

A Mixed-Method Analysis of Usability Study of Video and VR Safety Training: Towards Implementation of VR in Working at Height Training

Pranil GC, Ratvinder Grewal

Laurentian University, School of Engineering and Computer Science
935 Ramsey Lake Rd, Sudbury, Ontario, Canada P3E 2C6
pg_c@laurentian.ca; rgrewal@laurentian.ca

Abstract - Falling from height is considered one of the top causes of workplace injuries and fatalities in the construction industry. The regulatory WAH training, conducted in-class and lecture-based, has been successfully implemented; however, its effect is modest. This study aims to find the relationship between the traditional method and VR simulation in terms of user perception. A crossover design was adopted where participants experienced video and VR training in different sequences. Widely used SUS to measure perceived usability and a VR perception questionnaire was implemented. The two-factor analysis of SUS resulted in new usability and learnability. The result shows no significant difference in perceived usability between the training methods. However, on further analysis, one group found video easier to learn. Similarly, there was a significant inclination of users towards VR training in terms of preference, engagement, and ease of remembering. The Spearman correlation revealed older participants perceived the VR interface as less usable. It was also observed that the training order with video first followed by VR perceived the overall system better as compared to the other group. Further suggestions using qualitative data analysis are proposed.

Keywords: Virtual Reality, Working at height training, Usability, Learnability

© Copyright 2024 Authors - This is an Open Access article published under the Creative Commons Attribution License terms (<http://creativecommons.org/licenses/by/3.0>). Unrestricted use, distribution, and reproduction in any medium are permitted, provided the original work is properly cited.

1. Introduction

The construction industry is considered one of the most hazardous industries globally, consistently registering a high number of accidents [1], [2]. It has been reported that falling from height is one of the primary reasons for occupational injuries and fatalities [3], [5]. According to the report of CPWR [6], the United States construction sites constitute 38% of the fatalities due to falls. There were 1,013 fatalities on construction sites in 2017, 389 of which were workers who fell from a height. More than 40,000 injuries due to falls have been reported in Canada every year. According to the report by the Ministry of Labour [7], in 2022 there were 22 fatalities recorded on construction sites in Ontario Province.

Due to the high number of injuries and fatalities, the Ministry of Labor has regulated a mandatory Working-at-height (WAH) training program in 2015 under Ontario's O. Reg. 297/13 [8]. The training is compulsory for the workers working over a hazardous surface, 3 meters in height, or wearing fall protection equipment. Robson et. al. [9] evaluated the effectiveness of the regulatory WAH training program in Ontario between 2012 to 2019. It was concluded that the regulatory training has been adopted widely, and the impact was significant but modest.

The standard method for delivering WAH training includes in-person or online theory classes and an instructor-led practical class. The traditional methods fail to simulate high-risk real-world scenarios due to several limitations like time, cost, and safety [10]. The meta-analysis was conducted to evaluate the

effectiveness of different mediums of safety training [11]. Burke and his group classified the training into 3 different categories in terms of engagement.

1. Least engaging training,
2. Moderately engaging training, and
3. Most engaging training.

The regulatory Work at Height (WAH) training is predominantly delivered through lectures or videos. These traditional methods rank among the least engaging on the spectrum of training approaches. Research indicates that more interactive training methods, such as Virtual Reality (VR), are significantly more effective due to their higher level of engagement [11]. VR enhances safety training with an increased level of presence, ability to fail safely, and context-aware training [12].

Continuous improvement in Occupational Health and Safety (OHS) training is essential to enhance learning effectiveness, reduce preventable incidents on worksites, and foster a safer work environment. As experiential learners, construction workers often lose interest in memorizing safety regulations when traditional training methods are used [13]. The previous study evaluated the usability of VR safety training in comparison to traditional methods, such as PowerPoint slides, for working-at-height training [14]. A System Usability Scale (SUS) was used to assess usability, alongside participants' demographic data. A moderate association was found between age and system usability, with older participants rating the VR system as less usable.

Usability is defined as a user's ability to interact with a system in an efficient, comfortable, and intuitive way [15]. ISO 9126 defines usability as a blend of factors, including understandability, operability, and learnability [16]. Similarly, ISO 9241 describes usability as the ability of a product, service, or system to enable users to meet their goals efficiently, effectively, and with satisfaction within a specific context [17]. Al-Khiami and Jaeger studied the usability of VR for working at height training among Blue-Collar workers in Kuwait [14]. They adopted SUS to associate usability with the demographics of the worker.

