



OPEN ACCESS

EDITED BY
Timo Kortsch,
IU Internationale Hochschule, Germany

REVIEWED BY
Sarah Depenbusch,
University of Paderborn, Germany
Rabea Böttling,
Bielefeld University, Germany

*CORRESPONDENCE
Olga Vogel
✉ olga.vogel@rub.de

RECEIVED 21 July 2025
REVISED 02 November 2025
ACCEPTED 06 November 2025
PUBLISHED 10 December 2025

CITATION
Vogel O, Conein S, Gabriel A-L and Kluge A
(2025) What should I do now? Evaluation of
two refresher trainings in the chemical
industry to maintain (non)technical skills.
Front. Organ. Psychol. 3:1670117.
doi: 10.3389/forgp.2025.1670117

COPYRIGHT
© 2025 Vogel, Conein, Gabriel and Kluge. This
is an open-access article distributed under the
terms of the [Creative Commons Attribution
License \(CC BY\)](#). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with these
terms.

What should I do now? Evaluation of two refresher trainings in the chemical industry to maintain (non)technical skills

Olga Vogel^{1*}, Stephanie Conein², Anna-Lena Gabriel¹ and
Annette Kluge¹

¹Work, Organizational and Business Psychology, Faculty of Psychology, Ruhr-University Bochum,
Bochum, Germany, ²Structure and Organisation of Vocational Education and Training, Federal Institute
for Vocational Education and Training, Bonn, Germany

Introduction: In the context of Industry 4.0, High-Reliability Organizations (HROs) are increasingly shaped by automation and procedural standardization, which limits opportunities for employees to practice critical skills. In non-routine situations such as system failures, this may lead to delayed responses and elevated safety risks. While refresher trainings are well established in aviation and maritime domains, no comparable standard exists in the chemical industry.

Methods: Two experimental field studies were conducted to evaluate short-format refresher trainings targeting technical and non-technical competencies. Study 1 (iN = 24) used a virtual reality (VR) simulation to refresh emergency-related technical skills of trainee chemical technicians. Performance, error rates, and knowledge retention were assessed across three measurement points. Study 2 (iN = 45) implemented a gamified, on-the-job non-VR intervention over three sessions to reinforce team resource management (TRM) skills. Pre-post assessments captured the use of TRM competencies and TRM awareness.

Results: Both interventions were positively evaluated and demonstrated potential to counteract skill decay under real-world conditions. The VR training resulted in fewer errors, improved performance in emergency scenarios, and faster response times. The TRM intervention showed increases in TRM-related behaviors and awareness.

Discussion: The findings indicate that short-format refresher trainings can effectively support the maintenance of technical and non-technical competencies in safety-critical industrial environments. Integrating such targeted refresher formats into operational training strategies may help sustain competence and mitigate risks associated with automation-driven reductions in practice opportunities.

KEYWORDS

refresher training, skill-decay, operator skills, team resource management, high reliability organizations, virtual reality, chemical industry, pharmaceutical industry

1 Introduction

The transition toward Industry 4.0 has fundamentally transformed industrial production systems (Frank et al., 2019). As processes become increasingly automated and interconnected, the role of human operators is shifting toward supervisory control (Longo et al., 2017). In routine operations, employees primarily monitor self-regulated systems and intervene only in exceptional cases. This development reduces the frequency of hands-on engagement, despite operators being trained for significantly more complex tasks (Pfeiffer and Suphan, 2015). Numerous studies have addressed the new skill requirements that arise from the increasing automation of industrial value chains and operational processes

(Conein, 2020; Hammermann and Stettes, 2016; Schmidt et al., 2016). These efforts aim to anticipate future demands placed on skilled workers and align vocational education and continuing training with the technological changes associated with Industry 4.0.

On the contrary, the question of how digitalization affects existing skills has received considerably less attention. Skills that were previously considered essential, especially those grounded in manual experience and workplace experience, continue to represent a core component of professional practice (Tenberg and Pittich, 2017). Even in the context of increasing automation and academic specialization, such skills are regarded irreplaceable within vocational work settings. In this context, the development of new skills must be accompanied by systematic strategies aimed at preserving and strengthening existing occupational skills that remain essential for effective performance.

This issue becomes particularly salient in non-routine situations, which continue to occur even in highly automated environments (Bainbridge, 1983; Weyer, 1997). In such moments, operators have only a few seconds to recognize the situation and take effective action to contain the incident (Klostermann et al., 2022). This requires rapid access to both technical expertise and experiential knowledge under high cognitive and emotional demands. Since these skills are rarely practiced in day-to-day operations, their accessibility and reliability may decrease over time. This phenomenon is referred to as skill decay (Arthur et al., 1998; Arthur and Day, 2013).

This challenge is especially relevant in High-Reliability Organizations (HROs), where rare but safety-critical events must be managed with high precision to avoid serious consequences (Weick and Sutcliffe, 2001). Domains such as aviation, defense, nuclear energy, and the maritime sector have addressed this risk through structured refresher training and Crew/Team Resource Management (CRM/TRM) programs that sustain non-technical skills supporting safe performance (Barnett et al., 2006; Flin et al., 2002; Hagemann and Kluge, 2013; Klostermann et al., 2022). However, systematic approaches to mitigate skill decay remain largely absent in sectors such as the chemical and pharmaceutical industries.

To address this gap, the present study evaluates two refresher training formats developed for chemical operators. One format targets the reinforcement of technical-operational knowledge, while the second focuses on non-technical skills.

1.1 Non-routine situations in high-reliability organizations

HROs operate in environments characterized by inherent complexity and a high potential for catastrophic failure, yet manage to maintain consistently safe and effective performance over time (Dingee, 2019; Hagemann and Heinemann, 2024). Despite facing conditions of uncertainty, time pressure, and operational risk, they achieve this reliability by adhering to organizational principles that foster consistency and resilience (Lekka, 2011). These principles include continuous anticipation and management of potential problems, a strong orientation toward organizational learning,

transparent bottom-up communication, and the cultivation of a “just culture” that supports open dialogue about mistakes and promotes a constructive approach to error management (Roberts and Bea, 2001; Sutcliffe, 2011). Moreover, critical decisions are based on professional expertise rather than formal authority or hierarchical position. Nevertheless, such organizations remain aware that non-routine situations can occur. These are defined as infrequent and unexpected events, such as system failures, and may result in significant risks for humans, the organization and the environment (Klostermann et al., 2022). In order to resolve non-routine situations without adverse consequences, operators must apply relevant skills under acute time pressure, engage in dynamic decision making, and solve complex problems (Hagemann and Kluge, 2017; Kluge, 2014).

1.2 Technical and non-technical skills

Technical skills are defined as task-specific abilities that are observable, measurable, and often acquired through formal education and on-the-job training (Nasir et al., 2011). They encompass the capacity to carry out concrete work processes using specialized tools, equipment, or systems and are closely tied to a particular occupational context. In the context of the chemical industry, technical skills encompass a wide range of operational procedures, including process monitoring, plant control, maintenance routines, and safety-related interventions (Conein and Felkl, 2023). They are largely bound to specific tools, equipment, or technologies, and their execution requires adherence to standardized protocols and regulatory requirements. Empirical findings from Conein and Felkl (2023) confirm the continued relevance of these skills in chemical production, particularly when it comes to handling plant disturbances. In a survey of experienced operators, knowledge of plant-specific process flows, the ability to operate the plant manually in case of automation failure, and an understanding of the process control system (PLS) were rated especially important. Likewise, chemical and technical fundamentals as well as routine-based practical experience were considered essential for responding appropriately to non-routine situations.

Non-technical skills refer to cognitive, social, and personal abilities that support the effective and safe execution of task-related behavior, especially under demanding or dynamic conditions (Kluge, 2014). In contrast to technical skills, which refer to task-specific procedures or tools, non-technical skills relate to how tasks are performed and include behavioral, interpersonal, and decision-related dimensions.

In the aviation sector, non-technical skills are commonly integrated within the framework of CRM, a training approach designed to improve safety by strengthening communication, leadership, cooperation, and error management among cockpit crews (Flin et al., 2002). This concept has since been adapted to other high-reliability domains, where it is referred to as TRM (Arthur et al., 2003; Hagemann and Kluge, 2013). TRM emphasizes team-based competencies that are essential for managing complex operational challenges and includes core areas such as maintaining situational awareness, addressing human error, making decisions

under uncertainty, and managing acute stress (Kluge, 2014; Flin and Maran, 2015).

1.3 The impact of refresher trainings on skill-decay

To mitigate skill decay in both technical and non-technical domains, organizations increasingly rely on refresher trainings (Kluge and Frank, 2014). These interventions are structured and time-limited learning opportunities that aim to restore previously acquired skills which have deteriorated due to periods of non-use. Refresher trainings are implemented when task proficiency no longer meets predefined operational standards and are more intensive than maintenance or booster sessions, as they focus on reactivating complex task behavior through deliberate repetition in realistic training environments (Sullivan et al., 2019). For example, Sullivan et al. (2019) demonstrated that a 2-hour simulation-based refresher training administered 1 year after initial instruction was sufficient to restore participants' performance in emergency medical procedures to post-training levels. In high-reliability sectors such as aviation, refresher training is commonly used to maintain task performance across a range of operational requirements (EASA, 2015). Simulation-based formats are frequently employed to recover both technical and non-technical skills within realistic scenarios (Woollard et al., 2006).

1.4 Virtual reality and game-based learning as formats for refresher programs

Virtual reality encompasses several technologies that immerse users in a computer-generated environment (Stefan et al., 2023). These technologies exhibit unique characteristics of presence, immersion, and interactivity, with the degree of each varying between systems (Mütterlein, 2018). Immersion refers to the objective properties of the VR system, such as sensory fidelity, inclusiveness, and responsiveness, that determine the extent to which users are enveloped by the virtual environment (Lee, 2025). Nilsson et al. (2016) conceptualize immersion as comprising system, narrative, and challenge-based forms, highlighting that technological fidelity, storyline engagement, and task-related challenges can each contribute to the immersive experience. Presence is defined as the subjective psychological state in which users experience a sense of "being there" in the mediated environment, shaped not only by technological fidelity but also by cognitive and affective responses (Triberti et al., 2025). Interactivity describes the degree to which users can act within and influence the virtual environment in real time, receiving contingent feedback, and is considered a design element that supports immersion and presence (Radianti et al., 2020).

VR is highly suitable for use in safety-related training (Kanade and Duffy, 2022; Scorgie et al., 2024; Stefan et al., 2023). It enables safe simulation of hazardous scenarios that cannot be practiced in real settings, increases learning engagement and motivation, ensures consistency in training delivery, and allows repeated

practice under controlled conditions (Stefan et al., 2023). Most VR safety trainings rely on head-mounted displays in combination with handheld controllers, which provide high levels of immersion and presence through visual fidelity and head tracking, while enabling interactivity by allowing users to manipulate the virtual environment (Scorgie et al., 2024).

Game-based learning (GBL) is defined as the use of complete games as the primary instructional medium and is commonly implemented as serious games designed for purposes beyond entertainment (Zyda, 2005; Plass et al., 2015). It differs from gamification, which introduces game-design elements into existing instruction without creating a stand-alone game (Deterding et al., 2011; Landers, 2015). Game-based learning is characterized by the integration of rules and goals, assessment and feedback, challenge, learner control, and environmental features that support immersion (Bedwell et al., 2012). Rules and goals specify the tasks that must be achieved, assessment and feedback provide immediate information on performance, challenge regulates task difficulty in relation to learner proficiency, learner control allows active decision-making and exploration, and environmental features such as narrative or simulated context support immersion and transfer (Plass et al., 2015; Wouters et al., 2013).

GBL has been applied in safety-related domains, including the identification of construction hazards, emergency evacuation, and laboratory safety (Brown and Poulton, 2018; Feng et al., 2018; Ribeiro et al., 2012). Relative to non-game instruction, studies report higher hazard detection rates, better evacuation preparedness, and improved adherence to laboratory safety protocols (Feng et al., 2018; Ribeiro et al., 2012). Brown and Poulton (2018) note that conventional safety training often lacks context, practical application, and meaningful consequences, and propose that game-based designs can address these gaps by embedding scenarios that mirror real environments, allowing practice in risk-free settings, and linking player decisions to in-game outcomes.

Although most VR and game-based safety trainings have been evaluated as one-off interventions rather than periodic refreshers, their characteristics align well with refresher purposes. Refresher training aims to reactivate complex task behavior through realistic practice, and VR provides high-fidelity rehearsal of hazardous, low-frequency scenarios that cannot be trained near the job (Stefan et al., 2023). Game-based formats incorporate rules, feedback, and contextualized challenges that induce active retrieval and correction of errors, thus counteracting skill decay (Plass et al., 2015; Wouters et al., 2013). Evidence indicates that both approaches enhance knowledge retention and procedural performance compared to traditional training, suggesting their potential to restore proficiency when operational standards are no longer met (Scorgie et al., 2024; Feng et al., 2018; Ribeiro et al., 2012).

