Journal of Construction Engineering, Management & Innovation Volume 9 Issue 3 (2026) 683

Contents lists available at www.goldenlightpublish.com/journals/jcemi/2-toc

Journal of Construction Engineering, Management & Innovation

Journal homepage: www.goldenlightpublish.com/journals/2/jcemi

Evaluating students as surrogates in virtual reality-based construction safety training: A comparative performance analysis

Elif Gürbüz Taştekin, Umutcan Kılıç, Semra Çomu*

Boğaziçi University, Department of Civil Engineering, İstanbul, Türkiye

* Corresponding author: S. Çomu (semra.comu@bogazici.edu.tr)

https://doi.org/10.31462/jcemi.2026.683

Received 27 February 2026; Revised 24 June 2026; Accepted 13 July 2026; Available online 22 September 2026

Keywords

Virtual reality (VR)
Construction safety training
Student surrogates
Occupational health and safety (OHS)

Abstract

High rates of fatal and non-fatal occupational accidents worldwide have made construction safety a global concern. Despite the proven effectiveness of construction safety training in reducing these rates, current safety implementations often fail to enhance construction workers' safety awareness and knowledge. Recently, virtual reality (VR) has been increasingly employed in safety training to enhance its effectiveness and contribute to occupational safety. While construction workers and engineers are usually the intended target audience for safety training, students are frequently used as surrogates during the development of training tools. However, the appropriateness of this practice has not been thoroughly investigated in the literature. This study utilizes the V-SAFE v.2 training tool in the experiments to investigate this phenomenon by comparing the performances of students, construction workers, and engineers. The paper-based test scores of students, construction workers, and engineers before and after receiving safety training in the virtual environment are analyzed using appropriate statistical analysis approaches, including paired T-test, Wilcoxon signed-rank test, and two-factor mixed ANOVA. Results indicate that V-SAFE v.2 has significantly improved the learning outcomes of all groups. Additionally, the study reveals that students can serve as surrogates for construction workers during the development of virtual safety training tools. However, students' performances significantly differ from those of engineers, indicating they are not suitable surrogates for this subject group.

1. Introduction

The construction sector is complex and dynamic, involving many risk factors and facing a high number of accidents and fatalities worldwide. In terms of human safety, the construction industry is one of the most dangerous industries and is one of the sectors where occupational risks most frequently occur [1]. For instance, the statistical authority for the European Union (Eurostat) reported that the construction sector was the leading sector in terms of fatal accidents at work [2]. Construction workers face fatalities at work three to four times more than other workers in developed countries, and this ratio becomes three to six times higher in developing countries [3]. Because of the high rates of occupational accidents and fatalities, construction safety is seen as a global concern, with increasing attention being paid to health and safety management on construction sites [4].

Numerous research studies have been conducted to clarify the underlying reasons for the high rates of accidents and fatalities in the construction sector [5-7]. Proper worker training and education, worker awareness, and managerial procedures were found to be three main areas to be considered to effectively minimize accidents on construction sites [7]. Manzoor et al. [5] identified that worker-related problems, specifically not having proper protective equipment, constitute approximately two-thirds of the fall accidents in high-rise building projects. Hwang et al. [6] showed that ineffective hazard identification, usually resulting from a lack of safety education, and insufficient safety measures created unsafe site conditions and construction processes for workers on small-scale construction sites. On the other hand, several studies showed that unsafe behavior among workers, primarily originating from poor hazard identification and safety risk perception, can be effectively improved through safety training [8, 9]. Unfortunately, although companies apply safety prevention measures as corrective actions, this does not prevent workers from performing unsafe acts [10].

Umeokafor et al. [11] indicated that it is possible to prevent some of these accidents with certain procedures and regulations facilitating construction safety. In addition, worker training and professional development were identified as two of the key factors in enhancing the safety performance of workers on construction sites [12]. Construction companies typically implement safety regulations by adhering to mandatory safety training. They offer their personnel safety training through several channels, including online classes, videotapes, books, and lectures [13]. However, considering the high rates of fatalities and accidents at construction sites, the effectiveness of current safety training tools has been questioned. The key reasons for this issue are inadequate and unengaging training, which is designed without considering the target groups [14]. Although there are numerous alternative safety training methods, workers still lack essential safety knowledge, indicating a deficiency in current safety training practices adopted in the construction industry [15]. Current safety training methods are theoretical rather than practice-based and hands-on [16]. In their meta-analysis study, Burke et al. [17] categorized training methods into three groups based on the learners' participation in the training sessions: (i) least engaging (lectures and videos); (ii) moderately engaging (programmed instruction and feedback interventions), and (iii) highly engaging (training in behavioral modelling and hands-on demonstrations). They also found out that the high-engagement training was roughly two times more effective than the low-engagement training [17].

Workers' involvement and interest in health and safety issues can be increased with the use of computer technologies [18]. Consistent with this finding, Sacks et al. [10] also found that virtual reality safety training tools are superior to conventional methods in terms of construction workers' learning on construction safety risks, which indicates a positive correlation between the effectiveness of the training method and the level of engagement. In another study, Namian et al. [9] compared high-engagement and low-engagement training methods in terms of their association with the level of hazard recognition and risk perception, and the results showed that high-engagement trainings provide higher levels of hazard recognition and risk perception compared to low-engagement training methods. In addition, Perlman et al. [8] showed that trainees can identify hazards better in a virtual environment compared to traditional documents and visuals.

Over the last decade, the commercial interest in VR technologies and the use of VR environments for educational purposes have significantly increased [19]. Numerous research studies investigated the use of VR technology for enhancing occupational safety in a variety of domains, including aviation, construction, and mining, and have made significant contributions to these sectors in terms of occupational safety [20]. Using virtual reality enhances safety training by enabling trainees to experience it in a safe environment [13]. For instance, several studies in construction domain adopted virtual environments in the development of safety tools such as a virtual construction environment for steel erection [21], a multi-user virtual reality system with motion tracking function [22], a personalizable virtual reality-based safety training system that accounts for participant's prior safety knowledge [23], and a VR-based safety training system for heavy machinery operations using head-mounted VR displays and flat screen displays [24].

Involving numerous changes and revisions, the development of a safety training tool is a laborious procedure that takes a long time. As a consequence of this iterative approach, the experiments might be repeated several times. Even though engineers and construction workers are the intended target group for safety training, arranging experiments with them might be difficult because of scheduling, location, and time constraints. Therefore, the produced safety training tools should be an end product and already tested with an appropriate experiment group before setting up the tests with the construction workers and engineers. Researchers have long regarded the employment of students as surrogates as a convenient and expedient resource due to the advantages it offers in terms of time, cost, and distance. For instance, the experiments involving the use of virtual reality safety training tools usually use students as the subject group instead [8, 13, 18, 21-23, 25, 26]. Although the use of student subject groups in the development of new safety training tools is very common, limited research has questioned whether the students are good surrogates for the construction workers. For instance, Trottier and Gordon [27] questioned the appropriateness of using students as surrogates, but the results were inconclusive, and they advised that more research to be performed on this subject to achieve a comprehensive result. Shortly, there have been multiple studies on safety training tools that use students as subject groups, yet there is still a gap in the literature about the acceptability of employing students as surrogates for construction workers and engineers when developing safety training tools. This paper aims to investigate the appropriateness of using students as surrogates for construction workers and for engineers in the development of virtual environment-based safety training tools to fill this gap in the literature. The effectiveness of the tool on the learning of the subject groups will be assessed by comparing the paper-based test results of each group before and after receiving the virtual safety training with V-SAFE v.2, which covers the basic scaffolding tasks and relevant health and safety implementations. In other words, the safety performance of students will be compared with those of construction workers and engineers by using their paper-based test results in order to investigate the appropriateness of using students as surrogates for construction workers and for engineers, respectively.