A systematic review was conducted to evaluate the effectiveness of VR in safety training. The authors recommended, in terms of experimental design, that participants should be exposed to both training methods for a more accurate comparison. Regarding usability, it was suggested to use standard usability metrics, such as the System Usability Scale (SUS), and to explore their

relationship with other relevant metrics. Additionally, open-ended questionnaires were advised to gain deeper insights into trainees' perceptions [12].

SUS termed a 'quick and dirty usability scale' consists of ten items- five positive experiences and five negative experiences [18]. SUS is a unidimensional scale with a score ranging from 0 to 100. Bangor et al. [19] found the coefficient alpha of SUS to be .91 suggesting high reliability. Furthermore, they tested the one-dimensionality of SUS by examining the eigenvalues and factor loadings for only one significant factor solution where they found 5.5 implying to use of SUS having one overall single score. Sauro and Lewis [20] conducted the possibility of multi-factor analysis, which Bangor et al. disregarded. Varimax rotation, a common factor analysis, was adopted with multiple factors. The extent of convergence of the two factors was striking with a total variance of 56-58%. They suggested decomposing the SUS score into Usability and Learnability components. First factors 1,2,3,5,6,7,8, and 9 aligned as the first factor "a new scale Usability" while 4 and 10 as another factor "a new scale Learnability".

This study aims to develop and expose trainees to both video-based and VR interfaces for safety training. System usability was analyzed with trainees' demographics and perceptions after each exposure. Additionally, a multifactor analysis of the SUS was conducted to better understand the relationship between usability and learnability. Finally, the association between each usability aspect and trainees' perceptions was evaluated.

2. Materials and Method

The main objective was to create a learning medium that would educate novice workers to identify hazards, safe procedures, and elimination of hazards while installation of skylights. The learning content is extracted from Ontario's official guide for working at height safety and ministry approved learning center's reference manual. Regulation 213- Section 26 from the Construction project in Ontario was directly mentioned as it provides clear guidelines about who may be subjected to fall protection [21]. The introduction of worksite-related equipment and procedures like lanyards, full-body harnesses, and fall restriction equipment were introduced to the trainees before illustrating the safe procedure for installing the skylight.

Two instructional mediums were developed: a traditional lecture/presentation format and a virtual environment simulation. The video presentation featured slides, as illustrated in Figure 1, with an AI-

generated voiceover narrating the content aloud. Participants were allowed to take notes during the video presentation; however, they could not stop or rewind the video to simulate a real-time lecture training experience. Relevant images were used to illustrate the scenarios for the video presentation. The overall training in both mediums consisted of 4 modules. They are:

1. General rules
2. Safety equipment
3. Hierarchy of fall protection
4. Installation of skylight

The same learning content was used in the scenarios of VR applications. The iterative interaction design process was adopted to develop VR applications.



Figure 1. Screenshots of video interface of 4 modules.



Figure 2. Screenshots of the VR interface of 4 modules.

The 3D model of an under-construction multistorey building, full-body harness, and safety equipment in the Filmbox (FBX) format for Unity was incorporated into the VR environment. The screenshot of the VR interface is illustrated in Figure 2. Unity was used as a development platform and C# as a programming

language. Meta’s Oculus Quest 2 was employed as a VR device for the participants.

2.1 Experimental methods

This study was approved by the Laurentian University Research Ethics Board (LU REB) under approval number [6021449]. The study adopted a 2x2 crossover experimental design for two reasons: 1) to observe participant’s reactions to both training interfaces and 2) to allow participants to compare and evaluate based on their perceptions. The experiment involves the recruitment of 26 students from Laurentian University. It has been reported that the within-subject design introduces the potential of carryover or residual

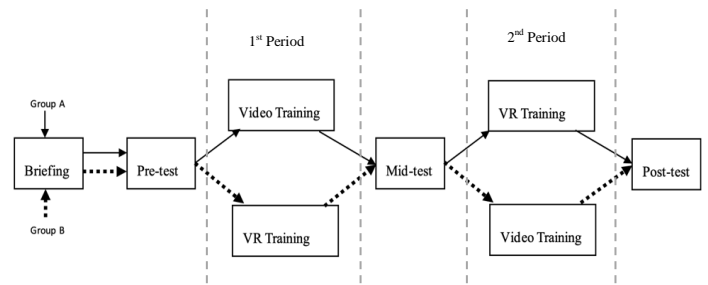


Figure 3. Experimental method.

effects from the previous to the subsequent period. The crossover experiment allows for control over sequence and carryover effects, as the order of treatments is alternated across participants which is also known as counterbalancing [22]. Participants were assigned to two groups: Group A and Group B. As shown in Figure 1, Group A also known as the video-VR sequence received video-based training followed by VR, while Group B experienced the reverse order. Inclusion criteria required participants to be over 18 years old. Participants with medical conditions like migraines and epilepsy were excluded from the study due to the risk of exposure to VR.