Despite the efficacy of refresher training in domains such as aviation and emergency medicine, there is a lack of comparable research and implementation in the chemical industry. As Klostermann et al. (2022) emphasize in their scoping review, research on mitigating skill decay is concentrated in high-risk domains with established simulation-based training traditions, while chemical production remains comparatively

under examined. This gap is not due to lower demands on operators: chemical and pharmaceutical plants are high-risk environments where non-routine events require immediate, error-free action under time pressure. In contrast to aviation, refresher formats in chemical operations are not institutionally embedded, and training cycles tend to be irregular or *ad hoc* (Conein and Felkl, 2023). At the same time, digitalization and the shift toward Industry 4.0 reduce opportunities for hands-on practice in day-to-day work, intensifying the risk of skill decay (Conein, 2020). Although evidence from other domains demonstrates the general efficacy of refresher trainings, direct transfer cannot be assumed. Sector-specific research is needed to determine how refresher concepts can be adapted to chemical working conditions, regulatory frameworks, and occupational profiles.

To address this research gap, two field studies were conducted. Each study focused on a distinct skill domain relevant to operator performance in non-routine situations. The first study investigated the effects of a VR refresher training on the retention of technical skills, while the second study evaluated a game-based on-the-job intervention aimed at maintaining non-technical, TRM skills. Both trainings were empirically tested with chemical technicians in real-world industrial settings.^{1,2}

2 Study 1: technical skills

2.1 Overview

The first study was conducted in cooperation with the chemical and pharmaceutical company Merck KGaA (2023). The aim was to evaluate an existing training virtual reality (VR) simulation in the company regarding its suitability as a refresher for dealing with non-routine situations. The following hypotheses were proposed:

H1: Refresher training results in higher processing speed.

H1a: Participants who received the refresher resolve the emergency situation under time pressure faster than those without the refresher.

H1b: Participants who received the refresher complete Stages III and IV of the simulation training faster than those without the refresher.

H2: Refresher training results in improved task performance.

H3: Refresher training increases awareness of rare and non-routine situations.

¹ The studies were approved by the local ethics committee of the Faculty of Psychology (Approval No. 768, dated 20 January 2022). Before participation, all subjects received standardized information on the study's structure and procedure of the study and were informed of their right to withdraw at any time, according to ethical guidelines for informed consent.

² These studies were not pre-registered. All hypotheses, instruments, exclusion criteria, and analysis decisions are reported in full in the manuscript. Analysis scripts and anonymized raw data can be shared upon justified request. Due to company confidentiality and intellectual property restrictions, the training materials cannot be shared.

2.2 Materials and methods

2.2.1 Sample

The sample consisted of $N = 27$ chemical technicians in their second year of training at the Darmstadt and Frankfurt sites. To determine the extent to which a refresher intervention leads to better performance of operational skills, participants were divided into two groups in a pseudo randomized manner. The experimental group was formed based on availability at the respective session because some apprentices were attending training at a different site during the refresher program. The experimental group (EG; $N = 13$) received a refresher intervention 4 weeks after initial training, while the control group (CG; $N = 14$) did not participate in the refresher session. During the course of the study, three participants from the control group dropped out. As a result, data from $N = 24$ participants ($N_{EG} = 11$; $N_{CG} = 13$) were included in the final analysis. In agreement with the cooperating partner, no demographic data were collected to ensure participant confidentiality. *A priori* power analyses using G*Power 3.1 (Faul et al., 2009) indicated a target sample size of $N = 102$ for between-group Mann-Whitney-*U*-Tests tests and $N = 158$ for the repeated-measures designs. The achieved sample did not reach the target sample size.

2.2.2 Study design

The quasi-experiment was conducted in a two-factor mixed factorial design with three measurement time points (T1–T3; Figure 1). All training sessions were conducted in a virtual reality simulation and were provided using Pico Neo 3 Link VR goggles.

Before the first training session (T1), participants received a 15-min introduction to the research project and the topic of “dealing with rare and unusual situations.” All participants then participated in the basic training (stages I–IV) and completed several questionnaires before and after the training. On the day of refresher training 4 weeks later (T2), the EG participants underwent stage III of VR simulation. The third session (T3) took place 2 weeks after the refresher intervention and was used to record performance during training. Both groups (EG and CG) answered several questionnaires and then performed stages III and IV of the VR simulation. All questionnaires were administered online using the Qualtrics survey platform.

A total of four VR goggles were available for each test session, which meant that four participants took part in the test in the same room at the same time (e.g., Figure 2). Each test session was accompanied by at least three instructors. They provided participants with technical support during training and ensured safety and comfort when using VR goggles. During the initial training, the trainers also supported the content of the training to ensure comparable learning success between the participants.

2.2.3 Virtual reality training “operate your own reactor”

The “Operate your own Reactor” training course was developed by Tehreem et al. (2022) in cooperation with the company Merck KGaA. The aim of the training is to develop procedural skills for the production of the organometallic compound n-butyllithium

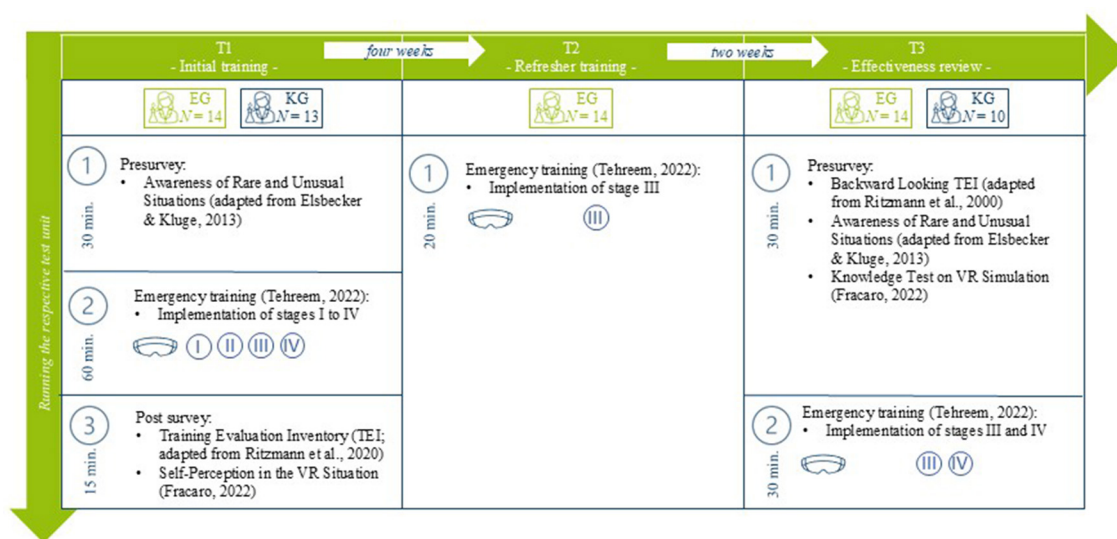


FIGURE 1

Overview of the study I design—Evaluation of a technical refresher training in the chemical and pharmaceutical industry; all questionnaires and items not documented by citations were invented by the research team.

(BuLi) in a realistic way. The virtual environment represents three floors of a reconstructed former chemical plant at Merck KGaA's Darmstadt site. The training is divided into four stages that are built on each other. In each stage, participants receive instructions on the main screen located in the center of the room and execute them by interacting with the process control system and the instruments of the chemical reactor (for a detailed training description, see Fracaro et al., 2024 and Tehreem et al., 2022).

In stages I and II, the training is closely accompanied by instructions on the main screen (Figure 3). Stage III contains two emergency scenarios. The first emergency scenario (*Emergency SEE*) is caused by an excessive dosage of the reactant BuCi and cannot be prevented, so participants observe the course of the accident. In the second emergency scenario (*Emergency SEE and SOLVE*), which is caused by the temperature of the filled production container being too high (40°), the participants have 2 min to prevent an explosion. In stage IV, BuLi production is continued, but without any additional help. Instead, participants can take advantage of textual hints in the form of standard operating procedures on the main screen.

2.2.4 Variables

2.2.4.1 Independent variable

To examine the effectiveness of the refresher intervention, the EG and CG groups were compared in terms of training performance and transfer to day-to-day work. The refresher intervention served as an independent variable which only the experimental group received.

2.2.4.2 Dependent variables

Table 1 provides an overview of all measurement instruments, grouped by measurement time points.

2.2.4.2.1 Training processing speed and performance

The performance assessment was divided into two parts. In the first part, participants completed Stages III and IV of the simulation training. These stages were selected due to their higher complexity compared to Stages I and II, as they involved emergency scenarios and minimal guidance.

The dependent performance and processing speed variables were the number of errors (E) and execution time (T). For Stage IV, E and T were measured at T1 and T3. In Stage III, the error counts were also assessed at both time points. To evaluate the processing speed, two time-based measures were collected in Stage III after the training (T3): (1) the time in seconds required to resolve the emergency situation, and (2) the total execution time for completing the stage. The error counts were displayed on the simulation's main screen upon completion of each stage. Execution time was recorded manually using stopwatches. To ensure accuracy in the time-based performance data (T), the entire VR simulation was also screen-recorded during the effectiveness test to account for any potential system failures.

The second part of the performance survey consisted of the "Operate your own reactor—Knowledge Test" (Fracaro, 2022). This was divided into one multiple-choice item, four one-choice items, and one sorting item. The participants were able to achieve a total of 13 points by answering all the questions correctly.

2.2.4.2.2 Awareness of rare and unusual situations

Transfer to daily working life was surveyed using the "Awareness of Rare and Unusual Situations" questionnaire (adapted from Elsbecker and Kluge, 2013) in the pre-post comparison.

2.2.4.2.3 Descriptive training evaluation

The "Training Evaluation Inventory" (TEI) adapted to simulation training (Ritzmann et al., 2014) was used in the initial



FIGURE 2
Example image of a participant during the training "operate your own reactor".

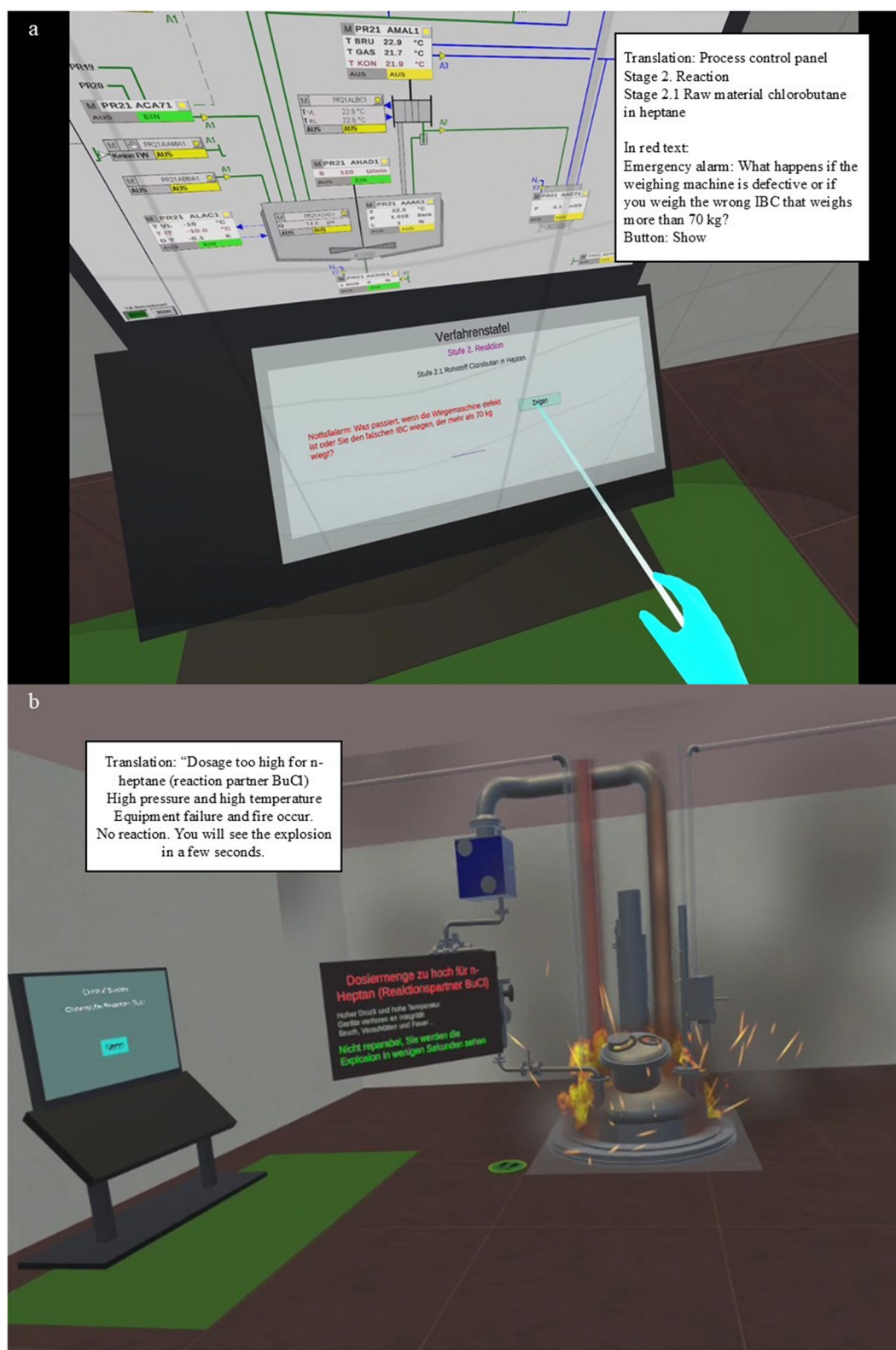


FIGURE 3

Examples of the VR-training. (a) Main screen and process control system from the perspective of the participant. (b) Production container shortly before the explosion in the emergency scenario "SEE" from the perspective of the participant.

