https://doi.org/10.20378/irbo-52667

Comparing Immersive Virtual Reality and PowerPoint as Methods for Delivering a Safety Training: Impacts on Learning, Risk Perception, and Decision Making

Johannes Leder[a*], Tina Horlitz[a], Patrick Puschman[b], Volker Wittstock[b], Astrid Schütz[a]

[a]Department of Personality Psychology and Psychological Assessment, University of Bamberg, 96047 Bamberg, Germany

[b]Institute for Machine Tools and Production Processes, Technische Universität Chemnitz, Reichenhainer Str. 70, 09126 Chemnitz, Germany

*Corresponding author. Email address: johannes.leder@uni-bamberg.de. Postal address: Department of Personality Psychology and Psychological Assessment, Markusplatz 3 (Room M3/3.02), 96047 Bamberg, Germany

Declarations of interest: none

Abstract

In two experimental studies, we compared a safety training given via immersive virtual reality with one given via PowerPoint in their effects on risk perception, learning, and risky choices. In Study 1, we compared the two methods in a sample of apprentices ($N = 53$) and also investigated whether participants' conscientiousness and locus of control moderated the effects of the safety training. In Study 1, we found an effect of training method on the change in risk perception in terms of probability judgments and on risky decisions but not on learning. In Study 2 ($N = 68$), we sought to replicate Study 1 and also tested whether domain-specific risk attitudes affected risk perception and choice. Furthermore, long-term effects of the safety training on information recall and risk perception after a 6-month interval were assessed. The effects found in Study 1 could not be replicated in Study 2. Neither study found an interaction between presentation medium and personality. We conclude that the costly procedure of immersive VR does not seem justified for safety training because the less costly PowerPoint procedures with vivid film scenes did not fare significantly worse with respect to changes in risk perception, learning outcomes, or decision making.

Keywords: VR, risk perception, risky choices, accidents, hazardous machines, learning

Comparing Immersive Virtual Reality and PowerPoint as Methods for Delivering a Safety Training: Impacts on Risk Perception, Learning, and Decision Making

1. Introduction

Would you consider a pillar power drill a hazard to your health? Would you carefully study safety information before using such a power drill? Most likely the answer to both questions is no even though using a pillar power drill can lead to severe accidents. What are effective ways to change risk perceptions and behavior in the face of such seemingly harmless machines?

Instruction methods can differ with respect to their degree of immersion; there is no immersion when text-based instruction is employed, low immersion when media involving video and sound is used, and high immersion when the learner actually experiences the situation in question. In the present paper, we test whether trainings that are low or high in immersion differ in their effects on risk perception, learning, and decision making. We studied immersive VR technology as an innovative medium that can be employed to convey safety information, and we tested whether a VR-based safety training led to more learning, increased risk perception, and changes in decision making in comparison with a PowerPoint presentation conveying the same information.

A recent analysis of workplace injuries across work domains in the 10-year period from 1998 to 2008 in Ontario, Canada showed that absence rates due to injury were three times higher in novices than in workers who had at least 1 year of job experience in the present job (Morassaei et al., 2013). Thus, preventive measures that focus on novices might prove especially efficient for reducing the absolute number of workplace injuries. Safety trainings are one standard means of prevention, and recently, immersive VR has been suggested as an innovative way to present such a training (e.g., Sacks et al., 2013; Zaalberg and Midden, 2013). But are such innovative methods in fact more effective than more traditional approaches in raising awareness and promoting safe conduct?

Out of an array of constructs from the empirical literature, a meta-analysis by Christian et al. (2009) identified three predictors as repeatedly standing out as particularly relevant for ensuring safe performance: *safety knowledge* and *safety motivation* as proximal person-related factors and *general risk-taking propensity* as a distal person-related factor. Safety knowledge reflects an individual's knowledge of how to perform safely. Safety motivation reflects "an individual's willingness to exert effort to enact safety behaviors and the valence associated with those behaviors" (Neal & Griffin, 2006, p. 947; cf. Christian et al., 2009). Safety knowledge and safety motivation are important predictors of safety performance, which directly influences the likelihood of accidents and the overall number of workplace injuries (Christian et al., 2009). Furthermore, according to protection motivation theory (Maddux and Rogers, 1983) and the health belief model (Janz and Becker, 1984), risk perception in a given situation is an important prerequisite for safety motivation and results in health-protective behavior. Thus, risk perception, knowledge about hazards, and knowledge about safety measures are assumed to predict protective behavior in general.

Weinstein (1993) explained that risk perception as a precondition for safety motivation is commonly regarded as consisting of two components: the perceived likelihood that an accident will occur (probability judgments) and the perceived severity of the consequences of such an accident (severity judgments). By contrast, knowledge about safety measures impacts beliefs about the efficacy of protective behavior (response efficacy) as well as the belief that one can use these measures (self-efficacy). The relevance of these factors was shown in a meta-analysis across 65 studies ($N \approx 30{,}000$) and 20 domains (Floyd et al., 2000): The findings showed moderate positive effects of severity judgments, probability judgments, response efficacy, and self-efficacy on adaptive intentions and behaviors.

When considering the problem of reducing the risk of accidents among novices at work, theory and empirical evidence have suggested that safety interventions addressing *risk perception* on the one hand and *improvements in safety-related knowledge* on the other (for an overview, see Laughery & Wogalter, 2006) should be effective.

1.1. Methods in Safety Training: The Importance of Learner Engagement

The methods used to communicate safety information can be distinguished according to the extent of learner engagement during the dissemination of the information. Safety training conveyed through written descriptions can be considered to be on the lower end of engagement, whereas simulations and experiential learning are on the higher end. Past research has emphasized the importance of experience: For example, warnings in the form of written instructions, even when they are clearly understood, are not sufficient for instigating safe conduct (Zeitlin, 1994). Although a lack of negative events often leads to unrealistic optimism (Weinstein, 1984), the experience of a negative event is positively correlated with the accuracy of risk perceptions and the likelihood of showing preventative behavior (Weinstein, 1989). Burke et al. (2006) aggregated 95 studies with over 20,000 participants and showed that the safety trainings that were more engaging and experiential were also more effective.

Drawing on the reasoning above and supported by meta-analytic findings (Christian et al., 2009), we argue that safety trainings utilizing experiential learning should be more successful in conveying information and changing behavior than passive methods. Thus, in order to be effective, interventions should vividly depict hazards and the potential negative outcomes of certain behaviors as well as safety behavior.

Immersive VR provides a presentation medium in which people can gain experience in situations that are rare and dangerous and thus cannot be staged. Immersive VR increases the impression of realness, an experience that has been termed *the sense of being there* (Steuer, 1992) or more generally *immersion* (Bystrom et al., 1999). In fields such as behavioral therapy, immersive VR has successfully been used to reduce fear of flying (Mühlberger et al., 2006) and fear of spiders (Peperkorn et al., 2015).

Past research on the use of immersive VR in the context of safety-related research has investigated behavior during fire emergencies (Gamberini et al., 2003), aviation safety (Buttussi and Chittaro, 2017; Chittaro and Buttussi, 2015), safety behavior related to construction sites (Sacks et al., 2013), flooding (Zaalberg and Midden, 2013) and individual behavior in tunnel accidents (Kinateder et al., 2015, 2013; Mühlberger et al., 2015). Most of these scenarios were extreme situations that were obviously dangerous, and it has yet to be shown whether such safety trainings utilizing immersive VR are also effective in workplace settings that may erroneously be perceived as harmless.

1.2. The Use of Immersive VR as a Medium in Safety Training

Due to the properties of immersive VR, it can be assumed that safety trainings using this option can have an impact that is similar to the impact of personal experiences and will be more successful than formats that do not involve immersion. We will first review the small body of research that has suggested that immersive VR technology may impact learning, risk perception, and protective behavior more effectively than non-immersive technologies such as a PowerPoint presentation. Then, we will argue that previous studies comparing immersive VR with less immersive presentation mediums that have been utilized to convey safety information are inconclusive. We then propose that a fair test of the effects of immersive technology should compare the VR format with a less immersive format that still conveys the same information. For example, the PowerPoint presentation should include pictures and filmed sequences rather than just written materials.

Previous studies on the perception of the risk of flooding have found that risk information provided with immersive VR in comparison with traditional methods (film and slides) increased people's motivation to search for safety-relevant information and affected some coping measures (i.e., buying insurance) but not others (i.e., evacuation intentions) and that the emotions that were elicited did not differ between conditions. The increase in coping intentions was mediated by the sense of presence, albeit the effect was only marginally significant (Zaalberg and Midden, 2013). In a study that compared a safety training given to prospective construction workers using immersive VR and traditional classroom instruction, the immersive VR condition resulted in better recall of safety-related knowledge, but there was no general change in risk perception. It is important to mention that the long-term effects in this study were not reliable because only 30% of the participants participated in the follow up after 1 month (Sacks et al., 2013), and this low rate may have led to strong selection effects. Comparing an aviation safety training given via a traditional method (i.e., safety cards) and an immersive method, knowledge retention was found to be higher when the immersive method was used (Chittaro and Buttussi, 2015).

Other studies did not find that a general training delivered though immersive VR offered an advantage over a control condition when both conditions involved interaction and the material was vivid (Gavish et al., 2015; Moreno and Mayer, 2002). In fact, interactive and non-interactive serious games had similar positive outcomes for learning and changes in risk-severity perception

(Chittaro and Sioni, 2015), thus suggesting that vividness might play a particularly important role. Consistent with this finding, another study showed that procedural training with serious games led to more knowledge retention after 2 weeks than a non-interactive method, and there was no difference between immersive VR and an interactive desktop presentation (Buttussi and Chittaro, 2017). Thus, the findings from the studies that have suggested the greater effectiveness of immersive VR could also be due to the use of a serious game or to the greater vividness of the material in the immersive VR condition, which was not the case in the control condition.

Regarding the question of whether immersive VR can be more effective than a traditional safety training, previous research as summarized above has shown clear evidence for an increased sense of presence in immersive VR, but the advantage of that sense of presence for learning, risk perceptions, and decisions is unclear for the following reasons. First, previous studies have typically used *situations that are generally considered dangerous* by the public, such as tunnel emergencies, terrorism, flooding, and airplane crashes. In these studies, there was an advantage of immersive VR regarding knowledge retention but not with regard to changes in risk perception. It is possible that the fact that risk perception did not change can be explained by considering that such situations (e.g., tunnel emergencies) are obviously dangerous, and thus, risk perception is already high. Therefore, the effects of immersive VR should be studied in situations in which the level of risk may be underestimated.

Second, there is a lack of studies on the effects of a safety training on actual decisions and their relation to risk perceptions and safety knowledge. So far, studies have addressed intentions (Zaalberg and Midden, 2013) and perceptions (Chittaro and Buttussi, 2015; Sacks et al., 2013) but not *decisions*.

Third, only two studies have addressed the *long-term effects* (1 and 2 weeks) of presentation medium (Chittaro and Buttussi, 2015; Sacks et al., 2013), but one of these studies had a dropout rate of 70% (Sacks et al., 2013), whereas the other (Chittaro and Buttussi, 2015) used a serious game in the immersive VR condition and compared it with the traditional pictorial method, which was non-interactive and non-immersive. Thus, the results were confounded by the degree of interaction. It is not yet clear whether the effects were actually due to VR or whether they may have been due to the level of interaction.

Fourth, so far, *interindividual differences* in risk-taking have not been considered in research on the effects of a safety training, even though such dispositions, such as locus of control, conscientiousness, and individuals' risk attitude are important predictors of protective behavior (Christian et al., 2009).

Finally, the studies that provided a direct test between immersive VR and traditional methods of presentation have often had only a few participants in each cell (Moreno and Mayer, 2002; Sacks et al., 2013; Zaalberg and Midden, 2013) or only marginally significant results when covariates were included in the analysis (Zaalberg and Midden, 2013). These methodological limitations result in deflated p-values when covariates are included (Simonsohn et al., 2014) and overestimated effect sizes (Button et al., 2013; Nieuwenstein et al., 2015). Also, due to small samples and a lack of power, the small effects that exist might not be detected.

1.3. The Present Studies

We tested whether an immersive VR-based safety training would be found to be more effective than a safety training presented via PowerPoint (PPT). We compared the effectiveness of these two formats in a sample of novices, in our case, high school students. The safety trainings were presented either in immersive VR or in a PPT presentation that included filmed

VR sequences and slides. Thus, the PPT presentation was also vivid and differed from VR only with respect to the fact that it was not immersive. The VR condition was immersive but not interactive to allow us to test for the effects of immersion only. As dependent variables, we measured (a) risk perception by assessing participants' judgments of the probability of accidents and accident severity, (b) learning by assessing participants' recall of safety knowledge and the hazards that had been identified, and (c) decision making on the basis of a choice between a safe and an unsafe machine. We included personality variables as covariates or potential moderators. Thus, we put VR to an especially hard test and aimed to understand whether the unique features of immersive VR actually make a difference.

