Speiser, K., Golovina, O., Teizer, J. (2025). Behavioral safety performance assessment in construction: a data-driven approach in virtual reality. In O. Seppänen, L. Koskela, & K. Murata (Eds.), *Proceedings of the 33rd Annual Conference of the International Group for Lean Construction* (IGLC33) (pp. 1114–1125). https://doi.org/10.24928/2025/0169

BEHAVIORAL SAFETY PERFORMANCE ASSESSMENT IN CONSTRUCTION: A DATA-DRIVEN APPROACH IN VIRTUAL REALITY

Kilian Speiser[1], Olga Golovina[2], and Jochen Teizer[3]

ABSTRACT

The construction industry suffers from high incidence rates compared to other sectors. Lean and safety training is one effort to increase work productivity while preventing accidents, but the effectiveness of various existing training methods relies on intention-based evaluations. These can fail to capture behavioral lean and safety performance. This paper presents a novel Virtual Reality (VR)-based assessment method for quantifying (lean and) safety performance in construction, addressing the critical gap between intention and behavior. By simulating real-world hazards, a simplified VR learning environment allows for objective, data-driven assessment of workers' behavior using metrics that capture, for example, the interaction with poor construction site layout and hazards. This study created a virtual environment and tested the capabilities of objectively assessing aspects of lean and safety performance in an experiment with 60 subjects. Splitting them into groups of experts and novices and exposing them to different training methods allowed a qualitative discussion on the implications of such an assessment method. The results demonstrate its applicability for a data-driven metric to assess lean and safety behavior, given that the environment includes relevant hazards and represents lifelike tasks. Construct validity requires further considerations when using this concept for virtual performance assessment. A more nuanced metric could further strengthen the study design. This study provides an explorative framework for further research to investigate the effectiveness of training methods based on behavior and overcoming the intention-behavior gap.

KEYWORDS

Active learning, behavioral assessment, education and training, lean and safe construction, safety performance, virtual reality, lean and safety indicators.

INTRODUCTION

Construction sites are inherently dynamic and high-risk environments where poor productivity and incidents occur more frequently than in other working environments (Eurostat, 2024). Empirical studies have shown that safe and lean construction go hand-in-hand, and respective concepts are mutually beneficial to each field (Li et al., 2020). A central aspect of lean construction is the reduction of waste. Originally defined in seven types (Ohno, 1988), recent literature identifies the following eight types of waste: defects, overproduction, Idling, not utilized talent, transportation, inventory, motion, and extra processing. Construction safety management can contribute to reducing some of these waste types. Adequate safety plans and

[1] Ph.D. Candidate, Department for Civil and Mechanical Engineering, Technical University of Denmark (DTU), Kongens Lyngby, Denmark, kilsp@dtu.dk, https://orcid.org/0000-0001-8428-8053

[2] Ph.D., Consultant, Kongens Lyngby, Denmark, kj7_i@web.de, https://orcid.org/0009-0009-5021-3139

[3] Professor, Department for Civil and Mechanical Engineering, Technical University of Denmark (DTU), Kongens Lyngby, Denmark, teizerj@dtu.dk, https://orcid.org/0000-0001-8071-895X

safe construction site layouts can decrease unnecessary motion and transport of workers. Reduced incidents as a consequence of effective safety training decrease the waste of unutilized talent. Moreover, effective safety training can enhance understanding of lean principles and reduce waste as a consequence of more awareness among the workforce. However, different training methods often do not result in a transfer of learning (Pham et al., 2023).

Traditional training methods, such as video tutorials, classroom lectures, and toolbox talks, are decoupled from lean construction concepts. They mostly focus on raising hazard awareness and improving knowledge. These methods focus on training intentions - what workers plan to do in theory - rather than their behavioral safety and lean performance - what they do in practice (Albert et al., 2014). This intention-behavior gap (Ajzen, 1985) poses a significant challenge, as effective, safe practices require workers to transfer intentions into real-world actions. Workers may demonstrate an understanding of safety practices in assessments but struggle to translate these intentions into safe behaviors on-site, creating an intention-behavior gap that remains a critical challenge in construction safety training and planning (Johansen et al., 2025).

The inability to measure behavioral safety performance becomes especially problematic when evaluating the effectiveness of different training methods. In particular, safety performance relies on workers' ability to recognize and respond to hazards. Several efforts compare the effectiveness of different training methods, often suggesting that active training is more effective (Burke et al., 2006). However, such studies often rely on intention-based measures, such as tests or interviews (Salinas et al., 2022; Adami et al., 2023). For instance, a worker may correctly state in a test that they would use protective equipment, but intention-based assessments cannot verify if the same worker adheres to this practice during actual tasks. Without directly observing and quantifying behavior, the true impact of training methods on safety performance remains unclear. The same thought relates to the impact on lean behavior.

Simulated environments like in virtual reality (VR) offer a promising solution. While VR has been studied for safety (Salinas et al., 2022; Abich et al., 2021; Zoleykani et al., 2023) and lean (Teizer & Melzner, 2019; Jacobsen et al., 2021; Wolf et al., 2022) training, the inherent characteristics that enable it for training may also make it suitable for behavior-based performance assessment. Such VR environments enable the creation of lifelike construction environments where workers can interact in suboptimal construction workplaces with hazards without real-world risks. Compared to on-the-job assessments, VR environments collect objective data that could be used to quantify safety and lean performance. Such a direct measurement of performance can overcome the intention-behavior gap.