TABLE 1 Measurement instruments separated by phases of the experiment.

Measurement instrument	Example item	No. items	Range	ω
T1: Pretest—before the first simulation training session				
Awareness of Rare and Unusual Situations (adapted from Elsbecker and Kluge, 2013)	The topic of rare and unusual situations at Merck is important to me personally.	4	5-point likert scale 1 = does not apply at all; 5 = applies very much	0.762
T1: Training evaluation “Emergency SEE”—after the first execution of the simulation training				
Training Evaluation Inventory (TEI) adapted to simulation training (Ritzmann et al., 2014)	I find the simulation training useful for my job.	35	5-point likert scale 1 = strongly disagree; 5 = strongly agree	0.967
VR experience (Fracaro, 2022)	To what extent do you believe that this type of virtual reality can help you learn?	6	5-point likert scale 1 = very little to 5 = very much	0.884
Learning objectives (self-invented item)	In your opinion, which learning objectives were necessary to master the situation(s)? Answer option, e.g.: “Dealing with stress”	1	Sorting Item	.
T3: Overall evaluation—before the execution of stages III and IV of the VR simulation				
Backward-looking-TEI (adapted from Ritzmann et al., 2014)	In the last 6 weeks... “the simulation training has proved useful for my job”	11	5-point likert scale 1 = does not apply at all; 5 = applies very much	0.922
Awareness of Rare and Unusual Situations (adapted from Elsbecker and Kluge, 2013)	The topic of rare and unusual situations at Merck is important to me personally.	4	5-point likert scale 1 = I agree much less than before; 5 = I agree much more than before	0.757
Operate your own Reactor—Knowledge Test (Fracaro, 2022)	To produce Butyllithium, I need the following reactants... ... metallic lithium, acetone, and cyclobutene ... metallic lithium, n-heptane, chlorobutan in heptane ... metallic lithium, toluol, isobutane	6	One multiple choice item, four one-choice items and one sorting item	0.688
T3: Performance Evaluation—during the execution of stages III and IV of the VR simulation				
Performance in the “SEE & SOLVE” training session	Speed in resolving the emergency situation	.	Time in seconds	.
Performance in stage III	Performance in the entire stage	.	Number of errors made	.
Performance in stage IV	Performance and speed in the entire stage	.	Number of errors made	.

ω , McDonald's Omega as measure for internal consistency. ., no information provided.

training to record the participants' ratings on the scales *subjective fun*, *perceived usefulness*, *perceived difficulty*, *subjective knowledge growth*, *attitude toward training*, *problem-based learning*, *activation*, *demonstration*, *integration* and *application*.

In addition, a shortened form (Backward-Looking-TEI) was used to evaluate the sustainability of the training after 6 weeks. The subscales *perceived usefulness*, *subjective knowledge growth*, *attitude toward training*, *application*, and *integration* were included. For the short form, only subscales that explicitly assess transfer of training to day-to-day work since training were selected. Within these subscales, items suitable for retrospective assessment were used (e.g., for perceived usefulness: “In the past 6 weeks, the ‘Emergency SEE & SOLVE’ simulation has proven useful for my job.”).

Self-perception in VR simulation was measured using 6 items developed by [Fracaro \(2022\)](#) to evaluate simulation training.

As no explicit learning objectives concerning the handling of non-routine situations were defined during the development of the training, a sorting item was used to assess which of the following learning goals participants perceived as most relevant: (1) gaining an overview of the situation, (2) setting priorities, (3) handling stress, and (4) making decisions.

2.3 Results

2.3.1 Descriptive results

The descriptive results to evaluate the training are displayed in [Table 2](#). The mean values shown are comparable to the norming values of similar samples of high-risk organizations (junior flight attendants, firefighters; [Ritzmann et al., 2020](#)), thus the training can be regarded as useful. It is noteworthy that the results of the Backward-Looking TEI are worse in comparison 6 weeks after training.

In addition, participants indicated which learning objectives they considered most important for successfully completing the simulation. [Figure 4](#) displays the absolute frequency of learning objectives that participants ranked in first place, separated by experimental and control group. The objectives “Getting an overview of the situation” and “Coping with stress” were most frequently rated as most important.

2.3.2 Inferential statistics

H1: Refresher training results in higher processing speed.

TABLE 2 Descriptive results of the training evaluation and internal consistencies of the TEI subscales—Study I.

Training evaluation inventory (TEI) adapted to simulation training (Ritzmann et al., 2014)						
Scale	Example item	EG N = 13		CG N = 14		ω
		M	SD	M	SD	
Subjective fun	The learning was fun	4.67	0.27	4.26	1.01	0.890
Perceived usefulness	I derive personal use from this training	4.06	0.60	4.29	0.80	0.923
Perceived difficulty	The time was sufficient for the themes covered	4.44	0.36	4.41	1.05	0.914
Subjective knowledge growth	I will be able to remember the new themes well	3.87	0.54	3.76	0.97	0.887
Attitude toward training	I would recommend this training to my colleagues	4.18	0.59	4.12	0.82	0.845
Problem-based learning	First of all, problems were addressed, and by working on them, I consequently learned the themes	4.00	0.60	3.87	0.98	0.918
Activation	I was able to bring my previous professional experiences	4.31	0.48	4.17	0.80	0.797
Demonstration	Contents were illustrated with concrete examples	4.17	0.41	3.82	0.61	0.711
Application	I was able to practice what I had learned in the training	3.62	0.65	3.60	0.73	0.716
Integration	Contents were consolidated in discussions	4.44	0.46	4.54	0.31	0.584
Backward-looking TEI (adapted from Ritzmann et al., 2014)						
		EG N = 11		CG N = 13		
		M	SD	M	SD	
Perceived usefulness	In the past 6 weeks...—...the simulation “Emergency SEE & SOLVE” has proven useful for my job	3.12	0.95	3.18	0.81	0.785
Subjective knowledge growth	In the past 6 weeks...—...I was able to remember the insights related to the learning objectives well	3.82	0.46	3.72	0.56	0.491
Attitude toward training	In the past 6 weeks...—...I have recommended the simulation “Emergency SEE & SOLVE” to my colleagues	2.55	0.69	3.00	0.98	.
Application	In the past 6 weeks...—...the feedback from the simulation “Emergency SEE & SOLVE” has helped me to further work on my learning objectives	3.27	0.90	3.31	0.85	.
Integration	In the past 6 weeks...—...I have applied the contents related to the learning objectives in my daily work	3.00	0.81	3.08	1.19	.
VR experience questionnaire (Fracaro, 2022)						
		EG N = 13		CG N = 14		
		M	SD	M	SD	
Overall scale: The VR simulation...	... was very well-received by the participants.	4.38	0.65	4.57	0.51	.
	... can help participants to learn.	3.77	1.09	4.21	0.89	.
	... s a suitable tool for learning how to make N-BuLi.	4.00	0.82	4.29	0.61	.
	... is better than traditional learning methods.	4.08	0.86	4.21	0.89	.
	... can help to prepare for real emergency situations.	3.85	0.80	4.43	0.65	.
	... is helpful to learn how to behave in emergency situations.	3.85	0.69	4.50	0.65	.

M, Mean; SD, Standard Deviation; ω , McDonald's Omega, where ω is not reported, the subscale comprised < 3 items (.).

H1a: Participants who received the refresher resolve the emergency situation under time pressure faster than those without the refresher.

The Mann-Whitney U-test was calculated to compare the time taken to solve an emergency situation under time pressure between the experimental group and the control group. The EG ($M_{post} = 16.43$, $SD_{post} = 4.22$) required significantly less time to resolve the emergency situation than the CG ($M_{post} = 38.30$, $SD_{post} = 31.22$), $U = 40$, $z = -1.83$, $p_{one-tailed} = 0.036$, $r = 0.37$ (see Figure 5), thus confirming the hypothesis.

H1b: Participants who received the refresher complete Stages III and IV of the simulation training faster than those without the refresher.

To assess whether the experimental group that received refresher training performed more efficiently, an univariate repeated-measure analysis of variance was conducted on the duration in Stage IV of the VR application. A significant main effect of time was observed, with both groups demonstrating faster processing at posttest compared to pretest, $F_{(1,22)} = 11.52$, $p = 0.003$, $\eta_p^2 = 0.34$ (see Table 3).

The two groups were also compared with respect to completion time for Stage III at T3. The experimental group ($M_{post} = 605.91$, $SD_{post} = 87.78$) completed Stage III significantly faster than the control group ($M_{post} = 979.85$, $SD_{post} = 318.05$), $U = 10$, $z = -3.56$, $p_{one-tailed} < 0.001$, $r = 0.73$.

H2: Refresher training results in improved task performance after a 2-week period.

The number of errors in Stages III and IV were also compared between the two groups using repeated-measures ANOVAs. No significant effects were found. An overview of the results is provided in Table 3.

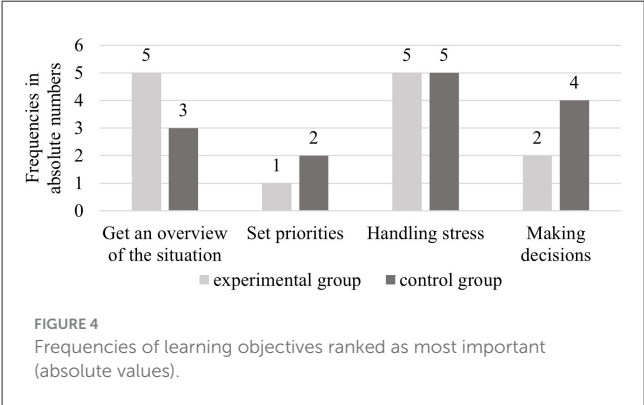


FIGURE 4
Frequencies of learning objectives ranked as most important (absolute values).

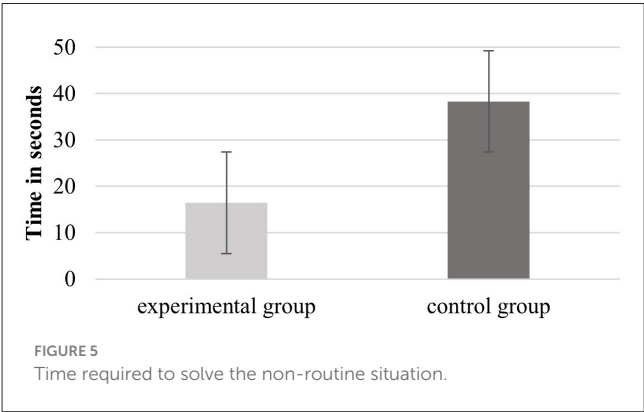


FIGURE 5
Time required to solve the non-routine situation.

TABLE 3 Results of the test of the second hypothesis.

Variable	Group	Pre <i>M</i> (<i>SD</i>)	Post <i>M</i> (<i>SD</i>)	Main effect	Interaction effect	η_p^2 Main effect (interaction effect)
Stage III—errors	CG	0.57 (0.65)	0.92 (0.86)	$F_{(1,22)} = 0.02, p = 0.885$	$F_{(1,22)} = 3.06, p = 0.094$	0.02 (0.12)
	EG	0.62 (0.65)	0.27 (0.47)			
Stage IV—errors	CG	0.85 (0.66)	1.08 (0.76)	$F_{(1,22)} = 0.02, p = 0.890$	$F_{(1,22)} = 0.30, p = 0.592$	0.00 (0.01)
	EG	0.54 (0.66)	0.45 (0.93)			
Stage IV—time	CG	761.57 (258.56)	662.77 (213.05)	$F_{(1,22)} = 11.52, p = 0.003^{**}$	$F_{(1,22)} = 0.15, p = 0.699$	0.34 (0.01)
	EG	677.15 (195.20)	507.36 (117.65)			

M, Mean; *SD*, Standard Deviation. $^{**}p < 0.01$; Values reported as 0.00 indicate $\eta_p^2 < 0.01$.

No significant difference was found between the two groups in knowledge test performance, $U = 87.50$, $z = 0.93$, $p_{one-tailed} = 0.185$, $r = 0.19$. The control group scored a mean of $M_{post} = 6.62$ ($SD_{post} = 2.18$, $range = 3-10$), while the experimental group achieved a slightly higher mean of $M_{post} = 7.27$ ($SD_{post} = 1.90$, $range = 3-10$).

H3: Refresher training increases awareness of rare and non-routine situations.

To test this hypothesis, participants' awareness of non-routine situations was compared between the two groups using a univariate repeated-measures analysis of variance with one within-subjects factor. No significant differences were found in the awareness of rare situations between groups, $F_{(1,22)} = 2.74$, $p = 0.112$, $\eta_p^2 = 0.11$, or within groups over time, $F_{(1,22)} = 0.86$, $p = 0.364$, $\eta_p^2 = 0.04$. The EG reported awareness scores of $M_{pre} = 3.54$ ($SD_{pre} = 0.80$) and $M_{post} = 3.85$ ($SD_{post} = 0.75$), while the CG showed $M_{pre} = 4.02$ ($SD_{pre} = 0.51$) and $M_{post} = 3.90$ ($SD_{post} = 0.42$).