The current studies therefore extend previous research by focusing on workplace safety in a context that is usually considered relatively safe and by comparing a medium that is high in immersion (VR) with one that is low in immersion (PPT) in their effects on risk perception, learning, and decision making. We further scrutinized the moderating effect of instruction medium immediately after the training as well as 6 months later and examined the potential influence of personality differences (locus of control, conscientiousness, and risk attitude) on the effectiveness of the safety training.

1.4. Study overview

In two experiments, we tested whether a VR-based safety training would be more effective than a PPT presentation in affecting risk perception, knowledge acquisition, and decision making. Participants received a safety training about a pillar drill delivered in an immersive VR format or as a PPT presentation.

On the basis of previous research, we expected that participants in the immersive VR condition would experience a higher sense of presence (Hypothesis 1). Because immersive VR provides a relatively realistic experience of the possible negative events, we predicted that the VR training would have a stronger impact than PPT training on risk judgments (Hypothesis 2), resulting in a greater change of severity and probability judgements of accidents, and learning, i.e., recall more safety-relevant information and detecting more hazards when facing a real machine after being instructed in the VR training (Hypothesis 3). Because experiential learning results in more complex knowledge networks and knowledge integration (Glaser, 1984), we expected that in the immersive VR condition, risk-related decisions would be more strongly affected by participants' recalled safety knowledge and identified hazards than in the PPT condition, reflecting a higher degree of knowledge integration in the immersive VR condition (Hypothesis 4).

As an exploratory question, we investigated the effect of three personality dispositions: (a) locus of control, (b) conscientiousness in Study 1, and (c) risk attitude in Study 2 on the relations between the independent and dependent variables. It is usually assumed that individuals' risk-taking behavior is mediated by risk perception and moderated by situational variables (Sitkin and Weingart, 1995). Thus, to change risk-taking, a change in risk perception is necessary in the respective domain. The relevance of situational factors also leads to the conclusion that risk-taking and risk perception might not be the result of a general trait but could be domain specific (Figner and Weber, 2011). For example, skydivers, who show a high degree of risk seeking in the recreational domain, tend to be risk averse in the financial domain (Hanoch et al., 2006).

A relevant domain for our studies is the health and safety domain, which, for example, reflects risk attitudes toward seatbelt use or sunbathing without sunscreen and therefore should be related to workplace safety risks. Hence, we expected that risk perceptions in the health and safety domain would predict risk perceptions and decisions in the workplace. Furthermore, we aimed to explore whether the type of safety training (VR vs. PPT) would moderate the effect of domain-specific risk attitudes on risk perception and decisions and whether an immersive approach would lead to greater integration between habitual and specific perceptions because the trainings are more experience-based. We expected that because immersive VR seems real, the previously unknown risk associated with the machine would become integrated in the general domain (in this case, the health/safety domain), and thus, health and safety domain risk attitudes would predict risk perceptions and decisions in the immersive VR condition but not in the PPT condition (Hypothesis 5).

1.5. General Approach

Both studies were designed as intervention studies, and risk perceptions were assessed before and after the safety training, whereas the sense of presence was assessed during the safety training. Recall of safety knowledge and decision making were measured only after the safety training. Study 1 and Study 2 differed in the personality variables that were assessed. In Study 1, we assessed locus of control, conscientiousness, and domain-specific risk perception as possible predictors of risk perception, safety knowledge, and decision making. In Study 2, we assessed domain-specific risk attitudes. Furthermore, in Study 2, we tested the sustainability of the training by employing a follow-up assessment after 6 months. As VR technology is expensive and time intensive, only effects that were at least medium in size were identified as practically important. For this reason, we tested for medium effects, and post hoc power calculations using G*Power 3.1.9.2 (Faul, Erdfelder, Buchner, & Lang, 2009) showed that the sample sizes in both studies were sufficient for detecting such effects (Cohen's $f = .2$) with a power of .90 in Study 1 and .95 in Study 2.

2. Study 1

2.1. Method

Participants and Design

The study was carried out at the Center for Virtual Reality at Chemnitz University of Technology with $N = 53$ participants. Female participants comprised 16.98% of the sample, the mean age was 18.42 years, and 64.15% of the participants had previously experienced VR or 3D movies. Participants were students at a local high school. They were randomly assigned to either the PPT condition (PPT) or the immersive VR condition (VR) to undergo a safety training concerning the use of pillar drills. The safety training began with a description of the parts and general functioning of the pillar drill. Then, the safety equipment and safety procedures were explained, and the animation was depicted. Finally, the participants saw accidents that occurred due to human error or due to the hazards associated with the pillar drill. The accidents were animated, and a commentary accompanied them.

The experimental conditions did not differ in their experience with VR or 3D movies, $\chi^2(1) = 0.948$, $p = 0.33$. Participants were asked to rate the probability of accidents and the severity of accidents of a real pillar drill presented to them twice, once before (T1) and once after they received the safety training (T2), which resulted in a 2 (condition: PPT vs. VR) x 2 (time: pre vs. post safety training) mixed design. Furthermore, at T2, learning outcomes were measured, and participants were confronted with a decision-making task.

Measures

Risk judgments were measured by asking participants to rate the severity and the probability of accidents with a real pillar power drill. The rating of the likelihood of accidents and the severity of accidents employed so-called risk ladders (Harrison and Rutström, 2008). The measures of severity and probability were on individual sheets and are depicted in Fig. 1. The ladder for the probability measure was presented in 10% increments with 0% as the lowest value and 100% as the highest value. To measure the subjective severity of accidents, the same ladder was used, but it was presented in increments of 1 with the lowest point labeled *not dangerous* (1) and the highest point labeled *life threatening* (10).

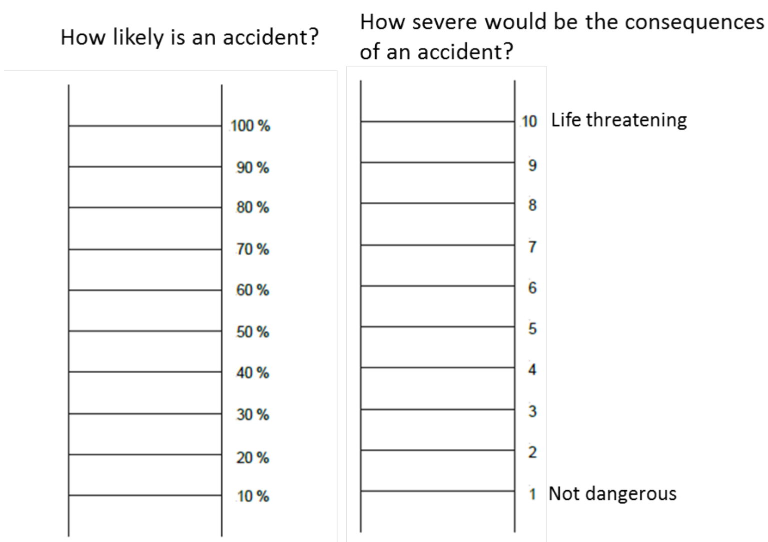

Fig. 1. Assessment of risk judgment using a risk ladder for probability and severity judgments.

Participants were instructed to mark the point on each ladder that reflected their judgment with respect to the real machine in front of them.

Learning outcomes were assessed with two variables: recall of safety information and knowledge of hazards with respect to a real pillar power drill. Recall of safety information was assessed with a multiple-choice test, which assessed how much of the content conveyed in the safety training participants remembered (i.e., safety procedures, signs, and general machine information), for example, “What should you not do in case of an emergency?” (a) stop the machine, (b) take pictures and draw a sketch of the area where the accident occurred, (c) move the victim away from the area of threat, or (d) place an emergency call, inform your immediate superior. Ten questions were asked, and the number of correct responses was summed so that the score had a range of 0 to 10.

The knowledge of hazards present in pillar power drills was assessed by asking participants to identify hazards pertaining to a real pillar power drill in front of them. The pillar power drill was identical to the one in the safety training and yielded a maximum of nine possible hazards. The score on this task was the proportion of correctly identified hazards present in the real machine.

Decision-making to assess risk-taking behavior was measured through choices between working on the real machine in front of them (i.e., the risky choice) or a fictitious safe machine (i.e., the safe machine). Whereas the real machine yielded safety hazards, the fictitious safe machine did not. The safe machine always yielded a sure payoff of 10 €. The real machine yielded two prospects: One had a payoff of 0 € if there was an accident, and the other prospect had a non-zero payoff that increased in steps of 5 € with each decision. Participants were told that the likelihood of an accident with the real machine was 50%. Participants had to make five choices, and the non-zero prospect of the safe machine increased from 10 € in the first decision to 30 € in the final one (see Table 1).

Table 1. Decision task presented to participants.

	Safe machine	Real machine		Decision
Decision	for sure	Accident (50%)	No accident (50%)	
1	10 €	0 €	10 €	A|B
2	10 €	0 €	15 €	A|B
3	10 €	0 €	20 €	A|B
4	10 €	0 €	25 €	A|B
5	10 €	0 €	30 €	A|B

In the decision task, the number of choices in favor of the safe machine was the indicator of risk aversion.

Sense of presence was measured with a 14-item scale (Schubert, 2003). The scale consists of three subscales plus one additional item that captures a general factor. The first subscale, spatial presence, assesses the sense of being physically present in the situation. The second subscale, involvement, measures the attention devoted to the situation and the involvement experienced. The third scale, experienced realism, measures the subjective experience of realism in the situation. The items had originally been phrased to be used in virtual environments, and we adapted them so they could also be used in the PPT condition.

The subscales yielded satisfactory reliability indicated by Cronbach's alpha for spatial presence ($\alpha = .76$), involvement ($\alpha = .78$), and experienced realism ($\alpha = .76$). When all items were aggregated into one sense of presence scale, this overall scale yielded a satisfactory reliability of $\alpha = .87$.

Locus of control was measured with the Locus of Control scale, German Version (FKK, Krampen, 1991). The secondary subscales created by summing the answers yielded a good internal consistency reliability indicated by a Cronbach's alpha for self-efficacy of $\alpha = .87$ and for externality of $\alpha = .78$. High values in self-efficacy indicate trust in one's own abilities and control, and low values indicate passivity and uncertainty in ambiguous situations. High values in externality indicate the belief that events are the result of situational forces and are often associated with fatalism, whereas low values indicate the belief that events are not due to situational forces and a low degree of dependence on others.

Conscientiousness was assessed using the NEO-FFI subscale (Borkenau and Ostendorf, 2008). The scale consisted of 12 items with a Cronbach's α of .88. High values indicate a high degree of conscientiousness.

Domain-specific risk perception was assessed using the Domain-Specific Risk-Taking (DOSPERT, Blais and Weber, 2006) scale. The DOSPERT assesses risk perception, benefit perception, and risk-taking in five domains: health/safety, finance, recreation, social, and ethical. In the current study, all domains were assessed but only with the risk perception items. It is important to note that we focused on the health/safety domain.

The health/safety domain yielded six items (Item 2 of the health/safety domain was left out because it referred to private matters of sexuality) and yielded good internal consistency reliability ($\alpha = .73$). High values in risk perception indicate risk aversion in the sense that the person judges the acts in this domain as risky.

Apparatus

The VR condition utilized the 3D-multisensory Cave Automatic Virtual Environment (CAVE) from the Technical University of Chemnitz. The system was cube-shaped and provided a five-sided projection in which participants could stand and move. Participants wore 3D glasses and used a wireless controller with which they were able to navigate in the VR environment. The participant's head and the controller's position could be used to adapt the visualization of the 3D images.

The PPT condition utilized a standard presentation projector and was presented in a computer lab classroom. The projection surface was 2 x 3 meters, and participants sat approximately 4 meters away from the projection.

Procedure

Upon their arrival, participants were greeted by the researchers and were given a general overview of how the day would go and the facilities they would see. Each participant received a personal schedule for the day, and each schedule contained a participant ID, which was randomly assigned to the VR or PPT condition. The timeline is presented in Fig. 2.