2. Methodology

In this study, we adopt a quantitative experimental research methodology to investigate the suitability of using students as surrogate participants in VR-based construction safety training research. Controlled VR-based training experiments were conducted with construction workers, engineers, and students, and the collected pre-test/post-test and survey data were statistically analyzed to evaluate differences in learning outcomes among the participant groups. The overall methodological framework, experimental design, VR training environment, participant selection process, and statistical analysis procedures are presented in the following sections.

2.1. Research questions and hypothesis statement

A total of five Research Questions have been formulated to accomplish the aim and objectives of this study. The first three questions seek to examine whether there exists a substantial disparity in the participants' comprehension of fundamental scaffolding activities and associated health and safety measures before and after their utilization of a virtual safety training tool. In a similar study conducted by Sacks et al. [10], a series of experiments was conducted to assess the safety test scores of the subject group before and after training. Pedro et al. [28] evaluated the participants' performance using a paper-based test to see how the participants' education scores differed before and after completing the developed Virtual Construction Safety Education System framework. Furthermore, the study conducted by Adami et al. [29] compared the workers' knowledge before and after completing a VR-based robotic teleoperation training. Based on all these relevant studies, the first three Research Questions are established to examine the changes in the safety knowledge and skills of each subject group after VR-based training as a basis for evaluating the suitability of using students as surrogate participants in VR-based construction safety training research.

On the other hand, the last two Research Questions are aimed at testing the appropriateness of the use of students as surrogates. The use of students as surrogates has been studied in various research areas. For example, Perlman et al. [8] adopted a research methodology that compares the hazard identification and assessment skills of three subject groups: construction superintendents, civil engineering students, and safety directors. In another study, Song et al. [30] conducted an experiment to evaluate whether graduate students could be used as surrogates for managers in business intelligence and analytics research. Considering these comparable research studies, Research Questions 4 and 5 aim to investigate whether the safety performances of students on the paper-based test results are equivalent to those of construction workers and engineers, respectively.

Table 1 displays null and alternative hypotheses constructed based on the aforementioned five Research Questions.

2.2. V-SAFE v.2 training tool

The study employs V-SAFE v.2, a VR-based training tool developed with the objective of minimizing occupational accidents during the installation of H-type scaffolding, to test the proposed Research Questions [31]. On top of introducing safe working procedures, the tool helps with the development of hazard identification and collaboration skills, which are essential for health and safety on construction sites.

Table 1. Hypotheses to be tested for each research question

Research Questions	Hypotheses	
1	H_0:	There is not a significant difference between the paper-based test scores of students before and after receiving VR-based training.
	H_1:	There is a significant difference between the paper-based test scores of students before and after receiving VR-based training.
2	H_0:	There is not a significant difference between the paper-based test scores of construction workers before and after receiving VR-based training.
	H_1:	There is a significant difference between the paper-based test scores of construction workers before and after receiving VR-based training.
3	H_0:	There is not a significant difference between the paper-based test scores of engineers before and after receiving VR-based training.
	H_1:	There is a significant difference between the paper-based test scores of engineers before and after receiving VR-based training.
4	H_0:	There is not a significant difference between the paper-based test results of the construction workers and students after receiving VR-based training.
	H_1:	There is a significant difference between the paper-based test results of the construction workers and students after receiving VR-based training.
5	H_0:	There is not a significant difference between the paper-based test results of the engineers and students after receiving VR-based training.
	H_1:	There is a significant difference between the paper-based test results of the engineers and students after receiving VR-based training.

H-type scaffolding and concrete formwork installation activities were specifically selected because they represent high-risk and frequently performed construction operations that are strongly associated with falls from height and scaffold-related accidents in real construction practice. In addition, these activities involve collaborative and sequential tasks requiring intensive interaction among workers, making them particularly suitable for VR-based safety training scenarios. The tool involves a scenario with 33 basic steps for the installation of H-type scaffolding in line with health and safety requirements. The scenario was developed to highlight the tasks in which occupational accidents could occur in cases of unsafe acts and the instructions for H-type scaffolding and concrete formwork installation. The tool also includes the appropriate visuals of the equipment used in the installation of H-type scaffolding and representative players that users can control to complete the tasks in the virtual environment. There are three roles available in the tool, and up to three users can participate simultaneously. Users can utilize the tool as a team or on their own, depending on their choice, and if they choose to use the tool on their own, the autonomous agents take over the remaining two roles. V-SAFE v.2 consists of two modules, namely the training and testing modules. The basic task sequence of H-type scaffolding and relevant health and safety implementations are covered in the training module. Throughout this module, participants are shown various information boxes that cover safety guidelines that they can follow to complete the training without any accidents. On the other hand, the testing module is designed to evaluate the performance of the participants based on the knowledge they gain from the training module. In both modules, the scenario flow remains the same. However, unlike the training module, the participants are not directed by the information boxes and flashing objects in the testing module. The participants are expected to learn the sequence of the work and the basic health and safety procedures during the training module and to perform the tasks based on what they have learned during the testing module. In case a single or multiple users make a mistake during the testing module, an accident will occur, and all users will fail the module. The simulation is also designed to be as close to reality as possible, which enables participants to experience real accident scenarios.

2.3. Experimental settings

A set of experiments was designed to collect the required data to evaluate the learning outcomes of different participant groups and assess the appropriateness of using students as surrogates for construction workers and engineers in virtual reality–based safety training research. Initially, the participants were categorized into three distinct groups: students, construction workers, and engineers. The studies were conducted using a sample of 20 undergraduate civil engineering students, 20 construction laborers, and 20 engineers. Student experiments were undertaken with a group of senior students in the Civil Engineering Department at Boğaziçi University. On the other hand, the rest of the experiments were carried out in an appropriate construction field. The second group of experiments was arranged at two technical offices on the construction site and involved a total of 20 construction workers. Lastly, an engineering subject group consisting of 20 engineers with mechanical, civil, and environmental engineering backgrounds, working in construction companies, was formed. Half of the experiments with engineers were conducted at the construction sites, and the rest of the experiments were carried out at the office of a construction consultancy company.

The experiment consists of six steps and begins with a concise introduction of V-SAFE v.2 to the participants. The initial two steps comprise a brief questionnaire designed to collect demographic data from participants, such as age, work experience, and role in the construction industry, and a paper-based test consisting of 10 questions is administered to assess participants' baseline knowledge of basic scaffolding installation tasks and associated health and safety protocols. This evaluation is conducted prior to their enrollment in the V-SAFE v.2 training program. In the third step, participants are required to finish the training module of V-SAFE v.2. Subsequently, they are instructed to go to the testing module, which is step four. During the final two steps, participants are instructed to repeat the paper-based test that was administered in the initial stage and are required to fill out a 13-question survey that is based on Kirkpatrick's training evaluation methodology [32] in order to provide feedback on V-SAFE v.2. Fig. 1 illustrates the overall workflow of the experimental methodology described above.

