table 4. For a downloadable version, please refer to the supplemental file.

After revision, we ran the chi-square test again and obtained the result $\chi^2 = 65.491$, df = 52, $p = 0.099$; which indicates a good model fit. The Comparative Fit Index ($CFI = 0.987$) and Tucker-Lewis Index ($TLI = 0.983$) are both higher than the threshold 0.9 and indicate a good model fit. The Root Mean Square Error of Approximation (RMSEA) for the standard model was 0.046, suggesting a good model fit ($p = 0.553$), while the scaled RMSEA was even lower at 0.034, indicating an excellent fit ($p = 0.744$). The robust RMSEA was 0.036, also indicating an excellent model fit ($p = 0.694$). The Standardized Root Mean Square Residual (SRMR) indicated a good fit with a value of 0.068.

Table 3: The confirmed factors (CFA results).

	VF	FF
VF 1. Dimension	0.437	
VF 2. Detail Level	0.782	
VF 3. Color/Pattern	0.600	
VF 4. Opacity/Transparency/Reflectivity	0.751	
VF 5. Physical Appearance	0.484	
VF 6. Identity Match	0.651	
FF 1. Reaction		0.907
FF 2. Ease of use		0.930
FF 3. Effectiveness		0.912
FF 4. Interaction		0.690

Scale Reliability and Standardized Path Coefficients. Table 3 shows the standardized path coefficients of the CFA. The reliability of the scale was assessed using Cronbach's alpha, indicate an acceptable reliability for *Visual Fidelity* ($\alpha = 0.79$), and an excellent reliability of *Functional Fidelity* ($\alpha = 0.92$) [9].

Study 2. No significant results were detected in terms of the SUS score among the different interaction modalities ($Hand : MEAN = 70.79, SD = 17.99; Controller : MEAN = 74.47, SD = 13.27;$) with Mann-Whitney U Test ($U = 160, p = 0.559$)), since the data were not normally distributed which examined by Shapiro-Wilk test. We further investigated the relationship between FF score (the cumulative sum of each item), interaction technology, objects, and their interactions on SUS score. For every unit increase in FF, the SUS score increases by 2.1324, holding all other variables constant with statistical significance ($p = 0.002$), suggesting that when participant rated a higher FF score will also rated a higher usability.

Table 4: Updated SVS questionnaire in VF and FF

VF: Visual Fidelity *The degree to which the physical equivalent and the virtual object look similar in terms of visual display.*
VF 1. Dimension I felt that the dimensions (size/thickness) of the virtual object were the same as the real-world equivalent. **VF 2. Detail Level** I felt that the virtual object was as detailed as the real-world equivalent (in terms of minor structure, decorations, markings, and other types of geometric detail that can be perceived visually). **VF 3. Color/Pattern** I felt that the virtual object surface color/pattern appeared the same as the real-world equivalent. **VF 4. Opacity/Transparency/Reflectivity** I felt that the virtual object was as opaque/ transparent/reflective as the real-world equivalent. **VF 5. Physical Appearance** I felt that the virtual object looked physically the same as the real-world equivalent (in terms of perceived rigidness, ability to deform or bend itself, etc.) **VF 6. Identity Match** I felt that the virtual object looked identical to the real-world equivalent.
FF: Functional Fidelity *The degree to which the virtual simulation acts like the physical equivalent in reacting to the operation executed.*
FF 1. Reaction I felt that the virtual object reacted to my interactions/operations in the same way as the real-world equivalent. **FF 2. Ease of use** I felt that the virtual object was easier to use than the real-world equivalent. **FF 3. Effectiveness** I felt that the virtual object was harder to use than the real-world equivalent. **FF 4. Interaction** I felt that I could interact with the virtual object in the same way as with the real world equivalent.

7 DISCUSSION

In this section we first discuss the dropped terms, the found correlation between FF and SUS scores; Then we discuss limitations and future work; At the end we introduce how to involve SVS in IVREs.

7.1 Dropped Terms

Term 2. The exploration originally associated with Term 2 aimed to assess the similarity between objects in terms of their shapes. This assessment notably incorporated an evaluation of shape when participants were asked to consider the characteristics of Term 1 in relation to the spatial dimension.