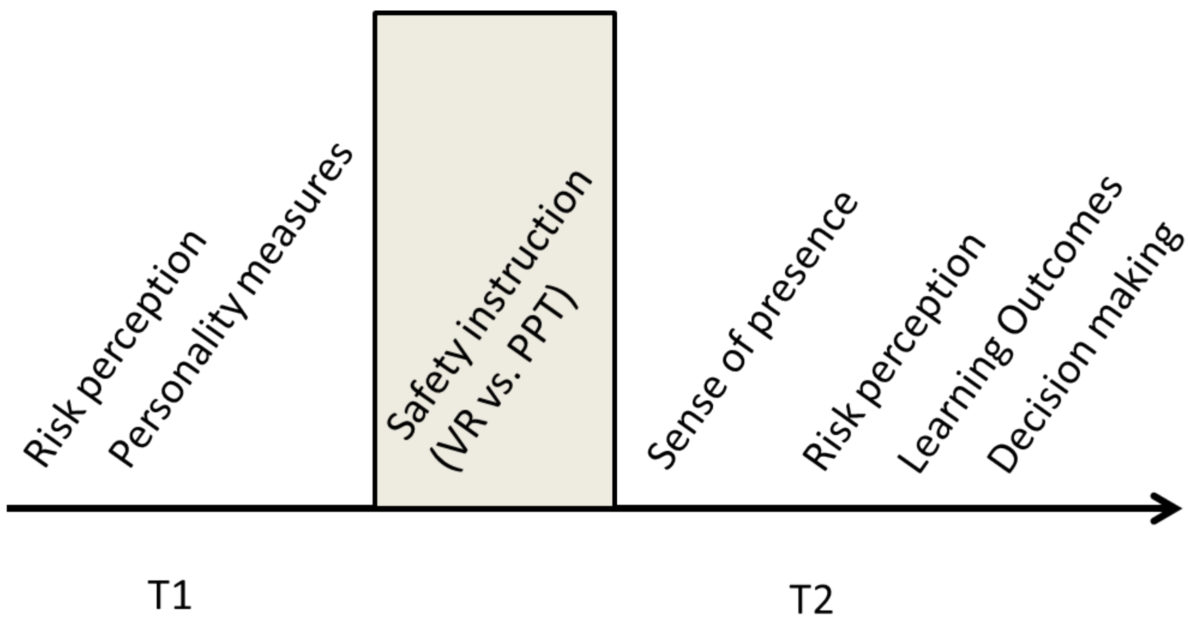

Fig. 2: Sequence of events for Study 1.

The participants were then guided to where they faced a real pillar power drill. Each participant received a clipboard on which the first risk perception measure was placed face down. Participants were instructed not to talk or interact with their neighbors while completing the questions. Then participants were told to fill out the paper on the clipboard. Two research assistants were present to ensure compliance. Then participants were guided to a separate room where they were asked to take a seat and to complete a questionnaire assessing demographic information, locus of control, and conscientiousness.

Participants were then individually led to either the CAVE or the computer lab to be given the safety training. During the waiting time, the participants took a tour of the premises. Participants in the VR condition were given the pillar drill safety training while standing in the CAVE. The safety training was presented as a VR animation with a corresponding voice recording. Participants in the PPT condition were given the pillar drill safety training in a PPT presentation in which screenshots from the VR safety training were shown. The audio comments were identical to those given in the VR training. Thus, the two conditions differed only with respect to immersion.

Participants then completed the sense of presence questionnaire and were guided to the main hallway. In the main hallway, they faced the real pillar drill and were asked to complete the risk perception measure. Afterwards, they were confronted with the decision-making task. Finally, after all participants had completed all stages of the study, they were all seated in the computer lab, had the opportunity to ask questions, and were thanked for their participation.

2.2. Results

Table 2 presents the descriptive results for all measured variables. The sense of presence was 3.01 in the VR condition compared with 2.13 in the PPT condition, $t(49.71) = -3.60$, $p < .001$, $d = 0.53$. A regression model predicting the sense of presence by condition was significant, *adj.* $R^2 = .18$, $F(1, 51) = 12.88$, $p < .001$. As expected, participants in the VR condition showed an increased sense of presence.

Table 2. Descriptive statistics and correlations.

	Total	PPT	VR	1	2	3	4	5	6	7	8	9	10	11	12	13
Age (1)	18.42 (±1.77)	18.93 (±2.14)	17.79 (±0.88)													
G (♀) (2)	0.17 (±0.38)	0.14 (±0.35)	0.21 (±0.41)	-.16												
PJ T1 (3)	3.30 (±2.16)	3.90 (±2.40)	2.56 (±1.60)	.27	.17											
PJ T2 (4)	4.02 (±2.18)	4.21 (±2.32)	3.79 (±2.03)	.01	-.03	.65***										
SJ T1 (5)	4.30 (±1.97)	4.83 (±2.19)	3.67 (±1.47)	.21	.11	.65***	.36**									
SJ T2 (6)	4.63 (±2.00)	4.83 (±2.02)	4.40 (±1.99)	-.03	-.08	.47***	.63***	.64***								
RI T2 (7)	6.25 (±2.49)	5.76 (±2.67)	6.83 (±2.16)	-.08	.12	-.13	-.16	-.11	-.05							
IH T2 (8)	0.29 (±0.19)	0.32 (±0.18)	0.25 (±0.20)	-.03	-.13	.31*	.37**	.16	.28*	.17						
RC (9)	0.58 (±0.99)	0.41 (±0.91)	0.79 (±1.06)	-.19	-.12	-.08	.01	-.05	-.01	-.22	-.12					
SofP (10)	2.53 (±0.99)	2.13 (±0.91)	3.01 (±0.88)	-.16	.03	-.16	.11	.08	.17	.2	.23	.19				
SE (11)	63.92 (±11.26)	64.28 (±10.59)	63.50 (±12.25)	.19	-.50***	-.1	.02	-.04	.12	-.03	.14	-.14	.13			
PC (12)	50.47 (±9.77)	50.97 (±10.09)	49.88 (±9.55)	-.22	.45***	-.11	-.07	-.18	-.2	.07	-.05	.12	-.12	-.67***		
C (13)	44.28 (±7.94)	44.79 (±6.08)	43.67 (±9.84)	-.04	-.1	-.14	-.08	-.01	.03	.11	.2	-.04	.32*	.43**	-.28*	
RP (14)	4.40 (±1.11)	4.29 (±1.04)	4.52 (±1.20)	-.1	.13	< .01	.15	.13	.03	-.01	.12	-.04	.23	-.07	.12	.21

	Total	PPT	VR	1	2	3	4	5	6	7	8	9	10	11	12	13

Note. G = Gender, PJ = Probability judgment, SJ = Severity judgment, RI = Recall of safety information, IH = Identified hazards, RC = Risky choice, SofP = Sense of presence, SE = Self-efficacy, PC = Personal control, C = Conscientiousness, RP = Risk perception health/safety domain. *$p < .05$. **$p < .01$. ***$p < .001$.

We used R (TeamR, 2013) and lme4 (Bates et al., 2014) to perform a linear mixed effects analysis of the effects of presentation format (VR vs. PPT) on the safety training outcome variables severity judgment, probability judgment, recall of safety information, and identified hazards. We carried out the analysis for each dependent variable and first tested the effect of the experimental treatment and participants' health. We then added safety-domain-specific risk perception, conscientiousness, self-efficacy, and personal control as covariates to control for person-specific effects.

In the models, we entered experimental condition, time of measurement, health and safety-domain-specific risk perception, conscientiousness, self-efficacy, and personal control as fixed effects. As random effects, we assumed random intercepts for participants, but we did not assume random slopes. We assumed a random intercept model because deviations from the grand mean (e.g., probability judgments at different time points) were of interest, and we had no predictors on Level 1 that would suggest random slopes. We obtained the *p*-values with the lmertest package (Kuznetsova et al., 2014). The intercept in the baseline model (i.e., model without predictors on Level 2) represents the grand mean. Intercepts in the subsequent models, including the predictors on Level 2, always represent the mean of the reference category, which assumes a value of 0 for all entered predictors. We tested multiple models in order to address our research questions. First, we tested models to investigate treatment effects (Models 1 to 3), and then we tested whether the effects depended on personality (Model 4).

Risk judgments

As can be seen in Fig. 3, the safety training led to an increase of judged severity and judged probability of accidents, and these variables were affected by the experimental condition.

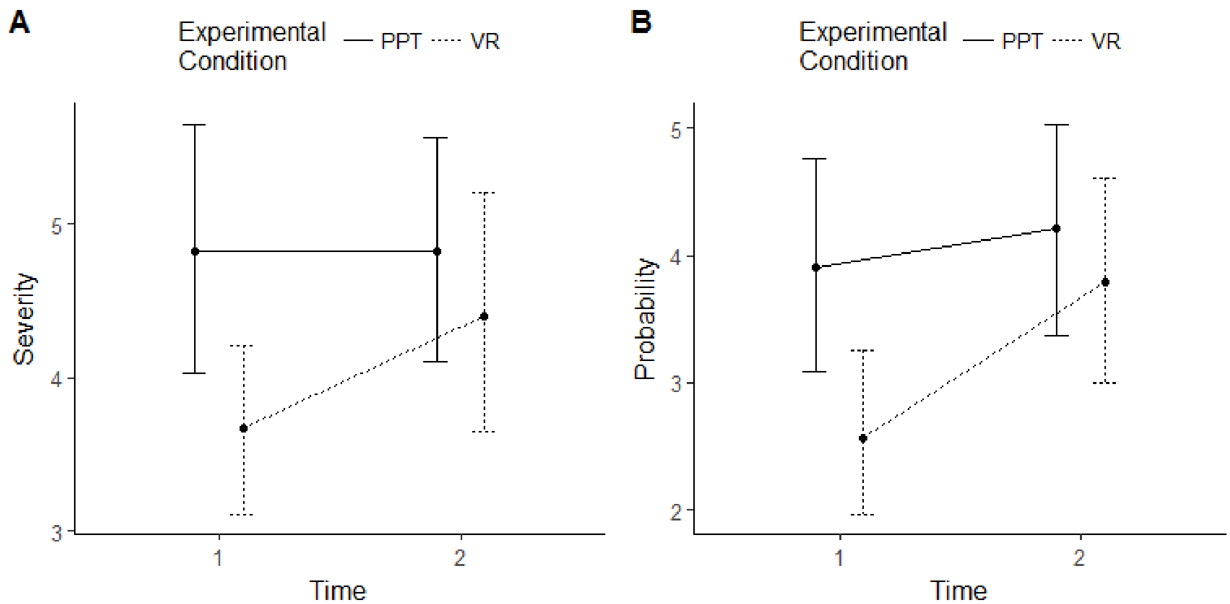

Fig. 3. Risk judgments before and after the safety training. Plot A depicting the severity judgment and Plot B depicting the probability judgment.

We found that the probability judgment of accidents was affected by the safety training as indicated by Models 1 and 2 in which time was a significant predictor, $b = 0.72$, $t(52) = 2.89$, $p = .01$. It is important to mention that Model 2 showed that the difference between experimental groups was not significant, $b = -0.88$, $t(51) = -1.64$, $p = .11$. It is also important to mention that the interaction between time and experimental condition indicated that the change in probability judgments from before to after the safety training (i.e., the slope) depended on the experimental condition, $b = 0.93$, $t(51) = 1.89$, $p = .06$. We found no main effects for any personality variables (all ps $\geq .10$, see Table 3) and no interactions between personality and the rate of change or the rate of change in each experimental group (all ps $>.18$). Simple effects were tested for the interaction between time and experimental condition and their effect on the probability judgment. In the VR condition, before the safety training was given, the mean judged probability of accidents was 2.56, but after the safety training, it increased to 3.79, which was a significant increase, $d = 0.67$, $t(23.00) = 3.39$, p = .003. This difference was not observed in the PPT condition (judged accident probability was 3.90 before the safety training and 4.21 afterwards, which was not a significant increase, $d = 0.13$, $t(28.00) = 0.92$, p = .37.

Table 3. Coefficients and model parameters for multilevel regression predicting the probability judgment.

Predictors	Probability judgment						
	Model 1 Est (95% CI)	Model 2 Est (95% CI)	Model 3 Est (95% CI)	Model 4 Est (95% CI)	Model 5 Est (95% CI)	Model 6 Est (95% CI)	Model 7 Est (95% CI)
Fixed Parts							
Int.	3.30 (2.71, 3.88) ***	3.69 (2.95, 4.44) ***	3.90 (3.13, 4.68) ***	3.10 (0.86, 5.34) **	4.73 (1.29, 8.17) **	7.45 (2.52, 12.38) **	9.97 (2.39, 17.55) *
T	0.72 (0.23, 1.21) **	0.72 (0.23, 1.21) **	0.30 (-0.34, 0.95)	0.31 (-0.35, 0.98)	0.31 (-0.35, 0.98)	0.31 (-0.35, 0.98)	0.31 (-0.35, 0.98)
VR		-0.88 (-1.93, 0.17)	-1.34 (-2.50, -0.19) *	-1.40 (-2.58, -0.22) *	-1.47 (-2.65, -0.28) *	-1.57 (-2.74, -0.39) *	-1.60 (-2.78, -0.42) *
VRxT			0.93 (-0.04, 1.89)	0.91 (-0.06, 1.89)	0.91 (-0.06, 1.89)	0.91 (-0.06, 1.89)	0.91 (-0.06, 1.89)
RP				0.19 (-0.30, 0.68)	0.26 (-0.24, 0.76)	0.34 (-0.16, 0.84)	0.32 (-0.18, 0.83)
C					-0.04 (-0.11, 0.03)	-0.06 (-0.13, 0.01)	-0.05 (-0.13, 0.03)
PC						-0.04 (-0.10, 0.01)	-0.06 (-0.14, 0.01)
SE							-0.03 (-0.10, 0.04)
Random Parts							
σ^2	1.659	1.659	1.581	1.612	1.612	1.612	1.612
τ_{00, vp_code}	3.057	2.934	2.973	3.059	3.02	2.926	2.947
N_{vp_code}	53	53	53	52	52	52	52
ICC_{vp_code}	0.648	0.639	0.653	0.655	0.652	0.645	0.646
Observations	106	106	106	104	104	104	104
R^2 / Ω_0^2	.827 / .792	.824 / .792	.835 / .807	.835 / .808	.834 / .808	.832 / .808	.832 / .809

Note. T = Time, VR = Experimental condition, RP = Risk perception health/safety domain, C = Conscientiousness, SE = Self-efficacy, PC = Personal control.
* p<.05. ** p<.01. *** p<.001.

