



Article

Influence of Virtual Reality on User Evaluation of Prototypes in the Development Process—A Comparative Study with Control Rooms for Onshore Drilling Rigs

Niels Hinricher ^{1,*} , Simon König ¹, Chris Schröer ¹  and Claus Backhaus ²

¹ Center for Ergonomics and Medical Engineering, FH Münster University of Applied Sciences, Bürgerkamp 3, 48565 Steinfurt, Germany; simon.koenig@fh-muenster.de (S.K.); chris.schroeer@fh-muenster.de (C.S.)

² Institute of Psychology and Ergonomics, Technical University Berlin, Fasanenstraße 1, 10623 Berlin, Germany; claus.backhaus@fh-muenster.de

* Correspondence: niels.hinricher@fh-muenster.de

Abstract: User evaluations of prototypes in virtual reality (VR) offer high potential for products that require resource-intensive prototype construction, such as drilling rigs. This study examined whether the user evaluation of a VR prototype for controlling an onshore drilling rig produces results comparable to an evaluation in the real world. Using a between-subject design, 16 drilling experts tested a prototype in VR and reality. The experts performed three different work processes and evaluated their satisfaction based on task performance, user experience, and usability via standardized questionnaires. A test leader evaluated the effectiveness of the work process execution using a 3-level rating scheme. The number of user interactions and time on task were recorded. There were no significant differences in the effectiveness, number of interactions required, perceived usability, and satisfaction with respect to task performance. In VR, the drilling experts took significantly more time to complete tasks and rated the efficiency of the VR prototype significantly higher. Overall, the real-world evaluation provided more insights into prototype optimization. Nevertheless, several usability issues have been identified in VR. Therefore, user evaluations in VR are particularly suitable in the early development phases to identify usability issues, without the need to produce real prototypes.

Keywords: virtual prototype; usability test; virtual reality (VR); oil and gas industry; human–technology interaction; control room design



Citation: Hinricher, N.; König, S.; Schröer, C.; Backhaus, C. Influence of Virtual Reality on User Evaluation of Prototypes in the Development Process—A Comparative Study with Control Rooms for Onshore Drilling Rigs. *Appl. Sci.* **2023**, *13*, 8319. <https://doi.org/10.3390/app13148319>

Academic Editor: João Marcelo Teixeira

Received: 15 June 2023

Revised: 12 July 2023

Accepted: 13 July 2023

Published: 18 July 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The development of new products with complex human–machine interfaces ideally follows a user-centered design process [1]. In this process, prototypes are iteratively developed based on user requirements, as evaluated by users in usability tests. However, prototyping is time-consuming and expensive [2]. To reduce these costs, companies use virtual prototypes at the beginning of their development process [3].

Virtual reality (VR) technology offers the possibility of visualizing and experiencing virtual prototypes in detail. Users can test and evaluate the prototypes early in the development process by simulating the anticipated real-world environment in which the prototypes will eventually be utilized [4]. A literature review by Freitas et al. [5] shows that the application areas of user evaluation in VR are the automotive (37%), engineering (26%), and academic or unspecified (37%) sectors. In all three fields, VR is primarily applied to review design. In VR design reviews, the prototype is presented to the development team, experts, and users in three dimensions, thereby revealing the optimization potential and improving the efficiency of the development process [6–13].

User testing is one of the most reliable methods for evaluating the usability of a prototype. Future users operate the new product and perform typical work processes [14]. Initial studies showed that the results of user testing in VR can be transferred to reality. However,

in these studies, the prototypes and products were not tested in a pure VR environment, but in mixed reality environments [15–18]. In these tests, virtual environments were mixed with real controls, such as a steering wheel, because the lack of haptic feedback can limit the prototype evaluation [19,20].

However, purely virtual prototypes are advantageous, particularly during the early development phases. Different concepts and combinations of human–machine interfaces can be tested, independent of control elements or other real content. In addition, user tests can be performed regardless of location, allowing users anywhere in the world to be included in testing.

Virtual user tests are particularly useful for complex machines and devices with a high demand for operational safety, such as control rooms, drilling rigs, and medical devices. To minimize possible usability issues and the resulting hazards to the environment and humans, numerous prototypes need to be tested, which results in high development costs and long development times. However, studies examining complex devices or machines in VR are scarce. Aromaa et al. [21] conducted a user test in VR using a tunnel-boring machine. In this study, the influence of two different transparency levels of a machine boom on the work performance was investigated. The findings of the study enabled the identification of the preferred transparency level among the operators. Bergroth et al. [22] investigated the suitability of VR for evaluating the control rooms in nuclear power plants. The test subjects rated VR as a suitable means for evaluating the control rooms. However, neither study validated their results with an evaluation in the real world.

User tests in VR offer high potential for the development of offshore or onshore drilling rigs, which feature human–machine interfaces comprising various controls and displays. Onshore drilling rigs are used for deep drilling operations in the exploration for oil, gas, and geothermal resources on land. A driller is responsible for the drilling process. This person controls the drilling process from a driller cabin, monitors several displays, and operates the technical equipment of the drilling rig. Catastrophes such as the Deepwater Horizon explosion in the Gulf of Mexico show the importance of the user-oriented design of human–machine interfaces in this industry [23].

Using VR, the number, suitability, and layout of displays and controls can be examined first in purely virtual tests before real controls are mixed with the virtual content, or before real prototypes are produced. Prototypes in VR can be changed more easily; therefore, more prototypes can be tested compared to the traditional user-centered development process.

Overall, there is a lack of studies investigating a product during the development process to determine whether a purely virtual user test yields comparable findings to a user test with a physical prototype. The literature review by Gutemberg Junior et al. [24] on the application of VR in product development in the oil and gas industry shows that no studies exploring user testing in VR are available for this industry.

Therefore, this study is accompanied by a multiyear project to develop a new prototype for controlling onshore drilling rigs. Figure 1 shows the prototype model. The front screen displays the process-relevant data that must be monitored by the driller. Consoles, such as joysticks and rotary controls, are controls for operating the various machines on a rig. Additional controls and functions are available in a newly designed user interface that is operated with touchscreens. Typical operations performed with the prototype can be found in the Supplementary Materials (Figures S1–S3).

In this study, user tests were conducted in VR and reality using the prototype shown in Figure 1. The goal was to investigate whether there were significant differences between the tests in terms of the number and types of user errors, user experience, user acceptance, and the time required to perform the work.

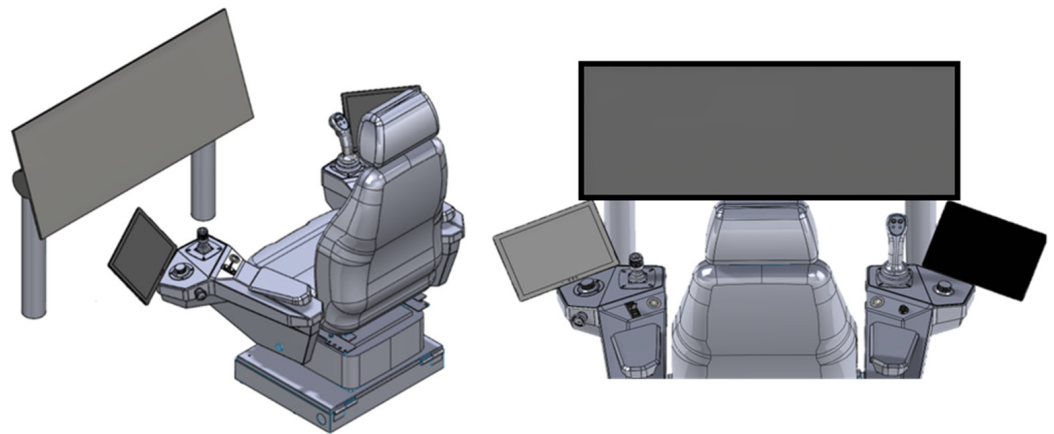


Figure 1. Model of the prototype developed for operating a drilling rig. Displays for process-relevant data are shown on the front screen. The side consoles contain touchscreens and operating elements for controlling machines and equipment of the drilling rig.

2. Materials and Methods

2.1. Participants

To investigate whether user evaluation in VR is comparable to that in real-world settings, the prototype was tested by drilling experts in a between-subjects design in reality and VR. Thus, the drilling experts tested either the real prototype or the virtual one.

The construct validity of VR simulations is typically assessed by comparing the work processes of experts and novices [25,26]. Therefore, in this study, the VR prototype was also tested by novices. The novices were students from a university environment. Table 1 presents the data of the participants. All novices were enrolled in a bachelor's or master's degree program with a technical focus at the time of the study, and they reported using computers daily. None of the participants had any prior experience with VR systems at the time of the study. Work experience refers to the operation of an onshore drilling rig.

Table 1. Subject data: gender, age, and work experience with operating drilling rigs.