2.2 Procedure

The two treatments, video, and VR were administered in two periods and alternating sequences as shown in Figure 3. The experimental procedure is illustrated in **Error! Reference source not found.** Participants were briefed and then randomly assigned to one of two groups: Group A (sequence video-VR) and Group B (sequence VR-video). A pre-test was conducted to record the participant’s background, which includes age, gender, and previous VR experience. After the first period, a mid-test was conducted. The SUS scale was

administered to both groups, as represented by SUS-1 in Figure 2. In the second period, participants received the training in an alternative order to observe any differences in their perceived usability. The post-test also considered the VR perception scale and an open-ended question. Participants were seated in video and VR sessions. During the video session, the VR headset and hand controllers were replaced with a laptop for displaying the video-based training.

The SUS questionnaire by Brooke, along with the revised wording from Bangor et al., were considered for this study [18], [19]. SUS is a 10-item scale with 5 positive and 5 negative experiences ranging from 1 to 5. The following are SUS scale:

- U1. I think that I would like to use this system frequently
- U2. I found the system unnecessarily complex
- U3. I thought the system was easy to use
- U4. I think I would need the support of technical person to be able to use this system
- U5. I found the various function of the system well integrated
- U6. I thought there was too much inconsistency in the system
- U7. I would imagine that most people would learn to use this system very quickly
- U8. I found this system cumbersome/awkward to use
- U9. I felt very confident using this system
- U10. I need to learn lot of things before I could get going to use this system

The VR perception scale was adopted from Lovreglio et al.'s study [23]. It is a 3 scale-based question ranging from -3 to +3 (strongly disagree - strongly agree).

- P1. I found VR simulation more engaging than lecture-based training
- P2. It was easier to remember fall protection recommendations provided in VR simulation than those provided in lecture-based training
- P3. I prefer the VR simulation over lecture-based training

All three items compare the VR simulation with lecture-based (traditional) training.

The first item (P1) is associated with engagement, the second item (P2) is the ease of remembering, and the last item (P3) is the preference between the training

interfaces. Overall perception was calculated based on Lovreglio et al.'s method, which sums the scores of each question [23]. For instance, an overall perception score of +9 indicates strong agreement among participants that they perceived VR as better, -9 reflects a preference for video training, and 0 represents a neutral perception. At the end of the survey, there is one open-ended questionnaire. This question is about the overall experience of participants along with likes and dislikes and suggestions on both interfaces.

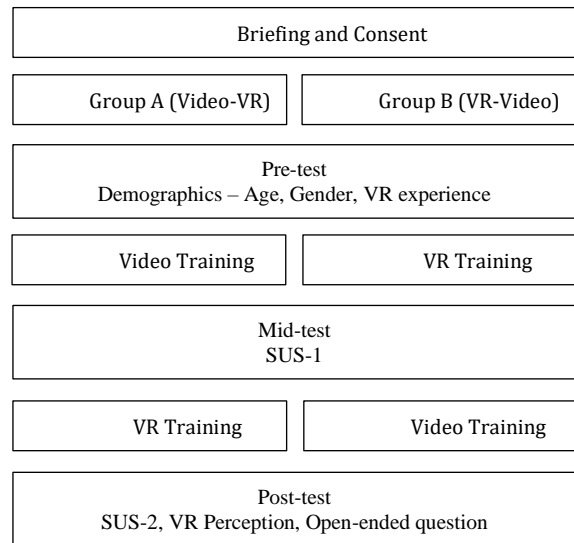


Figure 4. Experimental Procedure.

3. Results and Discussion

Group A (Video-VR) and Group B (VR-Video) refer to the sequence of training conditions. These terms are used interchangeably throughout the study to represent their respective sequences. The results of the Pre-test (Demographics), Mid-test (SUS-1), and Post-test (SUS-2, VR Perception) were recorded using an online survey administered on a tablet. The open-ended question at the end was collected on paper. Additionally, the completion of the VR training was observed and documented for each participant. Descriptive statistics and inferential statistics were utilized for this study. The data analysis was conducted using IBM SPSS version 29.

