

Article

HMD-VR Training for Confined-Space Operations: Evaluating Performance and Cognitive Demand in Real-World Execution

Francesca Massa ¹, Vincenzo Ronca ², Sara Buonocore ¹, Domenico Coccorese ¹, Stefano Moscatelli ¹, Luciano Di Donato ³, Giuseppe Di Gironimo ¹, Eduardo Palermo ⁴, Pietro Aricò ², Andrea Tarallo ¹, Simone Ercoli ², Luca Mattioli ⁴, Francesco Castelli Gattinara Di Zubiena ^{3,4} and Raffaele De Amicis ^{5,*}

- ¹ Department of Industrial Engineering, University of Naples Federico II, P.le Tecchio 80, 80125 Naples, Italy; francesca.massa@unina.it (F.M.); sara.buonocore2@unina.it (S.B.); domenico.coccorese@unina.it (D.C.); stefano.moscatelli.1@gmail.com (S.M.); giuseppe.digironimo@unina.it (G.D.G.); andrea.tarallo@unina.it (A.T.)
- ² Department of Computer, Control, and Management Engineering, Sapienza University of Rome, 00185 Rome, Italy; vincenzo.ronca@uniroma1.it (V.R.); pietro.arico@uniroma1.it (P.A.); ercoli.1919156@studenti.uniroma1.it (S.E.)
- ³ Department of Technological Innovations and Safety of Plants, Products and Anthropic Settlements, Istituto Nazionale per l'Assicurazione Contro gli Infortuni sul Lavoro (INAIL), 00144 Rome, Italy; l.didonato@inail.it (L.D.D.); f.castelligattinaradizubiena@inail.it (F.C.G.D.Z.)
- ⁴ Department of Mechanical and Aerospace Engineering, "Sapienza" University of Rome, 00185 Rome, Italy; eduardo.palermo@uniroma1.it (E.P.); luca.mattioli@uniroma1.it (L.M.)
- ⁵ School of Electrical Engineering and Computer Science, Oregon State University, Corvallis, OR 97331, USA
- * Correspondence: raffaele.deamicis@oregonstate.edu

Abstract

Training for confined-space operations presents significant challenges due to safety risks, limited accessibility, and the complexity of reproducing realistic conditions. Within this context, immersive technologies such as Head-Mounted Display Virtual Reality (HMD-VR) offer a promising alternative for procedural training. This study evaluates the effectiveness of a hands-free, bimanual HMD-VR training system for confined-space operations as an alternative to physical training when access to real environments is constrained. Two HMD-VR configurations were investigated and compared with a physical training condition using a shared multi-step procedural scenario. Training outcomes were assessed across repeated executions and through a real-world evaluation, focusing on procedural performance and cognitive demand. Results show that all groups achieved significant improvements in performance over repeated executions, with reductions in completion time and error rate. Importantly, no statistically significant differences were observed between groups during the final real-world evaluation, and non-inferiority analyses confirmed that HMD-VR training did not result in performance outcomes worse than those achieved through physical training within predefined operational margins. Similarly, no significant differences emerged in either EEG-derived Cognitive Load (CL) or subjective workload at the final execution, despite a reduction in perceived workload over time. Analysis of subjective immersive experience revealed consistently high levels of presence and engagement across HMD-VR conditions, with no significant differences between configurations. However, spatial presence and perceived realism decreased over repeated exposure, suggesting an adaptation effect rather than a continuous increase in immersion. Overall, the findings indicate that, under the controlled conditions of this study, HMD-VR training enables real-world procedural execution and cognitive demand profiles comparable to those observed following physical training. These results support the potential of HMD-VR as a viable approach for novice procedural training in confined-space operations, particularly in contexts where traditional training is limited by safety, cost, or accessibility constraints.



Academic Editor: Zhonghua Sun

Received: 31 May 2026

Revised: 3 July 2026

Accepted: 6 July 2026

Published: 13 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: virtual reality; HMD-VR training; confined-space operations; procedural preparation; experimental evaluation; cognitive load; presence; performance evaluation; EEG

1. Introduction

Training for confined-space operations is a critical component of workforce preparation in high-risk and spatially constrained industrial contexts, where workers are required to execute complex procedures under severe spatial, environmental, and safety constraints [1]. Restricted access, limited ventilation, reduced visibility, and the potential presence of hazardous atmospheric conditions substantially increase the operational risk associated with these activities [2,3]. Effective preparation, therefore, requires training approaches capable of reproducing realistic operational conditions while ensuring trainee safety and minimizing exposure to real-world hazards [4]. Conventional training approaches, including physical mock-ups and on-site drills, address these requirements only partially. Although such methods can reproduce selected aspects of the operational context, their cost, logistical complexity, limited repeatability, and residual safety risks constrain systematic training and hinder controlled experimental evaluation [5]. As a result, alternative solutions that enable realistic yet safe preparation for confined-space operations are increasingly needed.

From a training perspective, confined spaces closely align well with the DICE framework [6], as they involve activities that are Dangerous, Infrequent, Complex, and Expensive to reproduce in real-world settings. These characteristics make confined-space operations particularly challenging to prepare for using physical training modalities, reinforcing the need for training environments that can simulate complex procedures while maintaining experimental control and operational safety.

Within this context, HMD-VR has emerged as a promising technological solution for training in confined spaces [7–9]. By enabling controlled exposure to hazardous scenarios within a *safe-to-fail* environment, HMD-VR allows operators to rehearse complex procedures representative of real confined-space activities without physical risk, while supporting repeatable and standardized training experiences [10,11]. Despite growing evidence of the potential benefits of immersive technologies for training [12–14], empirical evaluation of HMD-VR training systems under standardized and operationally realistic conditions remains limited, particularly for procedurally complex confined-space scenarios [15].

In this study, we contribute to the research domain of HMD-VR training for confined-space operations by evaluating the effectiveness of a hands-free HMD-VR training system as an alternative approach in contexts where access to physical training environments is constrained. More specifically, the study provides three main scientific contributions.

First, from an engineering perspective, we introduce a hands-free HMD-VR training system that enables natural bi-manual interaction through inside-out tracking. The system supports the execution of a complete multi-step procedural workflow through direct interaction with virtual objects under constraints representative of confined-space operations. Interaction is further enriched through multimodal feedback, integrating visual and auditory cues with tactile feedback provided by haptic gloves and kinesthetic feedback delivered via a wearable exoskeleton.

Second, from a methodological perspective, we propose an evaluation framework that compares two HMD-VR training configurations (with and without multisensory feedback) against real-world physical training using a shared procedural scenario. The framework is grounded in an experimental protocol consisting of two training sessions followed by a final real-world evaluation performed by all groups under identical conditions. This design enables a controlled comparison of training outcomes across modalities and supports

the assessment of training transfer from virtual to real-world execution for the HMD-VR conditions. Training outcomes are evaluated in terms of performance and cognitive demand during execution. In this context, performance is defined as the effectiveness of real-world execution, operationalized through completion time and error rate, and complemented by subjective self-assessments.

Third, from an empirical perspective, we provide evidence on the effectiveness of HMD-VR training through a non-inferiority analysis of performance outcomes, showing that HMD-VR training can achieve results comparable to physical training under shared evaluation conditions.

Together, these contributions advance the understanding of how HMD-VR training systems can be effectively designed and evaluated in high-risk, spatially constrained operational contexts.

In light of this, the present work addresses the following research questions:

RQ1—Performance. *To what extent does HMD-VR training for confined-space operations result in real-world execution outcomes, in terms of final performance measures, relative to those obtained through physical training under identical evaluation conditions?*

RQ2—Cognitive Demands. *To what extent does HMD-VR training for confined-space operations result in cognitive demands during real-world execution relative to those observed following physical training under identical evaluation conditions?*

The remainder of the paper is organized as follows. Section 2 reviews the related literature, identifies the key research gaps, and introduces the contribution and significance of the present study. Sections 3 and 4 describe the design of the HMD-VR training system and the experimental methodology. The experimental results are presented in Section 5 and discussed in Section 6. Finally, Section 7 concludes the paper and outlines directions for future work.

2. Literature Review

Over the past five decades, immersive technologies, particularly HMD-VR systems, have been increasingly explored as tools for training across a wide range of application domains [14,16–20]. Such technologies have been adopted in several contexts, including aviation [21,22], construction [23–25], military [26] and helicopter operations [27], railway emergency management [28], firefighting [29], and medical training [30,31], where realistic and controlled operational environments are essential. These studies have evaluated HMD-VR training systems across a broad range of dimensions, including the EEG-based metric of Cognitive Load (CL), user performance, user experience, interaction quality, motion sickness, presence, usability, immersion, and educational outcomes, highlighting the potential of HMD-VR as an effective tool for skill acquisition and task-oriented training in complex operational environments.

However, research on evaluating HMD-VR for training in confined-space operations has only begun to emerge over the past decade. Recent studies, including those by Abbas et al. [14], Evangelista et al. [8], Dang et al. [24], and Lu et al. [32], typically adopt a comparative approach in which HMD-VR training is evaluated against traditional educational methods (e.g., lectures). While these studies demonstrate the advantages of HMD-VR, such as improved skill retention, enhanced hazard recognition, faster reaction times, and reduced human error, their validation remains focused on controlled comparisons rather than on real-world execution. Additionally, Shen et al. [33] investigated HMD-VR systems for training characterized by different levels of fidelity and interaction modalities, primarily based on controller-mediated input, showing that higher interaction fidelity and direct manipulation improve learning outcomes. However, the reliance on controller-based

interaction represents a limitation, as it differs from natural hand-based manipulation. This issue is particularly relevant in confined-space operations, where coordinated bi-manual actions are essential for performing real-world tasks such as inspection, maintenance, and system manipulation [34]. In such contexts, training systems should therefore support hands-free, bi-manual interaction that more closely reflects real operational behavior. A structured comparison of the most relevant studies is provided in Table 1, highlighting key differences in VR type, interaction modality, evaluation approach, transfer assessment, and outcome measures. The comparison emphasizes that existing studies predominantly rely on controller-based interaction and indirect or absent evaluation of real-world transfer, while the present study introduces hands-free bi-manual interaction and a direct comparison with physical training under shared real-world evaluation conditions.