The severity judgments of accidents were not affected by the safety training as indicated by Models 1 and 2 in which time was not a significant predictor, $b = 0.33$, $t(52) = 1.43$, $p = .16$. As Model 2 shows, the difference between experimental groups was not significant, $b = -0.8$, $t(51) = -1.63$, $p = .11$. It is important to mention that the interaction between time and experimental condition indicated that the change in probability judgments from before to after the safety training (i.e., the slope) did not depend on the experimental condition, $b = 0.73$, $t(51) = 1.59$, $p = .12$. Furthermore, we found that personal control (as measured with the FKK questionnaire) affected the overall severity judgment as indicated by Models 6 and 7 (see Table 4 for all coefficients), $b = -0.05$, $t(47) = -1.77$, $p = .08$. Participants with a high degree of internal locus of control generally perceived the severity of possible accidents to be less than participants with a low degree of internal locus of control.

Table 4. Coefficients and model parameters for multilevel regressions predicting the severity judgment.

Predictors	Severity judgment						
	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Est (95% CI)	Est (95% CI)	Est (95% CI)	Est (95% CI)	Est (95% CI)	Est (95% CI)	Est (95% CI)
Fixed Parts							
Int.	4.30 (3.77, 4.84) ***	4.66 (3.98, 5.35) ***	4.83 (4.12, 5.54) ***	3.99 (1.98, 6.01) ***	4.19 (1.06, 7.33) *	7.12 (2.66, 11.59) **	9.32 (2.45, 16.18) *
T	0.33 (-0.12, 0.78)	0.33 (-0.12, 0.78)	-0.00 (-0.60, 0.60)	0.04 (-0.58, 0.65)	0.04 (-0.58, 0.65)	0.04 (-0.58, 0.65)	0.04 (-0.58, 0.65)
VR		-0.80 (-1.75, 0.16)	-1.16 (-2.22, -0.10) *	-1.10 (-2.18, -0.03) *	-1.11 (-2.20, -0.03) *	-1.22 (-2.29, -0.15) *	-1.25 (-2.33, -0.17) *
VRxT			0.73 (-0.17, 1.63)	0.69 (-0.22, 1.60)	0.69 (-0.22, 1.60)	0.69 (-0.22, 1.60)	0.69 (-0.22, 1.60)
RP				0.17 (-0.27, 0.61)	0.18 (-0.27, 0.64)	0.27 (-0.19, 0.72)	0.25 (-0.21, 0.71)
C					-0.01 (-0.07, 0.06)	-0.02 (-0.09, 0.04)	-0.01 (-0.08, 0.05)
PC						-0.05 (-0.10, 0.00)	-0.07 (-0.13, 0.00)
SE							-0.03 (-0.09, 0.04)
Random Parts							
σ^2	1.413	1.413	1.372	1.39	1.39	1.39	1.39
$\tau_{00,\ vp\_code}$	2.531	2.432	2.452	2.434	2.498	2.361	2.382
N_{vp_code}	53	53	53	52	52	52	52
ICC_{vp_code}	0.642	0.632	0.641	0.637	0.643	0.63	0.632
Observations	106	106	106	104	104	104	104
R^2 / Ω_0^2	.822 / .783	.819 / .783	.827 / .794	.822 / .788	.822 / .790	.819 / .789	.818 / .790

Note. T = Time, VR = Experimental condition, RP = Risk perception health/safety domain, C = Conscientiousness, SE = Self-efficacy, PC = Personal control.
* p<.05. ** p<.01. *** p<.001.

However, this effect did not depend on the experimental condition because we found no evidence of an interaction, which was also the case for all personality variables (all *p*s < .16).

Learning

Participants' learning of the contents of the safety training was measured with a recall test and a test in which participants had to identify the hazards of a particular machine. Participants in the PPT condition had an average recall test score of $M = 5.76$ ($SD = 2.67$), whereas those in the VR condition had an average score of $M = 6.83$ ($SD = 2.16$). The difference was not significant, $b = 1.07$, $t(51) = 0.12$, $p = .12$. There were no main effects of the personality measures (for all coefficients, see Table 5) and no significant interaction.

Table 5. Coefficients and model parameters recall of safety information.

Predictors	Recall of safety information				
	Model 1 *Est (95% CI)*	Model 2 *Est (95% CI)*	Model 3 *Est (95% CI)*	Model 4 *Est (95% CI)*	Model 5 *Est (95% CI)*
Int.	5.76 (4.84, 6.67) ***	6.09 (3.22, 8.97) ***	4.50 (0.03, 8.96) *	2.31 (-4.21, 8.83)	2.62 (-7.50, 12.74)
VR	1.07 (-0.28, 2.43)	1.03 (-0.37, 2.42)	1.09 (-0.32, 2.50)	1.17 (-0.25, 2.59)	1.17 (-0.27, 2.61)
RP		-0.06 (-0.70, 0.57)	-0.13 (-0.78, 0.52)	-0.20 (-0.86, 0.47)	-0.20 (-0.87, 0.48)
C			0.04 (-0.05, 0.13)	0.06 (-0.04, 0.15)	0.06 (-0.04, 0.16)
PC				0.04 (-0.04, 0.11)	0.03 (-0.07, 0.13)
SE					-0.00 (-0.09, 0.09)
Observations	53	52	52	52	52
R^2 / adj. R^2	.047 / .028	.043 / .004	.060 / .001	.077 / -.002	.077 / -.023

Note. VR = Experimental condition, RP = Risk perception health / safety domain, C = Conscientiousness, PC = Personal control, SE = Self-Efficacy. * *p<.05.* ** *p<.01.* *** *p<.001.*

When presented with the real pillar drill, participants in the PPT condition identified $M = 0.32$ ($SD = 0.18$) of the given hazards, whereas those in the VR condition identified $M = 0.25$ ($SD = 0.20$), which was not a significant difference, $b = -0.06$, $t(51) = 0.22$, $p = .22$. There was no main effect of the personality measures (see Table 6 for all coefficients), but the number of identified hazards was affected by an interaction between condition and self-efficacy, $b = 0.01$, $t(48) = 2.15$, $p = .04$.

Table 6. Coefficients and model parameters predicting identified hazards.

Predictors	Identified hazards				
	Model 1	Model 2	Model 3	Model 4	Model 5
	Est (95% CI)	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*
Int.	0.32 (0.25, 0.39) ***	0.23 (0.01, 0.44) *	0.08 (-0.25, 0.41)	0.14 (-0.35, 0.63)	0.01 (-0.75, 0.77)
VR	-0.06 (-0.17, 0.04)	-0.08 (-0.18, 0.03)	-0.07 (-0.18, 0.04)	-0.07 (-0.18, 0.03)	-0.07 (-0.18, 0.04)
RP		0.02 (-0.02, 0.07)	0.02 (-0.03, 0.07)	0.02 (-0.03, 0.07)	0.02 (-0.03, 0.07)
C			0.00 (-0.00, 0.01)	0.00 (-0.00, 0.01)	0.00 (-0.00, 0.01)
PC				-0.00 (-0.01, 0.00)	0.00 (-0.01, 0.01)
SE					0.00 (-0.01, 0.01)
Observations	53	52	52	52	52
R^2 / adj. R^2	.029 / .010	.054 / .016	.080 / .022	.082 / .004	.086 / -.013

Note. VR = Experimental condition, RP = Risk perception health / safety domain, C = Conscientiousness, PC = Personal control, SE = Self-Efficacy. *p < .05. **p < .01. ***p < .001.

The simple slope analysis of the interaction between condition and self-efficacy for identified hazards showed that in the VR condition, participants with high scores in self-efficacy identified more hazards, whereas this was not the case in the PPT condition, *b* = 0.01, $t(48)$= 2.20, $p = .03$. The simple slopes are depicted in Fig. 4.

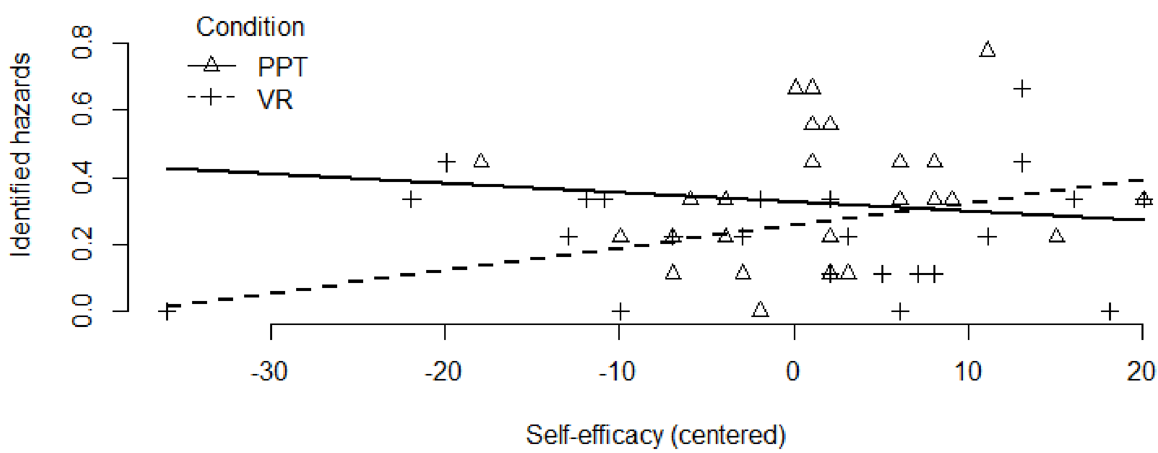

Fig. 4. Fitted simple slopes for the effect of self-efficacy on identified hazards moderated by condition.

Decision making

For the risky decision, participants had to repeatedly choose between the safe and unsafe option, and the payoff of the unsafe option increased with each decision. We were interested in identifying the level of the monetary reward at which the risky option would be preferred over the safe option. Fig. 4 shows that with the increasing payoff of the risky option, participants became more likely to choose the risky option.

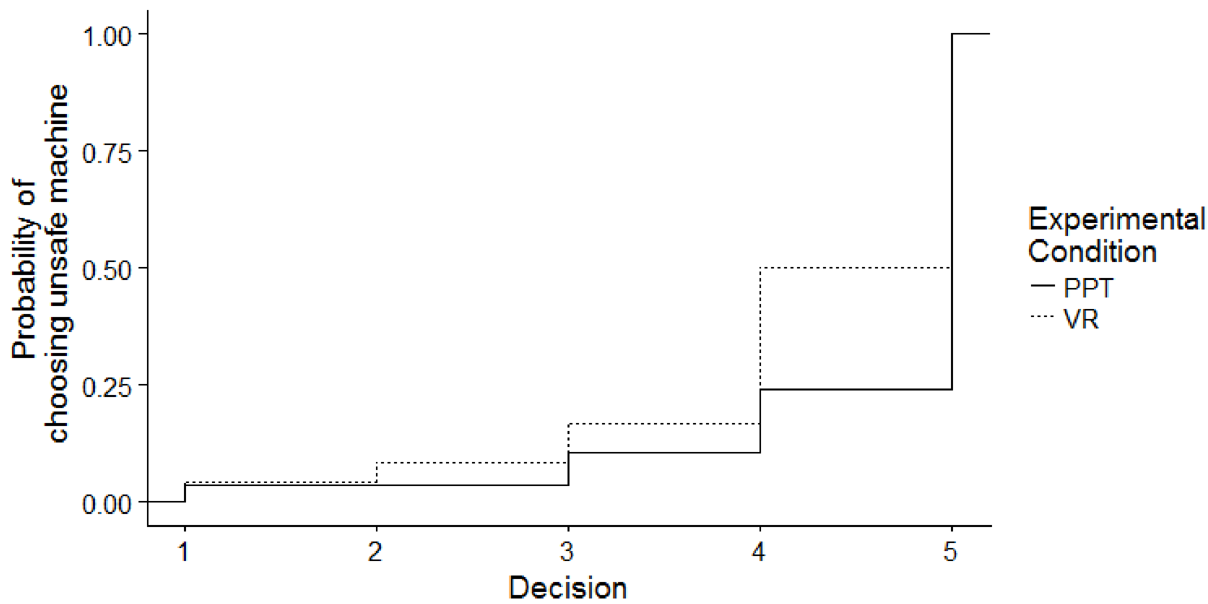

Fig. 5. Proportion of unsafe choices in each decision: Data averages of participants for each experimental condition.