Spearman correlation was employed to examine the relationship between demographics, SUS scores, and responses to the VR perception questionnaire. The consideration of the correlation coefficient is based on [24]. Correlation coefficients between 0.1 and 0.29 are considered indicative of a small association, those

between 0.3 and 0.49 represent a medium association, and values above 0.5 suggest a large association.

Both independent samples t-tests and dependent samples t-tests were conducted on the SUS scores, with a significance level of $\alpha = 0.05$, to assess differences in perceived usability. Additionally, a factor analysis of the SUS scores was performed. The relationships between the newly identified factors, demographics, and VR perception responses were analyzed. Qualitative data from open-ended questions were coded based on Rosala [25] and the thematic analysis was conducted to identify patterns and insights into participants' experiences.

3.1 Demographic data

The mean and standard deviation of demographic data collected from the pre-test are shown in Table 1. The completion time of VR training for Group A and Group B is expressed in seconds and shown in the third column. Participants' previous VR experience was categorized based on their prior exposure to virtual reality: 0 for no VR experience, and 1 for those with prior VR experience.

Table 1. Mean and Standard deviation of age, previous VR experience, and VR completion time based on different conditions.

Variable	Condition	Mean	SD
Age	video-VR	27.8	5.8
	VR-video	24.2	3.2
Previous VR	video-VR	0.54	0.52
	VR-video	0.23	0.44
VR-time	video-VR	462.5	87.9
	VR-video	429	87

Participants in Group A (Video-VR) were, on average, older (27.8 years vs. 24.2 years) than those in Group B (VR-Video) and exhibited greater age variability. Additionally, Group A had more prior VR experience (Mean = 0.54) compared to Group B (Mean = 0.23) and took longer to complete the VR task (462.5 seconds vs. 429 seconds), with both groups showing similar variability in completion time.

SUS was collected after each period of training. These are referred to as SUS-1 in 1st period and SUS-2 in the 2nd period. The SUS scores of participants were calculated as per Brooke [18]. The mean and standard deviation of the SUS score are shown in Table 2.

Table 2. Mean and standard deviation of SUS score in mid-test and post-test for video and VR interface.

Variable	Condition	Mean	SD
SUS-1	Video	61.54	21.32
	VR	70.57	20.52
SUS-2	Video	80.19	10.87
	VR	75.38	7.55

Participants in Group A, who received VR training in the second period, reported a higher usability score for the VR interface (SUS-2: Mean = 75.38) than in the first period with Video training (SUS-1: Mean = 61.54). Similarly, participants in Group B, who received Video training in the second period, rated the Video interface higher (SUS-2: Mean = 80.19) than the earlier VR training (SUS-1: Mean = 70.57). Overall, participants generally gave a higher SUS score to the interface they experienced later in the training sequence.

Table 3. Mean and standard deviation of individual and overall perception questionnaire for both groups.

Variable	Condition	Mean	SD
P1 (Engagement)	Video-VR	2.15	1.40
	VR-Video	1.77	1.09
P2 (Ease of remembering)	Video-VR	1.85	1.62
	VR-Video	0.15	2.03
P3 (Preference)	Video-VR	1.69	2.10
	VR-Video	0.38	2.18
Overall VR perception	Video-VR	5.70	4.70
	VR-Video	2.31	4.48

The scores of individual and overall VR perception questions are shown in Table 3. In all questions, participants in Group B (VR-Video sequence) rated lower than those in Group A (Video-VR sequence). For instance, Group B scored significantly lower in P2 (Ease of remembering) with a mean score of 0.15 compared to Group A's 1.85. Similarly, the overall VR perception score for Group B was 2.31, much lower than Group A's 5.70.

3.2 SUS score

The SUS scores were analyzed using the Shapiro-Wilk test for normality ($p < 0.05$), which indicated that the data were not normally distributed. Consequently, a non-parametric approach, the Mann-Whitney U test, was applied to compare SUS scores between the two groups across both periods. Table 4 presents the mean SUS scores and statistical significance for each group and medium across individual questions.

The results of the between-subjects analysis indicated no statistically significant differences for the majority of the questionnaire items. However, Question 1 from the first period revealed a significant difference, with a higher score for the VR training condition. Similarly, Questions 4 and 10 in the second period demonstrated statistically significant differences, with higher scores in the video training condition.