Term 19. Term 19 aimed to explore participants' perceptions of the similarity in operational mechanisms between virtual twins and their physical counterparts, particularly in the context of task completion where specific step sequences are expected. However, participant responses varied significantly. One noted,"since it's in a virtual environment, all work is just done by the program; they are completely different from the real world". Another commented,"I'm not even clear about the mechanism of the object in the real world, so I can't answer this question and chose N/A". These divergent interpretations highlight the complexity of the term and its influence on participant responses. Consequently, Term 19 was discontinued due to the challenge of reconciling these differing perspectives and ensuring consistent data collection and analysis.

Term 22. Participants indicated in multiple-choice that the virtual object was not easier (Term 21) and also harder to use (Term 22) than the real-world equivalent. This outcome can be interpreted as suggesting that the interaction modalities in VR, which differ from the real world, might impair the usability of virtual twins. Nevertheless, the interview provided contradictory results, because they explained that the virtual object was as easy to use as the real-world equivalent. However, should this interview statement be true, we would expect that participants would indicate that the virtual object was neither easier (Term 21) nor harder to use (Term 22), i.e.,"the virtual object was as easy and as hard to use as the real-world equivalent". Therefore, we hypothesise that this contradiction between multiple choice and interview answers may result from the nature of the "dual process theories" [5, 13], which facilitates quick thinking and may inadvertently lead participants to agree with Term 22's statement about the increased difficulty of using the virtual object, and vice versa.

However, it's essential to acknowledge that this reaction does not universally apply to all participants. During the user studies, 17 participants questioned whether the inclusion of these terms was intended to verify the sincerity of their participation. As a result of these inquiries and the potential for misunderstanding induced by the binary response framework, Term 22 was subsequently omitted from the questionnaire.

7.2 Correlation between FF and SUS

The analysis demonstrated a good fit for our model, alongside a notable positive correlation between FF scores (the cumulative sum of each item in FF) and SUS scores. The objects selected for replication in our studies are commonly utilised in everyday life, serving as ground truth. When interaction modalities in VR closely mirror real-world settings, the skills learned or mastered in VR are more readily transferrable to real-life contexts, and vice versa. Consistently, we found that when participants perceived a high degree of similarity between the VR system and real-world usability, they were more inclined to utilise the system and rate it with a higher usability score. This phenomena resonates with the heuristic evaluation principle of *Match between the system and the real world* as outlined by Nielsen [19]. This principle underscores the importance of designing systems that mirror the real world to enhance usability and user adoption.

7.3 Limitations and Future Work

In this work, we explored the visual and functional fidelity, the examination of physical fidelity and environmental fidelity remains an area that still needs to be explored.

In both studies, participants completed the SVS across all four fidelity dimensions. Specifically, within the physical fidelity category, which includes elements such as weight and temperature, responses varied significantly. Some participants rated these aspects based on the perceived colour [28, 34] , while others selected the "N/A" option, indicating non-applicability. This variation can be attributed to the study's implementation constraints, which did not facilitate the alteration of these modalities through additional haptic feedback. Future work should incorporate wearable haptic devices or additional physical proxies to enhance the experimental design.

In the study 1, utilising the Babylon.js-based web VR provides an avenue for showcasing VR environments with heightened accessibility, allowing designers to make iterative changes easily and quickly. Despite participants can effortlessly access this streamlined virtual environment without the need to install any applications, the restricted interactivity inherent to this method constrains content, making the implementation of complex interactions challenging. Hence, SUS score were not collected in the first user study, preventing an analysis of the relationship between VF and SUS scores.

Furthermore, in the second user study, although participants had the opportunity to observe reference images or the physical objects in both phases of the study, they were not tasked with evaluating the usability of the physical objects. We did not explore the difference in term of SUS scores between the virtual twin and its real world counterparts.

7.4 Applying SVS in the design process

This questionnaire is designed to quantify the degree of similarity between virtual twins and their real-world counterparts. Specifically, in this work we focused on visual and functional fidelity.