The experimental condition did not significantly affect decision making, $B = -0.57$, $t(48) = -1.71$, $p = .09$. It is important to mention that, when we controlled for all of the personality variables that had been assessed (Model 5), the participants in the VR condition were significantly less risk averse than those in the PPT condition, $B = -0.81$, $t(44) = -2.19$, $p = .03$. Furthermore, participants who recalled more safety information made safer choices (i.e., they exhibited more risk-averse behavior), $B = -1.04$, $t(42) = 2.03$, $p = .04$. All models and their respective coefficients with 95% confidence intervals are presented in Table 7.

Table 7. Ordered probit regression predicting safe choices.

	Safe choices						
Predictors	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6	Model 7
	Est (95% CI)	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*
VR	-.571 (-1.228, .085)	-.685* (-1.363, -.007)	-.678 (-1.365, .008)	-.805* (-1.525, -.086)	-.801* (-1.524, -.077)	-1.041** (-1.829, -.252)	-1.025* (-1.828, -.222)
RP		.027 (-.284, .338)	.022 (-.297, .342)	.069 (-.262, .400)	.070 (-.264, .403)	.119 (-.228, .466)	.113 (-.238, .465)
C			.003 (-.040, .045)	-.012 (-.058, .035)	-.017 (-.067, .033)	-.031 (-.084, .023)	-.032 (-.086, .023)
PC				-.033 (-.073, .006)	-.024 (-.074, .025)	-.033 (-.085, .019)	-.033 (-.085, .019)
SE					.014 (-.031, .058)	.013 (-.032, .058)	.013 (-.032, .058)
RI						.157* (.005, .310)	.155* (.001, .309)
IH							.228 (-1.874, 2.329)
Observations	53	52	52	52	52	52	52

Note. VR = Experimental condition, RP = Risk perception health/safety domain, C = Conscientiousness, PC = Personal control, SE = Self-efficacy, RI = Recall of safety information, IH = Identified hazards. $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

We tested the interaction of both learning outcomes with the condition and found no interaction of any predictor with the experimental condition. This was also the case for the personality variables (all *p*s > .18).

2.3. Discussion

Participants in Study 1 clearly experienced a greater degree of immersion in the immersive VR condition than in the PPT condition which corroborates hypothesis 1. But the effect of the experimental condition on other outcome measures was not as clear. We did find some evidence in support of the hypothesis that VR presents a more powerful tool for delivering a safety training than PPT. Participants who were informed about machine-related hazards and possible accidents in an immersive VR environment made higher risk judgments than those who had received the information via PPT. However, the effect was significant only for the judgment of probabilities but was not significant for the severity judgment. These finding are in partial support of Hypothesis 2. It is important to mention that the amount of safety information recalled and the knowledge of safety hazards did not differ between the two experimental conditions the findings refute Hypothesis 3.

We found that when participants were given information about machine-related hazards and safety procedures via an immersive VR environment, they were later less risk averse in their decisions than participants who had received the same information via a PPT presentation. Furthermore, the amount of safety information participants were able to recall predicted risky choice only in the VR condition such that, recalling more safety information led to more risk aversion. By contrast, in the PPT condition, the amount of safety information participants were able to recall was unrelated to participants' decisions. It seems that immersive VR leads to more integration between knowledge and behavior which is in line with Hypothesis 4. There was no evidence that the effects of the medium through which the information was delivered were affected by participants' conscientiousness or locus of control.

Overall, the findings are in line with the mixed results found in the literature (Gavish et al., 2015; Moreno and Mayer, 2002; Zaalberg and Midden, 2013), where a medium with higher immersion was not necessarily associated with larger changes of risk judgments or learning.

The results show that the safety training was effective in the immersive VR condition and apparently led to a change in risk perception. This assumption needs to be qualified. First, participants in the immersive VR conditions indicated very low pretest scores in the risk judgments; thus, the finding could also be due to a regression to the mean, since we did not find a difference in health and safety domain risk perception. While we did not find difference in the health and safety domain specific risk perception, the higher average level of risk seeking observed in the decision task in the immersive VR condition, might reflect the group's risk attitude, which was not affected by the training. The information concerning the unsafe machines properties might not have been completely clear when they stated that the probability of accidents was 50%, which could have led participants to assume that their decisions were in reference to a hypothetical machine. In the present study, we assessed only the domain-specific risk perception and not risk taking or domain specific risk attitudes, of all participants and found no relationship with their risk choice; however, domain-specific risk-taking might offer another explanation and was assessed in Study 2.

In order to rule out the explanation that risky choices were made about a hypothetical machine or that risky choices were the result of domain-specific risk-taking as well as to replicate the findings from Study 1 and investigate whether the effects were still present after a time lag of 6 months, we carried out Study 2.

3. Study 2

Study 2 employed the same procedure as Study 1. In Study 2, we altered the wording of the decision task to make the task clearer, and the full DOSPERT was assessed to include risk benefit perceptions, risk perceptions, as well as general risk-taking so that these could be compared with individuals' domain-specific risk attitudes. Furthermore, risk judgments and learning outcomes were assessed again in a follow up 6 months after the safety training had been given. Thus, the study allowed us to replicate Study 1, to test whether risk attitudes and risk-taking in the health and safety domain moderated the effect of the presentation medium, and to see whether the positive effects of the safety training persisted across time.

3.1. Method

Participants and Design

The study was carried out with $N = 68$ participants who participated at T1 and T2. Female participants comprised, 36.76% of the sample, the mean age was 17.18, and 67.65% of the participants had previously experienced VR or 3D movies. At T3, the sample size was reduced slightly to $N = 60$. Participants were randomly assigned to either the PPT condition or the VR condition, and the participants in the two conditions did not differ regarding their experience with VR or 3D movies, $\chi^2(1) = 0.06$, $p = .80$. Participants were asked to rate the probability of accidents and the severity of accidents of a real pillar drill that was presented to them before they were given the safety training (T1), after the safety training (T2), plus 6 months later (T3). Thus, we could compare their risk perception before (T1), directly after (T2), and 6 months after the safety training (T3). Learning was assessed directly after (T2) and 6 months after (T3) the safety training. Decision making was assessed directly after the safety training (T2). Furthermore, participants' sense of presence during the safety training and their domain-specific risk attitudes were measured.

Material and Measures

Study 2 was carried out at the same facility and utilized the same technology as Study 1. Furthermore, the safety training and the measures of risk perception and learning directly after the safety training and sense of presence questionnaire (Schubert, 2003) were identical to Study 1. Study 2 deviated from Study 1 in that we altered the decision-making task and employed the full DOSPERT (Blais and Weber, 2006). We also used a visual search task at T3, showing a picture of the machine taken at T2 instead of presenting a real machine to assess how many hazards associated with a pillar drill were still identified 6 months after the safety training had taken place.

Learning outcomes (i.e., knowledge about hazards) at T3, was assessed by presenting participants with a photograph of a pillar drill on a screen and asking them to mark the areas that contained hazards. Participants were informed that their score in this task would be computed as the ratio of correctly identified sources of hazards to existing hazards divided by total number of marks made. Participants were given a 3-min time limit for this task, and after the task, they were given feedback on their performance. The proportion of correctly identified hazards (i.e., identified hazards / all hazards present) was the dependent variable, as in the assessment that was administered directly after the safety training had taken place.

Decision-making was assessed by asking participants to repeatedly choose between working on the real machine, which yielded safety hazards (referred to as the risky choice), and a fictitious safe machine (referred to as the safe choice). The important difference from Study 1 was that in Study 1, participants were told that the likelihood of an accident was 50% when they chose the real machine. *In Study 2, participants were not given any probabilities*, so they were left uncertain, but they were told that the probabilities would resemble the real machine they had seen before. As in Study 1, participants again had to make five choices, and the non-zero prospect of the uncertain option increased from 10 € in the first decision to 30 € in the final choice. The number of choices in favor of the safe option was the measure of risk-taking behavior. Following best practice in experimental economics, the decision-making task was incentivized. Participants were told that at the end of the experimental session, 30% of all participants would be randomly selected, and one of their decisions would be randomly selected and played out (Harrison and Rutström, 2008).

Domain specific risk attitudes were assessed using the DOSPERT (Blais and Weber, 2006). The DOSPERT assesses risk perception, benefit perception, and risk-taking in five domains: health/safety, finance, recreation, social, and ethical. All domains and scales were assessed in Study 2. Risk perception measures how risky a certain act is perceived to be, risk-taking measures a person's likelihood of engaging in the described act, and benefit perception measures the assumed benefit of an act that is considered risky.

As in Study 1, we report only the results concerning the health/safety domain. The health/safety domain included six items (Item 2 from the health/safety domain was left out because it referred to private matters of sexuality) and yielded good internal reliabilities for risk perception ($\alpha = .76$), risk-taking ($\alpha = .63$), and risk benefit ($\alpha = .63$).

We used the risk benefit, risk perception, and risk-taking responses to assess individuals' risk attitude (Weber et al., 2002). Individuals' risk attitude was the result of a regression carried out for each individual, predicting their risk-taking based on the risk benefit and risk perception ratings. We regressed the preference for an act X on its expected benefit and perceived risk, that is,

$$Preference\ (X) = a(Expected\ Benefit(X)) + b(Perceived\ Risk(X)) + c \quad (1)$$

, for each participant. The regression coefficient b for the risk perception then becomes the measure of risk attitude: If $b < 0$, then this indicates risk aversion; if $b > 0$, then risk seeking.

Procedure

The procedure used in Study 2 differed from the one used in Study 1 in two ways: First, for logistical reasons, the domain-specific risk attitudes assessed with the DOSPERT were assessed after the safety training. Second, the assessment at T3 was computer-based and conducted in the computer rooms at the participating schools. Here, students were supervised by their teacher, and participants' email addresses were entered into a raffle for 10 Amazon gift certificates each worth 20 € as compensation for their participation. At T3, risk perception and learning outcomes were assessed. The procedure is presented in Fig. 6.

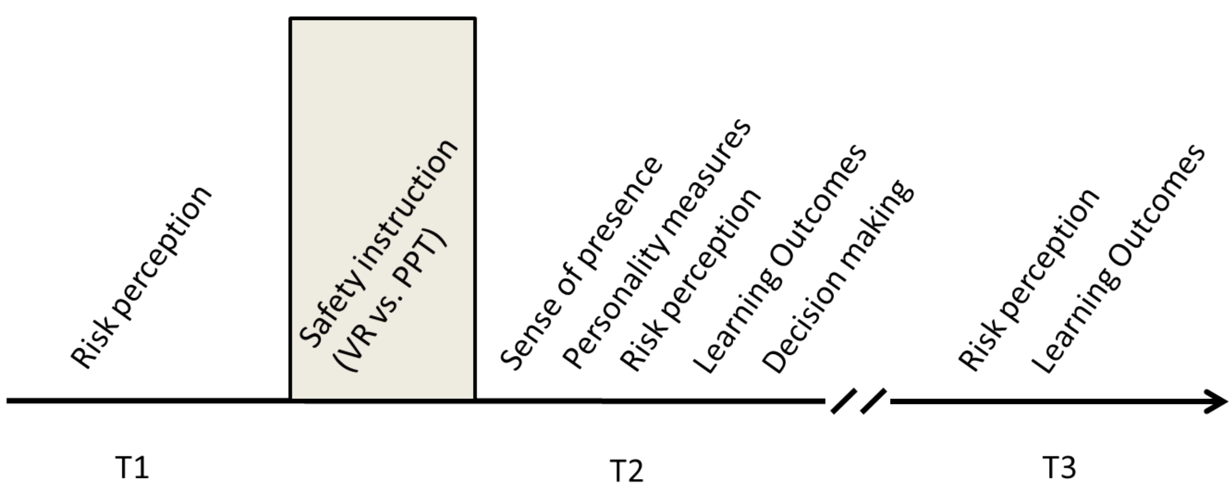

Fig. 6. Sequence of events during Study 2.