Although V-SAFE v.2 supports both individual use with autonomous agents and collaborative sessions with up to three simultaneous users, all experiments in this study were conducted individually, with each participant completing the training and testing modules independently while the remaining two roles were assumed by autonomous agents. This approach was adopted to eliminate peer interaction and team dynamics as a potential confounding variable, ensuring that the post-test scores reflect each participant's individual learning outcomes rather than the influence of collaborative gameplay. In parallel, the study exclusively considered paper-based pre-test and post-test scores rather than in-system performance metrics such as gameplay logs or correct/incorrect action counts within the VR environment. This choice is grounded in the distinction between learning outcomes and operational performance. The primary objective of this study is to measure learning outcomes, namely, the participants' knowledge acquisition and retention regarding basic scaffolding tasks and associated health and safety procedures. Learning outcomes are conceptually independent of real-time operational data generated inside the application.

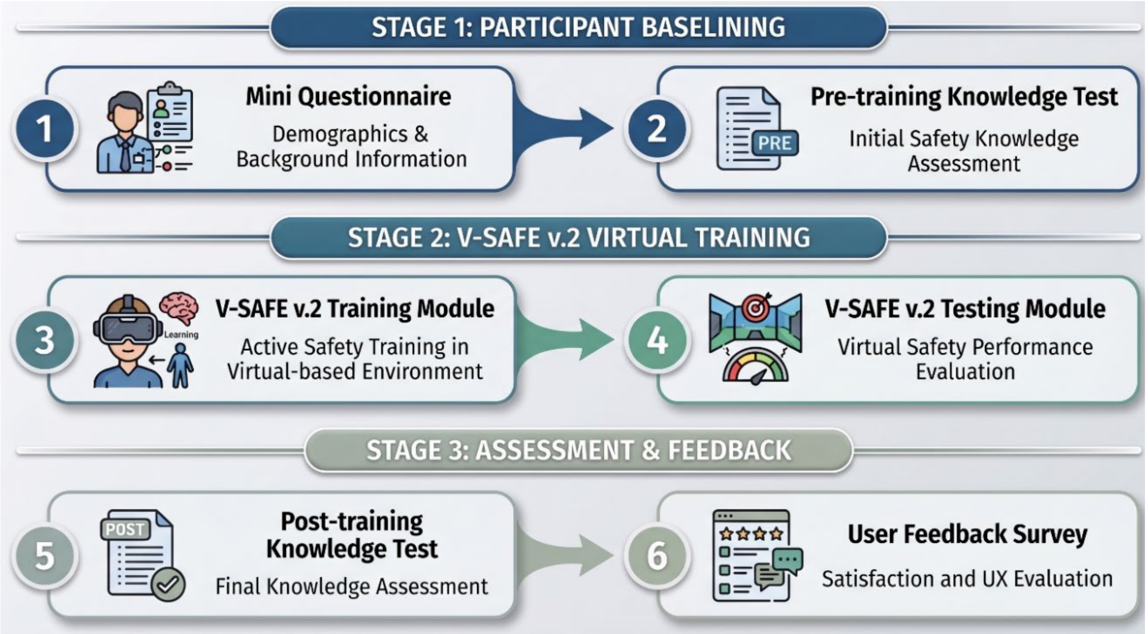

Fig. 1. Overview of the experimental workflow

A participant may make several mistakes or even trigger a simulated accident within the VR environment, yet still achieve a meaningful level of learning through exposure to those failures, since the testing module is designed to reinforce the knowledge gained during training regardless of in-session errors. Therefore, baseline knowledge and post-training retention are more accurately captured by the pre-test/post-test framework than by in-game mistake counters, which reflect momentary task execution rather than the durability of acquired knowledge.

2.4. Analysis methods

SPSS, the Statistical Package for the Social Sciences, is utilized to conduct the necessary statistical analyses in order to evaluate the hypotheses that have been previously established. The attributes of data and variables are the decisive elements for choosing suitable statistical analysis techniques [33]. Three tests are typically examined based on the nature of the data, whether it is parametric or nonparametric. These tests include the paired T-test, Wilcoxon signed-rank test, and the sign test [34]. A review of similar studies shows that paired T-tests are frequently employed to assess if participants' pre- and post-training knowledge differs significantly [10, 28, 35]. The statistical analyses conducted in this study were based exclusively on survey responses and paper-based pre-test/post-test scores, and simulation-generated in-system performance metrics were not included in the evaluation. The paired T-test enables the comparison of observations within the same group of participants, specifically the paired observations, such as the findings before and after an adjustment [34, 36]. The purpose of the test is to examine the disparities between paired observations in order to determine whether there is a noteworthy distinction in their averages [37]. Given that it is a parametric test, it is crucial to ensure that the assumptions of the paired T-test are satisfied prior to conducting the test. (1) The dependent variable must have a minimum level of measurement of interval. (2) The dependent variable should follow a normal distribution. (3) The variances of the samples should not differ significantly [37]. In order to conduct the analysis, the population means for the observations before and after are represented as $\mu 1$ and $\mu 2$, respectively. The difference between these means is indicated as μd, which is calculated as $\mu 1$ - $\mu 2$ [38]. After all the necessary preparations have been made, the paired T-test can be employed to examine the null and alternative hypotheses [38].

Given that Research Questions 1, 2, and 3 seek to assess the efficacy of V-SAFE v.2 on the knowledge of various subject groups before and after virtual training, it is logical to employ paired T-tests with a significance level of 95% to determine if there are any significant differences in test scores pre- and post-training. Upon verifying the assumptions for the paired T-test using SPSS, it becomes evident that only the data for Research Question 1 satisfies the requirements and is appropriate for the paired T-test. Since the test scores recorded on paper were continuous and used as the dependent variable, the initial assumption for the paired T-test was met for all three Research Questions. However, the normality assumption checks reveal that only the data for Research Question 1 satisfy the normality assumption. Since there were 20 observations for each subject group, normality was checked using the Shapiro-Wilk tests. The pre-test scores and post-test scores of students were found to be normally distributed based on the Shapiro-Wilk test results, which had p values greater than 0.05 ($p = 0.162$ for pre-test scores and $p = 0.147$ for post-test scores). However, the data for post-test scores of workers were found not to be normally distributed according to the Shapiro-Wilk test, which has a p value less than 0.05 ($p = 0.0496$). On the other hand, both pre-test scores and post-test scores of engineers were found not to be normally distributed according to the Shapiro-Wilk test, which has a p value less than 0.05 ($p = 0.019$ for pre-test scores and $p = 0.05$ for post-test scores). Although the concept is quite similar to Research Question 1, the paired T-test cannot be utilized for Research Questions 2 and 3, since the normality assumption is violated. Therefore, the Wilcoxon signed-rank test, which is the non-parametric equivalent of the paired T-test [34], seems to be more appropriate for the data for Research Questions 2 and 3.