Trial Group	Gender m/f [n]	Age \pm SD [a]	Work Experience [a]
VR Drilling Experts	8/0	42 \pm 5	12 \pm 10
VR Drilling Novice	10/0	26 \pm 3	0
Real Drilling Experts	8/0	40 \pm 7	9 \pm 6

2.2. Experimental Setup

2.2.1. Real Prototype

Figure 2 shows the experimental setup for the user test in real-world settings. The prototype shown in Figure 1 was manufactured in physical form. User interface mockups in the form of click dummies were displayed on the touchscreens on the side consoles of the prototype. The click dummies were created using Adobe XD software (version 36.0, Adobe XD, Adobe Inc., San Jose, CA, USA), and contained 321 interfaces.

On the front screen, the participants were presented with indicators of process-relevant data. The indicators, predominantly gauges and bar graphs, were also created using Adobe XD and could be modified in the background depending on the participant's interactions with Wizard of Oz [27]. The Wizard of Oz technique is an experimental technique used for simulating systems that are impossible or expensive to implement [28,29].

The participants were filmed using two cameras (GoPro Hero 5; GoPro Inc., San Mateo, CA, USA) during the tests. One camera was located above the front screen, and the participants were filmed from the front. The second camera was positioned behind the participants, and it filmed their interactions with the controls and touchscreens. The

cameras were connected via Bluetooth to a tablet (Samsung Note 10; Samsung Electronics Co., Suwon-si, Republic of Korea). The experiments were conducted in an empty office room.

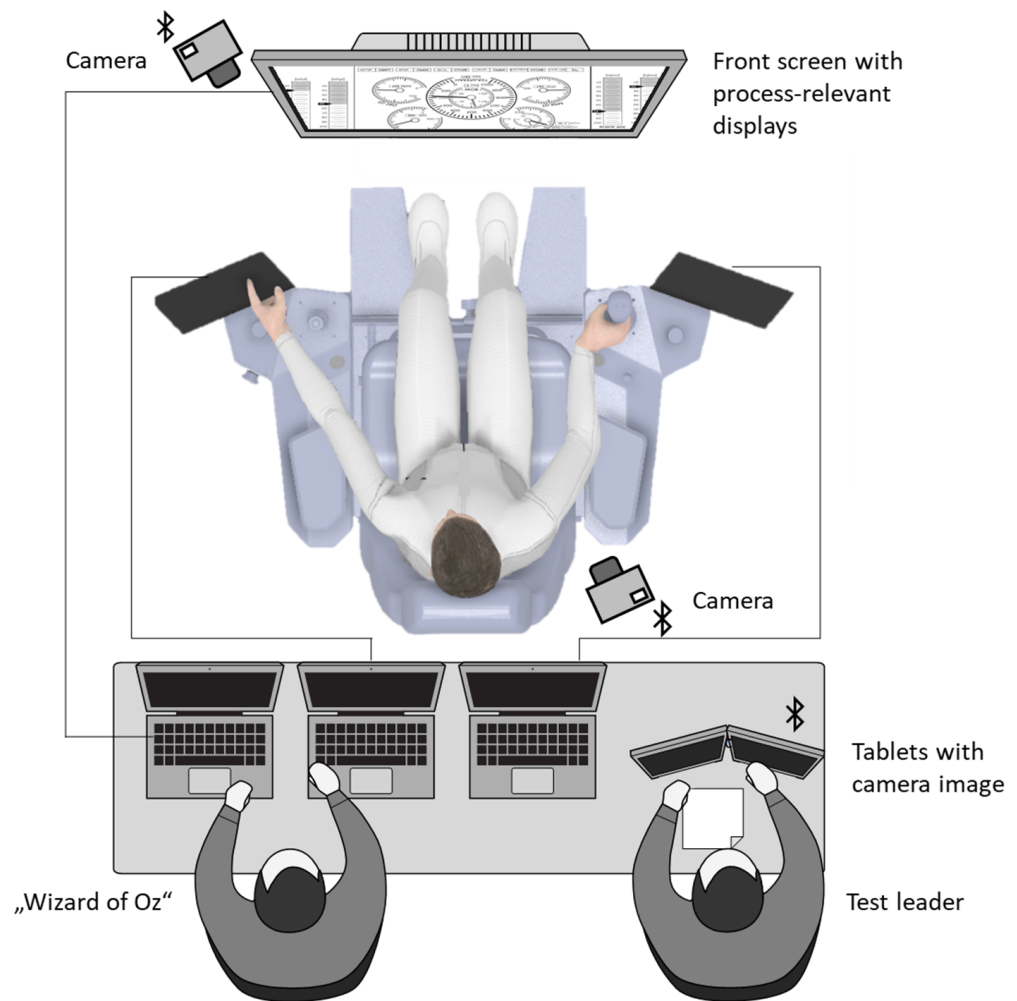


Figure 2. Experimental setup for user test in real-world setting: Participants performed tasks with the prototype. The “Wizard of Oz” changed the indicators on the front screen depending on the participant’s interactions. The test leader evaluated the execution of the tasks, using a 3-level scale (Table 2).

2.2.2. VR Prototype

The virtual prototype was created in the Unity 2020.2.3f1 development environment (Unity Technologies, San Francisco, CA, USA) and programmed using the C# language (Microsoft Corporation, Redmond, WA, USA). Visualization was performed using a Valve Index head-mounted display (HMD) (Valve Corporation, Bellevue, WA, USA), PC with an i7 processor, and a GeForce GTX 1070 graphics card (NVIDIA Inc., Santa Clara, CA, USA).

Figure 3 shows the experimental setup for the user test in VR. The participants sat in an industrial seat that is also used in onshore drilling rigs. An HMD was used to display the new prototype to the participants. The participants were able to freely interact with the prototype using controllers (Valve Index). The controllers used integrated sensors to detect the hand and finger positions, making it possible for the controls to be understood as in reality. In addition, natural interactions with the touchscreen were possible when the participants spread their index fingers away from the controller. This was detected by sensors, and the splayed finger was visualized using VR. Grasping in VR was performed by spreading all fingers away from the controller, guiding the controller to the joystick, and then grasping the controller again.

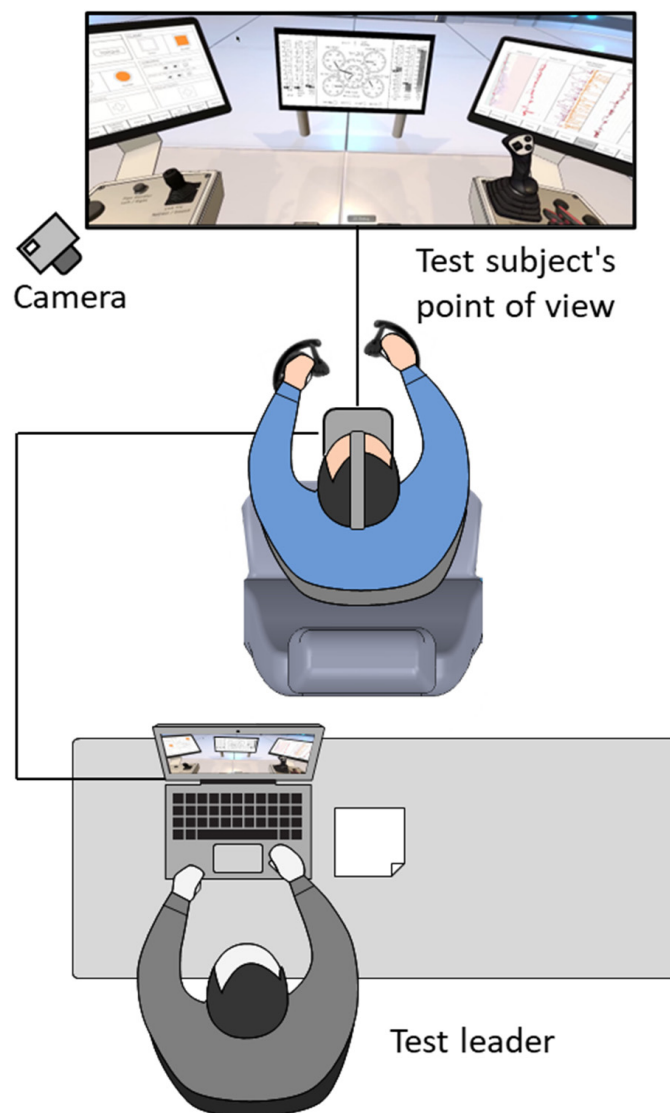


Figure 3. Experimental setup for the user test in VR: participants performed tasks using the VR prototype. The test leader evaluated the execution of the tasks, using a 3-level scale (Table 2).

The designs on the touchscreen and indicators on the front screen corresponded to those of the real test. A camera (GoPro Hero 5; GoPro Inc., San Mateo, CA, USA) documented the statements and interactions of the participants during the tests.

2.3. Experimental Procedure

2.3.1. Simulated Work Processes

The central work processes for geological drilling using the rotary drilling method are “drilling”, “connection making”, and “tripping” [30]. During drilling, a rotating drill bit mechanically crushes the rock to be drilled through. Subsequently, the crushed rock is conveyed to the surface using a drilling fluid pumped through a drill string. Torque is applied by a top drive, which is connected to the drill string, and the drive may be moved vertically in a mast. The drill string suspended in the mast is typically composed of several drill pipes with an approximate length of 9 m. After every 9 m (or more) of drilling, a new drill pipe is screwed onto the drill string. This process is called “connection making”. To continue the drilling process, the top drive must be reconnected to the drill string. This process is called “top drive connection”. When the drill bit needs to be changed (for wear reasons), the drill string is pulled out of the ground step-by-step to sequentially unscrew the individual drill pipes. This process is called “tripping” [31]. During the drilling process,

the driller must observe the indicators and adjust the target parameters as necessary. In the “top drive connection” and “tripping” processes, the driller must simultaneously observe indicators and operate controls located on the side consoles. Since this study focuses on interactions between the driller and prototype, the processes “top drive connection” and “tripping” were simulated.