Table 1. Systematic Comparison of HMD-VR training studies in confined-space operations.

Study	VR	Interaction	Transfer	Sample	Metrics	CL
Abbas et al. [14]	Mixed (VR/AR)	Not specified	Not directly assessed	Multiple studies	Performance, Safety, Engagement	Not analyzed
Evangelista et al. [8]	HMD-VR	Controller	No	Mixed	HRV, Knowledge, performance	Yes
Dang et al. [24]	HMD-VR	Controller	No	Workers	Knowledge, performance	No
Lu et al. [32]	HMD-VR	Controller	No		Knowledge, performance, retention	No
Shen et al. [33]	HMD-VR	Controller	Partial	Not specified	Behavior, knowledge, self-efficacy	No
This study	HMD-VR	Hands-free bi-manual	Yes	Non-experts	Performance, Cognitive Load	Yes

Moreover, increasing attention has been directed toward the evaluation of HMD-VR training outcomes beyond subjective self-reports alone. In the context of confined-space operations, where operators are exposed to physical strain, mental stress, and elevated CL, integrating objective and subjective measures is essential to obtain a comprehensive assessment of both the cognitive and physiological states of the user [35]. Both measures have been adopted by Evangelista et al. [36] in their study that investigated cognitive performance in confined spaces by combining subjective workload assessment with physiological measurements, using a modified task load questionnaire alongside heart rate variability analysis.

In light of this, and as summarized in Table 1, the reviewed literature highlights that, despite the demonstrated effectiveness of HMD-VR technologies for training purposes, significant gaps remain in their application to confined-space operations. In particular, as shown by the studies discussed above, the evaluation of HMD-VR training is predominantly based on controlled comparisons with traditional instructional methods, rather than on the assessment of training effectiveness through real-world execution. Furthermore, interaction paradigms are still largely based on controller-mediated input, with limited support for natural, hands-free bi-manual interaction, which is essential for accurately reproducing actual procedures in confined spaces. At the same time, the lack of integrated multimodal evaluation frameworks, combining objective and subjective measures to assess the same human factors, limits the ability to obtain a comprehensive understanding of users' cognitive and physiological states. Taken together, these limitations indicate that existing solutions do not yet fully capture the complexity of real-world operational performance, highlighting the need for evaluation approaches specifically designed to assess training effectiveness for confined-space operations under realistic conditions.

To the best of our knowledge, no hands-free HMD-VR training system for confined-space operations has been empirically evaluated through direct comparison with real-world physical training using a shared procedural scenario. In response to this gap, our work introduces a hands-free HMD-VR training system designed for the execution of complete, multi-step procedural workflows representative of real-world confined-space operations, and evaluates the extent to which training delivered through two configurations of this system translates to real-world execution in terms of performance and cognitive demand.

3. Design and Development of the HMD-VR Training System for Confined Spaces

The HMD-VR training system was designed to provide a realistic environment for preparation in confined spaces, prioritizing interaction fidelity, procedural coherence, and sustained cognitive engagement under constrained operational conditions. It enables natural, hands-free bi-manual interaction based on inside-out tracking, consistent with the tools and tasks required in real confined-space operations.

The system supports the execution of a complete multi-step workflow (Table A1) through direct interaction with virtual objects, allowing users to perform procedures under realistic constraints. Interaction is further enriched through multimodal feedback, combining visual and auditory cues with tactile inputs provided by haptic gloves (TactGloves DK2) and kinesthetic feedback delivered via a wearable exoskeleton (SIDE).

3.1. Design Objectives and Context

The proposed training system was conceived to support the execution and experimental validation of a complex, multi-step procedure representative of confined-space operations. Rather than reproducing an isolated task, the system targets a complete procedural workflow composed of sequential and interdependent operations performed under spatial, operational, and cognitive constraints.

3.2. Scenario Definition, Validation and Training Environment

The training scenario was identified in the real domain and translated into system-level requirements for the virtual training environment in collaboration with INAIL (Italian National Institute for Insurance against Accidents at Work), which provided domain expertise and validated the reference procedure adopted in the system. The resulting workflow consists of fifteen sequential tasks (Table A1) representing common operations in confined-space contexts, including system isolation, environmental verification, navigation in restricted area, and maintenance activity.

The same validated procedure was implemented in both the HMD-VR and real-world environments, ensuring a high degree of correspondence in terms of tools, control panels, pipelines, access points, spatial layout, and relative dimensions (Figure 1). This alignment enables training and evaluation to be conducted on an identical procedural scenario, with differences limited to the training modality.

The experimental environment incorporates key spatial and operational constraints characteristic of confined spaces, including restricted mobility, narrow passages requiring crouched navigation, and structured work areas. Elements of virtual embodiment, such as a high-visibility safety vest and virtual gloved hands, were introduced to enhance ecological validity by aligning the virtual representation with real-world conditions, in which participants were similarly equipped with corresponding protective garments (e.g., a high-visibility vest).

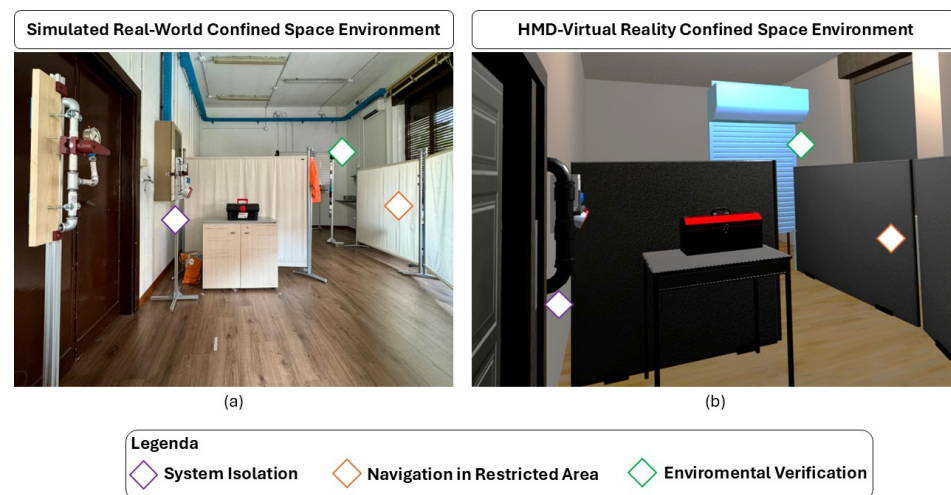


Figure 1. Simulated real-world and virtual confined-space environments used in the study. The left panel (a) shows the physical training environment, while the right panel (b) presents the corresponding HMD-VR environment reproducing the same procedural scenario. Spatial layout, tools, and operational elements (Legenda) were designed to ensure one-to-one correspondence between real and virtual conditions. Colored markers indicate representative task components within the procedural workflow, including system isolation (purple), navigation in restricted area (yellow), and environmental verification (green).

Video S1 illustrates the complete confined-space procedure implemented in the HMD-VR training system. The workflow begins with the safe isolation of the system through electrical power interruption and LockOut/TagOut (LOTO) procedures, followed by the closure and securing of steam valves to prevent accidental reactivation. The user then performs a series of safety verification activities, including pressure monitoring, venting operations, and environmental gas measurements to ensure safe working conditions. Once the system has been isolated and verified, the procedure progresses to the operational area, where additional safety checks are conducted before maintenance activities begin. The maintenance phase requires coordinated bi-manual interactions, including the manipulation of valves, the removal of a damaged flanged joint, and the installation of a replacement component using hand tools. Throughout the procedure, users interact with equipment, safety devices, and tools using natural hands-free manipulation, supported by visual, auditory, vibrotactile, and kinesthetic feedback depending on the experimental condition. The workflow concludes with the successful installation of the replacement component and confirmation that the system is ready for commissioning tests.

3.3. System Design Rationale

The design of the HMD-VR training system was guided by a set of expected capabilities aimed at addressing the specific constraints and demands of confined-space operations. In particular, it supports three core capabilities: (i) continuous navigation within spatially constrained environments; (ii) reliable identification, selection, grasping, and manipulation of virtual objects, including tool-mediated operations; and (iii) robust hands-free, hand-tracked interaction enabling coordinated bi-manual actions throughout the workflow. These design choices were motivated by the need to reproduce key characteristics of real-world confined-space tasks, where limited space, restricted visibility, and the use of tools demand precise coordination and reliable interaction. Particular attention was therefore devoted to ensuring interaction robustness and minimizing discrepancies between virtual affordances and their real-world counterparts. To address these considerations, the system was implemented as a wireless, standalone HMD-VR solution, enabling safe operation

within predefined boundaries and reducing environmental constraints. This configuration also allowed training sessions to be conducted without external connectivity, ensuring experimental consistency.

3.4. Complex Tool-Mediated Interaction

A defining characteristic of the proposed HMD-VR training system is its support for hands-free, tool-mediated interaction and dynamic object-object dependencies. Users interact directly with virtual objects through natural bi-manual hand movements, without the use of controllers, while the state of the environment evolves dynamically according to user actions and procedural progression.