In the early stages of development, the process begins with developers constructing a virtual environment. This construction is grounded on comprehensive documentation and enriched through collaborative discussions with designers, ensuring that both visual and interactive aspects are accurately replicated. To assess the fidelity of these virtual twins, additional designers and developers may be invited to engage with and assess the virtual environment, using the SVS as a comparative framework.

When it comes to enhancing existing products, the methodology involves creating a virtual twin and broadening the evaluation to encompass customers who have prior experience with the product. This expanded evaluation facilitates the gathering of detailed feedback on potential improvements from a wide array of users, employing the SVS for structured assessment.

During the similarity evaluation, researchers should set parameters for participants observation and interaction, such as a minimum observation duration, specific tasks, to ensure that participants engage sufficiently with the virtual twins. Following the completion of these tasks with one virtual twin, participants will fill in a SVS questionnaire, where they have the opportunity to comment on each item and provide reasons for their scores, either verbally or in writing. The cumulative sum of each item in the same category reflects the similarity in terms of this specific aspect; a higher number indicates a higher similarity. Feedback on the virtual twin is systematically collected through the SVS questionnaire and complemented by semi-structured interviews that explore the rationale behind the scores assigned to specific items. This comprehensive feedback mechanism enables the iterative refinement of the virtual twin, preparing it for subsequent phases of product evaluation involving the target user population.

Upon achieving refinement and approval, the virtual twin is disseminated across diverse regions and populations, serving as a vital tool for testing and implementing product improvements. This iterative, feedback-driven approach is critical for developing virtual twins that closely mirror the real product experience, thereby contributing to the enhancement of both new and existing products with precision and user-centred insights.

8 CONCLUSION

This work aims to develop a questionnaire that can quantify the similarity between a virtual twin and its physical counterpart in simple environments. This questionnaire consists of four parts: visual fidelity, physical fidelity, environmental fidelity, and functional fidelity. To validate the questionnaire in terms of visual and functional fidelity, we designed and conducted two user studies, collecting 173 valid responses. The results indicate that our model has a good fit. Future work should examine the physical and environmental fidelities and further investigate the relationship between perceived usability and the quantified similarity.

ACKNOWLEDGMENTS

This research is supported by *Internal Funds KU Leuven* (HFG-D8312-C14/20/078).

REFERENCES

[1] Godfred O Boateng, Torsten B Neilands, Edward A Frongillo, Hugo R Melgar-Quiñonez, and Sera L Young. 2018. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Frontiers in public health* 6 (2018), 149.

[2] Monica Bordegoni, Francesco Ferrise, and Joseba Lizaranzu. 2011. Use of interactive virtual prototypes to define product design specifications: A pilot study on consumer products. In *2011 IEEE international symposium on VR innovation*. IEEE, 11–18.

[3] Oculus Developer Center. [n. d.]. Interaction SDK Overview | Oculus Developers — developer.oculus.com. https://developer.oculus.com/documentation/unity/unity-isdk-interaction-sdk-overview/. [Accessed 14-02-2024].

[4] Lauren Christophers, Chi Tak Lee, and Brendan Rooney. 2023. Exploring Subjective Realism: Do evaluative realism and felt realism respond differently to different cues? *International Journal of Human-Computer Studies* 175 (2023), 103027. https://doi.org/10.1016/j.ijhcs.2023.103027

[5] Adele Diederich. 2023. A Dynamic Dual Process Model for Binary Choices: Serial Versus Parallel Architecture. *Computational Brain & Behavior* (2023), 1–28.

[6] Carmine Elvezio, Mengu Sukan, Ohan Oda, Steven Feiner, and Barbara Tversky. 2017. Remote Collaboration in AR and VR Using Virtual Replicas. In *ACM SIGGRAPH 2017 VR Village* (Los Angeles, California) *(SIGGRAPH '17)*. Association for Computing Machinery, New York, NY, USA, Article 13, 2 pages. https://doi.org/10.1145/3089269.3089281

[7] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M. Dollar, and Danica Kragic. 2016. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems* 46, 1 (2016), 66–77. https://doi.org/10.1109/THMS.2015.2470657

[8] Francesco Ferrise and Monica Bordegoni. 2011. Fast prototyping of virtual replica of real products. In *Virtual and Mixed Reality-Systems and Applications: International Conference, Virtual and Mixed Reality 2011, Held as Part of HCI*

International 2011, Orlando, FL, USA, July 9-14, 2011, Proceedings, Part II 4. Springer, 318–326.