2.4 Discussion

The first study examined the effectiveness of refresher training for technical skills in the context of chemical production, with a particular focus on performance under time pressure during an emergency scenario. The results indicate that the refresher training had a measurable impact on operational speed. Participants in the experimental group, who received the additional intervention, required significantly less time to resolve the simulated emergency than those in the control group and needed less time overall to complete Stage III. This supports Hypothesis 1, which posited that participants who received the refresher training would demonstrate faster task completion. The time advantage observed in the experimental group may be attributed to the reactivation of procedural routines, which enabled faster retrieval of task-relevant action sequences and more efficient decision-making. In high-reliability environments, temporal efficiency is critical, as even brief delays in emergency response can significantly increase the likelihood of system escalation and adverse outcomes (Kluge, 2014). Notably, we did not control for individual differences in prior technical experience or familiarity with emergency procedures, which may have affected baseline competence and

moderated responsiveness to the intervention. During Stage IV both groups showed improvement relative to their initial performance. During Stage IV both groups showed improvement relative to their initial performance. This effect likely reflects the sustained benefits of the original 6-week training and indicates that the refresher training effect was confined to the specifically trained Stage III and did not transfer to other training contents. Nevertheless, the simulated emergency scenarios may not have fully captured the complexity, urgency, or emotional load of real-world incidents, which could have limited ecological validity and constrained transfer effects (Hustad et al., 2019).

Hypothesis 2 addressed whether the refresher training would result in improved training performance 2 weeks after intervention. No significant group differences were found regarding the error rates during the training stages. In the knowledge test, no significant differences between the groups were observed. An explanation is that the training scenario focused on the production of n-butyllithium, a process that is not part of the formal curriculum for second-year apprentices. Consequently, the relevant knowledge was not sufficiently embedded in preexisting mental models, which likely limited its integration and retrieval. Knowledge acquisition is more successful when new information can be meaningfully linked to prior domain-specific schemas (van Gog and Sweller, 2015). In the absence of such cognitive structures, learners are more likely to engage in surface-level processing, which hinders the durability and application of the acquired content. Additionally, the short 2-week follow-up period may not have been sufficient to assess the long-term retention of procedural skills.

With regard to Hypothesis 3, which examined the transfer of competence to risk perception and awareness of non-routine situations, no significant differences were found between the experimental and control groups. Both groups rated the relevance of non-routine situations and the likelihood of accidents at similarly high levels, indicating that the intervention did not produce measurable changes in these attitudinal variables. One explanation may be that individuals working in high-risk operational settings already exhibit a heightened awareness of rare and hazardous situations, which tend to be stable over time and less amenable to short-term training effects (Flin et al., 2008). The lack of observed change may also be related to the fact that specific competence facets were not operationalized in the VR environment, which makes it difficult to determine which components of non-technical competence were actually addressed or reinforced.

The training was generally evaluated positively, though retrospective scores were slightly lower. Part of the initial enthusiasm may reflect a novelty effect associated with immersive VR, which can inflate user enjoyment and engagement during early exposures (Miguel-Alonso et al., 2024). This may also interact with perceived relevance, as the scenario was not part of the official curriculum.

3 Study 2: non-technical skills

3.1 Overview

The second study focused on the development and evaluation of a refresher training program for non-technical team resource

management (TRM) skills at Covestro AG Germany. The company had already implemented TRM training modules targeting non-technical, safety-relevant skills. Based on the existing program, a gamified refresher format called the TRM Box was developed in collaboration with the Qualification Management department.

The refresher training was implemented 6 months after the participants had completed the initial TRM training. The original quasi-experimental design followed a two-factorial structure with two pseudo-randomized groups. The second group was intended to serve as a control group, receiving the refresher intervention 2 weeks after the experimental group. This design aimed to assess whether the refresher training helped maintain TRM skills over time compared to participants who had not received the training within the same timeframe. Furthermore, a further measurement point immediately after the initial TRM training to assess potential declines in skills during the 6-month interval leading up to the TRM Box refresher.

The planned hypotheses were as follows:

H1: TRM skills, awareness of TRM, and perceived control over workplace safety significantly decrease over the 6-month period following the original TRM training and significantly increase again after the refresher intervention.

H2: Participants in the experimental group (EG) who receive the refresher training show significantly higher TRM skills and greater awareness of rare and non-routine situations than participants in the control group (CG), who have not yet received the training.

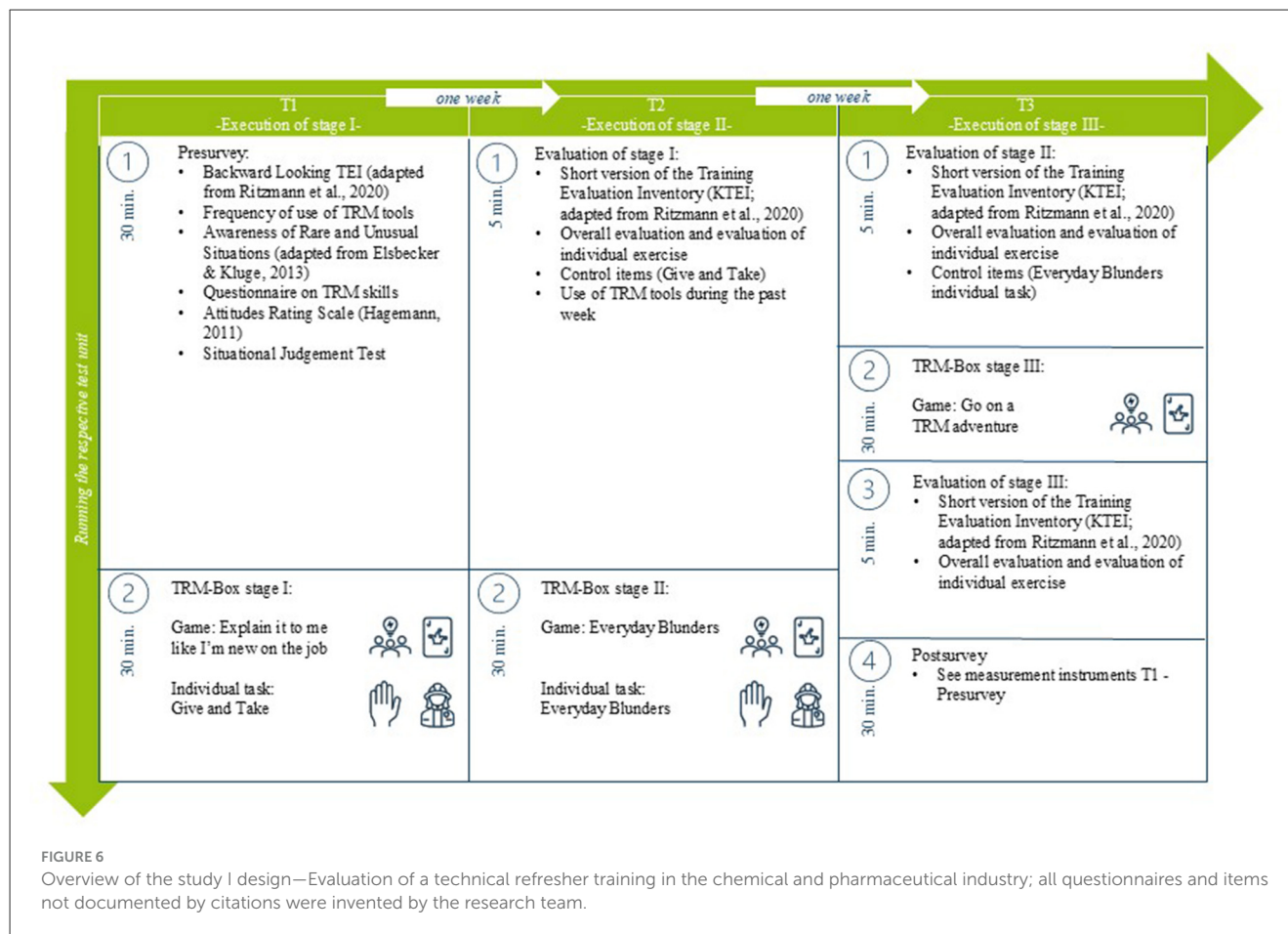
H2a: After the control group (CG) completes the refresher training, the differences between EG and CG in TRM skills and awareness of rare and non-routine situations are no longer significant.

The planning and implementation of the field study were associated with several practical challenges. Due to operational constraints related to shift work, the control group could not be maintained. Moreover, it was not possible to assess TRM skills immediately after the original TRM training. Measurements were implemented only during the refresher phase. In addition, the participants who took part in the refresher training were not the same individuals who had attended the original TRM training. As a result, no baseline measurement of TRM skills could be established, and no longitudinal comparison over the 6-month period was possible. Such limitations are common in field research conducted under real-world organizational conditions and can compromise both internal validity and the strength of causal inferences (Ditlmann and Paluck, 2015). Consequently, the evaluation was conducted as a one-factorial pre-post design with intermediate assessments after each training module.

Since the original hypotheses could not be tested, the following research question was examined in an exploratory manner:

RQ1: Do self-reported TRM skills and awareness of rare and unusual situations increase from pre- to post-measurement during the course of the refresher intervention?

In addition to refreshing selected TRM skills, the training was designed to include new content addressing additional aspects of safety-related behavior. Specifically, modules were developed to focus on managing human error and making decisions in



emergency situations. The following hypothesis was tested to evaluate these training elements:

H3: The TRM Box increases awareness of how to deal with human error and make decisions in safety-critical situations.

3.2 Materials and methods

3.2.1 Sample

The target group selected for the refresher training consisted of chemical technicians who work in the company's shift operations. A total of $N = 45$ individuals participated in the study, distributed across seven training groups of five to eight participants. Upon data inspection, it was revealed that engineers had also taken part in the training, although they were not part of the shift-based operational workforce. These discrepancies in the sample were identified through communication with company representatives. Since it was not possible to differentiate between specific participant groups in the data, all participants were included in the analysis. To ensure participant anonymity, no demographic data were collected. *A-priori* power analyses with G*Power 3.1 (Faul et al., 2009) yielded $N = 57$ for the calculation of paired *t*-tests and $N = 28$ for the ANOVA with repeated measures. In both cases, the sample size was below the required level.

3.2.2 Study design

The field study was implemented as a one-factorial pre-post design comprising four measurement points, each corresponding to a different stage of TRM Box training (see Figure 6). The first measurement was conducted immediately before the first training stage (T1). The second and third measurements were administered directly prior to the start of stages II (T2) and III (T3), respectively. After completion of the third stage, the participants completed additional questionnaires as part of the final evaluation. The TRM Box sessions were facilitated by in-house trainers at various Covestro sites throughout Germany. All evaluations were conducted using paper-pencil questionnaires, which were returned to the research team by mail.

3.2.3 Development of the gamification based training "TRM-Box"

To design the refresher training, a gamification-based approach was selected for the following reasons: (1) Gamification-based learning enables cooperative learning, which is particularly effective in fostering motivation and promoting positive attitudes toward learning content (Fonseca et al., 2023). (2) The short, modular structure of game-based elements is well-suited to shift work environments, in which operators have limited availability to participate in training activities. Accordingly, the training was designed to be completed on-the-job (Baldwin et al., 2009)

TABLE 4 Goals of the TRM box.

Stage	Cognitive goal	Affective goal	Behavioral goal	Game-based elements
(1) Refresher on TRM tools	Refresh knowledge of TRM tools	Deepen understanding through explanation perspective (group task)	Build motivation for the training through a gamification approach Develop commitment to the tandem partner and the group Actively give and receive feedback (individual mission) Reformulate TRM tools in one's own words	Competitive team play Card-based point system Immediate feedback by revealing correct answers on the card backs Clear rules and structured turn taking
(2) Impulse: Human error	Learn about the most common errors (unsafe acts) in one's own workplace Understand common preconditions that lead to unsafe acts	Accept that errors can occur and some are more frequent Recognize that team members make similar mistakes and it is acceptable to talk about them	Reflect on personal unsafe acts on the job and articulate them in a group discussion (individual mission)	Random card draws Symbolic tokens for hidden choices (white/black chips) Collective reveal of group results Immediate feedback triggering reflection and storytelling
(3) Impulse: Decision-making in emergency situations	Link preconditions to one's own decision-making strategies under uncertainty (group task) Connect TRM tools with preconditions and unsafe acts using a selected accident case (group task)	Appreciate one's own knowledge and that of the team Build motivation to engage with the materials beyond the training	Strengthen team cohesion Articulate learning experiences gained during the training	Narrative decision tree with "instant" and "path" cards Branching choices with visible consequences Role play (fictional operator Murat) Multiple endings and replayability

independently by small teams of two or more participants without the need for direct trainer supervision. However, during the evaluation phase, trainers were involved to ensure a higher degree of standardization in the delivery of the intervention.