The Wilcoxon signed-rank test requires fewer assumptions about the shape of the data, but it still assumes that the data are continuous [34] and quantitative [38]. Another assumption for this test is that there are at least six pairs of data [34]. Thus, the data for Research Questions 2 and 3 meet all the assumptions for the Wilcoxon signed-rank test. In contrast to the paired T-test, the Wilcoxon signed-rank test employs the absolute values of the differences between paired observations, ranking these absolute differences and comparing the sum of ranks for positive differences with the sum of ranks for negative differences [38]. The null and alternative hypotheses for this analysis are articulated as:

H_0: The probability of a positive difference is equal to the probability of a negative difference.

H_1: The probability of a positive difference is not equal to the probability of a negative difference.

Lastly, Research Questions 4 and 5 are aimed at testing the appropriateness of using students as surrogates for construction workers and engineers in the development of the virtual reality safety training tool, respectively. A number of other studies explored the use of students as surrogates for construction workers in the development of the safety training tools [8, 39]. For example, Perlman et al. [8] used two-way analysis of variance (two-way ANOVA) to compare the effectiveness of traditional test methods and virtual test methods over the subject groups. In a similar manner, two-factor mixed ANOVA is utilized in our study to test the hypotheses for Research Questions 4 and 5.

Two-factor ANOVA is a robust test and can be used when there are two independent ways of making observations of groups, and there is more than one observation per factor combination [34]. Features of an experiment are referred to as factors, and their actual values are defined as the levels of a factor [36]. Two-factor mixed ANOVA consists of three tests, one test for the interaction effect and two tests for the main effects, which are shown in an ANOVA table as F-tests where a low p-value (usually less than 0.05) is evidence to reject the null hypothesis [38]. Assumptions regarding the data to be checked before applying two-factor ANOVA are: (i) the data set must be continuous; (ii) the data set must be at least approximately normally distributed; (iii) the variation is the same [34].

Since the number of observations is 40 for both pre-test and post-test data, assessment of normality should be done by the analysis of skewness and kurtosis values, which measure symmetry and flatness of the data, respectively [34]. A skewness value of zero indicates that the data set is symmetrical, whereas a positive or negative skewness value indicates a right or left tail, respectively. On the other hand, the kurtosis value indicates how a symmetrical distribution differs from the normal distribution by being either leptokurtic (i.e., there are more observations closer to the mean and in the tails) or platykurtic (i.e., there are fewer observations closer to the mean and tails) [34]. To show that the distribution of the data does not significantly deviate from a normal distribution and is normal with at least 95% confidence, z-values ($Z_{kurtosis}$ and Z_{skew}) are calculated as the division of the kurtosis and skewness values by their respective standard errors (SE_{skew} and $SE_{kurtosis}$) as shown in (1) and (2).

$$Z_{skew} = Skew \,/\, SE_{skew} \tag{1}$$

$$Z_{kurtosis} = Kurtosis \,/\, SE_{kurtosis} \tag{2}$$

When the calculated z-value is within the range of ±1.96, it can be concluded that the distribution is normal within a 95% confidence interval [37]. These calculations will be performed for each dataset prior to conducting a two-factor mixed ANOVA to test the hypotheses related to Research Questions 4 and 5.

3. Results

The data were collected from all 60 participants, i.e., 20 civil engineering students, 20 construction workers, and 20 engineers. Hypotheses for the five Research Questions are tested with appropriate statistical analysis methods. Based on the collected data, relevant statistical analyses are carried out using SPSS.

Initially, the data for Research Question 1 were analyzed. The means (μ) and standard deviations (SD) were calculated as $\mu = 5.80$ and SD = 1.436 for the pre-test scores and $\mu = 8.05$ and SD = 1.146 for the post-test scores of students. The test results for Research Question 1, i.e., the paired T-test results for pre-test and post-test scores of students, are given in Table 2. The table includes statistics for paired differences such as the mean, standard deviation, the paired T-test result, the degrees of freedom, and the p-value. The mean difference between pairs of data (pre-test score and post-test score) is calculated as -2.25. The t-test statistic is -7.336. The significance value approximates 0, which is smaller than 0.05, meaning that there is a significant difference between the pre-test scores and post-test scores of the student subject group, and thus the null hypothesis of the paired T-test is rejected.

Using the data set for Research Question 2, the means (μ) and standard deviations (SD) were calculated as $\mu = 6.35$ and SD = 1.631 for the pre-test scores and $\mu = 7.85$ and SD = 1.387 for the post-test scores of construction workers. Wilcoxon signed-rank test results for Research Question 2 are given in Table 3. There are no cases of “Negative Ranks”, meaning that none of the post-test scores is less than the pre-test scores. “Positive Ranks”, representing that the post-test score is greater than the pre-test score, were observed in 16 cases.

Table 2. Paired T-test results for the student subject group's pre- and post-test results

	Paired Difference				
	Mean	Std. Deviation	t	df	Sig. (t-tailed)
Pre-Test Scores of Students - Post-Test Scores of Students	-2.25	1.371	-7.336	19	0.000

Table 3. Wilcoxon signed-rank test for Research Question 2

	Ranks	N	Mean Rank	Sum of Ranks
Pre-Test Scores of Workers - Post-Test Scores of Workers	Negative Ranks	0	0.00	0.00
	Positive Ranks	16	8.50	136.00
	Ties	4	-	-
	Total	20	-	-

Lastly, in 4 cases, post-test scores are equal to pre-test scores, resulting in a tie. Z-value and "Asymp. Sig. (2-tailed)" value, which represents the p-value [34], can be seen in Table 4. Since the p-value approximates zero, which is less than 0.05, the null hypothesis is rejected.

Table 4. Test statistics for paper-based test scores of construction workers

	Pre-Test Scores of Workers - Post-Test Scores of Workers
Z	-3.581
Asymp. Sig. (2-tailed)	0.000

Similarly, for Research Question 3, the means (μ) and standard deviations (SD) were calculated as $\mu = 7.35$ and SD = 0.988 for the pre-test scores and $\mu = 8.90$ and SD = 0.788 for the post-test scores of engineers. Wilcoxon signed-rank test results for Research Question 3 are given in Table 5. Z-value and "Asymp. Sig. (2-tailed)" value for Research Question 3 can be seen in Table 6. Since the p-value approximates zero and is less than 0.05, the null hypothesis is rejected.

Table 6. Test statistics for paper-based test scores of engineers

	Pre-Test Scores of Engineers - Post-Test Scores of Engineers
Z	-3.794
Asymp. Sig. (2-tailed)	0.000

To test Research Question 4, the within-subject factor is assigned as "Performance" with two levels, pre-test scores and post-test scores. Similarly, the between-subject factor is assigned as "Subject_Type" with two levels, which are students and workers. Table 7 gives a summary of the descriptive statistics of the data that will be used to test Research Question 4. Since there are fewer than three groups, post-hoc tests are not performed for subject type.

Before applying two-factor mixed ANOVA, the assumptions regarding the data are checked first. Skewness value and the standard error of Skewness (SE_{skew}) are calculated for pre-test and post-test score data using SPSS, and these values are plugged into (1). Z_{skew} values are calculated as 0.59 for pre-test score data and 0.045 for post-test score data, which are within the specified range of ±1.96 with 95% confidence.