This paper proposes a novel VR-based assessment method for quantifying (lean and) safety behavior. Compared to existing research, the novelty of this research lies in simulating real-world working conditions in VR to assess (primarily) the safety performance of workers rather than train them. The VR environment enables us to expose workers to hazards without posing them with actual risks. At the same time, we can observe how workers behave in certain situations and how they interact with suboptimal work site layouts and hazards. The VR environment furthermore provides functionality to quantify the performance of workers based on run-time data collection and analysis.

RELATED WORK

PASSIVE AND ACTIVE TRAINING

Traditional safety education and training in construction have primarily relied on classroom lectures, video tutorials, and on-site field instructions. These methods are often designed to convey specific knowledge, such as safety rules related to hazardous equipment, materials, or restricted work areas, as outlined in construction delivery methods (Teizer et al., 2024b). However, such approaches are passive, positioning participants as observers rather than active learners. This passive role fails to address the dynamic characteristics of real construction sites, where hazards can be unpredictable and require immediate, behavior-based responses.

VR offers a promising alternative by actively immersing trainees in lifelike hazard scenarios, providing a more engaging and impactful learning experience. Unlike traditional methods, VR enables participants to interact with realistic construction environments, confront hazards, and receive real-time feedback without exposure to actual risk. VR-based training has been widely adopted in professional fields like military, aviation, and medicine, and its application in construction is gaining traction. Several studies have reviewed the benefits of VR for construction safety training (Salinas et al., 2022; Abich et al., 2021; Zoleykani et al., 2023). According to Golovina et al. (2019), VR offers the following advantages. Users can be confronted by hazards without being at risk. VR training increases engagement and empowers them to explore scenarios that have not been anticipated. VR training can be personalized and can provide data-driven feedback due to its capability to collect performance data reliably.

On the opposite side, VR-based learning environments were criticized in research for being unrealistic compared to participants making real-world experiences (Hilfert et al., 2016). Speiser & Teizer (2023a) and Speiser & Teizer (2024b) proposed a data model to include more realism from BIM and IoT, and Speiser & Teizer (2024a) included these data sources into a digital twin context to enhance realism and include real-world hazards. Such integrations may reduce the time required to create VR training, which has been outlined as the main reason for preventing details and real-world context (Wolf et al., 2022). Similarly, Zhao & Lucas (2015) emphasize using geometric data from BIM to increase the visual and contextual accuracy of VR construction scenes. However, implementing VR-based training faces practical challenges.

Other research identified that up to 25% of the participants experienced motion sickness (Jacobsen et al., 2022), the need for teaching people unaccustomed to computerized environments and creating multi-user environments (needed when training larger groups) (Jacobsen et al., 2021). Finally, some studies have reported that safety knowledge gains achieved through VR training are comparable to those obtained through traditional learning approaches (Adami et al., 2023; Sacks et al., 2013). This finding suggests that while VR excels in engagement and hazard simulation, its ability to achieve superior learning outcomes may depend on the design and application of the training scenarios.

These studies commonly use pre- and post-training interviews or tests to assess the effectiveness of the training. Such assessments, however, do not address the behavior of the user but only the intention of such users. Workers may demonstrate an intention to follow safety practices during assessments but fail to translate these intentions into real-world behaviors. This intention-behavior gap represents a critical limitation of conventional training methods, especially in dynamic and high-risk construction environments where the ability to respond to hazards is essential for safety.

BEHAVIORAL SAFETY METRICS

Safety assessment can occur in different ways. Intention-based assessments rely on interviews, tests, or surveys to gauge workers' knowledge or intentions. While easy to implement, these methods do not reliably quantify the safety behavior of a subject. On the other hand, behavior-based assessments involve observing workers' actions during tasks, providing more accurate insights into safety behavior as the intention-behavior gap exists (Ajzen, 1985). However, such assessments are resource-intensive and difficult to scale in real-world settings as they require observation or sensors to monitor workers and collect reliable data.

Simulated environments offer a promising alternative for behavior-focused assessments. By simulating tasks and hazards in a controlled environment, such systems enable detailed observation of worker behavior and interaction with hazards. However, prior research has primarily focused on evaluating knowledge and intentions, leaving the behavioral component underexplored (Salinas et al., 2022; Speiser et al., 2023).

Safety performance relates "to the effectiveness of the prevention of injury and ill health of worker" (ISO 45001:2023). One of the most common indicators is the total recordable incident rate (TRIR) or the day away, restricted work, or transfer (DART) injury rate. The purpose of these indicators is to quantify their performance to have comparable metrics (Hinze et al., 2013), but they have a lagging character, making them unsuitable for predictive safety performance. More recently, Erkal & Hallowell (2023) introduced the high-energy control assessment (HECA) score as a leading indicator. The underlying theory of HECA is that the outcome of accidents correlates with hazardous energy and represents a leading indicator of safety performance (Hallowell et al., 2017). Built on the same theory, the same authors have presented a new lagging indicator that weighs outcomes of injuries based on the energy level but fails to accommodate leading characteristics (Hallowell et al., 2024). The number of close calls represents another leading indicator.