2.3.2. Real Prototype

At the beginning of the tests, the participants were provided a standardized introduction to the test procedure and prototype. The controls on the side consoles and indicators on the front screen were explained to the participants. Subsequently, the participants performed 20 different introductory tasks, such as “logging in”, “opening inside blowout preventer” (IBOP), and “displaying camera image of the mast”. These tasks were process-independent, and were used to test menu structures. All tasks and instructions for the participants can be found in the Supplementary Materials.

The tripping process was simulated following the introductory tasks. The participants were asked to pull the drill pipe, set it down, and move the top drive back onto the drill string. The use cases consisted of 17 tasks (Figure 4). After the tripping process, the “top drive connection” process was simulated. The use cases consisted of eight tasks (Figure 5). Finally, participants were asked to repeat the tripping process. The learnability of the prototype was investigated by comparing two tripping simulations.

After each process execution, the participants completed an after-scenario questionnaire (ASQ) (see Section 2.4.2). At the end of the experiment, the participants filled out the user experience questionnaire (UEQ) (Section 2.4.3) and system usability scale (Section 2.4.4). Finally, semi-structured interviews were conducted with the experts (Section 2.4.6).

2.3.3. VR Prototype

The drilling experts in the VR environment were instructed according to the same standardized procedure as the real-world drilling prototype. Subsequently, the test participants were shown how to use the controllers in VR to grip the control elements and operate a touchscreen. The drilling experts operated each control element once. The novices were shown a video of the basics of the rotary-drilling procedure and work processes to be carried out in the study, prior to receiving standardized instruction on the prototype and VR controllers. The subsequent procedure was identical to the user test in the real world. The test participants completed the questionnaires after removing the VR HMD.

2.4. Measures

2.4.1. Task Success Rate

A drilling expert and usability expert evaluated the performance of the tasks according to a 3-level rating scheme (Table 2). After the participants completed the test, their ratings were compared. In case of differences, the video material was reviewed and a rating was agreed upon.

Table 2. Criteria for assessing task success rate.

Evaluation	Description
Good	Fast operation without assistance Error-free execution
Medium	Prolonged hesitation before operation Errors are corrected without indications by the test leader
Poor	Execution of the task after assistance of the test leader

For analysis, the ratings were presented as stacked bar graphs (Figures 4 and 5). Each task was rated individually. The bars indicate the relative frequencies of the evaluation

levels (green = good, yellow = medium, red = poor). Subsequently, success rates were calculated using the following formula (Nielsen [32]):

$$\text{success rate} = \frac{\sum \text{good} + \sum \text{medium} \cdot 0.5}{\text{participants} \cdot \text{tasks}} \times 100$$

The calculated success rates were averaged and individually evaluated for each task and usage scenario.

2.4.2. Satisfaction with the Task Performance

After each usage scenario, the participants completed the ASQ. Participants rated the following questions with a seven-point Likert scale ranging from strongly agree to strongly disagree. The scale assigns values ranging from one (strongly agree) to seven (strongly disagree). For the evaluation, the arithmetic mean of all of the ratings per prototype was determined.

- “Overall, I am satisfied with the ease of completing the tasks in this scenario”.
- “Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario”.

2.4.3. User Experience

User experience was measured using the UEQ [33]. The UEQ consists of 26 bipolar items divided into the following six dimensions:

- Attractiveness: Describes the overall impression of the product.
- Perspicuity: Describes a user’s feeling that the interaction with a product is easy, predictable, and controllable.
- Efficiency: Describes how quickly and efficiently the user can use the product.
- Dependability: Describes the feeling of being in control of the system.
- Stimulation: Describes the user’s interest and enthusiasm for the product.
- Novelty: Describes whether product design is perceived as innovative or creative.

Participants rated the items using a seven-point Likert scale [34]. Each box on the Likert scale was assigned a value between −3 and +3. +3 corresponds to an adjective with positive connotation. The scores were averaged per dimension and reported as the UEQ score.

2.4.4. User Acceptance

User acceptance was measured using the system usability scale (SUS) [35]. The SUS is an effective and simple method for evaluating the user acceptance of a system, and consists of 10 alternating positive and negative statements. Each statement was given a point between one and five. Depending on the phrasing of the item (positive/negative), a 5-point score corresponds to either the statement “strongly agree” or “strongly disagree”. The results were expressed as a score between 0 (negative) and 100 (positive). This 100-point scale facilitated the comparison of different products [36].

2.4.5. Number of User Interactions and Time for Use Scenario

To check whether the participants interacted with the prototype comparably in VR and reality, the number of user interactions and completion times were recorded for each use scenario. Grasping the joystick, turning the control knob, and pressing a button on the touchscreen were considered as interactions. The processing time was defined as the time between reading the use scenario aloud and completing it.

2.4.6. Semi-Structured Interview on Satisfaction with the Prototype

Following the usability tests, semi-structured interviews were conducted with experts. The interview questions are presented in Table 3. The participants who tested the virtual prototype remained in the VR environment during the interviews. The interviews were

evaluated using content analysis, according to Mayring [37]. Hence, the interviews were transcribed, and categories were inductively formed. These categories were then quantified. Potential improvements to the prototype were identified via two workshops with usability experts ($n = 2$) and engineers developing drilling rigs ($n = 3$). Separate workshops were conducted for the VR and real prototypes.

Table 3. Interview questions.

Category	Question
General operation	How do you evaluate the operation?
	Which functions are you missing?
	Are there functions that would not be directly accessible during safety-critical moments?
	Would you prefer this concept to the previous operating concept?
Touchscreens	How do you like the touchscreens?
	Could you read everything on the touchscreens?
	Do the touch fields have an appropriate size?
Control elements	How do you like the control elements on the side consoles?
	Are there any physical control elements that you would prefer to have implemented as touch functions?
	Are there touch functions that you would prefer to have implemented as physical control elements?
Iron roughneck	How do you rate the control of the iron roughneck via the touchscreens?
Indicators	How do you rate the indicator screen?
	Could you read indicators correctly?

2.5. Statistical Analysis

Statistical analyses were performed using SPSS Statistics software (version 27, IBM, Armonk, NY, USA). A *t*-test for independent samples was used to examine whether the experts achieved significantly ($\alpha = 0.05$) different results in the user test when using the VR prototype compared to the real prototype. Evaluation parameter success rates, ASQ scores, UEQ scores, SUS scores, number of interactions, and task completion time presented in Section 2.4 were compared.

To examine construct validity, the *t*-test for independent samples was used to determine whether the evaluation parameters differed significantly between the VR drilling expert and novice groups ($\alpha = 0.05$). However, the differences between the novices in VR and experts in the real world were not examined.

All of the participant groups performed the tripping process twice. The *t*-test for independent samples was used to examine whether the parameter success rates, ASQ scores, numbers of interactions, and task completion times significantly improved when performed for the second time ($\alpha = 0.05$).

3. Results

3.1. Task Success Rate

With the real prototype, the drilling experts achieved a mean success rate of $88 \pm 12\%$ when performing the introductory tasks. With the VR prototype, the experts achieved a success rate of $87 \pm 14\%$, while the novices achieved a success rate of $87 \pm 17\%$. No significant differences were observed ($p > 0.05$). The success rates for each task are provided in the Supplementary Materials.

Figure 4 presents the success rates of the tripping process. The drilling experts achieved a mean success rate of $91 \pm 13\%$ with the real prototype. For the VR prototype, the experts scored a mean success rate of $91 \pm 12\%$. The novices achieved a mean success rate of $76 \pm 18\%$ with the VR prototype. The difference between drilling experts and novices in the mean success rate for the VR prototype was significant ($p = 0.006$). The definitions of the individual tasks can be found in the Glossary.

Moreover, Figure 4 shows that experts had issues with the same tasks in both the VR and real prototype environments. For example, experts were sometimes unable to extend the iron roughneck (IR) and adjust its height in both VR and reality (Tasks 7 and 8). An iron roughneck is a machine used for screwing and unscrewing a drill string. Several novices had issues in Tasks 9–15. The experts had no difficulty in performing these tasks.