Table 5. Wilcoxon signed-rank test for Research Question 3

	Ranks	N	Mean Rank	Sum of Ranks
Pre-Test Scores of Engineers - Post-Test Scores of Engineers	Negative Ranks	0	0.00	0.00
	Positive Ranks	18	9.50	171.00
	Ties	2	-	-
	Total	20	-	-

Table 7. Descriptive statistics of paper-based test scores used for Research Question 4

Variable	Subject Type	Mean	Std. Deviation	N
Pre-Test Score	Students	5.80	1.436	20
	Workers	6.35	1.631	20
	Total	6.08	1.542	40
Post-Test Score	Students	8.05	1.146	20
	Workers	7.85	1.387	20
	Total	7.95	1.260	40

In a similar manner, the Kurtosis values and the standard error values of Kurtosis ($SE_{kurtosis}$) for pre-test and post-test score data are generated in SPSS, and using (2), $Z_{kurtosis}$ values are calculated as -0.556 for pre-test score data and -1.296 for post-test score data. Both values are within the specified range of ±1.96 with 95% confidence. Lastly, to check the homogeneity of variance assumption for the data, Levene's Statistic is used for mixed ANOVA, which should be assessed at all levels of the within-subject variable. P-values generated in SPSS are greater than 0.05, meaning that the homogeneity of variance assumption is met. Thus, all assumptions are satisfied, and two-factor mixed ANOVA can be applied to test Research Question 4.

Table 8 shows the ANOVA results for the within-subject variable (performance) and any interactions with that variable. The values in the "Sphericity Assumed" rows are considered only. The p-value for the within-subject variable, performance, is less than 0.05, indicating that performance has a significant effect. On the other hand, the p-value for the interaction effect is greater than 0.05, meaning that the interaction effect is not significant. Table 9 demonstrates the ANOVA statistics for the between-subject variable (subject type). Since the p-value is greater than 0.05, there is no significant effect of subject type on the test results.

Table 8. Test results of performance effect in Research Question 4

	Sphericity Assumed				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Performance	70.312	1	70.312	97.603	0.000
Performance * Subject_Type	2.813	1	2.813	3.904	0.055
Error (Performance)	27.375	38	0.720	-	-

Similarly, for Research Question 5, "Subject_Type" is assigned as the between-subject factor with two levels, students and engineers, and "Performance" is assigned as the within-subject factor, also with two levels, pre-test scores and post-test scores. Table 10 gives a summary of the descriptive statistics of the data that will be used to test Research Question 5. Since there are fewer than three groups, post-hoc tests are not performed for subject type.

To check the assumptions of two-factor mixed ANOVA regarding the data, Skewness values, the standard error values of Skewness (SE_{skew}), Kurtosis values, and the standard error values of Kurtosis ($SE_{kurtosis}$) are calculated for pre-test and post-test score data using SPSS. These values are plugged into (1) and (2). Z_{skew} values are calculated as -0.56 for pre-test score data and -1.44 for post-test score data, which are within the specified range of ±1.96 with 95% confidence. On the other hand, $Z_{kurtosis}$ values are calculated as -0.24 for pre-test score data and -0.16 for post-test score data, which are also within the specified range of ±1.96 with 95% confidence. Lastly, the homogeneity of variance assumption for the data is checked using Levene's Statistic. The assessment is made at all levels of the within-subject variable. p-values generated in SPSS are greater than 0.05, meaning that the homogeneity of variance assumption is met. Thus, all assumptions are satisfied, and two-factor mixed ANOVA can be applied to test Research Question 5.

Table 11 shows the ANOVA results for the within-subjects variable (performance) and any interactions with that variable. The values in the "Sphericity Assumed" rows are considered only. The p-value for the within-subject variable, performance, is less than 0.05, showing that performance has a significant effect.

Table 11. Test results of performance effect in Research Question 5

	Sphericity Assumed				
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Performance	72.200	1	72.200	100.314	0.000
Performance * Subject_Type	2.450	1	2.450	3.404	0.073
Error (Performance)	27.350	38	0.720	-	-

Table 9. Test results of subject type effect in Research Question 4

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	3934.012	1	3934.012	1206.801	0.000
Subject_Type	0.612	1	0.612	0.188	0.667
Error	123.875	38	123.875	-	-

Table 10. Descriptive statistics of paper-based test scores used for Research Question 5

Variable	Subject Type	Mean	Std. Deviation	N
Pre-Test Score	Students	5.80	1.436	20
	Engineers	7.35	0.988	20
	Total	6.58	1.448	40
Post-Test Score	Students	8.05	1.146	20
	Engineers	8.90	0.788	20
	Total	8.48	1.062	40

However, the p-value for the interaction effect is greater than 0.05, indicating that the interaction effect is not significant. Table 12 portrays the ANOVA statistics for the between-subject variable (subject type). There is a significant main effect of subject type on the test results because the p-value is smaller than 0.05.

4. Discussion

Despite the increasing use of virtual environments in construction safety training, questions remain about the appropriateness of using students as surrogates for construction workers and engineers during the development phase of these tools. While students offer a convenient and accessible sample group, their representativeness in terms of learning outcomes and safety performance has yet to be validated. This study addresses this critical gap by evaluating whether student performance meaningfully reflects that of real-world counterparts when exposed to a virtual safety training tool—V-SAFE v.2. The following discussion explores the results through this perspective.

Research Questions 1, 2, and 3 investigate the effectiveness of V-SAFE v.2 on the learning outcomes of three subject groups regarding the topics covered in its training module. The data used to address these Research Questions consist of pre-test and post-test scores obtained by participants in each group through a paper-based test. This test includes 10 questions related to basic tasks in scaffolding installation and associated health and safety procedures, all of which are covered in the V-SAFE v.2 training module. Pre-test scores reflect the number of questions participants answered correctly before receiving the V-SAFE v.2 training. The purpose of the pre-test is to assess each participant's initial knowledge of the topics addressed in the training module. Post-test scores, on the other hand, represent the number of correct responses after the training, serving to measure learning gains following exposure to the V-SAFE v.2 tool.

Pre-test and post-test scores were analyzed in SPSS using appropriate statistical methods, and the results were interpreted accordingly for each research question. For Research Question 1, 20 civil engineering students from Boğaziçi University participated in the experiments, and we collected their pre- and post-test scores. The mean pre-test score was 5.80, while the mean post-test score increased to 8.05, demonstrating a 38.79% improvement in performance on the paper-based test after completing the V-SAFE v.2 training module. According to the paired t-test results, the difference between pre- and post-test scores for students ($t = -7.336$, $p < 0.05$) was statistically significant. This finding indicates that the V-SAFE v.2 training module had a significant positive effect on students' learning of basic scaffolding knowledge and related health and safety issues. For Research Question 2, we investigated the effectiveness of V-SAFE v.2 on the learning outcomes of the construction worker subject group. We conducted the experiments with 20 construction workers from two construction sites. The mean pre-test score was 6.35, while the mean post-test score rose to 7.85, indicating a 23.62% improvement in their performance on the paper-based test following the training. The Wilcoxon signed-rank test showed a statistically significant difference between pre- and post-test scores ($Z = -3.581$, $p < 0.05$). This result suggests that the V-SAFE v.2 training module significantly enhanced construction workers' understanding of basic scaffolding tasks and related health and safety procedures. For Research Question 3, we collected data from 20 engineers employed at two construction companies. Their average post-test score increased to 8.90 from a pre-test average of 7.35, reflecting a 21.09% improvement in performance. According to the Wilcoxon signed-rank test, the difference in scores was statistically significant ($Z = -3.794$, $p < 0.05$). This result further confirms that the V-SAFE v.2 training module positively impacted engineers' learning of scaffolding-related knowledge and health and safety issues. Based on the results of the first three research questions, we conclude that V-SAFE v.2 is an effective virtual safety training tool for improving the learning outcomes of students, construction workers, and engineers on basic scaffolding tasks and associated health and safety practices.