Procedural execution frequently requires coordinated manipulation of tools and operational components. In particular, users can grasp and manipulate virtual tools in order to interact with other objects within the environment, similarly to real-world operational procedures. For example, a user may grasp a wrench and use it to loosen the bolts of a flanged pipe joint. In this case, the interaction outcome depends not only on the user's hand movements, but also on the specific tool being used and on the current procedural state of the system.

As a result, virtual objects do not behave as isolated entities, but interact dynamically through user-driven manipulation. Object states, tool affordances, and procedural dependencies are continuously updated during task execution, enabling realistic multi-step workflows and supporting interaction fidelity under operationally constrained conditions.

This interaction model goes beyond simplified task abstractions by enabling actions that closely resemble real-world industrial operations, where users must coordinate tools, object manipulation, and sequential task dependencies throughout the workflow.

3.5. Design Rationale

The overall design of the HMD-VR training system reflects a deliberate balance between realism, usability, and experimental control. High interaction fidelity and procedural completeness were prioritized to reduce the gap between virtual and real execution, while robustness and usability constraints were introduced to prevent excessive cognitive or physical workload.

By grounding the system design in a procedure validated with INAIL and representative of common confined-space operations, the proposed approach establishes an experimental protocol for the evaluation of HMD-VR training systems for confined spaces.

4. Materials and Methods

4.1. Participants

Thirty-nine adult volunteers participated in the study. The experimental protocol was approved by the University Transdisciplinary Ethics Committee in accordance with the Declaration of Helsinki, and all participants provided written informed consent before participation. Participants were recruited through public announcements within the university and volunteered on an unpaid basis. The sample included master's students, doctoral candidates, researchers, and faculty members, with ages ranging from 20 to 35 years and an approximately balanced gender distribution. All participants reported normal or corrected-to-normal vision and no history of neurological or psychiatric disorders. None of the participants had prior professional experience or formal training related to industrial confined-space operations. Additionally, approximately 90% of the sample reported only rare interactions with immersive technologies.

A statistical summary of participant characteristics, including age, gender, VR experience, and manual task familiarity (measured on a 1–5 Likert scale), is reported in Table 2.

Table 2. Participant characteristics by experimental group.

Variable	G1	G2	G3
N	15	14	10
Age (years, mean $\pm$ SD)	26.7 $\pm$ 3.7	27.9 $\pm$ 3.1	28.7 $\pm$ 2.8
Age range	20–33	21–32	25–35
Gender (Male/Female)	7/8	8/6	10/0
VR experience (1–5, median)	1	1	1
Manual task confidence (1–5, median)	3	3	3

Participants were anonymized through alphanumeric identifiers and assigned to one of three experimental groups. A combined screening and random allocation procedure was adopted. To ensure comparability, gender distribution was balanced in G1 and G2. Eligibility for G3 was determined prior to allocation based on anthropometric compatibility with the SIDE upper-limb exoskeleton (Figure 2; [37]), resulting in a male-only group, which is acknowledged as a limitation of the study. No a priori power analysis was conducted to determine the required sample size. The target number of participants per group was defined based on practical feasibility and available resources.

- The *Physical Training Group* (G1, $n = 15$) performed two training sessions in the real-world environment.
- The *HMD-VR Training Group with visual feedback* (G2, $n = 14$) performed two training sessions using the HMD-VR system with visual feedback.
- The *HMD-VR Training Group with multimodal feedback* (G3, $n = 10$) performed two training sessions using the HMD-VR system with visual, auditory cues, tactile feedback provided by haptic gloves and kinesthetic feedback delivered via a wearable exoskeleton.

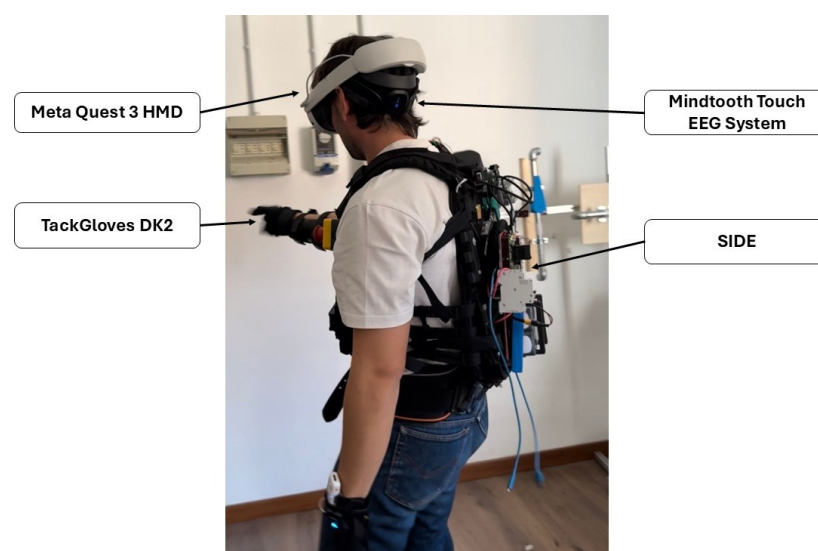


Figure 2. Experimental configuration for the multisensory HMD-VR condition (G3). A participant is equipped with a Meta Quest 3 HMD for immersive visualization, the Mindtooth Touch EEG system for CL data acquisition, TackGloves DK2 for tactile interaction, and the SIDE upper-limb exoskeleton providing force-related feedback.

4.2. Study Design

The study (Figure 3) follows a design comprising an experiment phase and a subsequent data analysis phase. Participants were first recruited and completed a pre-test procedure, including informed consent, and were then equipped according to the assigned experimental group. All participants executed the same multi-step workflow across three consecutive runs. The protocol included two training sessions (run 1 and run 2), followed by a final real-world evaluation (run 3). During the training phase, participants repeated the procedure either in the physical environment (G1) or in HMD-VR (G2 and G3).

The study adopts a combined within and between-subject design, enabling both the analysis of learning dynamics across repeated executions and the comparison of training modalities across experimental groups. Within this framework, the design is intended to investigate a *training transfer mechanism under controlled conditions*. Specifically, the combination of repeated exposure to an identical procedural workflow, high interaction fidelity enabled by hands-free bi-manual manipulation, and strict procedural equivalence between virtual and physical environments creates comparable action and perception conditions across training contexts. This alignment is expected to support the development of motor and cognitive patterns that can be applied consistently during real-world execution, thereby enabling the empirical evaluation of training transfer.

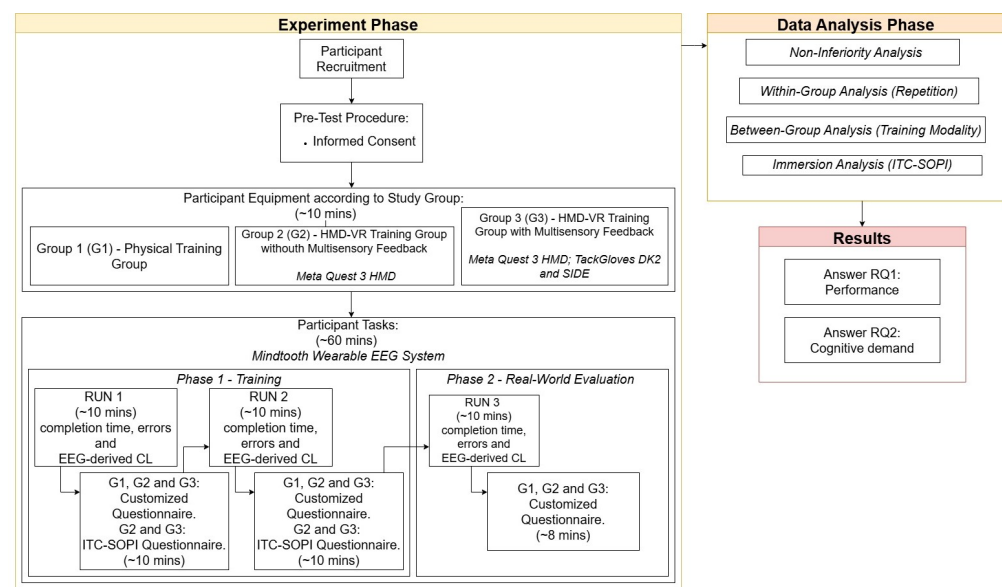


Figure 3. Study design and evaluation method illustrating the experiment phase (participant recruitment, pre-test, test, post-test), followed by data analysis and results addressing the research questions.

From a cognitive perspective, the design also allows the investigation of potential adaptation processes across repeated executions by jointly analyzing performance, cognitive load, and subjective measures.

The final execution was conducted exclusively in a real-world simulated confined-space environment for all groups in order to provide an evaluation context under identical physical conditions. This design allows for a direct comparison of training effectiveness across modalities by assessing procedural execution in a shared real-world setting, thereby isolating the effect of training modality from differences in execution environment.

Throughout the experiment, performance metrics (completion time, number and type of errors) and neurophysiological activity were recorded during each run. In particular, CL was continuously measured using the Mindtooth Touch EEG system and subsequently analyzed to derive CL indicators. Subjective measures were collected via a post-run questionnaire (Table 3). In addition, the ITC-Sense Of Presence Inventory (ITC-SOPI)

questionnaire [38] was administered to participants in the HMD-VR conditions (G2 and G3) after the training sessions to assess the immersive experience.

The collected data were analyzed using both within-group and between-group comparisons to evaluate changes across runs and differences between training modalities.