Tasks		Real Prototype Experts	VR Prototype Experts	VR Prototype Novices
1	Pull drill string out of PWR slips	100%	64%	65%
2	Open PWR slips	94%	100%	80%
3	Pull up drill string	88%	100%	100%
4	Close PWR slips	69%	79%	85%
5	Place drill string into PWR slips	94%	86%	55%
6	Call up IR menu	100%	86%	50%
7	Extend IR	63%	71%	55%
8	Set IR height	63%	71%	50%
9	Close button clamp	100%	100%	95%
10	Unscrew drill string	100%	100%	100%
11	Extend Link Tilt	100%	100%	90%
12	Lower Elevator	100%	100%	70%
13	Open Elevator	100%	100%	75%
14	Activate active float	100%	100%	95%
15	Lower top drive to rig floor	100%	100%	85%
16	Extend and retract Link Tilt	86%	93%	55%
17	Close Elevator	100%	100%	95%
Total:		91%	91%	76%

Figure 4. Comparison of the success rates for the simulation of the tripping process. The bars indicate the relative frequencies of the evaluation levels (green = good, yellow = medium, red = poor).

Figure 5 lists the success rates of the simulation of the top drive connection process. The experts achieved a mean success rate of $94 \pm 5\%$ with the real prototype. For VR, the experts achieved a success rate of $95 \pm 9\%$. The novices achieved a mean success rate of $84 \pm 19\%$. There were no significant differences ($p > 0.05$).

Tasks		Real Prototype Experts	VR Prototype Experts	VR Prototype Novices
1	Activate TD mode "Spin"	94%	100%	65%
2	Lower TD	100%	100%	50%
3	Close Clamp	94%	79%	95%
4	Activate TD mode "Torque"	94%	100%	90%
5	Open IBOP	88%	100%	95%
6	Activate TD mode "Drill"	100%	86%	95%
7	Open Clamp	88%	100%	100%
Total:		94%	95%	84%

Figure 5. Comparison of success rates for the simulation of the top drive (TD) connection process. The bars indicate the relative frequencies of the evaluation levels (green = good, yellow = medium, red = poor).

The success rates for the second tripping attempt are shown in the Supplementary Materials (Figure S6). The experts achieved higher success rates in the second tripping attempt, and consequently made fewer errors ($M_{\text{Real}} = 93 \pm 8\%$; $M_{\text{VR}} = 95 \pm 9\%$) in both VR and reality. This difference was not statistically significant ($p > 0.05$). The novices had significantly higher success rates in the second tripping process ($p = 0.048$; $M_{\text{VR-Nov}} = 88 \pm 14\%$).

3.2. Satisfaction with the Task Performance

Figure 6 shows the mean ASQ scores for individual test scenarios. In each scenario, the experts who evaluated the real prototype were most satisfied with the task performance. Novices were least satisfied with the task performance, except in the top drive connection test scenario. The differences between descriptions provided by the experts for both the VR and real prototypes were insignificant; moreover, the descriptive differences between the novices and experts for VR were insignificant ($p > 0.05$).

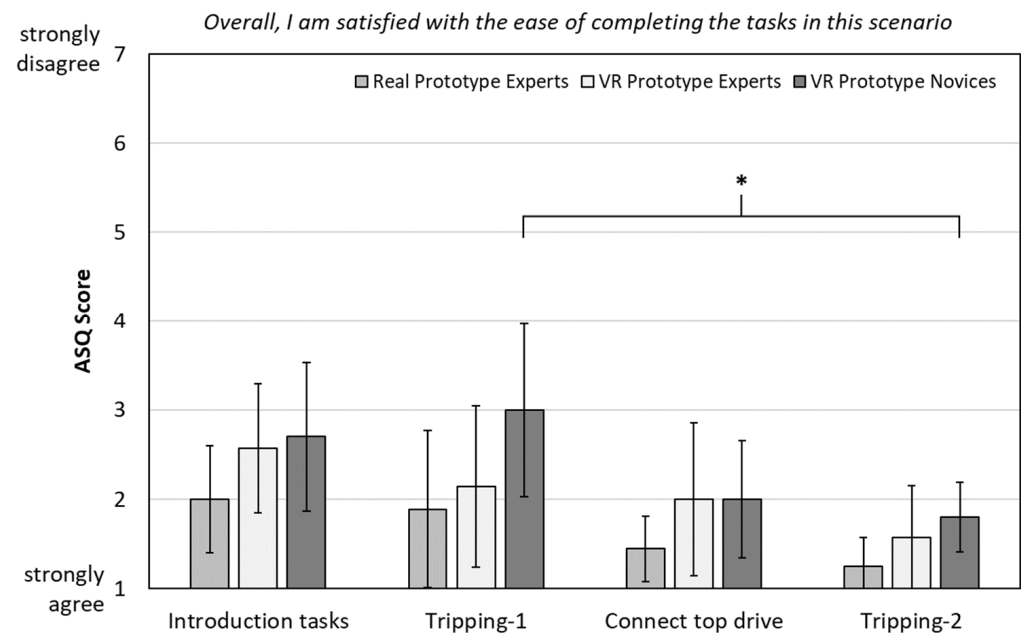


Figure 6. Mean ASQ scores and 95% confidence intervals for the statement “Overall, I am satisfied with the ease of completing the tasks in this scenario”. Significant differences ($p < 0.05$) are marked with an asterisk (*).

All of the subject groups were more satisfied with the task performance after the second simulation of the tripping process (Tripping-2) than after the first simulation. The novices were significantly more satisfied with the second execution ($p = 0.04$; $M = 3.0 \pm 1.0$; $M = 1.8 \pm 0.4$).

The drilling experts who tested the real prototype were also most satisfied with the time required to complete the tasks. The figure for this question can be found in the Supplementary Materials (Figure S4). Except for the test scenario “tripping-1”, the experts who tested the VR prototype were the most dissatisfied with the time needed to complete the tasks. Significant differences between the groups could not be identified ($p > 0.05$).

3.3. User Experience

Figure 7 shows the dimensions of the UEQ and UEQ scores achieved. The experts rated the VR prototype best; except for the *efficiency* dimension, the novices rated the prototype worst.

The experts who tested the real prototype rated the *efficiency* dimension to be significantly worse than those who tested the VR prototype ($p = 0.05$; $M_{\text{Real}} = 1.06 \pm 0.58$, $M_{\text{VR}} = 1.96 \pm 0.58$). The UEQ scores corresponded to good and very good user experiences in all dimensions.

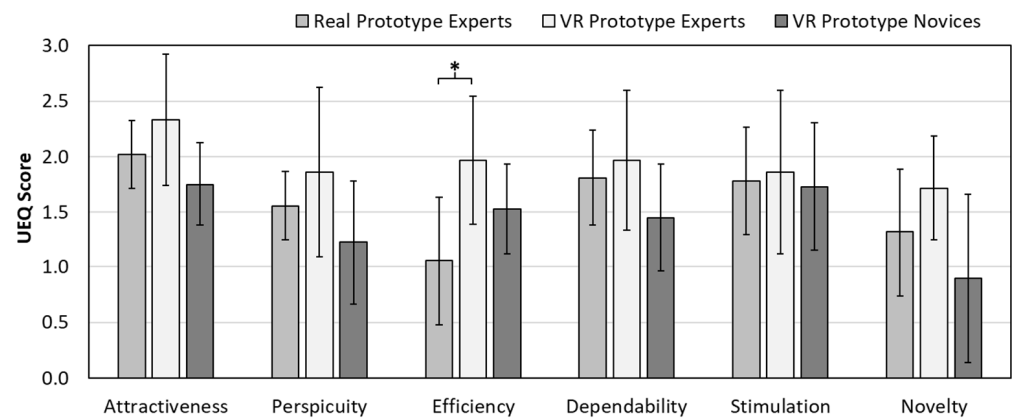


Figure 7. Mean scores and 95% confidence intervals of the individual UEQ dimensions achieved by the different prototypes. Significant differences ($p < 0.05$) are marked with an asterisk (*).

3.4. User Acceptance

The VR drilling experts group rated the prototype best, with a SUS score of 83.6 ± 15.8 . The real prototype achieved an SUS score of 81.1 ± 8.9 . The novices rated the VR prototype with a score of 77.3 ± 13.9 . Furthermore, the scores are in the good to very good range according to Bangor et al. [36]. Significant differences were not identified ($p > 0.05$).

3.5. Number of User Interactions and Time for Use Scenario

Figure 8 shows the time required by the participants to complete the individual test scenarios. For the “introduction tasks” scenario, the VR drilling experts group needed $548 \text{ s} \pm 185 \text{ s}$, which is approximately the same time as the group of real drilling experts with $543 \text{ s} \pm 191 \text{ s}$. The novices (t_{Nov-VR}) needed significantly more time than the experts (t_{Ex-VR}) for the “tripping-1” test scenario ($p < 0.001$; $t_{Nov-VR} = 622 \text{ s} \pm 140 \text{ s}$; $t_{Ex-VR} = 361 \text{ s} \pm 65 \text{ s}$). Using the real prototype, the experts were able to perform this scenario most quickly ($t_{Ex-Real} = 305 \text{ s} \pm 140 \text{ s}$). The “top drive connection” scenario was also performed fastest with the real prototype ($t_{Ex-Real} = 118 \text{ s} \pm 40 \text{ s}$). For the virtual prototype, the experts were significantly slower ($p = 0.02$; $t_{Ex-VR} = 182 \text{ s} \pm 47 \text{ s}$).