While the mentioned metrics are intended to be collected in real-world settings, they may also be relevant for virtual environments. In a simulated world, like in virtual safety training, data collection is simple to implement. The collected data can then act as a data-driven approach to quantify safety performance. For this reason, this study will build on close-call analysis to quantify safety performance.

LIMITATIONS AND RESEARCH GAPS

Existing construction safety training methods, including traditional classroom and passive video-based training, primarily assess knowledge or intentions rather than actual behavior. Behavior, however, may be more significant because safety performance in construction heavily relies on behavioral skills, such as adhering to safe paths or recognizing and avoiding hazards. Similarly, intention-based metrics are used when assessing training effectiveness.

Simulated environments and VR have increasingly been adopted for construction safety training due to their potential to expose users to realistic hazards in a controlled environment. Nonetheless, such environments could also function as an environment to assess workers' safety behavior. Given that the results in a virtual environment are valid for the real world, the safety performance of a worker in a simulated environment corresponds to the safety performance in the real world. For this reason, this paper proposes a simulation-based assessment method in VR. Using close calls as a quantifiable metric to express workers' safety performance, the presented method expresses the objective and data-driven safety performance of workers.

RESEARCH METHODS

BEHAVIOR-BASED SAFETY PERFORMANCE ASSESSMENT IN SIMULATED ENVIRONMENTS

This study presents a VR approach for assessing construction safety performance in an objective and quantifiable manner. The framework focuses on behavior-based assessment methods that evaluate workers' actions during simulated tasks, as opposed to intention-based methods that rely on written tests or interviews. The VR environment enables monitoring of worker behavior in scenarios that mimic real-world construction hazards, eliminating risks associated with physical tasks. Therefore, the VR environments are modeled according to lean/safe and not lean/unsafe practices, as workers might experience some of these in reality.

The behavior of the users in such a learning environment can be monitored to collect data representing the user's (lean and) safety performance (Aziz et al., 2019; Salinas et al., 2022). With this data-driven approach, the performance is objectively quantified. In virtual environments, collecting data is more reliable and simpler than assessing safety performance in a real-world scenario where several sensors are required to track a worker's actions.

However, there are two relevant considerations when using simulated worlds for behavioral analysis: (1) the results of the simulation need to be valid for the real world (construct validity) (Harris et al., 2020), and (2) algorithms analyzing the collected data must quantify the safety performance. Construct validity can be achieved through different approaches, with the expert-novice comparison being one of the most common ones (Harris et al., 2020). A simulated environment where experts perform better than novices indicates construct validity. To quantify lean and safety performances, the scope of the metric needs to be defined within the worker's task. The worker's ability to interact with concurrent construction-typical equipment operations is investigated in this study. Hence, the assessment framework shown in Figure 1 places the user in a VR environment where a skid steer loader is operating, exposing the user to struck-by hazards related to the equipment. Lean elements (ground markings) are sometimes placed in the VR scene to guide the worker's pathway. Without compromising on experimental design intent, the level of detail of the VR scene was kept intentionally low.

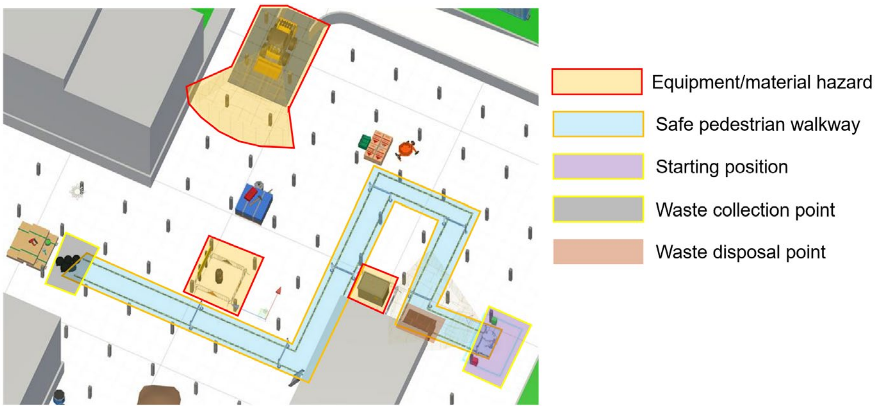

Figure 1: Virtual environment for safety performance assessment with lean and safe walkways and an operating skid steer loader.

The VR environment was developed using the Unity game engine due to its extensive asset library and suitability for creating construction scenarios. The environment includes a skid steer loader actively operating in the space, posing struck-by hazards. The environment also includes restricted zones and protected walkways designed to guide workers and minimize hazard exposure. The user in the scene has the mission of collecting and relocating trash bags while the loader is in operation, exposing them to concurrent hazards typical in construction.