Within this study design, the two research questions are addressed through distinct sets of outcome measures. RQ1 is informed by objective and subjective performance indicators, including completion time, error rate, and self-reported performance, capturing procedural efficiency and accuracy during real-world execution. RQ2 is addressed through measures of cognitive demand, combining EEG-derived CL indicators with subjective cognitive workload assessments, providing both physiological and perceived perspectives on cognitive effort. In addition, subjective immersive experience, assessed through ITC-SOPI, is analyzed separately and is not directly associated with either research question. Instead, it provides complementary insights into participants' perceived presence and engagement during HMD-VR training, supporting the interpretation of performance and cognitive outcomes.

Accordingly, the study adopts an evaluative perspective, aiming to determine whether HMD-VR training enables procedural execution and cognitive demand profiles that align with those achieved through physical training under real-world conditions.

Table 3. Customized questionnaire administered after each run, including a single-item Likert-scale measure of perceived performance and selected items from the NASA-TLX assessing different dimensions of CL.

#	Question	Likert Answer
Self-Performance Assessment		
1	How do you evaluate your performance?	1 (Very Low)– 5 (Very Fast)
NASA-Task Load Index		
2	How effortful was it to understand what to do in the simulation?	1 (Very Low Effort)–20 (Very High Effort)
3	How effortful was it to move and assume the necessary postures when interacting with objects?	1 (Very Low Effort)–20 (Very High Effort)
4	To what extent did you feel under pressure because of the time available to complete the activity?	1 (Very Low Effort)–20 (Very High Effort)
5	How satisfied are you with your performance during the training activity?	1 (Very Low Effort)–20 (Very High Effort)
6	How much overall effort did the training activity require?	1 (Very Low Effort)–20 (Very High Effort)
7	How much did the training activity discourage/stress/irritate/frustrate you?	1 (Very Low Effort)–20 (Very High Effort)

4.3. Hardware and Software

The HMD-VR training system was delivered using a Meta Quest 3 HMD [39] operated in standalone mode and without hand controllers, enabling natural, bi-manual hands-free interaction through inside-out tracking. The application was developed in Unreal Engine 5.3.2 [40], with Blueprint Visual Scripting (Unreal Engine 5.3.2) used to implement event-driven logic and procedural workflow execution. Native hand-tracking functionality was supported via OpenXR (version 62.0) [41], HandTracking plugin 0.6.3 beta [42] in combination with the Meta XR Plugin [43], while Meta Quest Casting (v74, Meta Platforms, Inc., Menlo Park, CA, USA) [44] was used to monitor the virtual environment in real time during experimental sessions. In one experimental condition, a multisensory feedback configuration was implemented using the SIDE upper-limb exoskeleton (prototype developed within the INAIL BRIC ID 37 project, Department of Mechanical and Aerospace Engineering, Sapienza University of Rome, Rome, Italy) [37] in combination with TactGlove DK2 (bHaptics Inc., Seoul, Republic of Korea) haptic gloves [45]. Neurophysiological activity

was recorded throughout all experimental runs using the Mindtooth Touch EEG system (BrainSigns, Rome, Italy; BrainProducts, Gilching, Germany) [46], equipped with eight passive sponge-based electrodes positioned at AFz, AF3, AF7, AF4, AF8, Pz, P3, and P4, with mastoid reference and ground electrodes. Real-world training and evaluation were conducted within a real-world simulated confined-space environment that replicated the same procedural workflow and operational elements implemented in the virtual environment.

4.4. Experimental Protocol

The experimental protocol was identical across all experimental groups and designed to ensure consistency, replicability, and comparability independently of the training modality. All participants executed the same scenario following a session structure.

Each session began with an initial briefing, during which participants received an introduction to the study and provided informed consent. Participants were then equipped with the Mindtooth Touch EEG system. For the HMD-VR conditions, participants in G2 were provided with a Meta Quest 3 HMD, while participants in G3 were additionally equipped with the SIDE and TactGloves DK2. Electrode placement was performed according to the device guidelines [46], ensuring stable contact between sensors and the scalp. Signal quality was verified in real time using the Mindtooth Neurometrics App, which provides indicators of signal stability and noise levels. System functionality, including data acquisition and synchronization, was checked before the experiment to ensure correct operation. Participants were asked to report any physical discomfort or instability of the device, and adjustments, when requested, were made as needed. Finally, a short EEG calibration procedure was conducted, consisting of 60 s with eyes open and 60 s with eyes closed, to establish baseline neurophysiological activity before the experimental tasks.

Participants executed the corresponding experimental protocol three consecutive times. The first two executions (run 1 and run 2) served as training sessions according to the assigned training modality, whereas the third execution (run 3) constituted the final evaluation and was performed by all participants in the real-world environment. During each execution, CL was monitored, completion time was measured, and both the number and type of errors were recorded.

After each execution, participants completed the customized questionnaire (Table 3). In addition, participants wearing the HMD-VR display (G2 and G3) completed the ITC-SOPI at the end of run 1 and run 2 to assess the immersive experience during the training sessions. At the end of the session, all equipment was removed, and participants underwent a brief debriefing to confirm well-being and the absence of adverse effects.

Overall, the protocol was designed to reproduce the operational constraints of confined-space activities while enabling controlled and comparable observation of performance and cognitive demands across conditions.

4.5. Evaluation Measures

Performance outcomes and cognitive demand were evaluated using objective and subjective measures. Objective performance was quantified through completion time and errors, while cognitive demand was assessed using EEG-derived CL. Subjective measures included a custom questionnaire for perceived performance and cognitive workload, and the ITC-SOPI for immersive experience.

Objective performance outcomes were recorded during each run by measuring completion time and errors. Errors were categorized by type and mapped to specific steps of the workflow to identify tasks more frequently associated with incorrect execution, with coding performed by a single trained observer present during all sessions who systematically recorded errors for each participant in each run. To support this analysis, a qualitative error

taxonomy was defined, comprising three main categories: (i) *confusion errors*, involving incorrect identification or misuse of objects required to complete a task; (ii) *omission errors*, corresponding to missed actions or failure to maintain required resources (e.g., forgetting to carry the toolbox across task locations); and (iii) *task execution errors*, related to difficulties in performing specific procedural actions, such as valve manipulation or the assembly and disassembly of components requiring coordinated bimanual interaction.

The taxonomy was primarily used to identify and interpret recurring error patterns throughout the procedural workflow. Error categories were descriptively analyzed and reported to support qualitative interpretation of procedural difficulties, rather than being statistically compared as independent outcome variables.

Subjective measures were collected at the end of each run using a customized questionnaire (Table 3), including both perceived performance and subjective workload. Perceived performance was used as a complementary indicator to objective measures, providing insight into participants' self-assessment of their execution rather than serving as a direct proxy for actual performance.

For participants in the HMD-VR conditions (G2 and G3), the ITC-SOPI was administered at the end of run 1 and run 2 to evaluate immersive experience during virtual training.

Neurophysiological activity was recorded continuously during each run using the Mindtooth Touch EEG system while participants executed the full procedural workflow. EEG signals were processed following the methodology described in [47], from which CL indicators were derived. More specifically, the EEG data were preprocessed offline using a pipeline optimized for wearable EEG recordings in ecological operational settings [48]. Signals were band-pass filtered in the 2–30 Hz range, ocular artifacts were corrected using a regression-based procedure, and residual movement- or muscle-related transients were excluded using amplitude- and probability-based artifact rejection criteria [49]. After pre-processing, EEG activity was analyzed in individualized frequency bands defined around each participant's individual alpha frequency. Cognitive load indicators were derived from Global Field Power measures computed over frontal and parietal electrode clusters, following previously validated neurometric formulations. In particular, EEG-derived cognitive load was estimated from the ratio between frontal theta activity and parietal alpha activity, reflecting increased executive control and reduced cortical idling during more demanding task execution. Values were normalized within participants to account for inter-individual differences in absolute EEG amplitude and spectral distribution.

4.6. Data Analysis

Statistical analyses were conducted to evaluate both performance outcomes and cognitive demands across repeated executions and between training conditions.

To address RQ1 from a comparative perspective, final performance outcomes (completion time and error rate) were first analyzed within a non-inferiority framework. Non-inferiority margins were defined *a priori* based on operational relevance and expert judgment from domain specialists involved in the design and validation of the training scenario. For completion time, the margin was specified as a maximum acceptable deviation of 15% from the expected duration of the physical procedure. Given a procedure duration of approximately 12 min, this corresponded to a threshold of about 1.8 min. This deviation was considered operationally acceptable because it would not affect task completion, procedural sequencing, or safety outcomes. For procedural accuracy, the margin was set at one additional error, corresponding to less than 7% of the total workflow steps. This threshold was considered the maximum acceptable deviation while maintaining procedural validity. The definition of these margins was independent of the observed data and completed prior to the analysis, ensuring that the non-inferiority assessment does not rely on post

hoc adjustments. Non-inferiority was assessed by estimating 95% bootstrap confidence intervals of the differences between G2-G1 and G3-G1, using 1000 bootstrap resampling iterations. Non-inferiority was concluded when the upper bound of the confidence interval remained below the predefined margin. Confidence intervals were computed using a bootstrap resampling procedure applied to pairwise differences between groups. As a result, the intervals reflect variability in the difference distribution and may appear relatively wide or asymmetric, particularly in the presence of small sample sizes and high within-group variability.

Within-group effects of repetition (run 1, run 2, run 3) were analyzed separately for each group (G1, G2, G3) using Friedman tests. This non-parametric test was selected as it is appropriate for repeated measures designs involving dependent samples and does not assume normality of the data. Where significant effects of repetition were observed ($p < 0.05$), post-hoc pairwise comparisons were performed using the Durbin–Conover procedure with appropriate correction for multiple comparisons, in order to identify differences between specific runs.