Similar studies examining the effectiveness of VR-based training showed similar findings [28, 29, 35, 40]. Pedro et al. [28] evaluated the students' performances on the developed virtual construction safety education system by comparing the paper-based test results of students before and after training. They concluded that the developed virtual construction safety education system had a positive impact on improving students' safety knowledge and hazard identification abilities [28]. Adami et al. [29] compared the knowledge assessment ratings of construction workers before and after attending their VR-based robotic teleoperation training tool and found that the proposed tool significantly increased the knowledge levels of construction workers. Similarly, Joshi et al. [35] compared the knowledge gain of participants before and after completing their VR-based precast/prestressed concrete training tool and underlined that the proposed tool significantly increased participants' performance.

Table 12. Test results of subject type effect in Research Question 5

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	4530.05	1	4530.050	2563.540	0.000
Subject_Type	28.80	1	28.800	16.300	0.000
Error	67.15	38	1.770	-	-

Lastly, Guo et al. [40] compared the effectiveness of VR-based and traditional safety training methods for different learner types and concluded that although their effectiveness varies depending on the learning style of the participant, VR training caused a great improvement in hazard identification performance. Based on the results of our study and other similar studies, the effectiveness of virtual safety training tools in increasing the safety knowledge and skills of students, workers, and engineers has been validated in many studies.

The appropriateness of using students as surrogates in the development of virtual safety training tools was investigated in Research Questions 4 and 5. In Research Question 4, the subject groups were civil engineering students and construction workers, and the aim was to investigate whether students are appropriate surrogates for construction workers in the development of virtual reality safety training tools. According to the ANOVA analysis, the within-subject variable called "Performance" had a significant effect on the paper-based test results ($F = 97.603$; $p < 0.05$). This indicates that the pre-test and post-test results of the two subject groups were significantly different, which proves the effectiveness of V-SAFE v.2. On the other hand, a two-factor mixed ANOVA analysis revealed that there was no significant main effect of the between-subjects variable called "Subject Type" on the test results ($F = 0.188$; $p > 0.05$). In other words, the two subject groups' safety performances were not significantly different. This result indicates that students and construction workers showed similar safety performances on the paper-based tests.

For Research Question 5, the safety performances of civil engineering students were compared with the safety performances of engineers based on their paper-based test results, with the aim of investigating the appropriateness of using students as surrogates for engineers in the development of virtual reality safety training tools. According to the two-factor mixed ANOVA results, the within-subject variable, "Performance," had a significant effect on the paper-based test results ($F = 100.314$; $p < 0.05$), again revealing the effectiveness of V-SAFE v.2. However, the two-factor mixed ANOVA analysis showed that there was a significant main effect of the between-subjects variable, which is "Subject Type," on the test results ($F = 16.298$; $p < 0.05$). Looking at the two-way mixed ANOVA results for Research Questions 4 and 5, it is observed that students can be used as surrogates for construction workers in the development of virtual safety training tools, but not for engineers. The safety performances of students, construction workers, and engineers in pre-test and post-test scores are summarized in Fig. 1.

The slopes of the lines presented in Fig. 2 can be compared to examine the three subject groups' learning outcomes. For the construction worker and engineer subject groups, the slopes were found to be 1.5 and 1.55x, respectively, while it was found to be 2.25x for the student subject group. The steepest line belongs to the student subject group, which indicates that the learning capacity of students is higher compared to that of the construction workers and engineers. Lokka and Çöltekin [41] also studied the effect of age on learning and spatial knowledge acquisition in virtual environments by using younger and older participants in their experiments and found that the younger participants performed better in virtual environments. Since the student subject group consists of younger participants compared to the engineer and construction worker subject groups, the trends in Fig. 1 are in line with the study by Lokka and Çöltekin [41]. In addition to this, the student subject group is more familiar with paper-based tests and the multiple-choice question format. Therefore, they might find some of the correct answers by eliminating other choices. As a result, their learning trend ended up with the steepest line among the others.

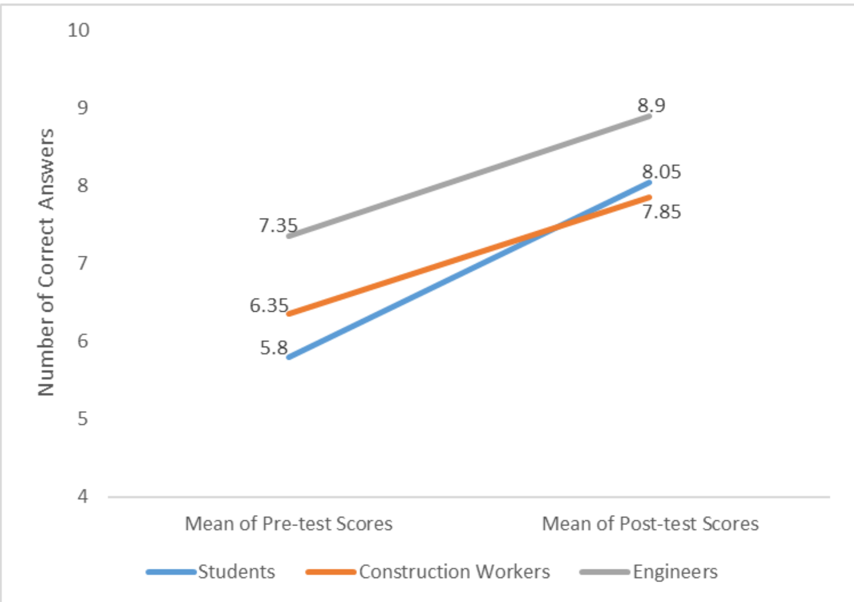

Fig. 2. Overview of the experimental

Numerous construction studies have utilized students as surrogates for construction workers in the development of virtual training tools [18, 21-23, 25, 26], yet the appropriateness of this practice has not been thoroughly examined. For example, Perlman et al. [8] studied the impact of work experience on the hazard recognition skills of construction superintendents by conducting experiments with civil engineering students with no work experience, construction superintendents, and safety managers, and by comparing the number of correctly identified hazards by each subject group in the virtual environment using a two-way between-subject ANOVA test. Even though their aim was considerably different from the aim of this study, their findings are remarkably similar to ours. In their study, almost all safety managers (85.7%) and construction superintendents (92.8%) reported that they had received safety instruction prior to the study, while only 34.7% of the students had received safety instructions before. However, it was found that students and construction superintendents had comparable performances in hazard identification tasks, while the hazard recognition performance of safety managers and students was not similar [8]. They proposed two possible reasons for these findings: (i) safety managers had taken more safety training than the construction superintendents and students, or (ii) since their main role is to maintain construction safety, they performed better [8]. In addition, even though construction superintendents had also received safety instructions beforehand, their performance was associated with their safety culture [8]. Thus, safety culture is another influential factor when comparing the safety performances of construction workers and engineers. Similarly, the reason behind the performance difference between the students and engineers in this study is that the engineer subject group consists of engineers who have had a number of safety trainings and who possess comprehensive knowledge of health and safety implementations prior to this study. The difference mainly originated from the fact that engineers already had better performance on the paper-based test before receiving V-SAFE v.2, compared to the students.