The developed setup allowed for the observation of user behavior and interaction with hazards under realistic construction-like conditions to express the user's safety performance. The primary safety performance metric is the total number of incidents, defined as events "arising out of, or in the course of, work that could or does result in injury and ill health" (Dansk Standard, 2023). Incidents were categorized into three classes: (a) 0m-incidents - direct collisions with the skid steer loader, (b) 1m-incidents where the worker enters the 1m bounding zone around the loader, and (c) blind-spot-incidents where workers entered the operator's blind spaces, increasing the likelihood of potential collisions. Additional metrics, such as velocity, orientation, and duration of proximity to hazards, were logged for potential future analyses as presented in previous approaches (Golovina et al., 2019). Speiser & Teizer (2023) integrated these metrics into a virtual learning platform. Nevertheless, a simplified metric expressing the number of incidents is chosen for this study.

To detect the user's interaction with the hazards, collider objects to the skid steer loader were added in Unity (Figure 2). These enable the system to detect when a worker entered a hazard zone (e.g., the 1m bounding box or blind spots). The concepts have been described in previous studies (Golovina et al., 2019; Speiser & Teizer, 2023). Once such collisions are detected, an additional script added to the skid steer loader logs the time, location, and type of the incident in a structured JSON format. Each blind space (e.g., front left, rear middle) was monitored individually to classify the type of incidents. The aggregation of these incidents allows us to objectively quantify the safety performance in the number of near misses and hits.

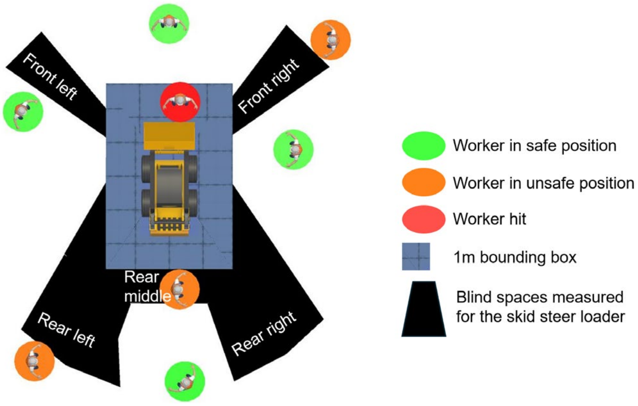

Figure 2: Incident detection based on collisions with blind spaces and proximity.

EVALUATION METHODS

The assessment environment was tested in a case study where two groups were assessed. As illustrated in Figure 3, Group 1 receives active training in the form of a “not lean/unsafe” scene, with no guidance towards safety issues besides instructing them in advance to work safely on the scene. This scene is the same as the previously described scene but without safety guidance. After training in this scene, they receive feedback on their performance. Group 2, on the other hand, receives passive training before being assessed in the simulated environment. The passive training consists of standard safety instructions outlining relevant construction hazard types and mitigation strategies. Furthermore, the instructions familiarized those participants with the site layout, the task, and the potential hazards involved. Therefore, Group 2 receives information common in daily toolbox talks in real life. All participants were introduced to the typical VR controls (moving, picking, and placing) in a virtual mock-up environment. Both groups then enter the assessment scene described in the previous section.

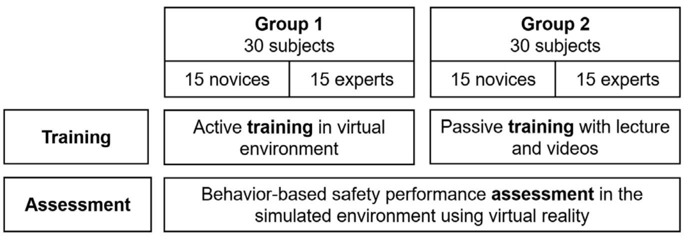

Figure 3: Experiment set up with two groups separated into experts and novices.

Based on a self-assessment and brief interview, both groups were equally split into two subgroups: experts and novices. We highlight that some experts may not be significantly more knowledgeable or experienced than some novices as we split the groups in half. However, we still use the terminology to distinguish clearly between the groups in the text. The distinction between experts and novices allows us to evaluate the quality of the simulated environment and the transferability of the results to the real world.

The design of the experiment enables the analysis of how differences in training impact safety performance. Besides, the splitting into experienced and less experienced groups shows how experience correlates to performance. This is a relevant feature for assessing the construct validity of the virtual environment. Lastly, we also collected the same data while group 1 was training in the VR environment. The training was in a similar training scene but without signage or safe walking spaces. For instance, the markers on the ground shown in Figure 1 were removed. It was only after the training in the simulated environment that the users were taught that such safe walkways would improve their safety. Collection of the safe data in both environments, the lean/safe and not lean/unsafe scene, furthermore allowed us to compare the safety performance of workers when having visual aid in construction scenes.

After the training sessions, the collected data is analyzed. During each user's session, the hazard interaction is monitored, and incidents are logged with different scripts in Unity written in C#. Log files are then analyzed using Python. The data structure follows simple structured lists where the user ID, the hazard type, and the incident location are stored. Despite not being required for this study, plenty of other data is logged, such as the location of the user, the duration of the incident, and the orientation of game objects. This data could be used for more nuanced data analysis, as illustrated in previous research, which provides a more nuanced distinction between performances. However, there is not yet a common indicator that allows for objective assessment on such a level, e.g., is it safer to enter the rear right several times with a larger distance than going into the front left once but very close to the 1m bounding box?