Between-group differences were assessed using Kruskal–Wallis tests. This non-parametric test was chosen as it allows comparison across independent groups without assuming normal data distribution. These analyses were conducted at each run (run 1, run 2, and run 3) to evaluate differences between training modalities under each experimental condition. When significant between-group effects were detected ($p < 0.05$), post-hoc pairwise comparisons were conducted using the Dwass–Steel–Critchlow–Fligner procedure. This method was selected as it is specifically designed for non-parametric multiple comparisons following Kruskal–Wallis tests and is robust to unequal sample sizes.

For both Friedman and Kruskal–Wallis tests, the χ^2 statistic represents the extent to which ranked observations differ across conditions. Larger χ^2 values indicate stronger deviations from the null hypothesis, reflecting greater differences across runs or between groups, whereas smaller values indicate more similar distributions. In this study, the null hypothesis for the Friedman test assumes no differences across repeated executions, while for the Kruskal–Wallis test it assumes no differences between training groups. Significant χ^2 values ($p < 0.05$) in the Friedman tests indicate changes across runs, reflecting learning effects, whereas non-significant values indicate stability of the measured variable. Similarly, for the Kruskal–Wallis tests, significant χ^2 values ($p < 0.05$) indicate differences between groups, while non-significant values indicate that no statistically detectable differences were found between training modalities.

To complement hypothesis testing and provide a clearer interpretation of the results, effect size measures were computed to quantify the magnitude of observed differences. Kendall's W was used for Friedman tests, and epsilon-squared (ϵ^2) was used for Kruskal–Wallis tests, providing an estimate of the practical relevance of statistically significant findings.

Subjective immersive experience, assessed through ITC-SOPI, was analyzed for the HMD-VR conditions (G2 and G3) using non-parametric methods. Between-group comparisons at each run were performed using Mann–Whitney U tests, while within-group changes across runs were analyzed using Wilcoxon signed-rank tests.

In addition, errors were examined through a qualitative approach based on the defined error taxonomy. Errors were systematically recorded and categorized, and the analysis focused on identifying recurring error types and their distribution across the different steps of the procedural workflow.

Descriptive statistics are reported as medians and interquartile ranges (IQR), in line with the non-parametric statistical approach and to provide a robust representation of central tendency and variability.

All statistical analyses were performed using Jamovi (version 2.6.44). The significance level was set at $\alpha = 0.05$ for all inferential tests.

5. Results

5.1. Non-Inferiority Analysis

Non-inferiority analyses were conducted on final performance outcomes measured during run 3, focusing on completion time and error rate (Table 4). Descriptive statistics showed similar median values across groups for both measures, with overlapping interquartile ranges, indicating comparable performance distributions.

For completion time, the differences between HMD-VR conditions and physical training were assessed using bootstrap-based confidence intervals. The estimated 95% confidence interval for the G2-G1 comparison ranged from -118.9 to 101.9 s, while the G3-G1 comparison ranged from -121.8 to -82.6 s. In both cases, the intervals remained entirely below the predefined non-inferiority margin ($+15\%$), indicating that execution times achieved after HMD-VR training were not inferior to those obtained through physical training. Although the confidence interval for the G2-G1 comparison is relatively wide, reflecting variability associated with the sample size, it remains entirely within the predefined non-inferiority margin. From an operational perspective, this margin corresponds to a deviation that domain experts considered acceptable without compromising task safety or procedural effectiveness. The width of the interval should therefore be interpreted as reflecting statistical uncertainty due to data variability and limited sample size, rather than as evidence of reduced training effectiveness.

Table 4. Descriptive statistics and non-inferiority analysis for completion time and error rate measured during Run 3.

Outcome	Group	Median	IQR	Comp.	95% CI	Conclusion
<i>Completion Time (s)</i>						
	G1	721	216	-	-	Reference
	G2	725	166	G2-G1	$[-118.9, 101.9]$	Non-inferior
	G3	706	175	G3-G1	$[-121.8, -82.6]$	Non-inferior
NI margin					$+15\%$	
<i>Procedural Error Rate</i>						
	G1	0	1.0	-	-	Reference
	G2	0	0.8	G2-G1	$[-0.47, 0.75]$	Non-inferior
	G3	1	0.8	G3-G1	$[-0.17, 0.80]$	Non-inferior
NI margin					$+1$ error	

A similar pattern was observed for error rate. Median values were low across all groups, and the 95% confidence intervals for the G2-G1 and G3-G1 comparisons ranged from -0.47 to 0.75 errors and from -0.17 to 0.80 errors, respectively. Both intervals remained below the predefined margin of one additional error, indicating that HMD-VR training did not lead to an increase in procedural errors during real-world execution.

Overall, these findings support the non-inferiority of HMD-VR training relative to physical training for both efficiency and accuracy outcomes during real-world task execution.

5.2. Effects of Repetition

Within-group analyses revealed a consistent effect of repetition across all measures (Table 5), reflecting progressive learning and adaptation throughout the experimental sessions. Large effect sizes were observed for completion time, error rate, and subjective

performance across all groups (Kendall's $W > 0.6$) in most cases, indicating substantial and practically meaningful improvements over repeated executions. Moderate effect sizes were observed for NASA-TLX, while EEG-derived CL showed negligible effect sizes ($W < 0.05$), suggesting stable neurophysiological engagement across runs.

Table 5. Summary of statistical analyses for within-group (Friedman test) and between-group (Kruskal–Wallis test) comparisons across performance and cognitive measures. For within-group analyses, results are reported separately for each group (G1, G2, G3), while between-group comparisons are presented for each run (run 1, run 2, run 3). Test statistics are expressed as χ^2 values with corresponding p-values. Effect sizes are reported as Kendall's W for Friedman tests and epsilon-squared (ϵ^2) for Kruskal–Wallis tests.

Measure	Condition	G1	G2	G3
Within-group comparisons (Friedman test)				
Completion Time	χ^2	23.3	19.9	18.2
	p -value	<0.001	<0.001	<0.001
	Kendall's W	0.78	0.71	0.91
Error Rate	χ^2	18.9	20.1	14.2
	p -value	<0.001	<0.001	<0.001
	Kendall's W	0.63	0.72	0.71
NASA-TLX	χ^2	8.37	10.9	10.1
	p -value	0.015	0.004	0.006
	Kendall's W	0.28	0.39	0.51
EEG CL	χ^2	1.20	1.00	0.60
	p -value	0.549	0.607	0.741
	Kendall's W	0.04	0.04	0.03
Subjective Performance	χ^2	26.1	20.9	7.13
	p -value	<0.001	<0.001	0.028
	Kendall's W	0.87	0.75	0.36
Between-group comparisons (Kruskal–Wallis test)				
Measure	Condition	Run 1	Run 2	Run 3
Completion Time	χ^2	3.20	2.66	0.003
	p -value	0.072	0.265	0.998
	ϵ^2	0.03	0.02	0.00
Error Rate	χ^2	1.39	7.75	2.33
	p -value	0.499	0.021	0.312
	ϵ^2	0.00	0.16	0.01
NASA-TLX	χ^2	7.05	8.36	0.645
	p -value	0.029	0.015	0.724
	ϵ^2	0.14	0.18	0.00
EEG CL	χ^2	1.67	0.673	1.13
	p -value	0.435	0.714	0.569
	ϵ^2	0.00	0.00	0.00
Subjective Performance	χ^2	0.476	0.633	0.476
	p -value	0.788	0.729	0.788
	ϵ^2	0.00	0.00	0.00

Completion time significantly decreased in all groups (Figure 4a; Table 5), indicating improved procedural efficiency. While G1 exhibited a rapid improvement followed by a plateau between run 2 and run 3, G2 and G3 showed a more gradual but continuous reduction across runs. Post-hoc analyses (Durbin–Conover) (Table 6) revealed that reductions in completion time were primarily driven by significant differences between Run 1 and both Run 2 and Run 3 ($p < 0.001$ across groups), while no significant differences were observed between Run 2 and Run 3 in most cases, indicating stabilization of performance.

A similar trend was observed for error rate, which significantly decreased in all groups (Figure 4b; Table 5), indicating improved accuracy over repeated executions. Error reductions occurred early in G1 and more progressively in the HMD-VR conditions, suggesting ongoing refinement of procedural execution. Post-hoc comparisons (Table 6) confirmed that reductions in error rate were primarily concentrated between Run 1 and subsequent runs, while differences between Run 2 and Run 3 were generally not significant, indicating stabilization of accuracy after the initial learning phase.

Subjective performance ratings increased significantly across runs (Table 5), reflecting improved self-perception of execution quality. In parallel, perceived CL (NASA-TLX) significantly decreased in all groups (Figure 4c; Table 5), suggesting progressive adaptation and reduced perceived effort. Post-hoc analyses (Table 6) indicated that reductions in perceived workload were mainly driven by significant differences between Run 1 and later runs, although some variability across groups was observed.

In contrast, EEG-derived CL did not show significant changes across runs (Figure 4d; Table 5), indicating stable neurophysiological demand despite improvements in performance and perceived workload.

Table 6. Results of post-hoc pairwise comparisons following significant Friedman tests (within-group) and Kruskal–Wallis tests (between-group) across performance and cognitive measures. Pairwise comparisons for within-group analyses were conducted using the Durbin–Conover procedure and are reported separately for each group (G1, G2, G3) across runs (Run 1–3). For between-group analyses, pairwise comparisons were performed using the Dwass–Steel–Critchlow–Fligner procedure and are reported for each run. Only comparisons associated with statistically significant omnibus tests are included.