The "TRM Box" consists of three 40-min modules that can be flexible implemented in a weekly rhythm during regular shift work on the job during passive monitoring periods. Each module includes a group task, conducted in sessions of five to eight participants, which is approximately 30 min long. Additionally, each participant receives an individual task designed to be completed independently or with a tandem partner between sessions, requiring approximately 10 min. All materials are provided in printed form and organized in a physical training box. All training stages were developed based on predefined learning objectives, structured according to Bloom's taxonomy of educational objectives (Table 4; Anderson and Krathwohl, 2001; Bloom, 1976). In the revised taxonomy, cognitive goals describe intended mental processes from remembering and understanding to applying, analyzing, evaluating, and creating, linked to factual, conceptual, procedural, and metacognitive knowledge (Anderson and Krathwohl, 2001). Affective goals concern changes in attitudes and values and range from attending and responding to valuing, organizing, and value-guided action (Bloom, 1976). Behavioral goals express these aims as observable performance by stating the expected action, the conditions, and the criterion for acceptable proficiency, aligned with the relevant cognitive processes (Anderson and Krathwohl, 2001).

The training begins with a general written introduction to the rules. For each stage, there is a separate envelope containing the instructions and all related training materials, which is to be opened only at the beginning of the respective stage. Each stage

starts with the reception of a one-page document outlining the implementation rules.

3.2.3.1 Stage I

The group task "Explain it to me like I'm new on the job" focused on refreshing the five core TRM tools that form the foundation of the TRM training at Covestro: (1) two-way communication, (2) briefing, (3) STAR (Stop-Think-Act-Review), (4) the four-eyes principle, and (5) feedback. Participants were given a set of cards, each displaying a tool on the front and its corresponding definition on the back. Participants were split into two groups that competed against each other to establish a game challenge. During play, participants drew cards and either defined the tool shown or identified the correct tool for a given definition. Correct definitions were awarded two points, while correctly naming the tool earned one point. Feedback on learning progress was provided immediately by revealing the solution printed on the reverse side of each card.

The individual task aimed to foster the ability to give and receive active feedback. At the end of the stage, participants chose a tandem partner and drew two cards, each describing a situation in which they were to provide feedback to their partner during the upcoming week (e.g., "Give feedback to a colleague who connected or disconnected a container or rail tank."). To better adapt the task to their workplace context, participants were allowed to exchange cards. As a support, they received an additional card containing "useful knowledge" about feedback, briefly summarizing research findings on the relevance of feedback in the chemical and pharmaceutical industries.

TABLE 5 Subscales, example items and internal consistencies of the “TRM skills” questionnaire.

Subscale	Example item	Number of items	ω Pre	ω Post
Situational awareness	Gaining an overview of the situation to prevent potential hazards	7	0.938	0.862
Decision making	Anticipating the consequences of decisions	4	0.926	0.671
Stress awareness	Identifying states and causes of mental fatigue	5	0.929	0.858
Awareness of human error	Taking ownership of your own mistakes at work within the team	4	0.888	0.783
Two-way communication	Repeating key points of instructions in one's own words	2	0.637	0.750
STAR	Checking whether the desired result has been achieved after a task	3	0.813	0.739
Four-eyes principle	Questioning a colleague's task execution to ensure safe working practices	3	0.911	0.912
STOP	Immediately interrupt work when anomalies or irregularities occur	4	0.937	0.854
Feedback	Checking whether the recipient has understood the feedback	10	0.946	0.938

ω , McDonald's Omega.

3.2.3.2 Stage II

The second stage began with a group reflection round based on guiding questions, focusing on the individual exercise of the first stage. This was followed by the group activity *Everyday Blunders*, a gamification-based exercise designed to raise participants' awareness of human error in the workplace. The game incorporated typical game-design elements such as rules, chance, symbolic tokens, and immediate feedback (Bedwell et al., 2012). Each participant received a fixed set of white and black chips representing their past experiences with errors. During the game, participants took turns drawing a card that displayed a short description of a common workplace mistake (e.g., “I once failed to use the handrail”). After each card was read aloud, all group members secretly placed a chip in a bag (white = this has never happened to me; black = this has happened to me). The bag was then emptied, and the aggregated group outcome was revealed to all players. This mechanism introduced uncertainty and curiosity about others' behavior, while the symbolic tokens and collective reveal created a playful but structured context for self-reflection. The embedded rule system, the use of tokens as points, and the randomness of card draws ensured that the activity could not be reduced to a mere discussion round. Immediate feedback was provided through the revelation of group responses, which in turn triggered storytelling, comparison, and joint reflection. These gamified elements served to normalize error experiences, encourage openness, and stimulate discussion in a safe and engaging format.

In the individual task, participants were asked to pay attention to their own “everyday blunders” over the following week and document them on a designated card. They also received an additional card containing “useful knowledge” about common types of human error namely attentional slips, memory lapses, and intentional violations (e.g., Marquardt, 2019; Reason, 1990).

3.2.3.3 Stage III

The introduction to the third stage included a reflection round based on selected questions concerning the individual task from stage II. To encourage participants to talk about their own perceptions of error, the guiding questions were determined by rolling a colored die to ensure that not all group members

had to answer all questions. An example was: “*What did you gain from talking to your tandem partner about mishaps?*” In the group task, participants collaboratively shaped the course of a fictional narrative by making decisions. The card game was conceptualized as a decision tree with different card types. There were “instant” cards, indicated by a lightning symbol, which required an immediate choice about how the game's protagonist should act as soon as the card was drawn. “Path” cards described the consequences of the preceding decision and were placed in the center of the table to visualize the unfolding story. The storyline was based on an internal incident related to TRM in the workplace. Participants took on the role of a fictional chemical operator named Murat and followed his experience of the incident. The game featured three possible outcome scenarios (best-case, worst-case, and middle-case scenario), allowing for repeated playthroughs and critical evaluation of decisions made.

The training ended with a forward-looking reflection on how the training content could be applied in the work context. For each game, blank cards were provided that the group to expand the training base and adapt it to the specific conditions of their workplace. All training materials remained at the site after completion of the training, allowing ongoing adaptation and use as needed.

3.2.4 Variables

The TRM Box served as the treatment variable. An overview of the dependent variables is presented below.

3.2.4.1 Dependent variables

3.2.4.1.1 TRM-skills

To assess the skills related to TRM, a questionnaire was developed based on the framework of Mariani et al. (2019). The instrument comprised 42 items distributed across nine subscales. Participants rated the ease or difficulty of applying these skills in their workplace behavior using a scale ranging from “1 = very easy” to “5 = very difficult.” An overview of the subscales and example items is presented in Table 5.

In addition to the competency-based assessment, participants were asked to retrospectively rate how frequently they had applied

each of the five TRM tools since the beginning of the training. This frequency was rated on a scale from “1 = very rarely” to “6 = very often.”

3.2.4.1.2 Awareness of TRM

Awareness of TRM in the workplace was evaluated using an adapted version of the questionnaire “Awareness of Rare and Non-Routine Situations” in a pre-post comparison (see [Elsbecker and Kluge, 2013](#); see [Table 1](#)). As the training did not focus on rare and unusual situations but rather on team resource management, the items were adapted accordingly (e.g., “*I believe we should address the topic of team resource management more intensively at Covestro.*”).

3.2.4.1.3 Dealing with human error

Attitudes toward human error were assessed using the subscales *realistic stress perception* (three items) and *error* (two items) from the “Attitude Rating Scale” for safety-promoting attitudes in high-responsibility teams ([Hagemann, 2011](#)). A total of five items were administered at time points T1 and T4. Beyond pre and post-assessments, the subscales were also employed in T3 as part of the evaluation of stage II, which focused on raising awareness for dealing with human error. Responses were rated on a scale from “0 = complete disagreement” to “4 = complete agreement.” An example item on the *error* subscale is: “*I am more likely to make mistakes in tense or critical situations.*”

3.2.4.1.4 Decision-making in safety-critical situations

To evaluate decision-making in safety-critical situations, four case vignettes were developed in collaboration with the organization in the format of a situational judgment test (SJT; [Christian et al., 2010](#)). Each vignette presented participants with a safety-critical scenario and four possible decision options. Each option had a predefined score: one point was assigned to the least safe decision and four points to the safest decision. The following example illustrates one of the test cases, titled “*Incorrect Sampling.*” Illustrative images were presented for all vignettes. These are not displayed for the example item, as they constitute internal company materials.

3.2.4.1.5 Descriptive training evaluation

To evaluate the original TRM training after a 6-month interval, Backward-Looking TEI (adapted from [Ritzmann et al., 2014](#); see [Table 1](#)) was administered at time point T1. This questionnaire was also used to evaluate the TRM Box at time point T4. Additionally, each training stage was descriptively evaluated. For this purpose, a short version of the Training Evaluation Inventory (KTEI; adapted from [Ritzmann et al., 2014](#)) was used. To ensure temporal efficiency, a representative item was selected for each of the *subjective fun*, *perceived usefulness*, *subjective knowledge growth*, and *attitude toward training* (e.g., subjective fun: “I generally liked the first stage of the TRM box.”). The subscale *integration* was omitted, as its items could not be meaningfully adapted to the short time frame of 1 week.

The usefulness of each stage and its individual exercises for workplace safety behavior was additionally assessed using individual items. For the overall evaluation, participants responded

BOX 1 Example SJT Scenario: “Incorrect Sampling.”

You observe a colleague taking a morning sample of a mixture containing caustic soda (sodium hydroxide) and bleach. At this time, renovation work is being carried out in the wastewater pump of the hydrogen treatment unit, and no water or caustic soda may enter the foundation. There is a designated sampling point and another sampling point near the tank. You are fairly certain that your colleague has taken the sample from the wrong location, namely the one near the tank. However, the colleague has already left to deliver the sample to the lab. How do you respond?

(A) I catch up with the colleague and inform him about the mix-up. No one else needs to know; after all, everyone makes mistakes from time to time.

(B) I immediately inform the shift supervisor and suggest that all labeling in that plant section be checked. In addition, I recommend that the mix-up be addressed in the next morning meeting to prevent recurrence.

(C) I catch up with the colleague, since the incorrect sample could contaminate the laboratory equipment or even pose a risk to the laboratory technician (e.g., due to incorrect PPE). I also inform the shift supervisor and ask that the issue be mentioned in the next morning meeting.

(D) I rush to the lab to try to stop any contamination in time. I know the colleague is currently under stress, because his basement was recently flooded, and is generally very experienced. I assume this is a one-time mistake, so I will not tell him/her about the mix-up.

Scoring: (A) = 1 point; (B) = 4 points; (C) = 3 points; (D) = 2 points

to the question “*How would you rate this stage of the TRM Box in terms of supporting the application of the TRM goals on the workplace?*” using a scale from “1 = very poor” to “10 = very good.”

The usefulness of specific exercises was rated on a scale from “1 = not useful at all” to “5 = very useful” (for example, “*How useful did you find the group exercise ‘Everyday Blunders’ in Stage 2 of the TRM Box to improve workplace safety behavior?*”).

Control items were used to verify whether the individual task had been carried out during the previous week (e.g., “*Did you actively record daily blunders as part of the individual task last week? YES/NO*”).

3.3 Results

3.3.1 Descriptive results

The results of the descriptive analysis are presented in [Table 6](#). To ensure comparability, only responses with complete data were included on all backward-looking TEI subscales and all subscales of the TEI short version ($N = 17$). Many of the paper-pencil questionnaires contained missing values, resulting in varying sample sizes between individual results.

Compared to the evaluation results of Study 1 and reference samples in the literature ([Ritzmann et al., 2014](#)), the results were lower. Therefore, the following additional exploratory research question was formulated to assess whether the newly developed TRM Box is rated equally well as the original TRM training and thus integrates effectively into the company’s qualification program.

TABLE 6 Descriptive results of the training evaluation—Study 2.

Subscale	Presurvey M (SD)	Stage I M (SD)	Stage II M (SD)	Stage III M (SD)	Postsurvey M (SD)	ω Pre	ω Post
Perceived usefulness	3.73 (0.79)	3.71 (0.72)	4.00 (0.63)	4.05 (0.85)	3.49 (0.71)	0.885	0.795
Subjective knowledge growth	3.68 (0.71)	3.43 (0.93)	3.69 (0.79)	3.47 (0.70)	3.63 (0.67)	0.836	0.738
Attitude toward training	3.33 (0.80)	3.23 (1.06)	3.69 (0.87)	3.53 (0.90)	2.87 (0.70)	.	.
Application	3.48 (0.87)	4.05 (0.67)	3.88 (0.62)	3.68 (0.58)	3.26 (0.87)	.	.
Integration	3.45 (0.55)	–	–	–	3.24 (0.75)	.	.

M, Mean; SD, Standard Deviation; ω , McDonald's Omega, where ω is not reported, the subscale comprised < 3 items (.).

TABLE 7 Ratings of the individual training stages.

Training unit	M	SD	Scale
Stage I	6.35	1.42	1–10
Explain it to me like I'm new on the job	3.15	0.88	1–5
Give and take	3.60	0.75	1–5
Stage II	6.80	1.51	1–10
Everyday blunders—group task	3.80	0.89	1–5
Everyday blunders—individual task	3.75	0.79	1–5
Stage III	6.65	2.06	1–10
Go on a TRM adventure	3.50	0.76	1–5

M, Mean; SD, Standard Deviation.