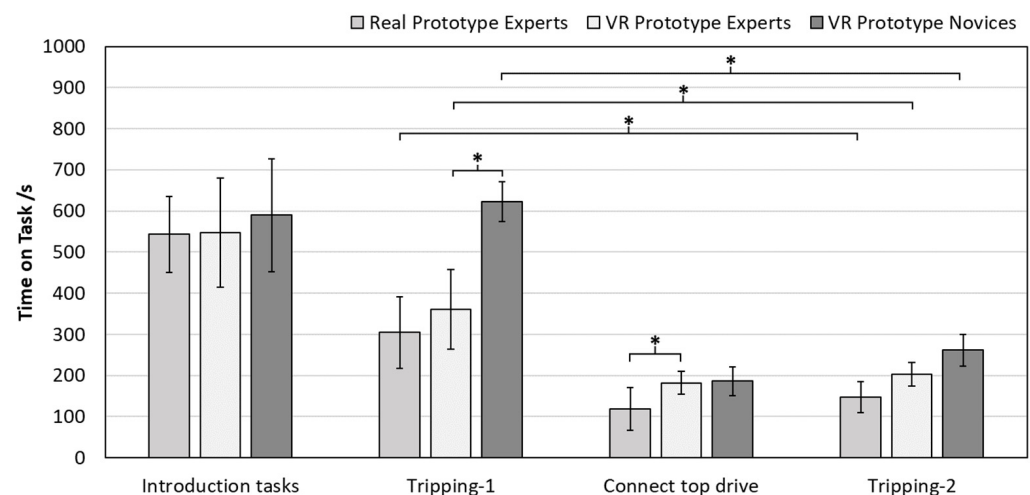


Figure 8. Mean time and 95% confidence intervals for completing the different test scenarios. Significant differences ($p < 0.05$) are marked with an asterisk (*).

A comparison of the two tripping simulations shows that all of the groups improved significantly ($p_{Real-Experts} = 0.01$; $p_{VR-Experts} < 0.001$; $p_{VR-Novices} < 0.001$). In the second test, the experts required only $147 \text{ s} \pm 42 \text{ s}$ with the real prototype. The experts who tested the VR prototype were able to perform the second tripping simulation in $203 \text{ s} \pm 53 \text{ s}$. Novices

showed the greatest improvement. The participants required $262 \text{ s} \pm 61 \text{ s}$ for tripping in the second trial using the VR prototype.

Figure 9 shows the mean number of interactions required by the participants to complete the different test scenarios. In all four scenarios, the experts with the real prototype required the fewest interactions to complete a task. The novices required an average of 55 ± 12 interactions, while the drilling experts completed the test scenario “Tripping-1” with an average of 42 ± 9 interactions using the VR prototype. The observed difference is statistically significant ($p = 0.035$).

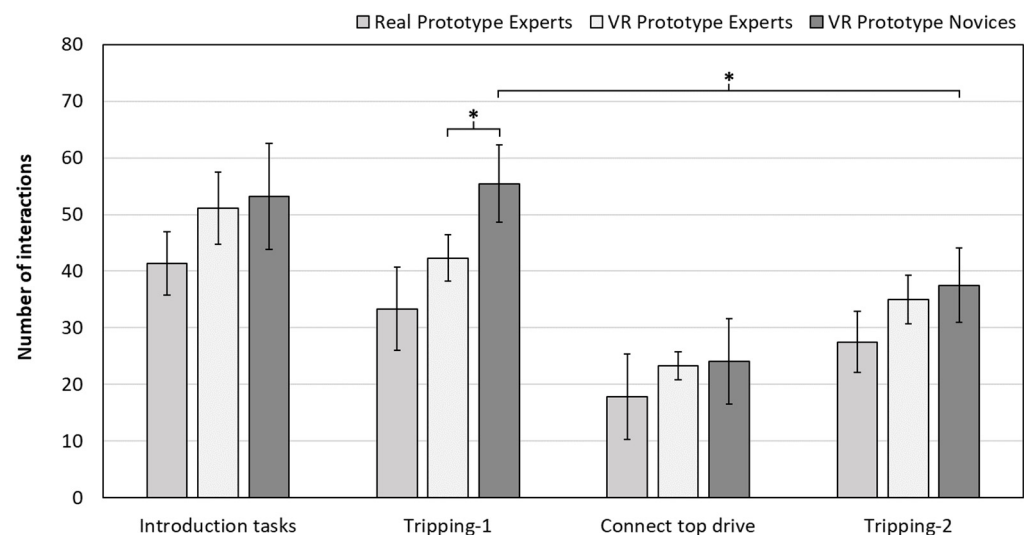


Figure 9. Mean number of interactions and 95% confidence intervals for completing each test scenario. Significant differences ($p < 0.05$) are marked with an asterisk (*).

Similar to the results for task completion times, all of the subject groups showed improvements in the second tripping process. For the second tripping trial, the novices needed significantly fewer interactions, specifically 38 ± 9 , compared to their initial performance ($p = 0.002$).

3.6. Semi-Structured Interview on Satisfaction with the Prototype

Both the VR drilling experts group and the real drilling experts group rated the operation positively overall (Question 1). Seven of the eight VR drilling experts preferred the new concept to the previous operating concept. The real prototype was preferred by six out of eight experts. One expert preferred the previous operating concept. One expert each in VR and in the real world preferred neither the new nor the old operating concept.

All the participants stated that the control elements had an appropriate size. In response to the question “Which functions are you missing?”, six participants who tested the VR prototype answered “none”. One participant wanted to see the blowout preventer controls. The participants who tested the real prototype missed a switch to turn the hoist on and off, an emergency stop button, and an indicator of the equipment status. Half of the participants in this group stated that no functions were missing or that they would only be able to answer the question once they had tested the system for a longer period.

Seven of the eight participants testing the real prototype complained that the IBOP was not available quickly enough in safety-critical moments. In the virtual test, this was criticized by only two participants.

Regardless of the prototype tested, all participants rated the touchscreen positively. The text and images could be read easily by all the participants in both the real and VR prototypes, and the size of the touchscreen was rated as sufficiently large.

There were differences in the response behaviors between the groups, particularly in questions regarding the control elements. The question “how do you like the control

elements on the side consoles?” was answered positively by all participants for the VR prototype. Participants liked the clarity of the consoles. Specific control elements and their positions were criticized in isolated cases. The participants who tested the real prototype were also satisfied with the design and clarity of the side consoles. However, individual control elements were discussed and criticized more intensively.

Control of the iron roughneck was rated good by all participants. In both the VR and physical prototypes, individual participants doubted the feasibility of the new control concept for the iron roughneck.

Regardless of the prototype tested, the front screen and indicators displayed on it were rated positive. In both VR and reality, the most frequent responses referred to the drillometer (a central indicator for monitoring a hook load). In both groups, 50% of the participants preferred a larger drillometer with more sensitive and larger pointers.

From the interviews with the participants who tested the real prototype, 13 optimization potentials were identified in expert workshops, such as an optimized layout of controls, indicators, and the design of menu structures. Eight optimization potentials were identified from interviews with the VR participants.

4. Discussion

4.1. Construct Validity

In this study, user tests were conducted using a new prototype for the control of onshore drilling rigs in VR and reality. We investigated whether there were significant differences between the tests in terms of the number and type of user errors, user experience, user acceptance, and the time required to complete the work processes.

The construct validity of the VR simulation was investigated by testing the VR prototype with novices. The novices had significantly more difficulty with the prototype than the experts (Figure 4). Considering the significantly higher number of interactions (Figure 9) and the longer processing time (Figure 8), we considered construct validity to be given. If there were no significant differences in the success rates and task times between experts and novices, this would indicate that the simulation was not close to reality. An oversimplified simulation of work processes, for instance, would result in the inability to transfer results, such as the subjective evaluation of the prototype, to real-world situations.

In addition to testing construct validity, the evaluation of the prototype by novices provided insights into the learnability of the new operating concept. Novices had significantly higher success rates in the second tripping simulation, and required less time and interaction. This improvement partly resulted from the training effects of the VR. Since novices had significantly higher success rates and fewer procedural errors, we assume that the prototype is easy to learn. In a follow-up study, we plan to investigate whether the learnability of a real prototype is comparable to that of a VR prototype.

4.2. Task Success Rate

Figures 4 and 5 indicate that the experts had comparable issues in VR and in the real world. In both VR and reality, the participants had difficulty in controlling the iron roughneck (Tasks 7 and 8) and closing power slips (Task 4).

There were clear differences in Task 1, “pull drill string out of power slips”. In this task, the participants pulled the joystick slightly, and the hook load increased. The hook load was displayed centrally in front of the user via a large round indicator (drillometer). In the VR environment, three participants forgot to perform the task. This task was not forgotten in the real prototype. Under certain circumstances, omitting this task can damage machines, resulting in high costs. We hypothesize that the VR environment led to a decrease in situational awareness. Situational awareness refers to the state of awareness of one’s environment, its objects, and possible changes [38,39].

Drillers must maintain a high level of situational awareness to ensure rig safety [40,41]. However, previous studies have not found significant differences in situational awareness between VR and reality [42,43]. Training simulators were tested in these studies. Studies

investigating the situational awareness in user tests for new prototypes have not yet been conducted. In future VR user tests, situational awareness should be considered and further investigated using a situation awareness rating technique (SART) [44].