Measure	Condition	G1	G2	G3
Within-group comparisons (Friedman test)				
Completion Time	Run 1 vs. Run 2	<0.001	<0.001	<0.001
	Run 1 vs. Run 3	<0.001	<0.001	<0.001
	Run 2 vs. Run 3	0.072	0.026	<0.001
Error Rate	Run 1 vs. Run 2	<0.001	<0.001	0.018
	Run 1 vs. Run 3	<0.001	<0.001	<0.001
	Run 2 vs. Run 3	0.281	0.003	<0.001
NASA-TLX	Run 1 vs. Run 2	0.045	0.405	1
	Run 1 vs. Run 3	0.003	<0.001	0.002
	Run 2 vs. Run 3	0.259	0.006	0.002
Subjective Performance	Run 1 vs. Run 2	<0.001	<0.001	0.361
	Run 1 vs. Run 3	<0.001	<0.001	0.006
	Run 2 vs. Run 3	<0.001	<0.001	0.046
Between-group comparisons (Kruskal–Wallis test)				
Measure	Condition	Run 1	Run 2	Run 3
Error Rate	G1 vs. G2	-	0.401	-
	G1 vs. G3	-	0.015	-
	G2 vs. G3	-	0.287	-
NASA-TLX	G1 vs. G2	0.339	0.131	-
	G1 vs. G3	0.024	0.016	-
	G2 vs. G3	0.369	0.560	-

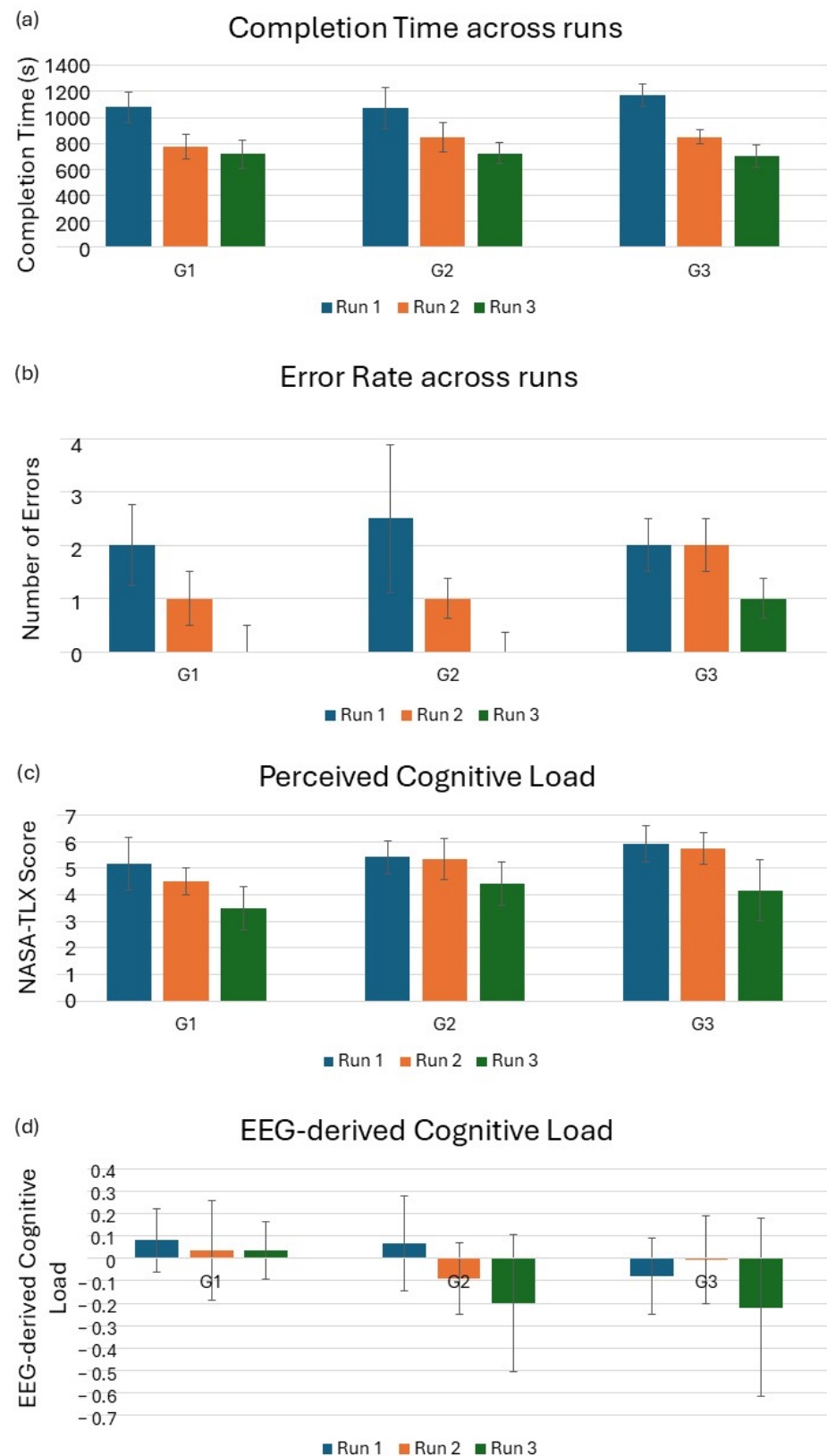


Figure 4. Evolution of performance and cognitive measures across runs for each experimental group. Completion time (a) and error rate (b) progressively decrease, indicating improvements in procedural efficiency and accuracy over repeated executions. Perceived CL (NASA-TLX) (c) decreases across runs, reflecting reduced perceived effort, while EEG-derived CL (d) remains relatively stable, suggesting consistent neurophysiological demand despite performance improvements. Error bars represent the interquartile range (IQR).

In addition, a qualitative analysis of error types was conducted based on the predefined error taxonomy (Table 7). Errors were systematically recorded during execution and categorized into confusion, omission, and task execution errors. The analysis examined their distribution across tasks and training conditions, enabling the identification of recurring error patterns. Confusion errors were predominantly observed in tasks involving lockout procedures, where participants were required to correctly identify and apply different locking mechanisms. Additional confusion emerged in valve management tasks, where participants struggled to identify the correct operational sequence. In the HMD-VR training conditions, approximately 60% of participants exhibited confusion during the initial run. Omission errors were mainly associated with failure to transport essential tools between task locations, requiring participants to interrupt the procedure and correct the omission. Task execution errors were primarily associated with complex manipulative actions, such as pipe disassembly and assembly, often requiring coordinated bi-manual interaction. Overall, approximately 70% of errors occurred during the first run in the HMD-VR conditions, followed by a substantial reduction in subsequent runs, with near-zero error rates in the final executions. This pattern indicates a strong learning effect and progressive procedural mastery.

Table 7. Distribution of error categories by task and experimental group across runs. Values are reported as Confusion/Omission/Task Execution errors (C/O/TE). T1–T15 correspond to the sequential tasks described in Table A1.

Group	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	Total
Run 1																
G1	-	-	7/0/2	2/0/0	3/0/2	-	0/5/0	0/3/0	-	0/3/0	-	-	0/1/3	0/1/2	0/0/4	38
G2	2/0/1	1/0/1	8/0/1	4/0/2	3/0/1	-	1/4/0	0/1/0	-	0/2/0	-	-	1/1/5	0/0/3	0/0/4	46
G3	1/0/0	-	3/0/1	-	4/0/2	-	-	-	-	-	-	-	0/1/4	0/0/2	0/0/5	23
Run 2																
G1	-	-	1/0/1	-	1/0/0	-	-	1/1/0	-	0/1/0	-	-	0/0/1	0/0/1	0/0/1	9
G2	1/0/1	1/0/0	3/0/1	1/0/0	1/0/0	-	0/0/1	-	-	-	-	-	1/0/3	0/0/1	0/0/2	16
G3	2/0/0	-	1/0/1	-	2/0/1	-	-	-	-	-	-	-	0/0/3	0/0/2	0/0/4	16
Run 3																
G1	-	-	1/0/0	-	1/0/0	-	-	0/1/0	-	0/1/0	-	-	0/0/1	0/0/1	0/0/1	7
G2	-	-	1/0/0	-	1/0/0	-	1/0/0	-	-	-	-	-	0/0/2	0/0/1	0/0/2	8
G3	-	-	1/0/0	-	2/0/0	-	-	-	-	-	-	-	0/0/2	0/0/1	0/0/2	8

Descriptive statistics (median and interquartile range, IQR) for all measures are reported in Figure 4, supporting the observed trends.

Overall, across all groups, significant improvements in procedural efficiency and accuracy were observed over repeated executions, as reflected by reductions in run completion time and error rate. Subjective performance increased, while perceived CL decreased, indicating progressive adaptation to the task. In contrast, EEG-derived CL remained stable across runs, suggesting that improvements in performance and perceived workload were not accompanied by measurable changes in neurophysiological demand. In conclusion, the inclusion of effect size measures enables a more robust interpretation of the magnitude of observed effects, providing additional insight beyond statistical significance, particularly in the presence of limited sample size. The post-hoc analyses consistently indicate that learning effects are concentrated between run 1 and run 2, while run 3 lead to performance stabilization rather than further statistically significant improvements.

5.3. Between-Group Comparisons

Between-group analyses were conducted to compare training modalities at each stage of the experiment.

At run 1 (Figure 4), no significant differences were observed between groups in completion time, error rate, EEG-derived CL, or subjective performance, indicating comparable initial conditions. A significant difference was observed only for NASA-TLX, with higher perceived CL in G3 compared to G1. Effect size analysis revealed moderate-to-large effects for NASA-TLX during early runs ($\epsilon^2 = 0.14\text{--}0.18$) (Table 5), indicating meaningful differences in perceived workload across training conditions.