RQ2: Does the evaluation of the TRM Box differ significantly from the evaluation of the original TRM training?

The overall ratings of each training stage as well as the evaluations of the individual exercises are presented in Table 7. Complete data were available from $N = 20$ participants.

Overall, all descriptive values are slightly above the midpoint of the respective rating scales and can therefore be interpreted as moderate to slightly positive.

The analysis of the control items showed that, in the week following Stage I, $N = 31$ out of $N = 35$ participants (88.57%) actively gave feedback, while $N = 10$ participants did not respond to the item. Of the $N = 35$ valid responses, $N = 21$ participants (60%) actively received feedback. After Stage II, $N = 28$ participants responded to the control item. Of these, only $N = 11$ participants (32.14%) actively documented their “everyday blunders.”

3.3.2 Inferential statistics

RQ1: Do self-reported TRM skills and awareness of rare and unusual situations increase from pre- to post-measurement during the course of the refresher intervention?

To examine the research question regarding changes in self-reported TRM skills and awareness of TRM, pre-post comparisons were conducted using paired-sample t -tests. Specifically, all subscales of the TRM competence questionnaire, the reported frequency of TRM tool use, and the awareness of TRM scale were analyzed. The results are summarized in Table 8.

Since none of the comparisons reached statistical significance, the assumption of a competence improvement resulting from the refresher training could not be supported.

H3: The TRM Box increases awareness of how to deal with human error and make decisions in safety-critical situations.

The analysis of the hypothesis was divided into two parts. To examine whether the TRM Box provided a new impulse for dealing with human error in the workplace, the item scales were first recoded from 0–4 to 1–5. A one-way repeated measures ANOVA with Greenhouse-Geisser correction revealed no significant main effect for *realistic stress perception* [$F_{N=25(1.62,38.84)} = 0.17$, $p = 0.798$, $\eta_p^2 = 0.01$] or *error* [$F_{N=25(1.72,41.24)} = 0.88$, $p = 0.406$, $\eta_p^2 = 0.04$].

Means in the presurvey were $M = 3.68$ ($SD = 0.79$) for *realistic stress perception* and $M = 3.32$ ($SD = 0.81$) for *error*. These values remained stable 1 week after the second stage of the intervention, with $M_{realisticstressperception} = 3.76$ ($SD = 0.98$) and $M_{error} = 3.28$ ($SD = 0.61$). After completion of the TRM-Box, the values remained close to baseline with $M_{realisticstressperception} = 3.76$ ($SD = 0.56$) and $M_{error} = 3.45$ ($SD = 0.54$), indicating no statistically significant change. Accordingly, the hypothesis that the TRM Box would significantly improve awareness of stress and error handling in safety-critical situations could not be confirmed.

The evaluation of the Situational Judgment Test ($M_{pre} = 2.36$, $SD = 0.42$; $M_{post} = 2.35$, $SD = 0.37$) showed no significant change in performance, $t_{one-tailed(22)} = 0.14$, $p = 0.445$, $r = 0.03$. However, item-level analysis revealed a significant improvement for the fourth test vignette ($M_{pre} = 3.36$, $SD = 0.70$; $M_{post} = 3.72$, $SD = 0.46$), $t_{one-tailed(24)} = 3.17$, $p = 0.004$, $r = 0.55$.

An analysis of item-specific change scores (Ferreira et al., 2013) indicated that 18.75% of participants improved and 17.02% declined in performance on the first vignette. For the second vignette, 17.54% showed improvement and 6.00% declined. In the third vignette, 4.92% improved and 1.69% declined. For the fourth vignette, only improvements were observed (13.24%). Overall, more gains than losses in performance were recorded, although item difficulty varied.

RQ2: Does the evaluation of the TRM Box differ significantly from the evaluation of the original TRM training?

To compare the two trainings using the Backward-Looking TEI, a paired-sample t -test was conducted on the overall mean scores. The original TRM training ($M = 3.57$, $SD = 0.64$) was rated slightly more positively than the TRM Box ($M = 3.24$, $SD = 0.66$). However, this difference did not reach statistical significance, $t_{(25)} = 1.90$, $p = 0.069$, $r = 0.35$.

TABLE 8 Pre-post comparison of TRM tool usage and TRM skills.

Instrument	Pre M (SD)	Post M (SD)	t	p	r	N
Use of TRM tools						
Two-way communication	4.10 (1.74)	4.25 (1.77)	0.53	0.603	0.12	19
Briefing	4.05 (1.43)	4.30 (1.81)	0.63	0.536	0.14	20
STAR	3.45 (1.23)	3.00 (1.56)	1.58	0.131	0.35	19
Four-eyes principle	4.30 (1.49)	4.30 (1.30)	0.00	1.00	0.00	20
STOP	3.05 (1.57)	2.40 (1.64)	1.66	0.114	0.36	19
Feedback	3.50 (1.15)	3.95 (1.39)	1.31	0.206	0.29	20
TRM skills						
Situational awareness	2.25 (0.61)	2.21 (0.66)	0.25	0.807	0.05	27
Decision making	2.25 (0.64)	2.15 (0.58)	0.79	0.436	0.15	27
Stress awareness	2.57 (0.83)	2.56 (0.74)	0.60	0.953	0.12	27
Awareness of human error	2.43 (0.66)	2.34 (0.53)	0.77	0.450	0.15	28
Two-way communication	2.48 (0.71)	2.44 (0.63)	0.37	0.711	0.07	27
STAR	2.02 (0.67)	1.96 (0.63)	0.61	0.545	0.12	28
Four-eyes principle	2.28 (0.80)	2.23 (0.24)	0.96	0.786	0.19	26
STOP	2.09 (0.71)	2.09 (0.68)	0.00	1.00	0.00	27
Feedback	2.55 (0.69)	2.49 (0.75)	0.49	0.629	0.10	27
Awareness of TRM	3.13 (0.63)	3.22 (0.60)	1.09	0.359	0.21	27

M, Mean; SD, Standard Deviation. r, Rosenthal effect size for paired-samples t-tests. Values reported as 0.00 indicate $r < 0.01$.

3.4 Discussion

The second study examined the effectiveness of a gamified refresher training (TRM Box) intended to mitigate the decay of non-technical skills among chemical technicians working in shift-based operations. The training was delivered in the form of three short, modular sessions integrated into everyday work routines.

The originally planned quasi-experimental design with a delayed control group could not be implemented due to operational constraints related to shift work. As a result, no untreated comparison group was available, and the evaluation was conducted using a simplified pre-post design. This reduces internal validity and limits the ability to draw causal inferences about the effectiveness of the intervention. Furthermore, the participants who took part in the refresher training were not the same individuals who had completed the initial TRM program. This disruption in continuity made it impossible to track skill retention or decay over the 6-month interval. Without access to baseline data from the original training, the extent to which the TRM Box restored or preserved previously acquired competencies remains unclear.

Regarding RQ1, which investigated whether TRM skills and awareness of rare and unusual situations increased during the intervention, no significant improvements were identified. This finding should be interpreted in relation to the objective of the training. The TRM Box was not designed to convey new content, but rather to support the continued application of previously acquired skills. The absence of measurable decline may indicate that existing competencies were preserved over time. However, this interpretation must be viewed with caution, as the sample included

individuals such as engineers who are not directly involved in operational processes. For these participants, the relevance of the TRM skills may have been limited, which could have weakened the overall effect of the intervention. This highlights the importance of aligning refresher trainings with the specific tasks and responsibilities of the target group.

H3 explored whether the TRM Box increased awareness of how to deal with human error and supported decision-making in safety-critical situations. The results did not show any significant changes in perceived stress or attitudes toward human error. These findings suggest that short and low-intensity training was insufficient to influence deeper cognitive or attitudinal dimensions. A similar pattern emerged in the results of the Situational Judgment Test. Only one item, which was closely aligned with the content of one of the training modules, showed a statistically significant improvement. This suggests that although the training may have had an impact on specific and targeted content, it did not lead to broader changes in decision-making performance.

Additional insights were provided by the control questions. Although most of the participants reported actively giving and receiving feedback after the first session, only a small proportion engaged with more reflective individual exercises, such as documenting their own daily errors. This limited participation likely reduced the potential of the intervention to support attitudinal change or promote deeper reflection on safety-related behavior. Although the TRM Box may have initiated short-term reflection, the lack of consistent participation in the individual tasks restricted its effectiveness in fostering lasting change. Additionally, incomplete or missing responses in the

paper-and-pencil questionnaires further reduced the usable sample size and may have introduced bias in the analyses.

In relation to RQ2, which compared the TRM Box to a more comprehensive, previously conducted TRM training, no significant differences in participant evaluations were found. Despite its shorter duration and reduced instructional scope, the TRM Box was rated equally useful and relevant. It should be noted that the previous training had been evaluated 6 months after completion, whereas the TRM Box was evaluated immediately after implementation. The similarity in the evaluation results suggests that the refresher format was perceived as an effective and efficient method of supporting existing competencies.

Overall, the TRM Box appears to have been successful in maintaining the application of non-technical skills, but did not lead to measurable improvements. Competencies such as situational awareness, decision making under uncertainty, and the management of human error can require more extensive, repeated, and immersive training formats to be sustainably developed (Flin et al., 2008; Salas et al., 2006). While brief refresher interventions can reinforce previously learned content, they are unlikely to support the acquisition of new competencies or bring about significant changes in attitudes and behavior.

4 General discussion

The present research article addresses a central challenge in the chemical industry as a high-reliability sector: how to preserve critical technical and non-technical skills under conditions of increasing automation and procedural standardization. As work environments become more structured and routine-based, the application of infrequent but safety-critical competencies such as managing emergency situations or addressing human error, becomes increasingly rare. In the absence of regular use, these skills are subject to gradual decay, both cognitively and procedurally, as described in prior research on skill retention in high-risk domains (Arthur et al., 1998). The two field studies presented here contribute to the growing body of research on skill retention by empirically testing low-threshold, operationally integrated refresher formats developed in collaboration with an industrial partner.

The findings suggest that brief refresher interventions can be both feasible and well-accepted in real-world production settings. Although the measurable effects of the interventions were limited, partly due to methodological constraints, the results indicate that the targeted reinforcement of technical and non-technical competencies is achievable even under the time pressure and structural limitations of shift-based operations. The studies also indicate that principles from high-risk domains such as aviation, especially crew resource management, can be effectively adapted to the process industry when content and delivery are aligned with operational conditions. This work expands the empirical foundation for the use of refresher training in industrial environments. Both virtual reality simulations and gamified on-the-job interventions proved to be suitable for maintaining key competencies, although they served different purposes. The interventions were not designed to replace full-scale qualification programs, but rather to function as maintenance tools that support the continued application of previously acquired knowledge

and routines. This approach aligns with established theories of skill decay, which emphasize that long-term retention depends on repeated, context-relevant activation, and practice (Arthur et al., 1998; Frank and Kluge, 2018). Even though no significant improvements were observed in the non-technical domain, the consistently positive participant feedback and high degree of integration into daily work routines highlight the potential of low-threshold formats for long-term reinforcement.

4.1 Limitations

The findings of the present research must be interpreted in light of several methodological and contextual limitations that apply across both studies. Conducting training evaluations directly within industrial environments ensured high ecological validity but came at the expense of experimental control. Organizational and operational constraints related to shift work, production schedules, and limited participant access restricted sample sizes and prevented random assignment. These contextual limitations also made it impossible to implement control groups or maintain participant continuity across training phases. As a result, both studies relied on quasi-experimental designs that allow for examining practical feasibility and immediate effects but do not support strong causal inference, a common challenge in field-based research where organizational realities constrain methodological rigor (Ditlmann and Paluck, 2015). The absence of untreated comparison groups and the lack of participant continuity between the initial and refresher trainings further limit the ability to determine whether observed differences can be attributed to the interventions or to external factors such as workplace learning or experience.

Another limitation concerns the nature of the measurement instruments. Both studies relied primarily on self-report data and scenario-based judgment tests, which provide valuable subjective perspectives but offer only indirect evidence of behavioral competence in real operational contexts. The absence of behavioral or performance-based indicators restricts the extent to which the findings can be generalized to actual workplace behavior.

Data quality also constrains the interpretability of the results. Incomplete or missing responses reduced the usable sample sizes, and demographic or role-specific information was not consistently collected, which limits the ability to assess representativeness and subgroup differences. Digital data collection tools with integrated completeness checks could mitigate such issues in future evaluations, but within the current research context these limitations necessarily reduce analytical precision.

Finally, the short temporal scope of both studies confines the interpretation of the observed effects to immediate post-training outcomes. Long-term retention and the sustainability of training effects cannot be inferred from the available data, and the conclusions must therefore be viewed as preliminary.

4.2 Implications for further research

Future research should examine the long-term effectiveness of refresher trainings in maintaining and restoring essential

competencies in the chemical process industry. In particular, longitudinal studies are needed to determine whether short targeted interventions can produce measurable effects on behavior during rare but safety-critical events, such as emergency shutdowns or process deviations. These investigations should be based on objective data, including structured behavioral observations and long-term performance tracking within operational environments.