Table 4. Mean SUS score for both groups after each training interface. (*=Significant difference at $\alpha=0.05$)

	Group A			Group B		
	Video	VR	Sig.	VR	Video	Sig.
U1	3.54	4.08	0.008*	2.16	1.30	0.25
U2	2.38	2.31	0.34	1.71	1.67	0.14
U3	3.46	4.00	0.14	2.21	2.50	0.83
U4	2.54	2.85	0.79	2.08	2.03	0.001*
U5	3.46	3.92	0.26	2.26	2.48	0.68
U6	2.38	2.15	0.78	1.71	1.68	0.39
U7	3.31	4.15	0.06	2.83	2.42	0.60
U8	2.54	2.31	0.39	1.47	1.67	0.46
U9	3.62	4.15	0.50	2.97	2.57	0.83
U10	2.92	2.46	0.18	1.35	1.86	0.01*

As discussed in Section 1, the revised SUS scale introduces a new dimension of "Learnability" by excluding Questions 4 and 10 from the original scale. To further assess the 2-dimensional nature of the SUS score, factor analysis of the SUS score was conducted which is discussed in the next section.

3.3 Factor-analysis of SUS

The factor analysis of SUS was conducted using the SUS score of both periods (1st and 2nd periods) with a total of 54 SUS scores. A 10-item correlation matrix with varimax rotation as the input. A similar method was conducted by Bangor et al. [19] and Sauro and Lewis [20]. The result of the factor analysis is illustrated in Table 5.

The eigenvalues for the one-factor and two-factor models are 4.61 and 1.52, respectively, explaining 46.139% and 15.233% of the variance, with a cumulative explained variance of 61.37%. According to the rule of thumb that suggests retaining factors with eigenvalues greater than one, these results support the two-factor structure of the SUS, aligning with the division into

usability and learnability dimensions as proposed by Sauro and Lewis [20].

The new SUS score was calculated based on Sauro and Lewis[20]. Table 6 illustrates the new SUS score across both the period and training medium.

Table 5. Component matrix using Principal Component Analysis.

Item	1	2
U1	0.611	0.50
U2	-0.67	-0.13
U3	0.72	0.04
U4	-0.29	0.85
U5	0.77	0.30
U6	-0.77	-0.06
U7	0.67	-0.17
U8	-0.74	0.06
U9	0.80	0.10
U10	-0.60	0.61

Table 6. Mean and standard deviation of two-factor SUS score across both period and conditions. (U= Usability, L = Learnability)

Variable	Period	Condition	Mean	SD
SUS-U	1 st	Video	62.74	21.56
		VR	76.68	9.17
	2 nd	Video	76.92	13.41
		VR	73.57	37.12
SUS-L	1 st	Video	56.73	39.07
		VR	70.19	17.19
	2 nd	Video	93.27	10.96
		VR	30.53	40.88

From the table, it can be observed that the mean score difference of the new Usability score (SUS-U) is generally smaller than that of the new Learnability score (SUS-L). Specifically, in the second period, there is a substantial gap of 62.74 in the learnability score between the Video interface (93.27, SD = 10.96) and the VR interface (30.53, SD = 40.88). This gap suggests that participants found the video interface easier to learn compared to the VR interface during the second period.

3.4 Statistical analysis

A statistical analysis was conducted on the new SUS score and the VR perception question. A within-subject study was conducted to test differences between Video and VR using the Wilcoxon Signed rank test. The test results are shown in Table 7.

Table 7. The test result of a within-subject study using Wilcoxon signed rank test between video and VR interface across both groups (U=Usability, L= Learnability).

Variable	Condition	Test (W,Z)	Sig.
SUS-U	Video vs VR - A	55, 1.25	0.209
	VR vs Video -B	40	0.70
SUS-L	Video vs VR - A	35, 0.178	0.858
	VR vs Video - B	0.0, -2.83	0.005*

For the Usability Score (SUS-U), the comparison between the Video and VR interfaces in both groups (Video-VR and VR-Video) yielded non-significant results. However, in Group B (VR-Video), the test revealed a statistically significant result with $W = 0.0$, $Z = -2.83$, and a p-value of 0.005. This significant result suggests that participants found a notable difference in learnability between the VR and video interfaces, with video likely being perceived as easier to learn.

A between-subject study was conducted to analyze the statistical significance of VR perception among the participants. Table 8 shows the Mann-Whitney t-test on perception items between Group A and Group B.