[9] Darren George and Paul Mallery. 2019. *IBM SPSS statistics 26 step by step: A simple guide and reference.* Routledge.

[10] Nathalie Henry, Howard Goodell, Niklas Elmqvist, and Jean-Daniel Fekete. 2007. 20 years of four HCI conferences: A visual exploration. *International Journal of Human-Computer Interaction* 23, 3 (2007), 239–285.

[11] Timothy R Hinkin. 1995. A review of scale development practices in the study of organizations. *Journal of management* 21, 5 (1995), 967–988.

[12] Hunter G Hoffman. 1998. Physically touching virtual objects using tactile augmentation enhances the realism of virtual environments. In *Proceedings. IEEE 1998 Virtual Reality Annual International Symposium (Cat. No. 98CB36180)*. IEEE, 59–63.

[13] Daniel Kahneman, Patrick Egan, et al. 2011. Thinking, fast and slow: Farrar. *Straus and Giroux* 1 (2011).

[14] Saskia Kuliga, Panagiotis Mavros, Martin Brösamle, and Christoph Hölscher. 2020. Comparing Human Wayfinding Behavior Between a Real, Existing Building, a Virtual Replica, and Two Architectural Redesigns. In *Spatial Cognition XII*, Jurǵis Šķilters, Nora S. Newcombe, and David Uttal (Eds.). Springer International Publishing, Cham, 160–179.

[15] Florian Mathis, Joseph O'hagan, Kami Vaniea, and Mohamed Khamis. 2022. Stay home! conducting remote usability evaluations of novel real-world authentication systems using virtual reality. In *Proceedings of the 2022 International Conference on Advanced Visual Interfaces*. 1–9.

[16] Michael McCurdy, Christopher Connors, Guy Pyrzak, Bob Kanefsky, and Alonso Vera. 2006. Breaking the fidelity barrier: an examination of our current characterization of prototypes and an example of a mixed-fidelity success. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 1233–1242.

[17] Ryan Patrick McMahan. 2011. *Exploring the effects of higher-fidelity display and interaction for virtual reality games.* Ph. D. Dissertation. Virginia Tech.

[18] Thomas Muender, Michael Bonfert, Anke Verena Reinschluessel, Rainer Malaka, and Tanja Döring. 2022. Haptic fidelity framework: Defining the factors of realistic haptic feedback for virtual reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–17.

[19] Jakob Nielsen. 1994. Enhancing the explanatory power of usability heuristics. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 152–158.

[20] Ohan Oda, Carmine Elvezio, Mengu Sukan, Steven Feiner, and Barbara Tversky. 2015. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. 405–415.

[21] Luca Piovano, Gonzalo Calderon, Guillermo Del Campo, Edgar Saavedra, Francisco Luque, and Asunción Santamaría. 2022. Towards a Digital Twin for Smart Street Lighting systems Using a Virtual Reality interface. In *2021 Joint Conference-11th International Conference on Energy Efficiency in Domestic Appliances and Lighting & 17th International Symposium on the Science and Technology of Lighting (EEDAL/LS: 17)*. IEEE, 1–6.

[22] Meta Quest. 12-10-2019. Hand Tracking Deep Dive: Technology, Design, and Experiences. https://www.youtube.com/watch?v=gpQePH-Ffbw [Accessed on 04-08-2024].

[23] Eric D Ragan, Doug A Bowman, Regis Kopper, Cheryl Stinson, Siroberto Scerbo, and Ryan P McMahan. 2015. Effects of field of view and visual complexity on virtual reality training effectiveness for a visual scanning task. *IEEE transactions on visualization and computer graphics* 21, 7 (2015), 794–807.