RESULTS

The results of the study are presented in three parts in the following section. First, the overall safety performance indicator as the total recorded incidents is presented and distinguished for each group and type of incident. Second, the location of the incidents is further discussed in relation to the blind spaces. The last part compares the expert and novice groups before the section finishes with a summary of the takeaways from the results.

OVERALL SAFETY PERFORMANCE

The safety performance of all participants shown in Figure 4 was quantified by the total number of recorded incidents, which included direct hits with the skid steer loader (0m incidents), near misses related to the 1m bounding box (1m incidents), and near misses related to the blind spaces. The performance was analyzed across the two training groups (Group 1 with active training and Group 2 with passive training) and the safe and unsafe scenarios within Group 1.

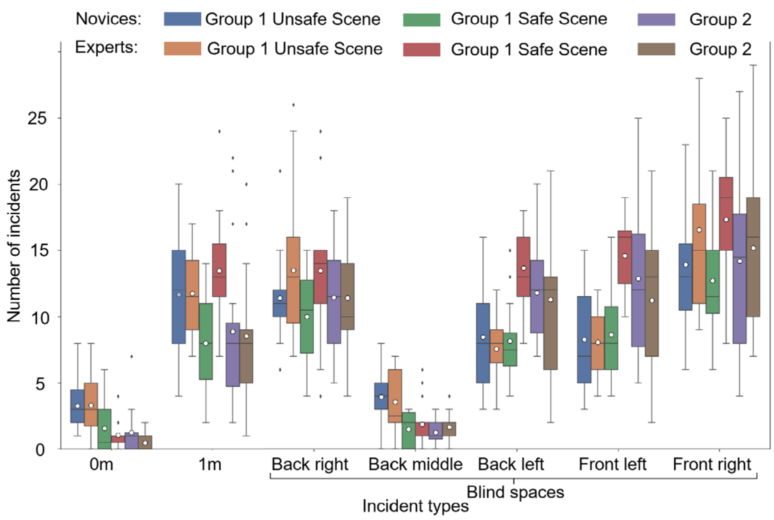

Figure 4: Number of incidents for each group and incident type (0m, 1m, or the blind space).

On average, each participant experienced at least one direct hit (0m incident) during the training experience. However, the number of 0m incidents significantly decreased when participants operated within the lean/safe scene compared to the not lean/unsafe scene. In the not lean/unsafe scene, participants averaged 3.3 direct hits. In the lean/safe scene (1u safe), this number was 1.3 collisions for Group 1 and 0.9 collisions for Group 2. The 1m incidents were higher than the 0m incidents across all scenarios. Group 1 averaged 11.8 incidents in the not lean/unsafe scene and 10.8 incidents in the lean/safe scene. Group 2 averaged 8.7 1m-incidents.

The results of the incidents relating to blind spaces are similar. On average, the participants entered blind spaces 48 times per experience. Group 1 averaged 46 incidents in the not lean/unsafe scene and 49 in the lean/safe scene. Group 2 averaged 50 incidents. Interestingly, the experts in Group 1 performed significantly worse in the lean/safe scenario, where the novices averaged 40 blind space incidents, and the experts averaged 58 incidents of the same type. While the lean/safe scene with the visual aid had significantly fewer direct hits and collisions with the 1m bounding box, the number of blind space incidents increased.

INCIDENT LOCATION AND BLIND SPACES

The distribution of incidents across different blind spaces was analyzed to assess the spatial patterns of safety behavior. Most incidents occurred in zones classified as front right and the least in the middle of the back. The middle blind space is also significantly smaller, which may be the reason for that. The blind spaces in the front right are significantly more frequent than in the front left. The back left and back right are similarly frequent.

Within the different groups, the results from the aggregation of the blind space incidents correlate with the number of incidents in the specific blind space. The only exception is the back middle, where users in the lean/safe scene perform significantly better than the users in the non-lean/unsafe scene. In the lean/safe scene, all four groups have a similar average incident number. Again, the space is much smaller, and as shown in Figure 3, the number is similar to the direct hits, indicating that the incidents related to the back middle resulted in a direct hit.

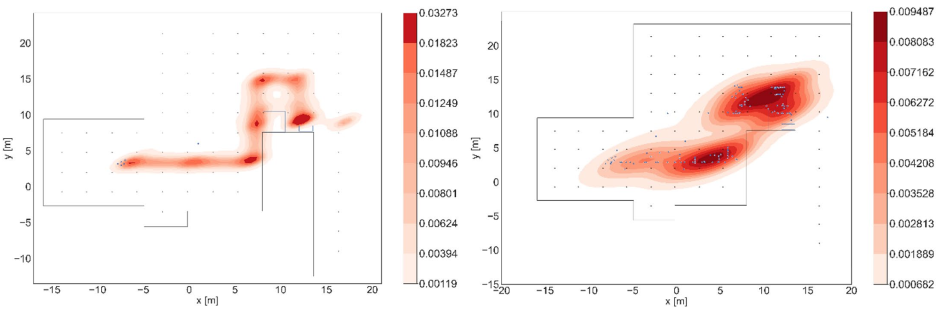

Figure 5: Comparing the incident frequency and location in the not lean/unsafe (left) and lean/safe (right) virtual learning environments for Group 1 (the darker a color is, the more incidents have occurred).