At run 2 (Figure 4), no differences were observed in completion time, EEG-derived CL, or subjective performance. However, significant differences emerged in error rate, with higher error rates in G3 compared to G1 ($p = 0.015$), and in perceived CL, again higher in G3. The effect size analysis revealed a moderate effect size ($\epsilon^2 = 0.16$) (Table 5). Post-hoc comparisons (Dwass–Steel–Critchlow–Fligner) (Table 6) indicated that this difference was primarily driven by higher error rates in G3 compared to G1, while no significant differences were observed between other group pairs.

At run 3 (Figure 4), corresponding to the real-world evaluation, no statistically significant differences were observed between groups for any measure, including completion time, error rate, EEG-derived CL, NASA-TLX, or subjective performance. This result is further supported by negligible effect sizes across all measures ($\epsilon^2 = 0$) (Table 5), indicating that differences observed during training did not persist and confirming convergence in both performance and cognitive demand across training modalities.

Median values were highly comparable across groups, with overlapping interquartile ranges, indicating strong convergence in procedural performance and cognitive demand regardless of training modality (Figure 4).

Overall, between-group analyses revealed no differences at run 1 for most measures, ensuring comparability, and no differences at run 3, indicating convergence towards similar performance and cognitive outcomes across training modalities. Temporary differences observed at run 2, particularly for error rate and perceived CL, did not persist in the final execution. In conclusion, post-hoc analyses indicate that between-group differences are limited to run 1 and run 2 and do not persist in run 3, confirming convergence across training conditions.

5.4. Subjective Immersive Experience (ITC-SOPI)

Immersive experience was evaluated for the HMD-VR conditions (G2 and G3) across the two training sessions using ITC-SOPI; values are reported in Table 8. No statistically significant differences were observed between groups for any dimension at either session, including spatial presence, engagement, realism, and negative effects. Within-group analyses revealed limited but selective changes across runs. In particular, a significant decrease was observed in spatial presence for G3 and in realism for G2, while no significant changes were found for the remaining dimensions. Across runs, spatial presence and realism generally showed a slight decrease, whereas engagement remained consistently high and negative effects remained low.

Table 8. ITC-SOPI results across runs for HMD-VR conditions (medians and statistical comparisons).

Dimension	Group	Run 1	Run 2	Within p	Between p (R1)	Between p (R2)
Spatial Presence	G2	5.57	5.29	0.152	0.977	0.725
	G3	5.64	4.93	0.042		

Table 8. Cont.

Dimension	Group	Run 1	Run 2	Within p	Between p (R1)	Between p (R2)
Engagement	G2	5.43	5.64	0.209	0.518	0.792
	G3	5.36	5.29	0.193		
Realism	G2	4.79	4.14	0.041	1.00	0.639
	G3	4.64	4.00	0.221		
Negative Effects	G2	3.00	2.57	0.306	0.557	0.725
	G3	2.79	3.00	0.919		

6. Discussion

This study investigated the effects of repetition and training modality on performance and cognitive demand in confined-space operations, with a particular focus on learning dynamics across training sessions and final real-world execution.

Accordingly, the study aims to evaluate whether HMD-VR training enables users to achieve performance and cognitive demand profiles that align with those obtained through physical training when assessed in a shared real-world environment.

6.1. RQ1: Performance and Learning Dynamics

RQ1 examined whether hands-free, bi-manual HMD-VR training results in real-world performance comparable to that achieved through physical training. Overall, the results indicate that HMD-VR training can support operationally comparable performance outcomes under the conditions investigated, without observable degradation in real-world execution.

Across all groups, significant improvements in procedural efficiency and accuracy were observed over repeated executions, as reflected by reductions in completion time and error rate (Figure 4a,b). These findings confirm that repetition supports skill acquisition regardless of training modality. This trend is further supported by the qualitative analysis of error types. The majority of errors occurred during the initial training phase (Table 7) and were primarily associated with confusion, omission, and task execution difficulties. The substantial reduction of these errors across subsequent runs suggests that repeated exposure not only improves overall performance but also enables participants to progressively resolve specific sources of misunderstanding and difficulty, leading to more accurate and consistent procedural execution. However, the analysis of learning dynamics (Figure 4a,b) highlights differences in the rate and progression of learning. In G1, performance improved rapidly between run 1 and run 2 and subsequently stabilized, suggesting early attainment of proficiency. In contrast, G2 and G3 exhibited a more gradual and continuous improvement across runs, indicating a more progressive learning trajectory, particularly in terms of procedural accuracy. These patterns suggest that physical training promotes faster initial performance gains, whereas HMD-VR training supports a more sustained and progressive learning process. Post-hoc analyses (Table 6) support this interpretation, showing that improvements occur primarily between run 1 and run 2, while the absence of significant differences between run 2 and run 3 indicates stabilization of performance. Despite these differences in learning dynamics, all groups reached similar levels of performance during the final real-world evaluation. This convergence is quantitatively supported by between-group effect sizes at run 3 ($\epsilon^2 = 0$), indicating that performance differences across training modalities disappear at the final stage. Temporary variations observed during the training phase, such as higher error rates in G3 at run 2, did not persist, indicating that early performance differences do not impact final procedural outcomes.

It is important to note that the present study was conducted on a non-expert population and within a controlled approximation of real confined-space conditions. As such, the

results primarily reflect learning dynamics under controlled experimental settings. Trained operators may exhibit different baseline skills, procedural strategies, and levels of familiarity with immersive technologies, which could influence training effectiveness. In addition, while key spatial and procedural constraints were reproduced, including restricted mobility, structured workflow, and physical interaction demands, other factors typical of real industrial environments, such as reduced visibility, thermal stress, and time pressure, were not explicitly simulated. These aspects may influence both performance and cognitive demand in real-world scenarios and should therefore be considered when interpreting the transferability of the present findings. However, these findings are particularly relevant in the context of confined-space operations, characterized by their Dangerous, Infrequent, Complex, and Expensive (DICE) nature, where access to realistic and repeatable training environments is inherently limited. In such settings, the primary objective of training is to provide safe, accessible, and controlled alternatives rather than to outperform traditional methods. Within this perspective, the observed non-inferiority of HMD-VR training relative to physical training represents a relevant result, as it indicates that HMD-VR can achieve performance outcomes within operationally acceptable limits without introducing additional risk. Despite the potential simplification of certain environmental aspects, the ability of HMD-VR to deliver controlled, repeatable, and risk-free training scenarios makes it a valuable tool for improving safety and preparedness in high-risk operations, even considering possible variability in user acceptance and interaction. This finding is reinforced by the observed convergence in performance and cognitive demand across groups during the final real-world evaluation (Figure 4), suggesting that HMD-VR may provide a preparatory training approach capable of achieving outcomes comparable to physical training under controlled conditions.

In addition, subjective performance ratings increased consistently across runs and aligned with objective improvements. This convergence indicates that participants were aware of their performance gains, reinforcing the validity of the observed learning effects.

Taken together, these findings provide evidence consistent with the interpretation of a training transfer mechanism underlying the observed results. The non-inferiority of HMD-VR training appears to be associated with the high correspondence between virtual and real task execution. In particular, the combination of hands-free bi-manual interaction, procedural equivalence, and repeated exposure is likely to facilitate the development of motor and cognitive patterns consistent with those required in the physical environment. This alignment provides a plausible explanation for the convergence in performance observed across conditions. The consistency of this pattern across multiple measures further suggests that the identified mechanism reflects a general training process rather than a sample-specific effect.

6.2. RQ2: Cognitive Demand and Adaptation

RQ2 examined whether HMD-VR training is associated with differences in cognitive demand during real-world execution. No significant between-group differences were observed at the final evaluation for either subjective workload (NASA-TLX) (Figure 4c) or EEG-derived CL (Figure 4d), indicating comparable cognitive demands across training modalities. This result is further supported by effect sizes, confirming that training modality does not influence neurophysiological demand during final task execution. These findings suggest that HMD-VR training does not appear to impose additional cognitive demands compared to physical training under the conditions investigated, supporting its suitability as a training approach that does not negatively affect cognitive resource allocation during real-world execution.

Analysis over time revealed a clear pattern of adaptation. Subjective workload decreased significantly across runs in all groups, indicating that participants experienced the procedure as progressively less demanding with repeated exposure and increasing familiarity.

In this perspective, stable EEG-derived CL should not be interpreted as the absence of neurocognitive adaptation, but rather as the absence of detectable changes in this specific EEG-derived index when averaged across the full procedural execution. One possible interpretation is that participants become more efficient in their use of cognitive resources. As procedural knowledge increases, task execution may require less conscious effort and reduced uncertainty, leading to lower perceived workload. At the same time, stable EEG-derived CL may reflect the sustained engagement of underlying cognitive processes necessary to correctly perform the procedure.

This pattern is consistent with the transition from effortful to more efficient processing, where improvements in performance are driven by better organization and automation of task-related actions rather than by a reduction in intrinsic task complexity. In this perspective, cognitive demand remains stable in terms of overall resource activation, while perceived effort decreases due to increased familiarity and procedural competence.

Differences observed during the training phase, particularly the higher perceived workload in G3 during the initial runs, were not present during the final evaluation. This indicates that the additional cognitive demand associated with the multisensory HMD-VR condition is primarily limited to early exposure and is progressively reduced through experience and task familiarization.

Overall, no evidence of additional cognitive burden was observed during the final real-world evaluation in participants trained with HMD-VR compared with those trained physically.

6.3. Subjective Experience and Immersion

In addition to performance and cognitive measures, subjective immersive experience was assessed through ITC-SOPI (Figure 5).