[24] Tenko Raykov and George A Marcoulides. 2011. *Introduction to psychometric theory.* Routledge.

[25] Katja Rogers, Sukran Karaosmanoglu, Maximilian Altmeyer, Ally Suarez, and Lennart E. Nacke. 2022. Much Realistic, Such Wow! A Systematic Literature Review of Realism in Digital Games. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 190, 21 pages. https://doi.org/10.1145/3491102.3501875

[26] Adalberto L. Simeone, Robbe Cools, Stan Depuydt, João Maria Gomes, Piet Goris, Joseph Grocott, Augusto Esteves, and Kathrin Gerling. 2022. Immersive Speculative Enactments: Bringing Future Scenarios and Technology to Life Using Virtual Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 17, 20 pages. https://doi.org/10.1145/3491102.3517492

[27] Maulshree Singh, Evert Fuenmayor, Eoin P Hinchy, Yuansong Qiao, Niall Murray, and Declan Devine. 2021. Digital twin: Origin to future. *Applied System Innovation* 4, 2 (2021), 36.

[28] Charles Spence. 2020. Temperature-Based Crossmodal Correspondences: Causes and Consequences. *Multisensory Research* 33, 6 (2020), 645 – 682. https://doi.org/10.1163/22134808-20191494

[29] Markus Tatzgern and Christoph Birgmann. 2021. Exploring input approximations for control panels in virtual reality. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. IEEE, 1–9.

[30] Unity Technologies. [n. d.]. 2021.3.16f1 — unity.com. https://unity.com/releases/editor/whats-new/2021.3.16. [Accessed 14-02-2024].

[31] Jimmy Kyaw Tun, Guillaume Alinier, Jessica Tang, and Roger L. Kneebone. 2015. Redefining Simulation Fidelity for Healthcare Education. *Simulation & Gaming* 46, 2 (2015), 159–174. https://doi.org/10.1177/1046878115576103 arXiv:https://doi.org/10.1177/1046878115576103

[32] Alexandra Voit, Sven Mayer, Valentin Schwind, and Niels Henze. 2019. Online, VR, AR, Lab, and In-Situ: Comparison of Research Methods to Evaluate Smart Artifacts. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3290605.3300737

[33] Peng Wang, Xiaoliang Bai, Mark Billinghurst, Shusheng Zhang, Sili Wei, Guangyao Xu, Weiping He, Xiangyu Zhang, and Jie Zhang. 2021. 3DGAM: using 3D gesture and CAD models for training on mixed reality remote collaboration. *Multimedia Tools and Applications* 80 (2021), 31059–31084.

[34] Benjamin Wright. 1962. The influence of hue, lightness, and saturation on apparent warmth and weight. *The American Journal of Psychology* 75, 2 (1962), 232–241.

[35] Xiaozhen Ye, Per Backlund, Jianguo Ding, and Huansheng Ning. 2020. Fidelity in Simulation-Based Serious Games. *IEEE Transactions on Learning Technologies* 13, 2 (2020), 340–353. https://doi.org/10.1109/TLT.2019.2913408

[36] Xiangyu Zhang, Xiaoliang Bai, Shusheng Zhang, Weiping He, Peng Wang, Zhuo Wang, Yuxiang Yan, and Quan Yu. 2022. Real-time 3D video-based MR remote collaboration using gesture cues and virtual replicas. *The International Journal of Advanced Manufacturing Technology* 121, 11-12 (2022), 7697–7719.

[37] Xuesong Zhang and Adalberto L. Simeone. 2022. Using the Think Aloud Protocol in an Immersive Virtual Reality Evaluation of a Virtual Twin. In *Proceedings of the 2022 ACM Symposium on Spatial User Interaction* (Online, CA, USA) *(SUI '22)*. Association for Computing Machinery, New York, NY, USA, Article 13, 8 pages. https://doi.org/10.1145/3565970.3567706

[38] Xuesong Zhang and Adalberto L. Simeone. 2023. *Using Think-Aloud Protocol in Immersive VR Evaluations.* Springer International Publishing, Cham, 197–226. https://doi.org/10.1007/978-3-031-05804-2_8