Incidents in the not lean/unsafe environment are more scattered than in the lean/safe environment. Figure 5 visualizes the location of the incidents in the not lean/unsafe scenario and the lean/safe scenario. The heatmap indicates that the incident locations are more predictable and concentrated near designated zones, in this case, the safe walking zones. This finding highlights the role of visual aids in influencing worker behavior and steering them toward safer paths. These results suggest that visual aids in a safe environment improved spatial awareness and guided participants toward safer behavior. This correlates well with lean-operated construction sites that use ground markers, signals, and barricades to separate pedestrian workers from construction vehicle traffic. It proves the effectiveness of such installations in lean and safe construction site layouts.

EXPERT-NOVICE COMPARISON

The distinction between novices and experts did not yield significant differences in safety performance. Both groups exhibited similar patterns in collision and proximity incidents, with novices occasionally outperforming experts. Figure 4 shows that the experts in Group 1 caused the most incidents. This finding raises questions about construct validity, as it suggests that expertise did not consistently translate into improved safety behavior in the VR environment.

On the other hand, the experts performed better in the categories of the 0m but not in the incidents with more distance. The lack of performance differences may also be attributed to a confidence bias. Experienced participants may have been overconfident and taken greater risks. Another reason could be that the design of the selection process, which involved self-assessment, may have led to misjudgment.

SUMMARY

The main results of the experiment show that the lean/safe scene with the visual aids reduced 0m- and 1m-incidents, demonstrating their effectiveness in guiding safer behavior. Besides, the occurrence of the incidents is more controlled in the designated spaces. Second, incidents were more predictable in a lean/safe environment, supporting the use of visual markers for improved lean/safety planning.

The study quantified safety behavior using objective metrics (incident counts), providing a foundation for data-driven safety assessments. The study showcased how virtual environments could be used in the future to assess training effectiveness using behavioral improvements. We can now objectively compare the user performance after the active and the passive training. In this specific case, the students performed worse after the active training, raising the question of whether passive training was effective. We need to discuss this carefully as training always depends on many factors, such as the teacher, the method, the content, and the tailoring of the training objectives. Within this study, construct validity remains an open question, as no significant performance differences were observed between novices and experts.

DISCUSSION

The lack of a significant performance difference between novices and experts raises questions about the construct validity of the virtual environment. According to Harris et al. (2020), experts are expected to perform better as novices in VR as they are also expected to perform better in the real world. In this study, the absence of clear performance distinctions between experts and novices may indicate that the self-assessment might not accurately reflect the actual levels of expertise of the subjects. Another reason for this result could lie in confidence bias among experts, where experts may have been less cautious, leading to riskier behavior. This overconfidence could neutralize any performance advantage they might have had over novices. Last, novices who are less familiar with construction safety might have been more careful.

Nevertheless, future studies should improve the selection process for finding construct validity. Investigating other concepts, such as training-induced learning, could also indicate construct validity, but these are not considered in this study. Although these findings do not strongly support the claim of construct validity, the results are still valuable for showcasing how virtual environments are able to assess the safety performance of workers.

The comparison between safe and unsafe scenes reveals that visual aids, such as designated safe walkways, play a crucial role in guiding workers toward safer behaviors. The reduction in 0m Incidents in the safe scene suggests that visual markers help workers maintain safer distances from the equipment. The decrease in 1m incidents indicates that visual aids may have influenced workers' spatial awareness, although the effect was less noticeable compared to direct hits. As the incident locations were less scattered, they were also more predictable, suggesting that visual aids can support the planning of a safe site layout. However, potential over-reliance on visual aids could negatively impact safety performance. Workers may have trusted the safety markers to the point of reducing their caution.

LEAN principles focus on minimizing waste across various aspects. Active training that quantifies workers' safety and lean performance can help decision-makers address different waste types. By objectively measuring safety behaviors, such as close-call frequency, the VR assessments offer insights into waste reduction. Unsafe behavior often leads to rework, idle time, or underutilized talent. Thus, improving safety through targeted VR training aligns directly with lean goals. Additionally, analyzing behavioral patterns through this approach enables construction managers to proactively identify inefficiencies before they result in waste. Consequently, this research improves safety outcomes and supports lean construction.

This study focused on safety performance with respect to interactions with a single hazard type - the skid steer loader. While this allows for targeted analysis, the findings may not generalize to other hazards commonly found on construction sites. According to Harris et al. (2020), the validity of a VR environment can only be evaluated within the objectives it is designed to achieve. To develop a more holistic understanding of safety behavior, future research should include a broader range of hazard types (e.g., falling objects, electrical hazards), investigate interactions between multiple dynamic hazards, and explore nuanced safety metrics, such as velocity, orientation, and task efficiency.