3.2. Results

We found that sense of presence in the VR condition was 2.98 (0.88) and was therefore higher than 2.57 (0.59) in the PPT condition, t = -2.20, df = 59.21, p = .032, d = 0.53. A regression model predicting the sense of presence by condition was significant, $adj.R^2$ = .05, F (1, 67) = 4.93, p = .03. As expected, participants in the VR condition had an increased sense of presence. Descriptive statistics and correlations of all measured variables are presented in Table 8.

Table 8. Descriptive statistics and correlations.

	Means and SDs			Correlations														
	Total	PPT	VR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Age (1)	17.16 (±1.98)	17.09 (±2.32)	17.24 (±1.60)															
G (♀) (2)	0.36 (±0.48)	0.34 (±0.48)	0.38 (±0.49)	-.11														
PJ T1 (3)	3.73 (±1.88)	3.34 (±1.59)	4.14 (±2.08)	-.1	.19													
PJ T2 (4)	4.40 (±2.10)	4.09 (±2.03)	4.72 (±2.16)	-.06	-.04	.35**												
PJ T3 (5)	6.63 (±2.66)	6.70 (±2.68)	6.55 (±2.69)	.15	-.11	-.14	-.39**											
SJ T1 (6)	4.01 (±1.81)	3.94 (±1.88)	4.07 (±1.76)	-.03	.09	.49***	.25*	-.17										
SJ T2 (7)	4.43 (±2.06)	4.43 (±2.03)	4.44 (±2.13)	.13	-.06	.27*	.75***	-.25	.40***									
SJ T3 (8)	5.88 (±2.03)	6.00 (±1.93)	5.76 (±2.15)	.03	.05	-.07	-.36**	.56***	-.30*	-.34**								
RI T2 (9)	7.64 (±1.32)	7.86 (±1.40)	7.41 (±1.21)	-.19	.07	.04	-.26*	.42**	-.03	-.17	.1							
RI T3 (10)	4.90 (±1.5	4.83 (±1.6	4.97 (±1.5	.02	-.31*	-.07	-.05	.03	-.13	-.12	.01	.14						

	Means and SDs			Correlations														
	Total	PPT	VR	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	5)	2)	0)															
IH T2 (11)	.39 (±.19)	.37 (±.18)	.42 (±.20)	.01	-.06	.02	.07	.08	-.09	.11	-.09	-.01	.09					
IH T3 (12)	.40 (±.19)	.37 (±.21)	.44 (±.16)	-.17	.08	-.14	.07	.18	-.01	.06	.15	.09	.1	.23				
RC (13)	1.88 (±1.37)	1.69 (±1.45)	2.09 (±1.26)	.09	-.09	.07	.09	.21	-.17	.08	.28*	.05	-.1	.05	.09			
SofP (14)	2.77 (±.78)	2.57 (±.59)	2.98 (±.89)	-.13	.08	.09	.11	-.03	-.06	-.07	.37**	-.07	.15	-.06	.13	-.02		
RA (15)	-.60 (±1.07)	-.74 (±.84)	-.46 (±1.27)	.08	.03	.02	.12	-.2	-.09	.02	-.21	-.33**	-.2	0	-.05	.06	-.01	
RT (16)	3.51 (±1.07)	3.45 (±.89)	3.57 (±1.24)	-.18	.17	.22	-.03	.03	.31*	.06	-.05	.28*	.2	-.03	.16	-.22	-.03	-.41***

Note. G = Gender, PJ = Probability judgment, SJ = Severity judgment, RI = Recall of safety information, IH = Identified hazards, RC = Risky choice, SofP = Sense of presence, RA = Risk attitude health/safety domain, RT= Risk-taking health/safety domain
**p* < .05. ***p* < .01. ****p* < .001.

We used R (TeamR, 2013) and lme4 (Bates et al., 2014) to perform a linear mixed effects analysis of the effects of presentation format (VR vs. PPT) on the safety training outcome variables severity judgment, probability judgment, recall of safety information, and identified hazards. We conducted the analysis for each dependent variable and first tested the effect of the experimental treatment and then added sense of presence, participants' health-and-safety-domain-specific risk attitude, as well as participants' health-and-safety-domain-specific risk-taking to the models as covariates to control for person-specific effects.

In the models, we entered experimental condition, time of measurement, sense of presence, health-and-safety-domain-specific risk attitude, as well as health-and-safety-domain-specific risk-taking as fixed effects. As random effects, we assumed random intercepts for participants, but we did not assume random slopes. We assumed a random intercept model because deviations from the grand mean (e.g., probability judgments at different time points) were of interest, and we had no predictors on Level 1 that, would suggest random slopes. We obtained *p*-values with the lmertest package (Kuznetsova et al., 2014). The intercept in the baseline model (i.e., the model without predictors on Level 2) represents the grand mean. Intercepts in the subsequent models including predictors on Level 2 always represent the mean of the reference category, which assumes a value of 0 for all entered predictors. We tested multiple models in order to address our research questions. First, we tested models to investigate treatment effects (Models 1 to 3) and then tested whether the effects depended on personality (Model 4).

Risk judgments

As can be seen in Fig. 7, the safety training led to an increase in risk judgment in terms of judged likelihood and judged severity of accidents, but the presentation format had no effect.

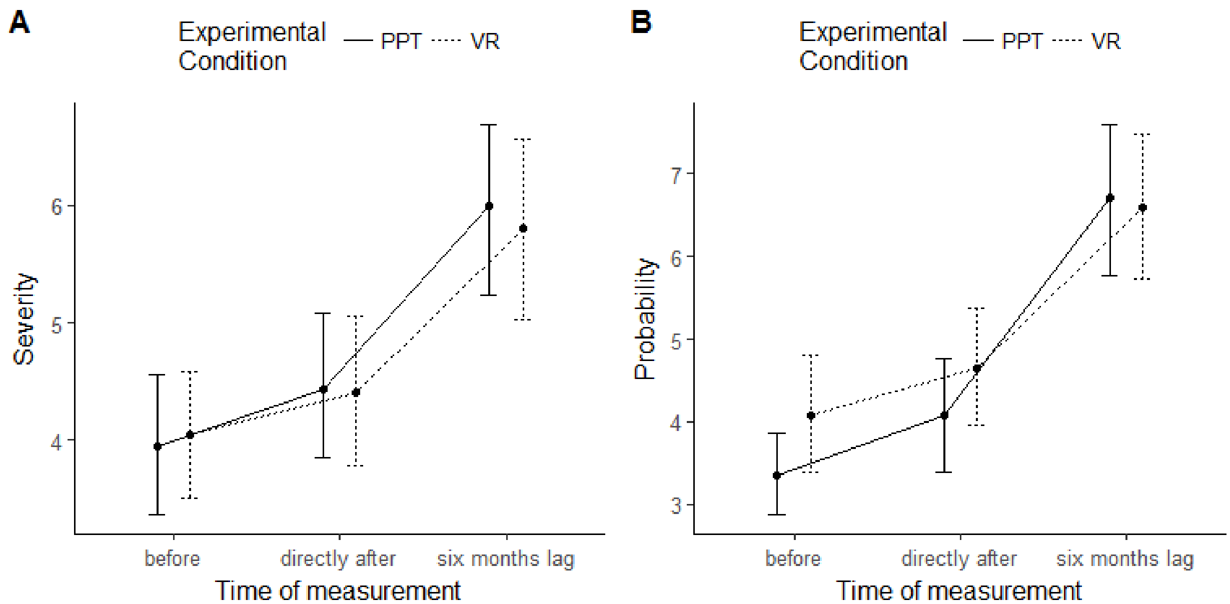

Fig. 7. Risk judgment before and after the safety training (Plot A depicting the severity rating and Plot B depicting the probability rating).

We found that the judgment of the likelihood of accidents was affected by the safety training across all models as indicated by time as a significant predictor, $b = 1.43$, $t(194) = 7.18$, $p < .001$. There was no effect of the condition or domain-specific risk attitudes or risk-taking (all coefficients, see Table 9).

Table 9. Coefficients and model parameters for multilevel regression predicting the probability judgment.

Predictors	Probability judgment			
	Model 1	Model 2	Model 3	Model 4
	Est (95% CI)	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*
Fixed Parts				
Int.	3.47 (2.99, 3.96) ***	3.25 (2.68, 3.83) ***	3.04 (2.36, 3.72) ***	2.65 (1.37, 3.92) ***
T	1.43 (1.04, 1.82) ***	1.43 (1.04, 1.82) ***	1.65 (1.11, 2.20) ***	1.65 (1.11, 2.20) ***
VR		0.45 (-0.18, 1.08)	0.89 (-0.08, 1.86)	0.92 (-0.06, 1.90)
VRxT			-0.46 (-1.24, 0.32)	-0.46 (-1.24, 0.32)
RA				-0.11 (-0.41, 0.19)
RT				0.09 (-0.21, 0.39)
Random Parts				
σ^2	5.025	5.000	4.991	5.022
$\tau_{00,\ vp\_code}$	0.000	0.000	0.000	0.000
N_{vp_code}	69	69	69	69
ICC_{vp_code}	0.000	0.000	0.000	0.000
Observations	196	196	196	196
R^2 / Ω_0^2	.210 / .210	.218 / .218	.223 / .223	.227 / .227

Note. T = Time, VR = Experimental condition, RA = Risk attitude health/safety domain, RT= Risk-taking health/safety domain. * *p<.05.* ** *p<.01.* *** *p<.001.*

As for the probability judgment, we found that the severity judgment of accidents was affected by the safety training across all models, as indicated by time as a significant predictor, $b = 0.92$, $t(195) = 5.27$, $p < .001$. There was no effect of the condition, domain-specific risk attitudes, or risk-taking (all coefficients, see Table 10).

Table 10. Coefficients and model parameters for multilevel regression predicting the severity judgment.

Predictors	Severity judgment			
	Model 1	Model 2	Model 3	Model 4
	Est (95% CI)	Est (95% CI)	Est (95% CI)	Est (95% CI)
Fixed Parts				
Int.	3.84 (3.42, 4.27) ***	3.85 (3.35, 4.36) ***	3.77 (3.16, 4.37) ***	3.17 (2.04, 4.29) ***
T	0.92 (0.58, 1.27) ***	0.92 (0.58, 1.27) ***	1.01 (0.53, 1.50) ***	1.02 (0.53, 1.50) ***
VR		-0.02 (-0.58, 0.53)	0.15 (-0.70, 1.01)	0.18 (-0.68, 1.04)
VRxT			-0.18 (-0.87, 0.50)	-0.17 (-0.86, 0.51)
RA				-0.17 (-0.43, 0.09)
RT				0.14 (-0.13, 0.40)
Random Parts				
σ2	3.908	3.928	3.942	3.932
τ00, vp_code	0.000	0.000	0.000	0.000
Nvp_code	69	69	69	69
ICCvp_code	0.000	0.000	0.000	0.000
Observations	197	197	197	197
R2 / Ω02	.125 / .125	.125 / .125	.126 / .126	.137 / .137

Note. T = Time, VR = Experimental condition, RA = Risk attitude health/safety domain, RT= Risk-taking health/safety domain. * p<.05. ** p<.01. *** p<.001.

These findings show that the safety training affected risk perception even after 6 months, but there was no difference in whether the training was carried out using PPT or VR. It is important to mention that the intra-class correlation of zero indicated that the participants' responses at each measurement point were not affected by inter-individual variation, and there was no variation at the intercept, which means the trajectory over time was similar for all participants' risk judgments and the variation is explained by the residual variance term alone.[1]

Learning

How well the participants had learned the contents of the safety training was measured with a recall test and a test in which participants had to identify the hazards of a particular machine. Recall of safety information was slightly higher in the PPT condition, $M = 7.86$ ($SD = 1.40$), than in the VR condition, $M = 7.41$ ($SD = 1.21$), directly after the safety training. However, the average of the recalled safety information decreased over time in both experimental conditions, and scores were similar to each other after 6 months; while the PPT condition had $M = 4.83$ ($SD = 1.62$), the VR condition yielded $M = 4.97$ ($SD = 1.50$). The

---

[1]We also tested a random slope model allowing for an cross-level interaction of time and subject, which allows for differences in change over time. It shows that the lack of variation on level 2 is explained by the perfect correlation of -1 between intercept and slope. Participants with lower than average risk judgments had a positive slope (i.e. they increased their risk judgment over time), while participants with a higher than risk judgment showed a negative slope. Importantly, the random slope models did not result in better fit – for this reason we only report the more parsimonious random intercept models.

amount of safety information that was recalled decreased over time, b = -2.75, t(65.36) = -11.6, p < .001. Furthermore, health-and-safety-domain-specific risk attitude, b = -0.26, t(67.08) = -2.16, p = 0.03, and risk-taking, b = -3.02, t(64.95) = -9.13, p < .001, affected the amount of information that was recalled. The lower participants' risk aversion, the less information they recalled. Furthermore, participants who are more risk seeking in their behavior in the health/safety domain were also more likely to recall less information. There was no effect of experimental condition, b = -0.19, t(66.04) = -0.71, p = 0.48 (for all coefficients, see Table 11). There was no interaction between domain-specific risk attitude or risk-taking with the experimental condition, all ps > .5, and no interaction between domain-specific risk attitude or risk-taking with time, all ps > .16.