4.3. Subjective Evaluation of the Prototype: ASQ, UEQ, and SUS

Although the experts with the VR prototype were less satisfied with task performance and the required time (ASQ) than the experts with the real prototype, both the user experience (UEQ) and user acceptance (SUS) were rated better. However, significant differences were only identified for the *efficiency* dimension of the UEQ. The *efficiency* dimension was rated significantly to be better by the VR drilling experts group. Since the VR prototype was operated with controllers and lacked important feedback components such as force feedback, the operation of the VR prototype was expected to be rated worse.

One reason for the overall higher subjective rating of the prototype could be the novelty effect [45,46]. In this effect, participants evaluated new products better, owing to the novelty of technology, and showed increased motivation. None of the participants had experienced VR before the test. Further studies should investigate whether the use of VR leads to bias in the results owing to the novelty effect. For example, a product can be evaluated by user groups with no VR experience, with little VR experience (<10 h), and with a lot of experience (weekly use). If the experience with VR systems has a significant effect on the subjective evaluation of a product, it is an indication of the novelty effect. Consequently, only participants with extensive VR experience should be recruited for future VR user tests. This could lead to problems in recruiting participants, and highlights the need for further studies to investigate whether prior VR training can attenuate the novelty effect.

Only a few studies have investigated user experience and acceptance of products in VR and compared them with real products. In a study by Zhou and Rau [2], users who were presented with a product on an HMD rated it significantly better than those who were shown the product on a monitor. Franzreb et al. [47] studied the user experience of three furnishing products in VR and real-world settings. For two of the three products, there were no significant differences in the user experience ratings. One product was rated as significantly better in VR. Overall, the existing literature and our study suggest that the user experience of products and prototypes tends to be rated better in VR [48]. This should be considered particularly for products where the user experience is crucial for product success, such as consumer products.

Sonderegger and Sauer [49] revealed that participants rated the usability of visually attractive products to be better. To prevent this, the VR and real prototypes used in our study were based on the same computer-aided design files. In addition, we ensured that the color schemes of the two prototypes were identical.

User acceptance of the new prototype was determined using the SUS. The SUS is used to measure the usability and acceptance of a product. Studies recommend a sample size of at least 30–35 participants to obtain reliable results [50,51]; however, to identify the differences between different products, 8–12 participants were sufficient [52]. Owing to the small sample size and high standard deviation of the questionnaires in terms of usability, significant differences were not expected between the SUS scores. Nevertheless, the SUS scores of the novices and experts were relatively close ($SUS_{Ex} = 83.6 \pm 15.8$; $SUS_{Nov} = 77.3 \pm 13.9$). Simoes et al. [53] revealed that novices have difficulties in estimating aspects such as the size, weight, and material of virtual prototypes. The novices operated the drilling rig for the first time; therefore, they could not compare the prototype with existing products. It is unclear whether the prototype convinced the novices, or if the novelty effect influenced the results.

4.4. Number of User Interactions and Time on Usage Scenario

With the real prototype, the participants required the least amount of time to complete the tasks. Novices required significantly more time to complete the tripping process. This

difference can be explained by the limited experience of the novices in controlling onshore drilling rigs. We attribute the differences between the VR drilling experts and real drilling experts groups to altered feedback modalities in the VR environment. To perform the tripping process, the participants had to operate joysticks and rotary controls. An important type of feedback when operating joysticks is the restoring force; the user experiences resistance when moving the joystick, and can better control the deflection [54,55].

Detent torque is an important feedback for the operation of rotary controls [56]. A user must generate a defined torque to move from one switching position to the next. These feedback components are missing in VR, which makes the operation of the control elements more difficult.

The fact that the prototype was more difficult to operate in VR than in real-world settings is also indicated by the number of interactions required. The drilling experts required fewer interactions with the real prototype than for VR for all test scenarios. In VR, experts had to correct the set values more often; however, these differences were not statistically significant.

4.5. Semi-Structured Interview about Satisfaction with the Prototype

Seven of the eight participants who tested the real prototype complained that the IBOP was not available quickly enough in safety-critical moments. In the virtual test, this was criticized by only two participants. The positions and types of control elements were also discussed more critically by the drillers who tested the real prototype. We assumed that the haptic experience of the control elements stimulated these discussions. These results are consistent with those of Vergara et al. [57], who found that a multisensory interaction (visual–haptic), compared to a purely visual interaction, affects the perceived ergonomics. Users who interacted with a product (hammer) visually noted fewer issues regarding ergonomics and interaction than those who interacted with the product haptically. Due to the small number of participants, no general conclusions could be drawn from the interview in our study. The participants maintained their VR HMD during the interviews. Further studies should investigate whether there are significant differences in the interview quality between VR and reality.

4.6. Limitations

A major limitation of this study is the small number of participants. Drillers were recruited for this study. Drillers who have several years of experience in drilling rigs are difficult to acquire because of their limited availability. Despite the small sample size, significant differences were observed between the VR and real environments. Further studies on the impact of VR on the evaluation of products should be conducted using prototypes that are easier to acquire in higher numbers. Suitable products include medical devices. Medical device manufacturers are required to demonstrate the usability of the product. Approval for the US market requires Americans to evaluate the products. European manufacturers can use VR to test initial prototypes independent of location. This could optimize the cost-intensive development of medical devices.

Another limitation is in simulating the work processes. To evaluate the physical prototype, the drilling processes were simulated using the Wizard of Oz method. This is a method frequently used when the simulation of use scenarios is only possible with significant effort [28,29]. No definite statement could be made in this study regarding the comparability with the evaluation of real-world use on a drilling rig. Evaluation of the prototype during actual use on a drilling rig can lead to different results. The objective of this study was to investigate the influence of VR on evaluation results. Therefore, the VR environment was adapted to the Wizard of Oz test environment, and the usage environment of the VR prototype was an empty room rather than a drilling rig.

A major advantage of VR is the creation of realistic virtual environments that simultaneously provide high control over experimental conditions and high ecological validity [58]. Therefore, the investigated prototype could be simulated directly in a later usage environ-

ment: an onshore drilling rig. To determine the influence of VR on the user evaluation, a real prototype must also be evaluated using a drilling rig. Elaborate drilling simulators already exist for driller training and educational purposes. However, the testing of new prototypes using these simulators is expensive. Further studies should investigate whether more complex tests are feasible in VR.

5. Conclusions

In this study, drilling experts made comparable errors in using the VR and the real prototypes. There were no significant differences in evaluating the execution of the work processes. There were no significant differences between VR and reality in terms of the number of required interactions. Thus, these partial results of the VR user test are transferable to the real world.

However, there were significant differences in the time required to complete the top drive connection process and rating the *efficiency* dimension of the UEQ. Drilling experts required more time for this process in VR, but still rated the VR prototype as more efficient. We assume that the better subjective evaluation in VR can be attributed to the novelty effect. Furthermore, there were no significant differences in the subjective evaluations of the prototype; however, the VR prototype received higher ratings. In the interviews, the drilling experts who tested the prototype in reality were more critical, and mentioned more optimization suggestions than the experts who tested the VR prototype.

The evaluation of the physical prototype provided more insights into the optimization of the prototype than the evaluation in VR. However, several usability problems could be identified in the VR environment. This study shows that, at the current stage of the development of VR systems, user evaluations in VR are suitable for identifying usability issues. An investigation of the user experience and acceptance should be conducted using real prototypes. If the VR experience of the participants increases in the future and interactions with virtual content are further improved, subjective parameters such as user experience can also be evaluated more accurately for VR environments.

Despite these limitations, VR user evaluations can optimize the development process. Based on this study and the accompanying development process, we recommend the following approach for the user-centered development of control rooms for onshore drilling rigs:

1. Development of an operating concept:
 - a. Creation of menu structures with click dummies.
 - b. Selection/development of suitable control elements and indicators.
2. Expert evaluation of the menu structures.
3. Expert evaluation of the control elements and indicators.
4. Development of a virtual prototype.
5. User evaluation of the prototype in VR with a focus on usability issues.
6. Optimization of the virtual prototype.
7. Re-evaluation of the virtual prototype in VR.
8. Optimization of the virtual prototype.
9. Manufacturing of the prototype.
10. User evaluation of a real prototype on a (drilling) simulator, focusing on user experience and usability issues.
11. Optimization of the real prototype.

Expert evaluations should be conducted with system users and usability experts. Steps 1–3 should be iterative. We recommend one to two user tests in VR. Depending on the number and severity of the identified usability problems, Steps 4 and 5 should be repeated. This approach reduces the number of real prototypes to be produced, which in turn, reduces the development costs and time. Future studies should investigate the extent to which this approach could be applied to other product types.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app13148319/s1>, Figure S1: Description of the tripping process for the participants; Figure S2: Description of the tripping process for the participants; Figure S3: Description of the top drive connection process for the participants; Figure S4: Mean ASQ scores for the statement “Overall, I am satisfied with the amount of time it took to complete the tasks in this scenario.”; Figure S5: Comparison of success rates for performing the introductory tasks; Figure S6: Comparison of the success rates in the simulation of the second tripping process.