Table 8. Between subject study using Mann-Whitney t-test for VR perception questionnaire. (P = perception)

Variable	Condition	Test (U,Z)	Sig.
P1	Group A vs B	60.5, -1.3	0.223
P2	Group A vs B	36.5, -2.53	0.012*
P3	Group A vs B	47.5, -1.96	0.057
P overall	Group A vs B	44.5, -2.07	0.039*

There is no significant difference in VR perception in terms of engagement and preference between the groups. However, in terms of ease of remembering, there is a statistically significant difference ($U=36.5$, $p=0.012$), with Group A (1.85) reporting a much higher ease of remembering than Group B (0.15). Furthermore, the overall VR perception is statistically significantly different ($U=44.5$, $p=0.039$), with Group A (5.70) scoring significantly higher than Group B (2.31).

It can be concluded that although the usability of both mediums (video and VR) does not differ significantly, the learnability of VR is perceived to be significantly lower, especially for the group that experienced VR first and video second (Group B). Additionally, the VR perception results have a similar trend. Group A (video-first) significantly scored higher in ease of remembering and overall perception as compared to Group B. This shows the order of training

matters and the sequence of training videos first, followed by VR, is easier to learn and easy to remember. Therefore, it is recommended to integrate VR after video training. This sequence of video introduction, familiarization with VR sets, and VR training was found to reduce stress levels and increase confidence in a previous study [14].

3.5 Correlation analysis

The relationship between different variables was analyzed. The correlation coefficients for demographic data (age, previous VR experience), new Usability score, Learnability score, and perception score are shown in Table 9.

Table 9. Spearman correlations of demographic data with new usability SUS score and perception questions. (Exp = Previous VR experience, U= new SUS usability, L = new SUS learnability, V = Video, Per= Perception).

	Age	Exp	U-V	U-VR	L-V	L-VR	P
Age	1						
Exp	-0.10	1					
U-V	0.20	-0.02	1				
U-VR	-0.42*	0.31	-0.21	1			
L-V	-0.12	-0.18	0.43*	-0.23	1		
L-VR	-0.08	-0.11	0.20	0.34	0.33	1	
P	-0.87	0.27	-0.20	0.57**	-0.14	0.26	1

The correlation results from Table 9 revealed several insights into how age, VR experience, new system usability (SUS) scores, and VR perception are related.

A significant negative correlation is observed between age and the usability score for VR, with $r=-0.421$, $p=0.032$. No significant relationship is found between age and learnability scores for video or VR. A positive medium association but non-significant relationship between previous VR experience and the usability score for VR (U_VR) ($r=0.314$, $p=0.118$). While not statistically significant, this trend suggests that participants with prior VR experience may find the VR interface more usable. There is a positive but nonsignificant correlation between usability scores in VR (U_VR) and learnability in VR (L_VR) ($r=0.339$, $p=0.090$), though this relationship is only marginally significant. There's a significant positive correlation between usability in video (U_V) and learnability in video (L_V) ($r=0.431$, $p=0.028$). The VR perception score

is strongly correlated with the usability score in VR ($r=0.570$, $p=0.002$). There is no significant relationship between the perception score and learnability in video or VR.

Some of the findings are summarized below:

1. Elder participants found the VR interface to be less usable.
2. Participants who rated the VR interface as more usable also found it easier to learn.
3. Participants who perceived the video interface as usable also found it easier to learn.
4. This suggests that participants who scored higher on the perception also rated the VR interface as more usable.

As with the findings of the previous study [14], this study also identified a moderate correlation between age and usability factors, suggesting the need for designing more inclusive VR interfaces that cater to all age groups. Furthermore, the study explored the relationship between usability scores and users' perceptions of VR, offering deeper insights into additional factors influencing user experience.

3.6 Qualitative analysis

At the end of the study, participants were asked an open-ended question regarding their overall experience. Keynotes were taken in written format. The research notes were coded and grouped to generate a thematic analysis. Thematic analysis of interview responses is presented in Table 10. The responses that occurred more than two times are presented in the table.

The following are the key observations in the qualitative feedback:

1. Inconsistency in UI (N=4) and distraction (N=4) were prominent issues for the VR interface
2. The main issue with the video was that participants found it boring (N=3), with additional complaints about the fast narration and overall inconvenience, even though the narration speed was identical in both the video and VR formats.
3. Participants favoring VR (N=7) was higher than video (N=2).
4. Participants found VR as a new experience (N=5), realistic (N=5), and knowledgeable and long-term (N=3).

5. The suggestions for improvements across both interfaces had a strong overlap. For the video interface, participants recommended adding a navigation button (N=3). Similarly, for the VR interface, participants suggested increasing interaction (N=4), adding more animations (N=2), and enabling the ability to take notes (N=2).