Further studies should also explore how the structure and delivery of refresher trainings influence their effectiveness. In particular, it would be useful to investigate how active forms of engagement, such as scenario-based problem solving or structured discussion of past events, support the reinforcement of technical and non-technical skills. Research in high-risk domains suggests that training formats that promote active participation and contextualized application are particularly beneficial for the development of safety-relevant competencies (Salas et al., 2006). A more precise understanding of how different formats align with task requirements and learner profiles could inform the design of interventions that are effective and context-appropriate.

Furthermore, future research should address how refresher training can be systematically integrated into existing qualification systems and operational routines in the chemical industry. Organizational conditions such as shift planning, the availability of internal facilitators and the degree of digital infrastructure are likely to shape both the quality of implementation and long-term sustainability. Clarifying these structural factors would support the institutionalization of refresher-based competency maintenance in high-reliability production environments.

4.3 Implications for practice

The findings of this research suggest that refresher training can be implemented effectively within the operational reality of the chemical industry, provided that their design is compatible with the constraints of shift-based production. Short and targeted interventions, such as those examined in the present studies, can be performed without significant disruption of existing workflows and were generally well-received by participants. Their integration into daily routines points to the practical feasibility of low-threshold approaches to competence maintenance.

In the case of the VR-based intervention evaluated in Study 1, the results indicate that technical skills relevant to emergency procedures can be refreshed successfully within a simulated training environment. Although it was carried out outside the regular work environment, the intervention was positively evaluated by the participants and proved flexible enough to be scheduled in conjunction with operational demands. The short duration of the VR sessions allowed for minimal time investment while still providing a realistic and immersive experience that supported the reactivation of safety-relevant routines. This efficiency is supported by meta-analytic evidence indicating that VR-based safety training significantly outperforms traditional methods in knowledge acquisition and retention across multiple industries (Scorgie et al., 2024). Studies in industrial contexts have shown that realistic audio-visual cues within immersive VR

environments enhance short-term memory retention of safety procedures (Lochmannová et al., 2022). This suggests that virtual simulation formats can serve as an efficient and scalable tool for maintaining technical competence, even in production contexts where time and resources are tightly managed.

For such interventions to be effective, their relevance to the specific responsibilities appears to be a crucial factor. When training scenarios are closely aligned with actual operational conditions and the concrete experiences of employees, they are more likely to support the retention and activation of knowledge and skills. This alignment becomes even more critical in heterogeneous teams, where members may vary in their exposure to hands-on operational tasks and, therefore, benefit differently from shared learning experiences (Salas et al., 2008).

In addition to supporting individual learning, refresher trainings can also contribute to the development of shared team norms around safety. The continuous availability of materials such as the TRM Box within the work environment created opportunities for informal exchange, peer learning, and content adaptation based on local practice. Such integration into daily communication structures can enhance the visibility and long-term relevance of non-technical skills in safety-critical settings. The low resource requirements of both interventions further underline their practical value. Neither the VR training nor the TRM Box format relied on external trainers or dedicated facilities, allowing for scalable implementation within existing organizational structures. Against the backdrop of increasing efficiency demands in high-reliability industries, this approach may offer a sustainable path toward preserving operational competence.

Beyond these implications, both field studies illustrate the structural trade-offs that characterize ecological training research in high-reliability industrial environments. Achieving ecological validity in such contexts inevitably involves compromises in methodological precision, as research must adapt to the constraints of real production systems. These contextual boundaries are not mere obstacles but part of what defines training in its natural environment. They shape the degree to which learning interventions can be embedded in ongoing operations and reveal how methodological rigor interacts with organizational feasibility.

Furthermore, the diversity of participant roles and the integration of training activities into daily production routines highlight that learning in industrial settings is rarely an isolated individual process. Instead, competence maintenance emerges from collective routines, informal exchanges, and local adaptations of safety practices. In the present studies, these factors became visible through small sample sizes, the limited continuity between training cohorts, and the necessity to align data collection with operational workflows. At the same time, they provided important insight into the organizational mechanisms that enable training to persist under real production conditions, such as management support, scheduling flexibility, and local ownership of learning processes.

These recurring challenges emphasize the need to design research strategies that consciously accommodate ecological complexity. The practical approaches summarized in Table 9, including flexible scheduling, digitalized data collection, and participatory scenario development, build directly on these

TABLE 9 Practical challenges in ecological training studies in the chemical industry and possible solutions.

Challenge	Description in chemical training context	Consequences for research and practice	Possible solutions and recommendations
Limited sample sizes	Quasi-experimental designs in field settings are constrained by operational structures and small groups per shift	Low statistical power—limited generalizability	Use repeated-measures designs, pool data across sites, complement with qualitative feedback
Lack of control groups and disrupted continuity	Shift work and staffing changes prevented delayed control groups and continuity across trainings	Reduced internal validity—limited causal inferences—no long-term retention tracking	Employ alternative comparison strategies such as historical controls, use within-subject baselines, plan flexible scheduling with plant management
Heterogeneous participant backgrounds	Samples included shift workers and engineers with very different exposure to operational practice	Unequal baseline competence—potential bias in results	Collect minimal demographic and role data, stratify analyses, adapt scenarios to the most relevant target group
Missing and incomplete data	Paper-and-pencil questionnaires led to high item non-response and incomplete sections	Reduced usable sample size—potential systematic bias	Use digital data collection with built-in completeness checks, simplify instruments
Limited ecological validity of scenarios	Simulated emergency cases did not fully capture complexity, urgency, or emotional load of real incidents	Limited transfer—underestimation of stress and decision-making demands	Co-develop scenarios with experienced operators, integrate realistic audio and visual cues, pilot test realism
Short follow-up intervals	Only 2 weeks between training and measurement in Study 1	No insights into long-term retention	Add follow-ups after 3–6 months, integrate booster tasks into daily routines

empirical experiences and offer concrete ways to strengthen methodological robustness while maintaining practical feasibility in comparable field settings.

5 Conclusion

This research highlights the practical value of targeted refresher interventions for maintaining critical skills in the chemical industry under conditions of increasing automation. The results demonstrate that both technical and non-technical competencies can be supported through brief, context-sensitive training formats that integrate into operational routines. To maintain workforce readiness in safety-critical environments, organizations should consider implementing low-threshold, scalable training strategies as part of their long-term competence management.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

The studies involving humans were approved by the Local Ethics Committee of the Faculty of Psychology (Approval No. 768, dated 20 January 2022). The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin. Written informed consent was obtained from the individual(s) for the publication of any potentially identifiable images or data included in this article.

Author contributions

OV: Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. SC: Funding acquisition, Project administration, Writing – review & editing. A-LG: Investigation, Writing – original draft. AK: Conceptualization, Funding acquisition, Methodology, Supervision, Writing – review & editing.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The work on which this publication is based was carried out on behalf of the Federal Institute for Vocational Education and Training (BIBB), thereby contributing to the promotion of vocational education and training. In accordance with the nature of scientific research, the views expressed by the author do not necessarily reflect those of the Federal Institute for Vocational Education and Training.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that Gen AI was used in the creation of this manuscript. ChatGPT (OpenAI) was used to support the writing process by generating suggestions for rewording and improving clarity in selected passages. Additionally, the tool assisted in converting tables to APA 7 formatting standards and adapting reference entries to comply with the Harvard citation

style. The final content was critically reviewed and approved by the author(s).

Any alternative text (alt text) provided alongside figures in this article has been generated by Frontiers with the support of artificial intelligence and reasonable efforts have been made to ensure accuracy, including review by the authors wherever possible. If you identify any issues, please contact us.

References

- Anderson, L. W., and Krathwohl, D. R. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York, NY: Longman.
- Arthur, H., Wall, D., and Halligan, A. (2003). Team resource management: a programme for troubled teams. *Clin. Governance Int. J.* 8, 86–91. doi: 10.1108/14777270310460001
- Arthur, W., Bennett, W., Stanush, P. L., and McNelly, T. L. (1998). Factors that influence skill decay and retention: a quantitative review and analysis. *Human Perform.* 11, 57–101. doi: 10.1207/s15327043hup1101_3
- Arthur, W. Jr., and Day, E. A. (2013). "Introduction: knowledge and skill decay in applied research," in *Individual and Team Skill Decay: The Science and Implications for Practice*, eds. W. Arthur Jr., E. A. Day, W. Bennett Jr., and A. M. Portrey (Routledge), 3–11. doi: 10.4324/9780203576076
- Bainbridge, L. (1983). Ironies of automation. *Automatica* 19, 775–779. doi: 10.1016/0005-1098(83)90046-8
- Baldwin, T. T., Ford, J. K., and Blume, B. D. (2009). "Transfer of training 1988–2008: an updated review and new agenda for future research," in *International Review of Industrial and Organizational Psychology*, Vol. 24, eds. G. P. Hodgkinson and J. K. Ford (Chichester: Wiley), 41–70. doi: 10.1002/9780470745267.ch2
- Barnett, M., Gatfield, D., and Pekcan, C. (2006). "Non-technical skills: the vital ingredient in world maritime technology," in *Proceedings of the International Conference on World Maritime Technology*.
- Bedwell, W. L., Pavlas, D., Heyne, K., Lazzara, E. H., and Salas, E. (2012). Toward a taxonomy linking game attributes to learning: an empirical study. *Simul. Gaming* 43, 729–760. doi: 10.1177/1046878112439444
- Bloom, B. S. (1976). *Human Characteristics and School Learning*. New York, NY: McGraw-Hill.
- Brown, L. D., and Poulton, M. M. (2018). "Improving safety training through gamification: an analysis of gaming attributes and design prototypes," in *Advances in Human Factors in Simulation and Modeling. AHFE 2018. Advances in Intelligent Systems and Computing*, Vol. 780 (Cham: Springer), 392–403. doi: 10.1007/978-3-319-94223-0_37
- Christian, M. S., Edwards, B. D., and Bradley-Geist, J. C. (2010). Situational judgment tests: constructs assessed and a meta-analysis of their criterion-related validities. *Pers. Psychol.* 63, 83–117. doi: 10.1111/j.1744-6570.2009.01163.x
- Conein, S. (2020). *Berufsbildung 4.0 – Fachkräftequalifikationen und Kompetenzen für die digitalisierte Arbeit von morgen: Der Ausbildungsberuf Verfahrensmechaniker/-in für Kunststoff- und Kautschuktechnik im Screening [Vocational Training 4.0 – Qualifications and Competencies for the Digitized Work of Tomorrow: The Occupation of Process Mechanic for Plastics and Rubber Technology in Screening]* (No. 210). Wissenschaftliche Diskussionspapiere. Bonn: Bundesinstitut für Berufsbildung (BIBB).
- Conein, S., and Felkl, T. (2023). "Competence retention for non-routine activities in digitized working environments (CONDITION) – studies based on the professions of chemical technician and pharmaceutical technician," in *Proceedings TecPsy 2023*, 95. doi: 10.2478/9788366675896-008
- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). "From game design elements to gamefulness: defining gamification," in *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, eds. A. Lugmayr, H. Franssila, C. Safran, and I. Hammouda (New York, NY: ACM), 9–15. doi: 10.1145/2181037.2181040
- Dingee, A. (2019). "The traits of a high reliable organization (HRO)," in *SPE Offshore Europe Conference and Exhibition* (Richardson, TX: Society of Petroleum Engineers), D021S009R005. doi: 10.2118/195714-MS
- Ditlmann, R., and Paluck, E. L. (2015). "Field experiments," in *International Encyclopedia of the Social and Behavioral Sciences*, Vol. 9, ed. J. D. Wright (Oxford: Elsevier), 128–134. doi: 10.1016/B978-0-08-097086-8.10542-2
- EASA (2015). *AMC and GM to Part-ORO — Issue 2, Amendment 4*. Available online at: <https://www.easa.europa.eu/en/downloads/20242/en> (Accessed November 2025).
- Elsbecker, G., and Kluge, A. (2013). "Können Arbeitssicherheitsseminare (allein) das Verhalten nachhaltig verändern?" in *Tagungsband Frühjahrskongress der GfA, Krefeld, 28 February–1 März 2013* (Dortmund: GfA-Verlag), 275–278.
- Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149
- Feng, Y., González, V. A., Amor, R., Lovreglio, R., and Cabrera-Guerrero, G. (2018). Immersive virtual reality serious games for evacuation training and research: a systematic literature review. *Comput. Educ.* 127, 252–266. doi: 10.1016/j.compedu.2018.09.002
- Ferreira, M. L. P., Almeida, R. M. V. R., and Luiz, R. R. (2013). A new indicator for the measurement of change with ordinal scores. *Qual. Life Res.* 22, 1999–2003. doi: 10.1007/s11136-012-0288-2
- Flin, R., and Maran, N. (2015). Basic concepts for crew resource management and non-technical skills. *Best Prac. Res. Clin. Anaesthesiol.* 29, 27–39. doi: 10.1016/j.bpa.2015.02.002
- Flin, R., O'Connor, P., and Crichton, M. (2008). *Safety at the Sharp End: A Guide to Non-Technical Skills*. Burlington, VT: Ashgate Publishing Company.
- Flin, R., O'Connor, P., and Mearns, K. (2002). Crew resource management: Improving teamwork in high reliability industries. *Team Perform. Manage. Int. J.* 8, 68–78. doi: 10.1108/13527590210433366
- Fonseca, I., Caviedes, M., Chantré, J., and Bernate, J. (2023). Gamification and game-based learning as cooperative learning tools: a systematic review. *Int. J. Emerg. Technol. Learn.* 18, 4–23. doi: 10.3991/ijet.v18i21.40035
- Fracaró, S. G. (2022). *Effectiveness of Immersive Learning Technologies for Training of Operators in the Chemical Industry*. Available online at: <https://lirias.kuleuven.be/retrieve/684851> (Accessed January 14, 2025).
- Fracaró, S. G., Tehreem, Y., Toyoda, R., Gallagher, T., Glassey, J., Bernaerts, K., et al. (2024). Benefits and impact of emergency training in a VR environment. *Educ. Chem. Eng.* 48, 63–72. doi: 10.1016/j.ece.2024.05.004
- Frank, A. G., Dalenogare, L. S., and Ayala, N. F. (2019). Industry 4.0 technologies: implementation patterns in manufacturing companies. *Int. J. Prod. Econ.* 210, 15–26. doi: 10.1016/j.iipe.2019.01.004
- Frank, B., and Kluge, A. (2018). Complex cognitive skill retention: the roles of general mental ability and refresher interventions in a simulated vocational setting. *J. Comput. Assist. Learn.* 34, 471–481. doi: 10.1111/jcal.12251
- Hagemann, V. (2011). *Trainingsentwicklung für High-Responsibility-Teams: Eine systematische Analyse von High-Responsibility Team-Arbeitskontexten und Ableitung der High Responsibility Team-spezifischen Situationen der Trainingsziele mit anschließender Trainingsevaluation [Training Development for High-Responsibility Teams: A Systematic Analysis of High-Responsibility Team Work Contexts and Derivation of Team-Specific Training Objectives with Subsequent Training Evaluation]*. Lengerich: Pabst Science Publishers. Zugl.: Universität Duisburg-Essen, Diss.
- Hagemann, V., and Heinemann, L. (2024). "Führung in Kontexten von Hochrisikooorganisationen [Leadership in high-risk organizational contexts]," in *Handbuch Mitarbeiterführung*, ed. J. Felfe (Berlin: Springer), 571–588. doi: 10.1007/978-3-662-68185-5_53
- Hagemann, V., and Kluge, A. (2013). The effects of a scientifically based team resource management intervention for fire service teams. *Int. J. Human Factors Ergon.* 2, 196–220. doi: 10.1504/IJHFE.2013.057617
- Hagemann, V., and Kluge, A. (2017). Complex problem solving in teams: the impact of collective orientation on team process demands. *Front. Psychol.* 8:1730. doi: 10.3389/fpsyg.2017.01730