Table 11. Coefficients and model parameters for multilevel regression predicting the amount of safety information that was recalled.

Predictors	Recall of safety information			
	Model 1	Model 2	Model 3	Model 4
	Est (95% CI)	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*
Fixed Parts				
Int.	7.64 (7.30, 7.97) ***	7.73 (7.31, 8.16) ***	7.86 (7.38, 8.33) ***	8.51 (7.56, 9.46) ***
T	-2.75 (-3.21, -2.28) ***	-2.75 (-3.21, -2.29) ***	-3.03 (-3.68, -2.37) ***	-3.02 (-3.67, -2.37) ***
VR		-0.19 (-0.72, 0.34)	-0.45 (-1.12, 0.23)	-0.32 (-0.97, 0.32)
VRxT			0.56 (-0.37, 1.49)	0.51 (-0.41, 1.44)
RA				-0.32 (-0.54, -0.09) **
RT				-0.26 (-0.49, -0.02) *
Random Parts				
σ^2	1.767	1.761	1.760	1.764
$\tau_{00,\ vp\_code}$	0.275	0.291	0.285	0.100
N_{vp_code}	69	69	69	69
ICC_{vp_code}	0.134	0.142	0.139	0.054
Observations	128	128	128	128
R^2 / Ω_0^2	.611 / .605	.617 / .611	.620 / .614	.589 / .588

Note. T = Time, VR = Experimental condition, RA = Risk attitude health/safety domain, RT= Risk-taking health/safety domain. * *p<.05.* ** *p<.01.* *** *p<.001.*

Regarding the number of hazards identified in the PPT condition, we observed the following means: for identified hazards t2: $M = 0.37$ ($SD = 0.18$) and t3: $M = 0.42$ ($SD = 0.20$), which was similar to the VR condition with t2: $M = 0.37$ ($SD = 0.21$) and at t3: $M = 0.44$ ($SD = 0.16$). Overall, the proportion of identified hazards differed between groups, and more hazards were identified in the VR condition than in the PPT condition, a difference that was marginally significant, $b = 0.06$, $t(65.59) = 1.69$, $p = 0.1$. The number of hazards that participants identified directly after the safety training was not different than 6 months later, which was indicated by the fact that there was no effect of time, $b = 0.01$, $t(64.61) = 0.32$, $p = 0.75$. Finally, there was no evidence that health/safety-domain-specific risk attitude or risk-taking influenced the likelihood of identifying hazards (for all coefficients, see Table 12).

Table 12. Coefficients and model parameters for multilevel regression predicting the proportion of identified hazards.

Predictors	Identified hazards			
	Model 1	Model 2	Model 3	Model 4
	Est (95% CI)	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*
Fixed Parts				
Int.	0.39 (0.35, 0.44) ***	0.36 (0.31, 0.42) ***	0.37 (0.30, 0.43) ***	0.35 (0.21, 0.49) ***
T	0.01 (-0.05, 0.07)	0.01 (-0.05, 0.07)	0.00 (-0.08, 0.09)	0.00 (-0.08, 0.09)
VR		0.06 (-0.01, 0.13)	0.06 (-0.03, 0.15)	0.06 (-0.03, 0.15)
VRxT			0.01 (-0.11, 0.13)	0.01 (-0.11, 0.13)
RA				-0.01 (-0.04, 0.03)
RT				0.00 (-0.03, 0.04)
Random Parts				
σ^2	0.028	0.028	0.029	0.029
$\tau_{00,\ vp\_code}$	0.008	0.007	0.007	0.008
N_{vp_code}	69	69	69	69
ICC_{vp_code}	0.224	0.206	0.198	0.211
Observations	128	128	128	128
R^2 / Ω_0^2	.629 / .369	.549 / .360	.542 / .352	.548 / .367

Note. T = Time, VR = Experimental condition, RA = Risk attitude health/safety domain, RT= Risk-taking health/safety domain. * *p<.05.* ** *p<.01.* *** *p<.001.*

Decision making

In the decision making task, participants had to repeatedly choose between the safe and unsafe options, and the payoff from the unsafe option increased with each decision. We were interested in identifying the level of monetary reward at which the risky option was preferred over the safe option and in whether this level differed between experimental conditions, which would indicate differences in risk aversion. Fig. 7 shows that as the payoff of the risky option increased, participants became more likely to choose the risky option.

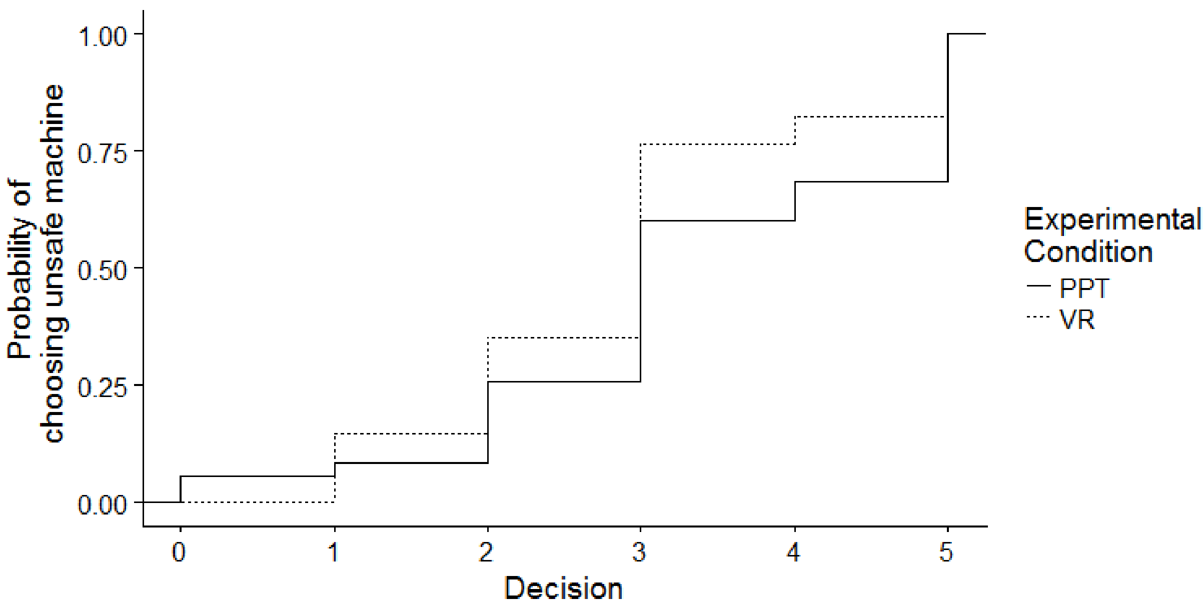

Fig. 8. Proportion of Unsafe Choices in Each Decision: Data Averages for participants in each experimental condition.

We used an orderd probit regression from the MASS package (Ripley et al., 2011) to test whether the points at which the participants changed their minds differed between groups and whether this was affected by domain-specific risk-taking and risk attitudes as well as the two learning outcomes (i.e., recalled safety information and identified hazards; see Table 13).

Table 13. Ordered probit model for the switching point in the risky choice task.

Predictors	Safe choices				
	Model 1	Model 2	Model 3	Model 4	Model 5
	Est (95% CI)	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*	*Est (95% CI)*
VR	-.301 (-.803, .201)	-.285 (-.792, .222)	-.254 (-.766, .258)	-.304 (-.823, .215)	-.288 (-.812, .235)
RA		-.053 (-.292, .185)	-.015 (-.258, .229)	-.066 (-.322, .191)	-.066 (-.323, .191)
RT			$-.438^{***}$ (-.689, -.186)	$-.461^{***}$ (-.716, -.206)	$-.464^{***}$ (-.719, -.209)
RI				-.137 (-.350, .075)	-.137 (-.350, .075)
IH					-.309 (-1.673, 1.055)
Observations	69	69	69	69	69

Note: VR = Experimental condition, RA = Risk attitude health/safety domain, RT= Risk-taking health/safety domain, RI = Recall of safety information, IH = Identified hazards. $^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$

The experimental condition did not significantly affect decision making, $B = -0.3$, $t = -1.18$, $p > .21$. Health/safety-domain-specific risk-taking was the only significant predictor of the risky decision, $B = -0.25$, $t = -3.41$, $p < .01$. No interaction of any predictor with experimental condition was observed.

3.3. Discussion

The second study found a positive effect of the safety training on risk perception even after 6 months. However, although the degree of immersion was higher in the VR condition, the presentation medium (VR vs. PPT) did not result in differences in risk perception. The VR condition did not increase the retention of safety information, but it led to a slightly higher ability to identify hazards in a pillar power drill. It is interesting that we found evidence that health-and-safety-domain-specific risk attitude and risk-taking were negatively associated with the recall of safety information. This indicates that participants who are more risk-seeking probably paid less attention to the safety information and thus remembered less. Furthermore, participants whose behavior was more risk-seeking in the health and safety domain were also more likely to work on the unsafe machine, and their decision making was less risk-averse than the decision making of others. It is important to mention that these effects did not depend on the experimental condition. Taken together, the findings suggest that while a safety training has the potential to affect risk perception, domain-specific risk attitudes and risk-taking influence risky choices and even the processing of safety information indicated by their effects on the recall of safety-relevant information.

4. General Discussion

On the basis of previous research, we expected that participants in the immersive VR condition would experience a higher sense of presence, a prediction that was corroborated in both studies. Because immersive VR provides a relatively realistic experience of the possible negative events, we predicted that the VR condition would have a stronger impact on risk perception and decision making than the PPT condition and that participants in the VR condition would recall more safety-relevant information and would detect more hazards when they faced a real machine later.

Contrary to these predictions, we found only weak evidence for the stronger impact of immersive VR than PPT. Only in Study 1 were the changes of risk perception more pronounced for the immersive VR condition than the PPT condition. Differences in learning, measured with recalled safety information and identified hazards, did not differ significantly between the experimental conditions. Surprisingly, we found that participants in the immersive VR condition made more risk-seeking choices than those in the PPT condition, a difference that was significant in Study 1 and still prevalent in Study 2, although not significant. Finally, because experiential learning results in more complex knowledge networks and knowledge integration (Glaser, 1984), we expected that risk-related decisions would be affected more strongly by participants' safety knowledge and identified hazards in the immersive VR condition than in the PPT condition, reflecting a higher degree of knowledge integration in the immersive VR condition. We found evidence in support of this hypothesis in Study 1, where higher recall of safety information was associated with more risk-averse decision making, but not in Study 2.

These two studies present an important step toward the understanding of immersive VR as a means for safety training. Whereas in Study 1, we found evidence for the prediction that immersive VR would result in a greater change in risk perception, we were not able to replicate this finding in Study 2. Similar to other studies (Sacks et al., 2013; Zaalberg and Midden, 2013), the differences between PPT and VR were small when looking at the outcome variables related to risk perception. It is important to mention that, in both studies, the sense of presence was higher in VR than in PPT, but this did not lead to differences in learning, a finding that is in line with previous studies (Gavish et al., 2015; Moreno and Mayer, 2002; Persky et al., 2009). Because we conducted the two studies using the same material, a strategy that has not been used before, we were able to show that the positive effects of immersive VR were not robust. It is important to mention that, whereas the experimental conditions differed in the two studies regarding their sense of presence, in Study 1, the variance in sense of presence explained by the manipulation was 20%, whereas in Study 2, it was only 5%. In Study 2, no knowledge differences were found at the later assessment. This finding is in contrast to a previous study (Sacks et al., 2013), but there, the dropout rate was very high (i.e., 70%), so it is possible that only motivated participants showed up for the final measure.

The two present studies were the first to assess decision making as a behavioral outcome after a safety training and found that participants in the immersive VR conditions did not differ from participants in the PPT conditions. It is important to mention that the results of Study 2 point to the relevance of personality because participants' domain-specific risk-taking was the most important predictor of their decisions and may therefore have overruled potential effects of the training. Domain-specific risk-taking and attitude did not influence perceptions but influenced participants' decisions and their recall of safety-relevant information. The two present studies extend knowledge in safety science by showing that it is not a technological effect of immersive VR that increases learning and changes in risk perception and decision making. When everything else was parallel, the differences between VR and PPT were small, a finding that is in line with previous studies that examined the effects of presentation mode on the effectiveness of training (Zaalberg and Midden, 2013). Thus, it may be interactive components that are the most important factors for explaining the differences found in previous studies between VR and the passive consumption of information in written form or with PPT (Buttussi and Chittaro, 2017; Chittaro and Buttussi, 2015).