The results demonstrate that the VR-based assessment framework can serve as a novel leading safety indicator by providing objective data on worker interactions with hazards. The total number of incidents and their spatial distribution offer valuable insights into safety behavior patterns. Decision-makers could use such an indicator to evaluate individual workers' safety awareness. The indicator could be used to identify relevant and required training for workers. Having such metrics is not only a leading indicator but would allow safety experts to evaluate safety performance before working in the field.

CONCLUSIONS

While previous studies primarily investigated simulated environments for active training methods, this study showcases how such learning environments can assess the construction safety performance of workers for increased lean performance. In contrast to intention-based assessments, this VR-based method enables objective quantification of safety behaviors. The data-driven metrics allow us to evaluate behavioral safety.

The findings emphasize the potential of VR environments to support comparative studies that investigate the effectiveness of different training methods. With this framework, future research can move beyond assessing the intentional effects of different training approaches to evaluating their actual impact on safety behavior. However, safety performance can always only be measured within the objectives of the study, limiting the scope of a performance indicator to a certain task or hazard type.

Besides demonstrating the advantages of simulated environments for performance assessment, the results suggest the need for more nuanced metrics to capture safety behavior. Future work should incorporate additional parameters and weighting to create a more holistic understanding of safety performance or include other tasks and hazard types. Another finding of the study suggests that visual aid and guidance for construction workers improves their safety performance. Such guidance also makes the occurrence of incidents more predictable. Future work could investigate how guidance in construction safety impacts workers' performance.

In conclusion, this research advances simulated environments from a training method to a performance assessment method. Researchers can now compare the effectiveness of training based on behavior, which contributes to safer practices in construction. While safety was this work's primary research objective, assessing lean behavior could follow a similar avenue. Nevertheless, it would be expected to return similar benefits as the earlier mentioned empirical finding stated that lean (i.e., productivity) and safety go hand-in-hand. Future work is needed to prove it in VR as part of Digital Twin for Construction Safety (DTCS) (Teizer et al., 2024a).

ACKNOWLEDGMENTS

This research has been funded by the European Union Horizon 2020 research and innovation program under grant agreement number 101058548 (BEEYONDERS).

REFERENCES

Abich, J., Parker, J., Murphy, J. S., & Eudy, M. (2021). A review of the evidence for training effectiveness with virtual reality technology. *Virtual Reality*, 25. https://doi.org/10.1007/s10055-020-00498-8

Adami, P., Singh, R., et al. (2023). Participants matter: Effectiveness of VR-based training on the knowledge, trust in the robot, and self-efficacy of construction workers and university students. *Adv Eng Inf*, 55. https://doi.org/10.1016/j.aei.2022.101837

Ajzen, I. (1985). From Intentions to Actions: A Theory of Planned Behavior. *Action Control*, Springer, pp 11–39. https://doi.org/10.1007/978-3-642-69746-3_2

Albert, A., Hallowell, M. R., et al. (2014). Enhancing Construction Hazard Recognition with High-Fidelity Augmented Virtuality. *Constr Eng Man*, 140.

Aziz, A., Ahmed, S., & Khan, F. I. (2019). An ontology-based methodology for hazard identification and causation analysis. *Process Safety and Environmental Protection*, 123, pp 220–229. https://doi.org/10.1016/j.psep.2018.12.008

Burke, M. J., Sarpy, S. A., et al. (2006). Relative effectiveness of worker safety and health training methods. *Am J Public Health*, 96. https://doi.org/10.2105/AJPH.2004.059840

Erkal, E. D. O., & Hallowell, M. (2023). Moving Beyond TRIR: Measuring & Monitoring Safety Performance With High-Energy Control Assessments. *Professional Safety*, 68.

Eurostat (2024), Accidents at work by NACE Rev. 2 activity and size of enterprise. https://ec.europa.eu/eurostat/databrowser/view/hsw_n2_02/default/table?lang=en&category=hlth.hsw.hsw_acc_work.hsw_n2. Retrieved January 13, 2025.

Golovina, O., Kazanci, C., Teizer, J., & König, M. (2019). Using Serious Games in Virtual Reality for Automated Close Call and Contact Collision Analysis in Construction Safety. *Proc 36th ISAR*C. https://doi.org/10.22260/ISARC2019/0129

Golovina, O., Perschewski, M., Teizer, J., & König, M. (2019). Algorithm for quantitative analysis of close call events and personalized feedback in construction safety. *Automation in Construction*, 99. https://doi.org/10.1016/j.autcon.2018.11.014

Hallowell, M. R., Alexander, D., & Gambatese, J. A. (2017). Energy-based safety risk assessment: Does magnitude and intensity of energy predict injury severity?. *Construction Management and Economics*, 35. https://doi.org/10.1080/01446193.2016.1274418

Hallowell, M. R., Deniz, E., & Erkal, O. (2024). Severity-based Lagging Indicator: An Alternative Measure of Safety Performance. *Professional Safety*, 69.