Author Contributions: N.H., S.K., and C.S. carried out the experiment. N.H. wrote the manuscript with support from S.K., C.S., and C.B. N.H. conceived the original idea. C.B. supervised the project. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded by Bentec GmbH Drilling & Oilfield Systems.

Institutional Review Board Statement: The work presented does not include any studies on humans or animals. Due to the study design, no formal vote of an ethics committee was required. The VR experiments performed do not result in any hazards that increase the general risk to life of the persons concerned. All participants gave their informed consent for inclusion before they participated in the study.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Glossary

Term	Meaning
Active float	Operation in which the link tilt is moved to the zero position.
Drill pipe	Pipe section with threaded ends that make up the drill string.
Drill string	Combination of various drill pipes and other tools used to turn the drill bit.
Clamp	Clamp for fixing the drill string.
Elevator	Hinge mechanism that can be closed around the drill string to raise or lower it.
IBOP	Inside blowout preventer: checks valve inside the drill string to prevent backflows.
Iron roughneck (IR)	Machine for connecting and disconnecting drill pipes.
Link-tilt	Device for horizontal movement of drill pipes.
PWR-slips	Hydraulic metal wedges above the borehole for fixing the drill string.
Rigfloor	Working area on a rig in which the rig crew conducts operations. Most dangerous location on the rig.
Top drive	Hydraulic or electric motor for rotating the drill string.

References

1. ISO 9241-220:2019; Ergonomics of Human-System Interaction—Part 220: Processes for Enabling, Executing and Assessing Human-Centred Design within Organizations. Beuth Verlag GmbH: Berlin, Germany, 2019.
2. Zhou, X.; Rau, P.-L.P. Determining fidelity of mixed prototypes: Effect of media and physical interaction. *Appl. Ergon.* **2019**, *80*, 111–118. [\[CrossRef\]](#)
3. Bullinger, H.-J.; Dangelmaier, M. Virtual prototyping and testing of in-vehicle interfaces. *Ergonomics* **2003**, *46*, 41–51. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Salwasser, M.; Dittrich, F.; Melzer, A.; Müller, S. Virtuelle Technologien für das User-Centered-Design (VR for UCD). Einsatzmöglichkeiten von Virtual Reality bei der Nutzerzentrierten Entwicklung. In *Mensch und Computer 2019—Usability Professionals*; Fischer, H., Hess, S., Eds.; Gesellschaft für Informatik e.V. Und German UPA e.V.: Hamburg, Germany, 2019.
5. De Freitas, F.V.; Gomes, M.V.M.; Winkler, I. Benefits and Challenges of Virtual-Reality-Based Industrial Usability Testing and Design Reviews: A Patents Landscape and Literature Review. *Appl. Sci.* **2022**, *12*, 1755. [\[CrossRef\]](#)
6. Adwernat, S.; Wolf, M.; Gerhard, D. Optimizing the Design Review Process for Cyber-Physical Systems using Virtual Reality. *Procedia CIRP* **2020**, *91*, 710–715. [\[CrossRef\]](#)

7. Ahmed, S.; Irshad, L.; Demirel, H.O.; Tumer, I.Y. A Comparison Between Virtual Reality and Digital Human Modeling for Proactive Ergonomic Design. In *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Human Body and Motion*; Duffy, V.G., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 3–21, ISBN 978-3-030-22215-4.
8. Aromaa, S. Virtual Prototyping in Design Reviews of Industrial Systems. In Proceedings of the 21st International Annual Academic Mindtrek Conference (AcademicMindtrek'17), Tampere, Finland, 20–21 September 2017; Turunen, M., Väättäjä, H., Paavilainen, J., Olsson, T., Eds.; ACM: New York, NY, USA, 2017; pp. 110–119, ISBN 9781450354264.
9. Aromaa, S.; Väänänen, K. Suitability of virtual prototypes to support human factors/ergonomics evaluation during the design. *Appl. Ergon.* **2016**, *56*, 11–18. [[CrossRef](#)] [[PubMed](#)]
10. Berg, L.P.; Vance, J.M. An Industry Case Study: Investigating Early Design Decision Making in Virtual Reality. *J. Comput. Inf. Sci. Eng.* **2017**, *17*, 011001. [[CrossRef](#)]
11. Chen, X.; Gong, L.; Berce, A.; Johansson, B.; Despeisse, M. Implications of Virtual Reality on Environmental Sustainability in Manufacturing Industry: A Case Study. *Procedia CIRP* **2021**, *104*, 464–469. [[CrossRef](#)]
12. De Clerk, M.; Dangelmaier, M.; Schmierer, G.; Spath, D. User Centered Design of Interaction Techniques for VR-Based Automotive Design Reviews. *Front. Robot. AI* **2019**, *6*, 13. [[CrossRef](#)]
13. Wolfartsberger, J. Analyzing the potential of Virtual Reality for engineering design review. *Autom. Constr.* **2019**, *104*, 27–37. [[CrossRef](#)]
14. Backhaus, C. *Usability-Engineering in der Medizintechnik*; Springer: Berlin/Heidelberg, Germany, 2010; ISBN 978-3-642-00510-7.
15. Bolder, A.; Grünvogel, S.M.; Angelescu, E. Comparison of the usability of a car infotainment system in a mixed reality environment and in a real car. In Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology (VRST '18), Tokyo, Japan, 28 November–1 December 2018; Spencer, S.N., Morishima, S., Itoh, Y., Shiratori, T., Yue, Y., Lindeman, R., Eds.; ACM: New York, NY, USA, 2018; pp. 1–10, ISBN 9781450360869.
16. Grandi, F.; Zanni, L.; Peruzzini, M.; Pellicciari, M.; Campanella, C.E. A Transdisciplinary digital approach for tractor's human-centred design. *Int. J. Comput. Integr. Manuf.* **2020**, *33*, 377–397. [[CrossRef](#)]
17. Ma, C.; Han, T. Combining Virtual Reality (VR) Technology with Physical Models—A New Way for Human-Vehicle Interaction Simulation and Usability Evaluation. In *HCI in Mobility, Transport, and Automotive Systems*; Krömker, H., Ed.; Springer International Publishing: Cham, Switzerland, 2019; pp. 145–160, ISBN 978-3-030-22665-7.
18. Pettersson, I.; Karlsson, M.; Ghiurau, F.T. Virtually the Same Experience? In Proceedings of the 2019 On Designing Interactive Systems Conference (DIS '19), San Diego, CA, USA, 23–28 June 2019; Harrison, S., Bardzell, S., Neustaedter, C., Tatar, D., Eds.; ACM: New York, NY, USA, 2019; pp. 463–473, ISBN 9781450358507.
19. Bruno, F.; Cosco, F.; Angilica, A.; Muzzupappa, M. Mixed Prototyping for Products Usability Evaluation. In *Volume 3: 30th Computers and Information in Engineering Conference, Parts A and B, Proceedings of the ASME 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (ASMEDETC)*, Montreal, QC, Canada, 15–18 August 2010; ASME: New York, NY, USA, 15–18 August 2011; pp. 1381–1390, ISBN 978-0-7918-4411-3.
20. Stamer, M.; Michaels, J.; Tümler, J. Investigating the Benefits of Haptic Feedback During In-Car Interactions in Virtual Reality. In *HCI in Mobility, Transport, and Automotive Systems. Automated Driving and In-Vehicle Experience Design*; Krömker, H., Ed.; Springer International Publishing: Cham, Switzerland, 2020; pp. 404–416, ISBN 978-3-030-50522-6.
21. Aromaa, S.; Goriachev, V.; Kymäläinen, T. Virtual prototyping in the design of see-through features in mobile machinery. *Virtual Real.* **2020**, *24*, 23–37. [[CrossRef](#)]
22. Bergroth, J.D.; Koskinen, H.M.K.; Laarni, J.O. Use of Immersive 3-D Virtual Reality Environments in Control Room Validations. *Nucl. Technol.* **2018**, *202*, 278–289. [[CrossRef](#)]
23. Roberts, R.C.; Flin, R.; Cleland, J.; Urquhart, J. Drillers' Cognitive Skills Monitoring Task. *Ergon. Des.* **2019**, *27*, 13–20. [[CrossRef](#)]
24. Gutemberg Junior, L.S.D.; Ferreira, C.V.; Winkler, I. Virtual Reality Applied to Product Development in the Oil and Gas Industry: A Brief Review. *JBTH* **2023**, *5*, 329–334. [[CrossRef](#)]
25. Bartlett, J.D.; Lawrence, J.E.; Stewart, M.E.; Nakano, N.; Khanduja, V. Does virtual reality simulation have a role in training trauma and orthopaedic surgeons? *Bone Jt. J.* **2018**, *100-B*, 559–565. [[CrossRef](#)] [[PubMed](#)]
26. Van Herzele, I.; Aggarwal, R.; Malik, I.; Gaines, P.; Hamady, M.; Darzi, A.; Cheshire, N.; Vermassen, F. Validation of video-based skill assessment in carotid artery stenting. *Eur. J. Vasc. Endovasc. Surg.* **2009**, *38*, 1–9. [[CrossRef](#)]
27. Dow, S.; Lee, J.; Oezbek, C.; MacIntyre, B.; Bolter, J.D.; Gandy, M. Wizard of Oz interfaces for mixed reality applications. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems, Proceedings of the CHI05: CHI 2005 Conference on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005*; van der Veer, G., Gale, C., Eds.; ACM: New York, NY, USA, 2005; pp. 1339–1342, ISBN 1595930027.
28. Kiss, M.; Schmidt, G.; Babbel, E. Das Wizard of Oz Fahrzeug: Rapid prototyping und usability Testing von zukünftigen Fahrerassistenzsystemen. In Proceedings of the 3 Tagung Aktive Sicherheit durch Fahrerassistenz, München, Germany, 7–8 April 2008.
29. Müller, A.I.; Weinbeer, V.; Bengler, K. Using the wizard of Oz paradigm to prototype automated vehicles. In Proceedings of the 11th International Conference on Automotive User Interfaces and Interactive Vehicular Applications: Adjunct Proceedings (AutomotiveUI '19), Utrecht, The Netherlands, 21–25 September 2019; Janssen, C.P., Donker, S.F., Chuang, L.L., Ju, W., Eds.; ACM: New York, NY, USA, 2019; pp. 181–186, ISBN 9781450369206.
30. Buja, H.-O. *Handbuch der Tief-, Flach-, Geothermie- und Horizontalbohrtechnik: Bohrtechnik in Grundlagen und Anwendung; mit 119 Tabellen*, 1st ed.; Vieweg + Teubner: Wiesbaden, Germany, 2011; ISBN 9783834812780.