Finally, the findings suggest that individual differences in risk-taking are highly relevant in affecting the processing of safety information and the recall of safety knowledge. As a consequence, we suggest that safety training programs should be tailored toward individuals with a tendency to take risks because safety knowledge is one predictor of safety performance (Christian et al., 2009).

4.1. Limitations and future research

The present studies were designed to provide a hard test of the effectiveness of immersive VR compared with a traditional method, in our case PPT. For this reason, videos were extracted from the immersive VR condition and presented in the PPT condition, and thus, the presentation format and the resulting immersion were the only differences between conditions. The extraction of the material from the immersive VR condition and its use in the PPT condition resulted in vivid and lively material, which may explain why even the control group increased their risk perception and had positive learning outcomes. Furthermore, the extent to which the ego-perspective alone is already beneficial for learning remains an open question because the depicted information is relevant for the self through its perspective. Future studies could use a third experimental group, which would receive the safety training with still pictures and text to determine the effects of vividness and perspective on attention and memory.

Another limitation of the present study was its reliance on a student sample. We chose a high school and apprentice student sample because they are representative of novice workers who are being taught to handle fast-moving machines in industry. However, particularly the statistical questions measuring risk perception could have been difficult for them because level of, education has been found to be positively correlated with risk numeracy (Cokely et al., 2012; Cokely and Kelley, 2009). This might explain why risk perception and decisions in our risky choice task were not correlated. It is important to mention that the relation between domain-specific risk attitude and risk-taking in the decision task points to the external validity of the design, whereas it calls into question the validity of our risk-perception measures. For this reason, future research should investigate whether, for samples where a lack of sufficient statistical understanding can be assumed, decisions might offer a better way to assess underlying beliefs.

Finally, in the present studies, participants were limited to the role of observers, even in the immersive VR condition. Whereas this was a necessary condition for investigating the effect of the presentation format on the effectiveness of a safety training, in future research, the degree of interaction should be varied along with the presentation format in order to identify a possible interaction between immersion and level of interaction.

4.2. Conclusion

In both studies presented here, the material and images displayed in the safety training were held constant across the experimental conditions. Thus, the two current studies provide evidence that the difference in safety training effectiveness observed in studies comparing immersive VR and traditional methods is most likely due to the displayed material (e.g., whether materials are more engaging or vivid) rather than the medium used for presentation per se.

Acknowledgments

This research project was funded by the Deutsche Forschungsgemeinschaft (ref. Wi 4053/5-1; Schu 1459/9-1). The authors are very grateful for the financial support.

References

Bates, D., Maechler, M., Bolker, B., Walker, S., 2014. lme4: Linear mixed-effects models using Eigen and S4 (Version 1.0-6)[R package] Available from http://CRAN. Rproject. org/package= lme4.

Blais, A.-R., Weber, E.U., 2006. A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. Judgm. Decis. Mak. 1, 33–47.

Borkenau, P., Ostendorf, F., 2008. NEO-FFI NEO-Fünf-Faktoren-Inventar nach Costa und McCrae (NEO-FFI) (2., neu normierte und vollständig überarbeitete Auflage), Hogrefe, Göttingen.

Burke, M.J., Sarpy, S.A., Smith-Crowe, K., Chan-Serafin, S., Salvador, R.O., Islam, G., 2006. Relative effectiveness of worker safety and health training methods. Am. J. Public Health 96, 315–324. https://doi.org/10.2105/AJPH.2004.059840

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., Munafò, M.R., 2013. Power failure: why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. 14, 365–376. https://doi.org/10.1038/nrn3475

Buttussi, F., Chittaro, L., 2017. Effects of different types of virtual reality display on presence and learning in a safety training scenario. IEEE Trans. Vis. Comput. Graph. 2626, 1–1. https://doi.org/10.1109/TVCG.2017.2653117

Bystrom, K.-E., Barfield, W., Hendrix, C., 1999. A Conceptual Model of the Sense of Presence in Virtual Environments. Presence Teleoperators Virtual Environ. 8, 241–244. https://doi.org/10.1162/105474699566107

Chittaro, L., Buttussi, F., 2015. Assessing knowledge retention of an immersive serious game vs.

A traditional education method in aviation safety. IEEE Trans. Vis. Comput. Graph. 21, 529–538. https://doi.org/10.1109/TVCG.2015.2391853

Chittaro, L., Sioni, R., 2015. Serious games for emergency preparedness: Evaluation of an interactive vs. a non-interactive simulation of a terror attack. Comput. Human Behav. 50, 508–519. https://doi.org/10.1016/j.chb.2015.03.074

Christian, M.S., Bradley, J.C., Wallace, J.C., Burke, M.J., 2009. Workplace safety: A meta-analysis of the roles of person and situation factors. J. Appl. Psychol. 94, 1103–1127. https://doi.org/10.1037/a0016172

Cokely, E.T., Galesic, M., Schulz, E., Ghazal, S., Garcia-Retamero, R., 2012. Measuring risk literacy: The Berlin Numeracy Test. Judgm. Decis. Mak. 7, 25–47.

Cokely, E.T., Kelley, C.M., 2009. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. Judgm. Decis. Mak. 4, 20–33.

Figner, B., Weber, E.U., 2011. Who Takes Risks When and Why? Curr. Dir. Psychol. Sci. 20, 211–216. https://doi.org/10.1177/0963721411415790

Floyd, D.L., Prentice-Dunn, S., Rogers, R.W., 2000. A Meta-Analysis. J. Appl. Soc. Psychol. 2000, 30, 2, pp. 407-429. 30, 407–429. https://doi.org/10.1111/j.1559-1816.2000.tb02323.x

Gamberini, L., Cottone, P., Spagnolli, A., Varotto, D., Mantovani, G., 2003. Responding to a fire emergency in a virtual environment: Different patterns of action for different situations. Ergonomics 46, 842–858.

Gavish, N., Gutiérrez, T., Webel, S., Rodríguez, J., Peveri, M., Bockholt, U., Tecchia, F., 2015. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. Interact. Learn. Environ. 23, 778–798. https://doi.org/10.1080/10494820.2013.815221

Glaser, R., 1984. Education and thinking: The role of knowledge. Am. Psychol. 39, 93–104. https://doi.org/10.1037/0003-066X.39.2.93

Hanoch, Y., Johnson, J.G., Wilke, A., 2006. Domain specificity in experimental measures and participant recruitment: An application to Risk-Taking-Behavior. Psychol. Sci. 17, 300–304.

Harrison, G.W., Rutström, E.E., 2008. Risk aversion in the laboratory, in: Cox, J.C., Harrison, G.W. (Eds.), Risk Aversion in Experiments. Emerald, Bingley, UK, pp. 41–196.

Janz, N.K., Becker, M.H., 1984. The Health Belief Model: A Decade Later. Health Educ. Q. 11, 1–47. https://doi.org/10.1177/109019818401100101

Kinateder, M., Gromer, D., Gast, P., Buld, S., Müller, M., Jost, M., Nehfischer, M., Mühlberger, A., Pauli, P., 2015. The effect of dangerous goods transporters on hazard perception and evacuation behavior - A virtual reality experiment on tunnel emergencies. Fire Saf. J. 78, 24–30. https://doi.org/10.1016/j.firesaf.2015.07.002

Kinateder, M., Pauli, P., Müller, M., Krieger, J., Heimbecher, F., Rönnau, I., Bergerhausen, U., Vollmann, G., Vogt, P., Mühlberger, A., 2013. Human behaviour in severe tunnel accidents: Effects of information and behavioural training. Transp. Res. Part F Traffic Psychol. Behav. 17, 20–32. https://doi.org/10.1016/j.trf.2012.09.001

Krampen, G., 1991. Struktur und statistische Kennwerte der Instrumente Fragebogen zu Kompetenz - und Kontrollüberzeugungen (FKK). Unveröffentliches Working Paper: Universität Trier 137–194.

Kuznetsova, A., Brockhoff, P.B., Christensen, R.H.B., 2014. lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package). R Packag. version. https://doi.org/http://CRAN.R-project.org/package=lmerTest

Laughery, K.R., Wogalter, M.S., 2006. Designing Effective Warnings. Rev. Hum. Factors Ergon. 241–271. https://doi.org/10.1177/1557234X0600200109

Maddux, J.E., Rogers, R.W., 1983. Protection motivation and self-efficacy: A revised theory of fear appeals and attitude change. J. Exp. Soc. Psychol. 19, 469–479. https://doi.org/10.1016/0022-1031(83)90023-9

Morassaei, S., Breslin, F.C., Shen, M., Smith, P.M., 2013. Examining job tenure and lost-time claim rates in Ontario, Canada, over a 10-year period. Occup Env. Med 70, 171–178. https://doi.org/10.1136/oemed-2012-100743

Moreno, R., Mayer, R.E., 2002. Learning science in virtual reality multimedia environments: Role of methods and media. J. Educ. Psychol. 94, 598–610. https://doi.org/10.1037//0022-0663.94.3.598

Mühlberger, A., Kinateder, M., Brutting, J., E, S., Muller, M., Gromer, D., Pauli, P., 2015. Influence of Information and Instructions on Human Behavior in Tunnel Accidents : A Virtual Reality Study. J. Virtual Real. Broadcast. 12.

Mühlberger, A., Weik, A., Pauli, P., Wiedemann, G., 2006. One-session virtual reality exposure treatment for fear of flying: 1-Year follow-up and graduation flight accompaniment effects. Psychother. Res. 16, 26–40. https://doi.org/10.1080/10503300500090944

Nieuwenstein, M.R., Blom, T.N., Morey, R.D., Wicherts, J.M., 2015. On making the right choice: A meta-analysis and large-scale replication attempt of the unconscious thought advantage. Judgement Decis. Mak. 10, 1–17.

Peperkorn, H.M., Diemer, J., Mühlberger, A., 2015. Temporal dynamics in the relation between presence and fear in virtual reality. Comput. Human Behav. 48, 542–547. https://doi.org/10.1016/j.chb.2015.02.028

Persky, S., Kaphingst, K.A., McCall, C., Lachance, C., Beall, A.C., Blascovich, J., 2009. Presence relates to distinct outcomes in two virtual environments employing different learning modalities. CyberPsychology Behav. 12, 263–268. https://doi.org/10.1055/s-0029-1237430

Ripley, R., Hornik, K., Gebhardt, a, Firth, D., 2011. MASS: support functions and datasets for Venables and Ripley's MASS. R Packag. version. https://doi.org/ISBN 0-387-95457-0

Sacks, R., Perlman, A., Barak, R., 2013. Construction safety training using immersive virtual reality. Constr. Manag. Econ. 31, 1005–1017. https://doi.org/10.1080/01446193.2013.828844

Schubert, T.W., 2003. The sense of presence in virtual environments: Zeitschrift für Medien. 15, 69–71. https://doi.org/10.1026//1617-6383.15.2.69

Simonsohn, U., Nelson, L.D., Simmons, J.P., 2014. P-curve: A key to the file-drawer. J. Exp. Psychol. Gen. 143, 534–547. https://doi.org/10.1037/a0033242

Sitkin, S.B., Weingart, L.R., 1995. Determinants of risky decision-making behavior: A test of the mediating role of risk perceptions and propensity. Acad. Manag. J. 38, 1573–1592.

Steuer, J., 1992. Defining Virtual Reality: Dimensions Determining Telepresence. J. Commun. 42, 73–93. https://doi.org/10.1111/j.1460-2466.1992.tb00812.x

TeamR, 2013. R Development Core Team. R A Lang. Environ. Stat. Comput.

Weber, E.U., Blais, A.-R., Betz, N.E., 2002. A domain-specific risk-attitude scale: Measuring risk perceptions and risk behaviors. J. Behav. Decis. Mak. 15, 263–290.

Weinstein, N.D., 1993. Testing four competing theories of health-protective behavior. Heal. Psychol. 12, 324–333. https://doi.org/10.1037/0278-6133.12.4.324

Weinstein, N.D., 1989. Effects of personal experience on self-protective behavior. Psychol. Bull.

105, 31–50.

Weinstein, N.D., 1984. Why it won't happen to me: Perception of risk factors and susceptibility. Heal. Psychol. 3, 431–457.

Zaalberg, R., Midden, C.J.H., 2013. Living behind dikes: Mimicking flooding experiences. Risk Anal. 33, 866–876. https://doi.org/10.1111/j.1539-6924.2012.01868.x

Zeitlin, L.R., 1994. Failure to follow safety instructions: faulty communication or risky decisions. Hum. Factors 36, 172–181. https://doi.org/10.1177/001872089403600111