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Hammermann, A., and Stettes, O. (2016). *Qualifikationsbedarf und Qualifizierung – Anforderungen im Zeichen der Digitalisierung [Qualification Needs and Training – Requirements in the Context of Digitalization]*. IW Policy Paper No. 3. Köln: Institut der deutschen Wirtschaft Köln.
- Hustad, J., Johannesen, B., Fossum, M., and Hovland, O. J. (2019). Nursing students' transfer of learning outcomes from simulation-based training to clinical practice: a focus-group study. *BMC Nurs.* 18:53. doi: 10.1186/s12912-019-0376-5
- Kanade, S. G., and Duffy, V. G. (2022). "Use of virtual reality for safety training: a systematic review," in *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management. Health, Operations Management, and Design (Lecture Notes in Computer Science)*, Vol. 13320, ed. V. G. Duffy (Cham: Springer), 364–375. doi: 10.1007/978-3-031-06018-2_25
- Klostermann, M., Conein, S., Felkl, T., and Kluge, A. (2022). Factors influencing attenuating skill decay in high-risk industries: a scoping review. *Safety* 8:22. doi: 10.3390/safety8020022
- Kluge, A. (2014). *The Acquisition of Knowledge and Skills for Taskwork and Teamwork to Control Complex Technical Systems: A Cognitive and Macroergonomics Perspective*. New York, NY: Springer. doi: 10.1007/978-94-007-5049-4
- Kluge, A., and Frank, B. (2014). Counteracting skill decay: four refresher interventions and their effect on skill and knowledge retention in a simulated process control task. *Ergonomics* 57, 175–190. doi: 10.1080/00140139.2013.869357
- Landers, R. N. (2015). Developing a theory of gamified learning: linking serious games and gamification of learning. *Simul. Gaming* 45, 752–768. doi: 10.1177/1046878114563660
- Lee, H. (2025). A unified conceptual model of immersive experience in extended reality. *Comput. Hum. Behav. Rep.* 18:100663. doi: 10.1016/j.chbr.2025.100663
- Lekka, C. (2011). *High Reliability Organisations: A Review of the Literature*. Health and Safety Executive. Available online at: <https://www.hse.gov.uk/research/rpd/rpdr899.pdf> (Accessed May, 2025).
- Lochmannová, A., Šimon, M., Horejší, P., Bárdy, M., and Reichertová, S. (2022). The use of virtual reality in training paramedics for a mass casualty incident. *Appl. Sci.* 12:11740. doi: 10.3390/app122211740
- Longo, F., Nicoletti, L., and Padovano, A. (2017). Smart operators in Industry 4.0: a human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Comput. Indus. Eng.* 113, 144–159. doi: 10.1016/j.cie.2017.09.016
- Mariani, M. G., Vignoli, M., Chiesa, R., Violante, F. S., and Guglielmi, D. (2019). Improving safety through non-technical skills in chemical plants: the validity of a questionnaire for the self-assessment of workers. *Int. J. Environ. Res. Public Health* 16:992. doi: 10.3390/ijerph16060992
- Marquardt, N. (2019). Situation awareness, human error, and organizational learning in sociotechnical systems. *Hum. Fact. Ergon. Manufact. Serv. Indus.* 29, 327–339. doi: 10.1002/hfm.20790
- Merck KGaA (2023). *We are Merck*. Available online at: <https://www.merckgroup.com/en/company.html> (Accessed January 14, 2025).
- Miguel-Alonso, I., Checa, D., Guillen-Sanz, H., and Bustillo, A. (2024). Evaluation of the novelty effect in immersive Virtual Reality learning experiences. *Virtual Real.* 28:27. doi: 10.1007/s10055-023-00926-5
- Mütterlein, J. (2018). "The three pillars of virtual reality? Investigating the roles of immersion, presence, and interactivity," in *Proceedings of the Annual Hawaii International Conference on System Sciences (HICSS) 2018-January* (Piscataway, NJ: IEEE), 1407–1415. doi: 10.24251/HICSS.2018.174
- Nasir, A. N. M., Ali, D. F., Noordin, M. K. B., and Nordin, M. S. B. (2011). "Technical skills and non-technical skills: predefinition concept," in *Proceedings of the IETEC'11 Conference, Kuala Lumpur, Malaysia*, 1–17.
- Nilsson, N. C., Nordahl, R., and Serafin, S. (2016). Immersion revisited: a review of existing definitions of immersion and their relation to different theories of presence. *Human Technol.* 12, 108–134. doi: 10.17011/ht/urn.201611174652
- Pfeiffer, S., and Suphan, A. (2015). *The Labouring Capacity Index: Living Labouring Capacity and Experience as Resources on the Road to Industry 4.0 [Der Labouring Capacity Index: Arbeitsvermögen und Erfahrung als Ressourcen auf dem Weg zu Industrie 4.0]*. Available online at: <https://www.iaf.eu/aktuell/veroeff/2015/pfeiffer15.pdf> (Accessed January 30, 2016).
- Plass, J. L., Homer, B. D., and Kinzer, C. K. (2015). Foundations of game-based learning. *Educ. Psychol.* 50, 258–283. doi: 10.1080/00461520.2015.1122533
- Radiant, J., Majchrzak, T. A., Fromm, J., and Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: design elements, lessons learned, and research agenda. *Comput. Educ.* 147:103778. doi: 10.1016/j.compedu.2019.103778
- Reason, J. (1990). *Human Error*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9781139062367
- Ribeiro, J., Almeida, J. E., Rossetti, R. J., Coelho, A., and Coelho, A. L. (2012). "Using serious games to train evacuation behaviour," in *7th Iberian Conference on Information Systems and Technologies (CISTI 2012)* (Piscataway, NJ: IEEE), 1–6.
- Ritzmann, S., Hagemann, V., and Kluge, A. (2014). The training evaluation inventory (TEI) – evaluation of training design and measurement of training outcomes for predicting training success. *Vocat. Learn.* 7, 41–73. doi: 10.1007/s12186-013-9106-4
- Ritzmann, S., Hagemann, V., and Kluge, A. (2020). *TEI. [Training Evaluation Inventory] [Verfahrensdokumentation, Inventar deutsch und englisch] [Procedural Documentation, Inventory in German and English]*. Leibniz-Institut für Psychologie (ZPID) [Leibniz-Institute of Psychology (ZPID)] (Hrsg.), Open Test Archive. Trier: ZPID. doi: 10.23668/psycharchives.4493
- Roberts, K. H., and Bea, R. (2001). Must accidents happen? Lessons from high reliability organizations. *Acad. Manage. Executive* 15, 70–79. doi: 10.5465/ame.2001.5229613
- Salas, E., Stagl, K. C., and Burke, C. S. (2008). Does team training improve team performance? A meta-analysis. *Human Factors* 50, 903–933. doi: 10.1518/001872008X375009
- Salas, E., Wilson, K. A., Burke, C. S., and Wightman, D. C. (2006). "Crew resource management training research, practice, and lessons learned," in *Review of Human Factors and Ergonomics*, Vol. 2, ed. R. C. Williges (Santa Monica, CA: Human Factors and Ergonomics Society), 35–73. doi: 10.1177/1557234X0600200103
- Schmidt, K., Winkler, B., and Gruber, B. (2016). *Skills for the future – Zukünftiger Qualifikationsbedarf aufgrund erwarteter Megatrends [Skills for the Future – Future Training Requirements Due to Expected Megatrends]*. Ibw-Forschungsbericht Nr. 187. Wien: IBN.
- Scorgie, D. G., Feng, Z., Paes, D., Parisi, F., Yiu, T. W., and Lovreglio, R. (2024). Virtual reality for safety training: a systematic literature review and meta-analysis. *Saf. Sci.* 171:106372. doi: 10.1016/j.ssci.2023.106372
- Stefan, H., Mortimer, M., and Horan, B. (2023). Evaluating the effectiveness of virtual reality for safety-relevant training: a systematic review. *Virtual Real.* 27, 2839–2869. doi: 10.1007/s10055-023-00843-7
- Sullivan, A., Elshenawy, S., Ades, A., and Sawyer, T. (2019). Acquiring and maintaining technical skills using simulation: initial, maintenance, booster, and refresher training. *Cureus* 11. doi: 10.7759/cureus.5729
- Sutcliffe, K. M. (2011). High reliability organizations (HROs). *Best Prac. Res. Clin. Anaesthesiol.* 25, 133–144. doi: 10.1016/j.bpa.2011.03.001
- Tehreem, Y., Fracaro, S. G., Gallagher, T., Toyoda, R., Bernaerts, K., Glassey, J., et al. (2022). "May I remain seated: a pilot study on the impact of reducing room-scale trainings to seated conditions for long procedural virtual reality trainings" in *Proceedings of the 2022 8th IEEE International Conference on Virtual Reality (ICVR), Lisbon, 5–7 July 2022* (Piscataway, NJ: IEEE), 62–71. doi: 10.1109/ICVR55215.2022.9848222
- Tenbergh, R., and Pittich, D. (2017). Ausbildung 4.0 oder nur 1.2? Analyse eines technisch-betrieblichen Wandels und dessen Implikationen für die technische Berufsausbildung [Training 4.0 or just 1.2? Analysis of a technical-operational change and its implications for technical vocational education and training]. *J. Tech. Educ.* 5, 27–46.
- Triberti, S., Sapone, C., and Riva, G. (2025). Being there but where? Sense of presence theory for virtual reality applications. *Human. Soc. Sci. Commun.* 12:79. doi: 10.1057/s41599-025-04380-3
- van Gog, T., and Sweller, J. (2015). Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. *Educ. Psychol. Rev.* 27, 247–264. doi: 10.1007/s10648-015-9310-x
- Weick, K. E., and Sutcliffe, K. M. (2001). *Managing the Unexpected: Assuring High Performance in an Age of Complexity*, Vol. 9. San Francisco, CA: Jossey-Bass.
- Weyer, J. (1997). Die Risiken der Automationsarbeit. Mensch-Maschine-Interaktion und Störfallmanagement in hochautomatisierten Verkehrsflugzeugen [The risks of automated work. Human-machine interaction and incident management in highly automated commercial aircraft]. *Zeitschrift Soziol.* 26, 239–257. doi: 10.1515/zfsoz-1997-0401
- Woollard, M., Whitfield, R., Newcombe, R. G., Colquhoun, M., Vetter, N., and Chamberlain, D. (2006). Optimal refresher training intervals for AED and CPR skills: a randomised controlled trial. *Resuscitation* 71, 237–247. doi: 10.1016/j.resuscitation.2006.04.005
- Wouters, P., van Nimwegen, C., van Oostendorp, H., and van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *J. Educ. Psychol.* 105, 249–265. doi: 10.1037/a0031311
- Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer* 38, 25–32. doi: 10.1109/MC.2005.297