In another study conducted by Trottier and Gordon [27] in the accounting research area, the appropriateness of using students as surrogates for managers was examined, and an extensive literature review on this issue was conducted. Their study also found that there was no robust conclusion on the appropriateness of students as surrogates for managers, and they found significant differences between the responses of students and managers [27]. However, there were similarities between the responses of students with work experience and those of the managers, and they suggested that the characteristics of students (e.g., work experience, gender, grade, or knowledge) should be taken into consideration [27].

The results indicate that the appropriateness of using students as surrogates depends on the research area, research scope, and the demographic information of the participants in the experiments, meaning that the results might change depending on the country and the country's safety training implementations. In the case of this study, the experiments were conducted in Turkey. Only 16% of workers in Turkey reported that they had received training paid for or provided by their companies over the last 12 months, whereas the percentages were 45% and 41% in the US and Europe, respectively [3]. On top of that, only 15% of workers in Turkey had been offered on-the-job training, while 59% of workers in the US and 34% of workers in Europe had received on-the-job training [3].

Based on these percentages, it can be said that the initial safety performances of construction workers before receiving the virtual safety training might have been better compared to the safety performances of construction workers observed in this study if the experiments had been conducted in the US or Europe. Nevertheless, the results of this study indicate that V-SAFE v.2 had a significant positive impact on the three subject groups' knowledge levels regarding basic scaffolding installation tasks and related health and safety implementations. Additionally, students were found to be suitable surrogates for construction workers in the development of virtual safety training tools, but they might not be appropriate surrogates for engineers.

5. Conclusions

This study examined the effectiveness of a VR-based safety training tool, V-SAFE v.2, and assessed the suitability of using students as surrogates for construction workers and engineers during the development of similar tools. Experimental results confirmed that the tool significantly enhanced the safety knowledge of all participant groups, namely, students, construction workers, and engineers, highlighting the potential of VR tools in construction safety training.

The study addresses a critical gap in the literature by testing the validity of using students as surrogates. The findings showed that while students performed comparably to construction workers before and after the training, their performance remained significantly different from that of engineers. This distinction suggests that students can be suitable surrogates for construction workers but not for engineers, whose prior professional knowledge plays a critical role in safety performance assessments during the development of VR-based training tools. These findings may also contribute to the digital transformation of the construction industry by supporting the more efficient development and evaluation of VR-based safety training systems. Demonstrating the potential use of students as surrogate participants in preliminary VR-based safety research may help reduce practical barriers associated with accessing industry professionals, thereby accelerating the testing and refinement of emerging digital training technologies. The results also demonstrated that V-SAFE v.2 was suitable and effective across groups with varying backgrounds, including students with no real construction site experience, workers with limited exposure to digital training tools, and engineers with prior theoretical knowledge. This supports the broader adoption of VR-based safety tools in real-world construction contexts.

Despite these contributions, the study is limited by its sample size and geographic scope. The participant pool consisted of 20 construction workers, 20 engineers, and 20 students recruited from two construction companies and one university in Turkey. Although this sample size is comparable to those used in many prior VR-based construction safety training studies and was sufficient for the statistical analyses conducted in this research, the relatively limited number of participants and the single-country context may restrict the generalizability of the findings to broader construction industry populations with different demographic, educational, and occupational characteristics. Furthermore, the training tool focused solely on basic scaffolding tasks. Future research should aim to validate these findings across diverse settings and expand VR content to cover broader safety topics. Furthermore, since students were found to be inappropriate surrogates for engineers in this study, future research directions can be narrowed accordingly. Specifically, larger-scale trials are recommended to focus exclusively on the student-worker dyad in order to further validate this surrogate relationship across different sample sizes, geographic contexts, and VR-based training tools. In conclusion, this study provides empirical support for the use of VR in construction safety training and offers evidence-based guidance on when student surrogates may or may not be appropriate for tool development.

Declarations

Conflict of Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK) under grant number(s) 315M186.

Author Contributions

E. Gürbüz Taştekin: Methodology, Investigation, Formal analysis, Writing-Original Draft. U. Kılıç: Writing-Original Draft, Writing-Review & Editing, Visualization. S. Çomu: Conceptualization, Resources, Project administration, Supervision, Methodology, Writing-Review & Editing.

Acknowledgments

Not applicable.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request, subject to restrictions.

Ethics Committee Permission

The authors received ethical approval for this study from the Boğaziçi University Human Research Ethics Committee for Science and Engineering (FMINAREK) on August 17, 2015 (registration number 2015/04).

Use of Generative AI and AI-assisted Technologies

The author(s) used OpenAI GPT-4 and GPT-5 models for language polishing and proofreading and have reviewed and take full responsibility for the final content.

References

[1] Dogbla L, Gouvenelle C, Thorin F, Lesage F-X, Zak M, Ugbolue UC, Charbotel B, Baker JS, Pereira B, Dutheil F (2023) Occupational risk factors by sectors: An observational study of 20,000 workers. Int J Environ Res Public Health 20(4):3632. https://doi.org/10.3390/ijerph20043632.

[2] Eurostat (2024) Accidents at work statistics. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Accidents_at_work_statistics. Accessed 25 Feb 2026.

[3] Eurofound, International Labour Organization (2019) Working conditions in a global perspective. Publications Office of the European Union, Luxembourg, and International Labour Organization, Geneva.

[4] Akinlolu M, Haupt TC, Edwards DJ, Simpeh F (2020) A bibliometric review of the status and emerging research trends in construction safety management technologies. Int J Constr Manage 22(14):2699–2711.

[5] Manzoor B, Othman I, Waheed A (2022) Accidental safety factors and prevention techniques for high-rise building projects – A review. Ain Shams Eng J 13(5):101723.

[6] Hwang J-M, Won J-H, Jeong H-J, Shin S-H (2023) Identifying critical factors and trends leading to fatal accidents in small-scale construction sites in Korea. Buildings 13(10):2472. https://doi.org/10.3390/buildings13102472.

[7] Shuang Q, Zhang Z (2023) Determining critical cause combination of fatality accidents on construction sites with machine learning techniques. Buildings 13(2):345. https://doi.org/10.3390/buildings13020345.

[8] Perlman A, Sacks R, Barak R (2014) Hazard recognition and risk perception in construction. Safety Sci 64:22–31. https://doi.org/10.1016/j.ssci.2013.11.019.

[9] Namian M, Albert A, Zuluaga CM, Behm M (2016) Role of safety training: Impact on hazard recognition and safety risk perception. J Constr Eng Manage 142(12):04015011. https://doi.org/10.1061/(asce)co.1943-7862.0001198.

[10] Sacks R, Perlman A, Barak R (2013) Construction safety training using immersive virtual reality. Constr Manage Econ 31(9):1005–1017. https://doi.org/10.1080/01446193.2013.828844.

[11] Umeokafor N, Evangelinos K, Windapo A (2020) Strategies for improving complex construction health and safety regulatory environments. Int J Constr Manage 22(7):1333–1344. https://doi.org/10.1080/15623599.2019.1707853.