Table 10. Thematic analysis of video and VR interface from user response.

Study Group	Video	VR
Issue	boring (3), fast narration (2), inconvenient (2),	inconsistency in UI (4), Distraction (4), inconvenience to use (2), environmental noise (2),
Suggestion	add navigation button (3)	more interaction (4), more animation (2), able to take notes (2)
Comments	Preferred Video (2)	preferred VR (7), new experience (5), realistic (5), knowledge and retention (3)

These observations suggest a desire for more interactive and dynamic content to enhance the learning experience in both mediums.

In a previous study, it was observed that participants reported difficulty in handling and interacting with the VR interface [14]. However, no such observations were reported in this study. Participants were found to skip the instruction (10 mentions), suggesting they tend to interact with the system rather than observe the text.

The main limitation of this study was the relatively small sample size, which may have affected the statistical power and the ability to generalize the findings. Additionally, the learning content for both video and VR sessions lasted only 5-6 minutes, limiting the participants' exposure to the training material. For future research, it is recommended to develop more comprehensive and extended learning modules in both mediums to ensure deeper engagement. Moreover, future studies should combine usability assessments with other metrics of training effectiveness, such as skill acquisition, behavioral changes, knowledge retention,

and overall training outcomes, to gain a holistic understanding of the effectiveness of VR and traditional training methods. More interaction should be incorporated into both training interfaces to enhance trainee engagement and motivation.

5. Conclusion

A mixed-method analysis of usability provided deeper insights into user perceptions by combining both quantitative and qualitative data. This approach allowed for a comprehensive understanding of how users interacted with and perceived the system. Furthermore, the Crossover design implemented in the study enabled an examination of both between-subject and within-subject relationships. We were able to expose participants to both training interfaces and analyze their responses across different conditions.

Both interfaces—video and VR—were generally perceived as marginal to acceptable, as reflected by the average SUS scores. The factor analysis of the SUS scores across both groups confirmed a two-factor structure, aligning with the findings of Sauro and Lewis, which distinguish between usability and learnability.

Older participants found the VR system less usable, indicating a need for future VR development in worker training to focus on creating more inclusive and accessible designs for all age groups.

Additionally, the study revealed medium correlations between the new usability and learnability factors, as well as strong correlations between participants' perceptions of the VR interface and their VR usability scores. The group that received video training first, followed by VR, showed higher scores in both usability and VR perception. This suggests that implementing VR as a supplementary tool after traditional forms of training can enhance usability and user perception, aligning with recommendations from previous studies.

References

- [1] "A systematic review of factors leading to occupational injuries and fatalities | Journal of Public Health." Accessed: Apr. 15, 2024. [Online]. Available: <https://link.springer.com/article/10.1007/s10389-020-01427-4>
- [2] X. Meng and A. H. S. Chan, "Current States and Future Trends in Safety Research of Construction Personnel: A Quantitative Analysis Based on Social Network Approach," *Int. J. Environ. Res. Public Health*, vol. 18, no. 3, Art. no. 3, Jan. 2021, doi: 10.3390/ijerph18030883.
- [3] H. F. van der Molen *et al.*, "Interventions to prevent injuries in construction workers," *Cochrane Database Syst. Rev.*, vol. 2018, no. 2, 2018, doi: 10.1002/14651858.cd006251.pub4.
- [4] T. A. Bentley *et al.*, "Investigating risk factors for slips, trips and falls in New Zealand residential construction using incident-centred and incident-independent methods," *Ergonomics*, vol. 49, no. 1, pp. 62–77, Jan. 2006, doi: 10.1080/00140130612331392236.
- [5] "Trends of fall injuries and prevention in the construction industry." Accessed: Apr. 15, 2024. [Online]. Available: <https://stacks.cdc.gov/view/cdc/81140>
- [6] "Trends of fall injuries and prevention in the construction industry." Accessed: Jan. 11, 2024. [Online]. Available: <https://stacks.cdc.gov/view/cdc/81140>
- [7] "MOL reports 22 construction fatalities in 2021 - constructconnect.com," Daily Commercial News. Accessed: Jan. 11, 2024. [Online]. Available: <https://canada.constructconnect.com/dcn/news/ohs/2022/02/mol-reports-22-construction-fatalities-in-2021>
- [8] "Program standard for working at heights training | ontario.ca." Accessed: Jan. 05, 2024. [Online]. Available: <http://www.ontario.ca/page/program-standard-working-heights-training>
- [9] L. S. Robson, V. Landsman, P. M. Smith, and C. A. Mustard, "Evaluation of the Ontario Mandatory Working-at-Heights Training Requirement in Construction, 2012–2019," *Am. J. Public Health*, vol. 114, no. 1, pp. 38–41, Jan. 2024, doi: 10.2105/AJPH.2023.307440.
- [10] D. Scorgie, Z. Feng, D. Paes, F. Parisi, T. W. Yiu, and R. Lovreglio, "Virtual reality for safety training: A systematic literature review and meta-analysis," *Saf. Sci.*, vol. 171, p. 106372, Mar. 2024, doi: 10.1016/j.ssci.2023.106372.
- [11] M. J. Burke, S. A. Sarpy, K. Smith-Crowe, S. Chan-Serafin, R. O. Salvador, and G. Islam, "Relative Effectiveness of Worker Safety and Health Training Methods," *Am. J. Public Health*, vol. 96, no. 2, pp. 315–324, Feb. 2006, doi: 10.2105/AJPH.2004.059840.
- [12] H. Stefan, M. Mortimer, and B. Horan, "Evaluating the effectiveness of virtual reality for safety-relevant training: a systematic review," *Virtual*