31. Reich, M. *Auf Jagd im Untergrund: Mit Hightech auf der Suche nach Öl, Gas und Erdwärme*, 3rd ed.; Springer: Berlin/Heidelberg, Germany, 2022; ISBN 978-3-662-64150-7.
32. Nielsen, J. Success Rate: The Simplest Usability Metric. Available online: <https://www.nngroup.com/articles/success-rate-the-simplest-usability-metric/> (accessed on 12 July 2023).
33. Laugwitz, B.; Held, T.; Schrepp, M. Construction and Evaluation of a User Experience Questionnaire. In *HCI and Usability for Education and Work*; Holzinger, A., Ed.; Springer: Berlin/Heidelberg, Germany, 2008; pp. 63–76, ISBN 978-3-540-89349-3.
34. Rauschenberger, M.; Schrepp, M.; Thomaschewski, J. User Experience mit Fragebögen messen—Durchführung und Auswertung am Beispiel des UEQ. In *Tagungsband UP13*; Brau, H., Lehmann, A., Petrovic, K., Schroeder, M.C., Eds.; German UPA e.V: Stuttgart, Germany, 2013; pp. 72–77.
35. Brooke, J. SUS: A quick and dirty usability scale. *Usabil. Eval. Ind.* **1995**, *189*, 189–194.
36. Bangor, A.; Kortum, P.; Miller, J. Determining what individual SUS scores mean: Adding an adjective rating scale. *J. Usabil. Stud. Arch.* **2009**, *4*, 114–123.
37. Mayring, P. Qualitative Inhaltsanalyse. In *Handbuch Qualitative Forschung in der Psychologie*; Mey, G., Mruck, K., Eds.; VS Verlag für Sozialwissenschaften: Wiesbaden, Germany, 2010; pp. 601–613, ISBN 978-3-531-16726-8.
38. Endsley, M.R. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum. Factors* **1995**, *37*, 32–64. [[CrossRef](#)]
39. Salmon, P.M.; Stanton, N.A.; Walker, G.H.; Jenkins, D.; Ladva, D.; Rafferty, L.; Young, M. Measuring Situation Awareness in complex systems: Comparison of measures study. *Int. J. Ind. Ergon.* **2009**, *39*, 490–500. [[CrossRef](#)]
40. Roberts, R.; Flin, R.; Cleland, J. How to recognise a kick: A cognitive task analysis of drillers' situation awareness during well operations. *J. Loss Prev. Process Ind.* **2016**, *43*, 503–513. [[CrossRef](#)]
41. Mehta, R.K.; Peres, S.C.; Shortz, A.E.; Hoyle, W.; Lee, M.; Saini, G.; Chan, H.-C.; Pryor, M.W. Operator situation awareness and physiological states during offshore well control scenarios. *J. Loss Prev. Process Ind.* **2018**, *55*, 332–337. [[CrossRef](#)]
42. Horsch, C.H.G.; Smets, N.J.J.M.; Neerinx, M.A.; Cuijpers, R.H. Comparing performance and situation awareness in USAR unit tasks in a virtual and real environment. In *Proceedings of the 10th International Conference on Information Systems for Crisis Response and Management 2013 (ISCRAM 2013)*, Baden Baden, Germany, 12–15 May 2013; Comes, T., Fiedrich, S., Fortier, S., Geldermann, J., Eds.; ISCRAM: Baden-Baden, Germany, 2013; ISBN 978-3-923704-80-4.
43. Read, J.M.; Saleem, J.J. Task Performance and Situation Awareness with a Virtual Reality Head-Mounted Display. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2017**, *61*, 2105–2109. [[CrossRef](#)]
44. Taylor, R.M. Situational Awareness Rating Technique (Sart): The Development of a Tool for Aircrew Systems Design. In *Situational Awareness*; Salas, E., Dietz, A.S., Eds.; Routledge: Oxford, UK, 2017; pp. 111–128, ISBN 9781315087924.
45. Clark, R.E.; Sugrue, B.M. Research on Instructional Media 1978–1988. In *Educational Media and Technology Yearbook*; Libraries Unlimited, Inc.: Englewood, CO, USA, 1988.
46. Karapanos, E.; Zimmerman, J.; Forlizzi, J.; Martens, J.-B. User Experience over Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*, Boston, MA, USA, 4–9 April 2009; Olsen, D.R., Arthur, R.B., Hinckley, K., Morris, M.R., Hudson, S., Greenberg, S., Eds.; ACM: New York, NY, USA, 2009; pp. 729–738, ISBN 9781605582467.
47. Franzreb, D.; Warth, A.; Futternecht, K. User Experience of Real and Virtual Products: A Comparison of Perceived Product Qualities. In *Developments in Design Research and Practice*; Duarte, E., Rosa, C., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 105–125, ISBN 978-3-030-86595-5.
48. Hinricher, N.; König, S.; Schröer, C.; Backhaus, C. Effects of virtual reality and test environment on user experience, usability, and mental workload in the evaluation of a blood pressure monitor. *Front. Virtual Real.* **2023**, *4*, 1151190. [[CrossRef](#)]
49. Sonderegger, A.; Sauer, J. The influence of design aesthetics in usability testing: Effects on user performance and perceived usability. *Appl. Ergon.* **2010**, *41*, 403–410. [[CrossRef](#)]
50. Chandran, S.K.; Forbes, J.; Bittick, C.; Allanson, K.; Brinda, F. *Sample Size in the Application of System Usability Scale to Automotive Interfaces*; SAE Technical Paper Series; SAE: Warrendale, PA, USA, 4 April 2017.
51. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. *Int. J. Hum. Comput. Interact.* **2008**, *24*, 574–594. [[CrossRef](#)]
52. Tullis, T.; Stetson, J. *A Comparison of Questionnaires for Assessing Website Usability*; Usability Professionals Association (UPA): Minneapolis, MN, USA, 2004.
53. Simoes, F.; Bezerra, M.; Teixeira, J.M.; Correia, W.; Teichrieb, V. A User Perspective Analysis on Augmented vs. 3D Printed Prototypes for Product's Project Design. In *Proceedings of the 2016 XVIII Symposium on Virtual and Augmented Reality (SVR)*, Gramado, Brazil, 21–24 June 2016; pp. 71–80, ISBN 978-1-5090-4149-7.
54. Bachman, P.; Milecki, A. Safety Improvement of Industrial Drives Manual Control by Application of Haptic Joystick. In *Intelligent Systems in Production Engineering and Maintenance*; Burduk, A., Chlebus, E., Nowakowski, T., Tubis, A., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 563–573, ISBN 978-3-319-97489-7.
55. Meera, C.S.; Sairam, P.S.; Veeramalla, V.; Kumar, A.; Gupta, M.K. Design and Analysis of New Haptic Joysticks for Enhancing Operational Skills in Excavator Control. *J. Mech. Des.* **2020**, *142*, 121406. [[CrossRef](#)]
56. Anguelov, N. Haptic and Acoustic Parameters for Objectifying and Optimizing the Perceived Value of Switches and Control Panels for Automotive Interiors. Ph.D. Thesis, Technical University Dresden, Berlin, Germany, 2009.

-
57. Vergara, M.; Mondragón, S.; Sancho-Bru, J.L.; Company, P.; Agost, M.-J. Perception of products by progressive multisensory integration. A study on hammers. *Appl. Ergon.* **2011**, *42*, 652–664. [[CrossRef](#)] [[PubMed](#)]
 58. Rebelo, F.; Noriega, P.; Duarte, E.; Soares, M. Using virtual reality to assess user experience. *Hum. Factors* **2012**, *54*, 964–982. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.