Harris, D. J., Bird, J. M., Smart, P. A., Wilson, M. R., & Vine, S. J. (2020). A Framework for the Testing and Validation of Simulated Environments in Experimentation and Training. *Frontiers in Psychology*, 11. https://doi.org/10.3389/fpsyg.2020.00605

Hilfert, T., Teizer, J., & König, M. (2016). First person virtual reality for evaluation and learning of construction site safety. *Proc. 33rd ISARC*. https://doi.org/10.22260/isarc2016/0025

Hinze, J., Thurman, S., & Wehle, A. (2013). Leading indicators of construction safety performance. *Safety Science*, 51. https://doi.org/10.1016/j.ssci.2012.05.016

Jacobsen, E. L., Solberg, A., Golovina, O., & Teizer, J. (2022). Active personalized construction safety training using run-time data collection in physical and virtual reality work environments. *Construction Innovation*, 22. https://doi.org/10.1108/CI-06-2021-0113

Jacobsen, E. L., Strange, N. S., & Teizer, J. (2021). Lean construction in a serious game using a multiplayer virtual reality environment. *IGLC*. https://doi.org/10.24928/2021/0160

Johansen, K.W., Schultz, C., & Teizer, J. (2025). Knowledge graph exploitation to enhance the usability of risk assessment in construction safety planning. *Adv Eng Inf*. https://doi.org/10.1016/j.aei.2025.103305

Li, B., Schultz, C., Melzner, J., Golovina, O., & Teizer, J. (2020). Safe and Lean Location-based Construction Scheduling. *37th ISARC*, https://doi.org/10.22260/ISARC2020/0195

Ohno, T. (1988). Toyota Production System: Beyond large scale production, Productivity Press, Portland, Oregon.

Pham, T. T., Lingard, H., & Zhang, R.P. (2023). Factors influencing construction workers' intention to transfer occupational health and safety training. *Safety Science. 167*. https://doi.org/10.1016/j.ssci.2023.106288

Sacks, R., Perlman, A., & Barak, R. (2013). Construction safety training using immersive virtual reality. *Construction Management and Economics*, 31. https://doi.org/10.1080/01446193.2013.828844

Salinas, D., Muñoz-La Rivera, F., & Mora-Serrano, J. (2022). Critical Analysis of the Evaluation Methods of Extended Reality (XR) Experiences for Construction Safety. *Environ Res Public Health*, 19. https://doi.org/10.3390/ijerph192215272

Speiser, K., Hong, K., & Teizer, J. (2023). Enhancing the Realism of Virtual Construction Safety Training: Integration of Real-Time Location Systems for Real-World Hazard Simulations. *23rd CONVR*. https://doi.org/10.36253/979-12-215-0289-3.15

Speiser, K., & Teizer, J. (2023a). An Ontology-Based Data Model to Create Virtual Training Environments for Construction Safety Using BIM and Digital Twins. *30th European Group for Intelligent Computing in Engineering Conference (EG-ICE)*, London.

Speiser, K., & Teizer, J. (2023b). An Efficient Approach for Generating Training Environments in Virtual Reality Using a Digital Twin for Construction Safety. *Proc. CIBW099W123*. https://doi.org/10.24840/978-972-752-309-2

Speiser, K., & Teizer, J. (2024a). Automatic creation of personalised virtual construction safety training using digital twins. *Management, Procurement and Law*. https://doi.org/10.1680/jmapl.23.00104

Speiser, K., & Teizer, J. (2024b). Formalizing virtual construction safety training: A schematic data framework enabling real-world hazard simulations using BIM and location tracking. *ITcon*, 29. https://doi.org/10.36680/j.itcon.2024.043

Speiser, K., Teizer, J., Ali, G.M. (2024). Smart signalperson for safe tower crane operations on a future digital construction site. *CIBW099W123*, Ghana, 353-365, ISBN: 9789988391386.

Teizer, J. & Melzner, J. (2019). Neue Wege zur Integration von BIM, IoT und Lean Construction in der Bauingenieurausbildung. VDI-Bautechnik, Jahresausgabe 2019/20, Bauingenieur, Springer, October 2019, 19-25, ISSN 0005-6650

Teizer, J., Johansen, K.W., Schultz, C.L., Speiser, K., Hong, K., & Golovina, O. (2024a). A Digital Twin Model for Advancing Construction Safety. Lecture Notes in Civil Engineering (LNCE 390), pp. 201–212, https://doi.org/10.1007/978-3-031-44021-2_22

Teizer, J., Migliaccio, G.C., Soliu, I., & Golovina, O. (2024b). Relationship between subcontracting models and adoption of innovative methods in construction safety. *J Legal Affair Disp Resol Eng Constr.* https://doi.org/10.1061/JLADAH.LADR-1050.

Wolf, M., Teizer, J., Wolf, B., Bükrü, S., & Solberg, A. (2022). Investigating hazard recognition in augmented virtuality for personalized feedback in construction safety education and training. *Advanced Engineering Informatics*, 51. https://doi.org/10.1016/j.aei.2021.101469

Zhao, D., & Lucas, J. (2015). Virtual reality simulation for construction safety promotion. *Int J Inj Contr Saf Promot,* 22. https://doi.org/10.1080/17457300.2013.861853

Zoleykani, M. J., Abbasianjahromi, H., et al. (2023). Extended reality (XR) technologies in the construction safety: Systematic review and analysis. *Construction Innovation*. 24(4). https://doi.org/10.1108/CI-05-2022-0131