[12] Mohammadi A, Tavakolan M, Khosravi Y (2018) Factors influencing safety performance on construction projects: A review. Safety Sci 109:382–397. https://doi.org/10.1016/j.ssci.2018.06.017.

[13] Xie H, Tudoreanu E, Shi W (2006) Development of a virtual reality safety-training system for construction workers. In: Digital Library of Construction Informatics and Information Technology in Civil Engineering and Construction, pp. 1–9.

[14] Bęś P, Strzałkowski P (2024) Analysis of the effectiveness of safety training methods. Sustainability 16(7):2732. https://doi.org/10.3390/su16072732.

[15] Nykänen M, Puro V, Tiikkaja M, Kannisto H, Lantto E, Simpura F, Uusitalo J, Lukander K, Räsänen T, Heikkilä T, Teperi A-M (2020) Implementing and evaluating novel safety training methods for construction sector workers: Results of a randomized controlled trial. J Safety Res 75:205–221. https://doi.org/10.1016/j.jsr.2020.09.015.

[16] Mohd NI, Ali KN, Bandi S, Ismail F (2019) Exploring gamification approach in hazard identification training for Malaysian construction industry. Int J Built Environ Sustainability 6(1):51–57. https://doi.org/10.11113/ijbes.v6.n1.333.

[17] Burke MJ, Sarpy SA, Smith-Crowe K, Chan-Serafin S, Salvador RO, Islam G (2006) Relative effectiveness of worker safety and health training methods. Am J Public Health 96(2):315–324. https://doi.org/10.2105/ajph.2004.059840.

[18] Becker P, Morawetz J (2004) Impacts of health and safety education: Comparison of worker activities before and after training. Am J Ind Med 46(1):63–70. https://doi.org/10.1002/ajim.20034.

[19] Marougkas A, Troussas C, Krouska A, Sgouropoulou C (2023) How personalized and effective is immersive virtual reality in education? A systematic literature review for the last decade. Multimedia Tools Appl 83:18185–18233. https://doi.org/10.1007/s11042-023-15986-7.

[20] Scorgie D, Feng Z, Paes D, Parisi F, Yiu TW, Lovreglio R (2024) Virtual reality for safety training: A systematic literature review and meta-analysis. Safety Sci 171:106372. https://doi.org/10.1016/j.ssci.2023.106372.

[21] Irizarry J, Abraham DM (2005) Application of virtual reality technology for the improvement of safety in the steel erection process. In: Computing in Civil Engineering (2005), pp. 1–11. https://doi.org/10.1061/40794(179)148.

[22] Shi Y, Du J, Ragan E, Choi K, Ma S (2018) Social influence on construction safety behaviors: A multi-user virtual reality experiment. In: Construction Research Congress 2018, pp. 174–183. https://doi.org/10.1061/9780784481288.018.

[23] Li W, Huang H, Solomon T, Esmaeili B, Yu L-F (2022) Synthesizing personalized construction safety training scenarios for VR training. IEEE Trans Visual Comput Graphics 28(5):1993–2002. https://doi.org/10.1109/tvcg.2022.3150510.

[24] Shringi A, Arashpour M, Golafshani EM, Rajabifard A, Dwyer T, Li H (2022) Efficiency of VR-based safety training for construction equipment: Hazard recognition in heavy machinery operations. Buildings 12(12):2084. https://doi.org/10.3390/buildings12122084.

[25] Hussain R, Zaidi SF, Pedro A, Lee H, Park C (2024) Exploring construction workers’ attention and awareness in diverse virtual hazard scenarios to prevent struck-by accidents. Safety Sci 175:106526. https://doi.org/10.1016/j.ssci.2024.106526.

[26] Szóstak M, Mahamadu A-M, Prabhakaran A, Caparros Pérez D, Agyekum K (2024) Development and testing of immersive virtual reality environment for safe unmanned aerial vehicle usage in construction scenarios. Safety Sci 176:106547. https://doi.org/10.1016/j.ssci.2024.106547.

[27] Trottier K, Gordon IM (2016) Students as surrogates for managers: Evidence from a replicated experiment. Can J Adm Sci 35(1):146–161. https://doi.org/10.1002/cjas.1377.

[28] Pedro A, Le QT, Park CS (2016) Framework for integrating safety into construction methods education through interactive virtual reality. J Prof Issues Eng Educ Pract 142(2):04015011.

[29] Adami P, Rodrigues PB, Woods PJ, Becerik-Gerber B, Soibelman L, Copur-Gencturk Y, Lucas G (2021) Effectiveness of VR-based training on improving construction workers’ knowledge, skills, and safety behavior in robotic teleoperation. Adv Eng Inform 50:101431. https://doi.org/10.1016/j.aei.2021.101431.

[30] Song Y, Arnott D, Gao S (2020) Graduate students as surrogates for managers in business intelligence and analytics research: A preliminary study. J Decision Syst 31(3):280–304. https://doi.org/10.1080/12460125.2020.1838037.

[31] Kazar G, Çomu S (2022) Developing a virtual safety training tool for scaffolding and formwork activities. Teknik Dergi 33:11729–11748. https://doi.org/10.18400/tekderg.711091.

[32] Kirkpatrick DL (1998) The four levels of evaluation. In: Brown SM, Seidner CJ (eds) Evaluating Corporate Training: Models and Issues, pp. 95–112. https://doi.org/10.1007/978-94-011-4850-4_5.

[33] Murrar S, Brauer M (2018) Mixed model analysis of variance. In: The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation, pp. 1075–1078.

[34] Dytham C (2011) Choosing and Using Statistics: A Biologist’s Guide, 3rd Edition. Wiley-Blackwell.

[35] Joshi S, Hamilton M, Warren R, Faucett D, Tian W, Wang Y, Ma J (2021) Implementing virtual reality technology for safety training in the precast/prestressed concrete industry. Appl Ergonomics 90:103286. https://doi.org/10.1016/j.apergo.2020.103286.

[36] Walpole RE, Myers RH, Myers SL, Ye K (2012) Probability and Statistics for Engineers and Scientists, 9th Edition. Prentice Hall, New York.

[37] Kerr AW, Kozub SA, Hall HK (2014) Doing Statistics with SPSS. SAGE Publications Ltd.

[38] Elliott AC, Woodward WA (2012) Statistical Analysis Quick Reference Guidebook: With SPSS Examples. SAGE.

[39] Guo H, Li H, Chan G, Skitmore M (2012) Using game technologies to improve the safety of construction plant operations. Accident Anal Prev 48:204–213. https://doi.org/10.1016/j.aap.2011.06.002.

[40] Guo X, Liu Y, Tan Y, Xia Z, Fu H (2024) Hazard identification performance comparison between virtual reality and traditional construction safety training modes for different learning style individuals. Safety Sci 180:106644. https://doi.org/10.1016/j.ssci.2024.106644.

[41] Lokka IE, Çöltekin A (2019) Perspective switch and spatial knowledge acquisition: Effects of age, mental rotation ability and visuospatial memory capacity on route learning in virtual environments with different levels of realism. Cartogr Geogr Inf Sci 47(1):14–27. https://doi.org/10.1080/15230406.2019.1595151.