- Real.*, vol. 27, no. 4, pp. 2839–2869, Dec. 2023, doi: 10.1007/s10055-023-00843-7.
- [13] H. T. K. Davies, H. J. M. Panko, and R. Kenley, “Activity-Based Teaching For Unitec New Zealand Construction Students,” *Emir. J. Eng. Res.*, vol. 12, pp. 57–63, Jan. 2007.
- [14] M. I. Al-Khiami and M. Jaeger, “Safer Working at Heights: Exploring the Usability of Virtual Reality for Construction Safety Training among Blue-Collar Workers in Kuwait,” *Safety*, vol. 9, no. 3, Art. no. 3, Sep. 2023, doi: 10.3390/safety9030063.
- [15] K. Stanney, R. Mourant, and R. Kennedy, “Human Factors Issues in Virtual Environments: A Review of the Literature,” *Presence*, vol. 7, pp. 327–351, Aug. 1998, doi: 10.1162/105474698565767.
- [16] “ISO/IEC TR 9126-4:2004(en), Software engineering — Product quality — Part 4: Quality in use metrics.” Accessed: Sep. 27, 2024. [Online]. Available: <https://www.iso.org/obp/ui/#iso:std:iso-iec:tr:9126:-4:ed-1:v1:en>
- [17] “ISO 9241-210:2019,” ISO. Accessed: Sep. 27, 2024. [Online]. Available: <https://www.iso.org/standard/77520.html>
- [18] J. Brooke, “SUS: A quick and dirty usability scale,” *Usability Eval Ind*, vol. 189, Nov. 1995.
- [19] “An Empirical Evaluation of the System Usability Scale: International Journal of Human–Computer Interaction: Vol 24 , No 6 - Get Access.” Accessed: May 14, 2024. [Online]. Available: <https://www.tandfonline.com/doi/full/10.1080/10447310802205776?scroll=top&needAccess=true>
- [20] J. R. Lewis and J. Sauro, “The Factor Structure of the System Usability Scale,” in *Human Centered Design*, M. Kurosu, Ed., Berlin, Heidelberg: Springer, 2009, pp. 94–103. doi: 10.1007/978-3-642-02806-9_12.
- [21] “Law Document English View,” Ontario.ca. Accessed: Sep. 27, 2024. [Online]. Available: <https://www.ontario.ca/laws/view>
- [22] D. Johnson, “Crossover experiments,” *Wiley Interdiscip. Rev. Comput. Stat.*, vol. 2, pp. 620–625, Sep. 2010, doi: 10.1002/wics.109.
- [23] R. Lovreglio, X. Duan, A. Rahouti, R. Phipps, and D. Nilsson, “Comparing the effectiveness of fire extinguisher virtual reality and video training,” *Virtual Real.*, vol. 25, no. 1, pp. 133–145, Mar. 2021, doi: 10.1007/s10055-020-00447-5.
- [24] J. Cohen, P. Cohen, S. G. West, and L. S. Aiken, *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences*, 3rd ed. New York: Routledge, 2002. doi: 10.4324/9780203774441.
- [25] “How to Analyze Qualitative Data from UX Research: Thematic Analysis,” Nielsen Norman Group. Accessed: Sep. 26, 2024. [Online]. Available: <https://www.nngroup.com/articles/thematic-analysis/>