No significant differences were observed between G2 and G3 across ITC-SOPI dimension suggesting comparable perceived immersion across HMD-VR configurations. This result should be interpreted with caution, as the limited and non-diverse sample, due to hardware constraints, may have reduced the statistical power of the comparison. In this context, the multisensory configuration (G3) should be considered exploratory, aimed at investigating the potential contribution of additional haptic and kinesthetic feedback, which did not appear to substantially influence subjective experience under the conditions investigated. Importantly, the absence of significant differences should not be interpreted as evidence that the additional feedback was imperceptible. Participants in G3 were able to perceive tactile and kinesthetic cues during interaction; however, these effects remained localized to the interaction experience and were not reflected in measurable differences in overall performance, cognitive demand, or immersive experience under the conditions investigated. Observed changes across repeated exposure suggest an adaptation effect, where familiarity with the virtual environment may reduce perceived novelty and slightly affect dimensions such as presence and realism, without impacting overall engagement. While the G3 condition provides initial insights into the role of multisensory feedback, its interpretation remains limited by sample characteristics. Nevertheless, the main findings of the study are not dependent on the comparison between HMD-VR configurations, as they are primarily supported by the overall convergence observed across training modalities. Overall, these findings indicate that HMD-VR training provides a stable and well-tolerated

experience, characterized by consistently high engagement and low negative effects, regardless of the presence of additional multisensory feedback.

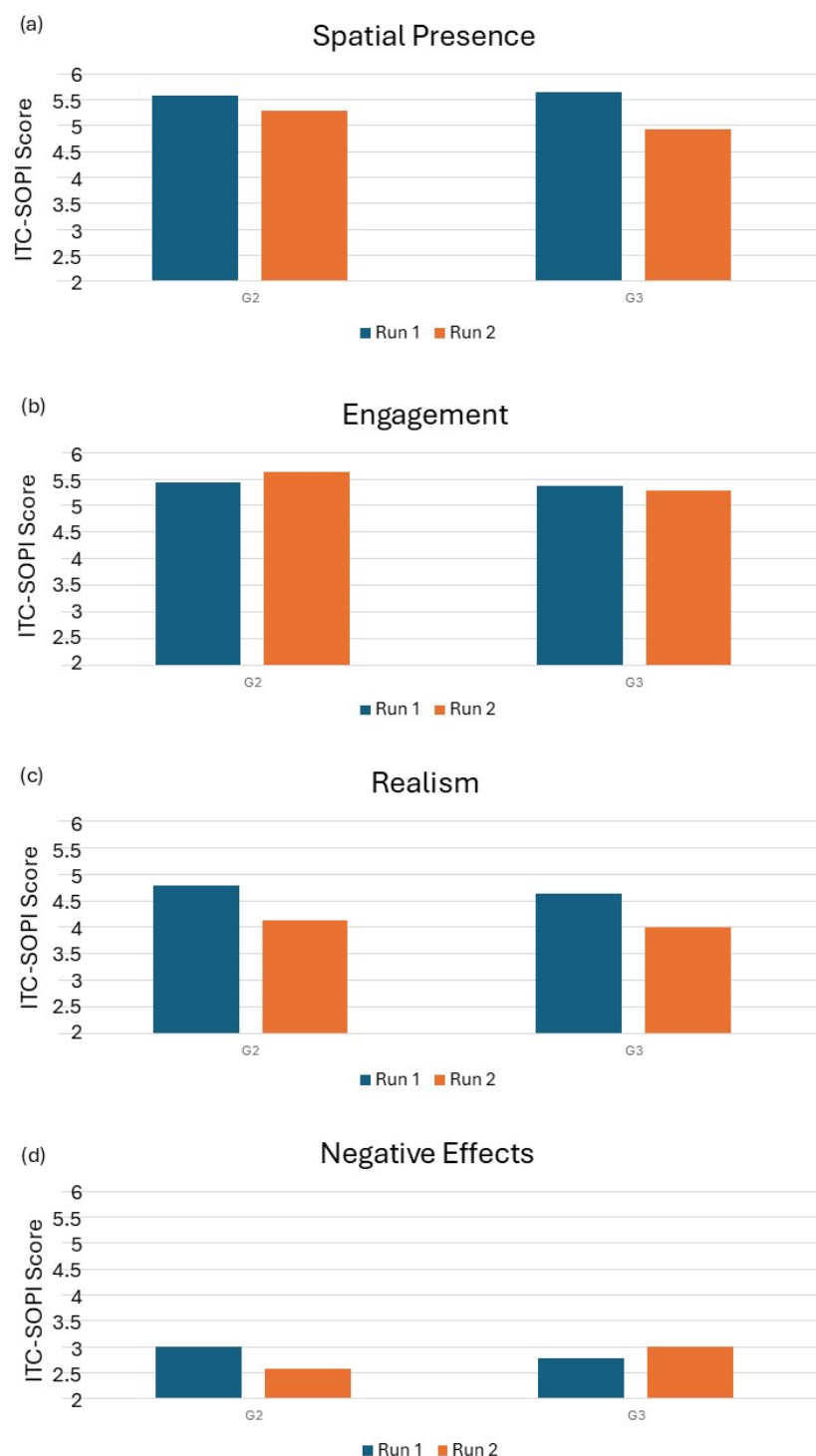


Figure 5. Evolution of ITC-SOPI dimensions across runs for the HMD-VR conditions (G2 and G3). Spatial presence (a) and perceived realism (c) show a decreasing trend between run 1 and run 2, whereas engagement (b) and negative effects (d) remain relatively stable across runs, indicating sustained involvement and consistent user experience over time.

6.4. Integrated Interpretation

Taken together, the results provide a coherent picture of learning, performance, and adaptation (Table 9).

Repeated exposure leads to significant improvements in performance and reductions in perceived workload, reflecting progressive task familiarization. While different training modalities produce distinct learning trajectories, all groups ultimately converge toward comparable levels of performance and cognitive demand.

The combination of stable EEG-derived CL, decreasing subjective workload, and consistent immersive experience suggests that training enhances efficiency primarily at the experiential level, without reducing the overall cognitive demands of the task. While subjective immersion remained comparable across conditions, minor within-group changes suggest a progressive adaptation effect over repeated exposure.

Overall, these findings support the potential of HMD-VR as a controlled tool for novice procedural familiarization in simulated confined-space training scenarios.

Table 9. Integrated interpretation of results across measures and research questions.

Measure	Learning Dynamics	Final Outcome	Implication
Completion Time	Rapid (G1) vs. gradual (G2/G3)	Convergence at Run 3	Different learning paths, same outcome
Error Rate	Progressive reduction	Convergence at Run 3	Procedural accuracy acquired across modalities
NASA-TLX	Decreases across runs	No differences at Run 3	Reduced perceived effort with training
EEG CL	Stable across runs	No differences	Cognitive demand maintained
ITC-SOPI	High and stable with selective decreases	No differences (G2 vs. G3)	Immersion stable across conditions, with minor adaptation effects over time

7. Conclusions and Future Work

This study evaluated the effectiveness of a hands-free HMD-VR training system for procedural preparation in confined-space operations under constrained training conditions.

An experimental protocol compared two HMD-VR configurations with real-world physical training using a shared multi-step scenario and a real-world evaluation.

Results showed significant performance improvements across repeated executions, with reduced completion time and error rate in all groups. No between-group differences were observed during the final evaluation, and non-inferiority analyses indicated that HMD-VR training did not result in poorer performance compared to physical training within predefined operational margins. Similarly, no differences emerged in cognitive demand, although perceived workload decreased over time.

Taken together, these findings suggest that, under the controlled conditions investigated, HMD-VR training enables real-world execution and cognitive demand profiles comparable to those obtained through physical training. These results support its potential as a safe and accessible training approach for confined-space operations, particularly in DICE scenarios where real-world training opportunities are limited. However, these conclusions should be interpreted within the scope of the study, which involved a controlled setup and a non-expert population, and therefore primarily reflect novice procedural familiarization.

No differences were observed between HMD-VR configurations, suggesting that hands-free immersive interaction may be sufficient to support procedural training under the conditions investigated. Nevertheless, this interpretation remains preliminary due to the exploratory nature of the multisensory condition and the limited sample size, and further research is required to clarify the role of additional sensory feedback.

Some limitations should be acknowledged. The use of a non-expert and relatively small sample, as well as group imbalance and hardware constraints, may limit generalizability. No a priori power analysis was conducted, and the sample size was determined based

on practical feasibility, equipment availability, and recruitment constraints. Consequently, the study may have been underpowered to detect small or moderate between-group effects, particularly in the G3 condition. The experimental environment, while controlled, does not fully reproduce highly constrained confined spaces, and error coding was performed by a single observer, which may introduce potential subjectivity in error classification. In addition, cognitive load was analyzed at the run level, potentially masking phase-specific variations, and no delayed assessment was conducted, limiting conclusions about long-term retention. Future work should extend the study to larger and more diverse populations, including trained operators, and investigate long-term retention, phase-specific cognitive dynamics, and training transfer across different procedural scenarios. Further research is also needed to explore the role of multisensory feedback and adaptive training strategies in more ecologically valid conditions.

Overall, this study provides empirical evidence supporting the potential of HMD-VR as a viable approach for novice training and procedural familiarization in confined-space operations under controlled conditions, while highlighting the need for further validation with professional operators in real industrial contexts.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/app16147005/s1>, Video S1: The complete confined-space procedure implemented in the HMD-VR training system.

Author Contributions: Conceptualization, L.D.D., G.D.G., E.P., P.A., A.T., L.M., F.C.G.D.Z. and R.D.A.; Methodology, F.M., R.D.A. and G.D.G.; Software, D.C., S.M., R.D.A. and G.D.G.; Validation, F.M., V.R., S.B., G.D.G., E.P., P.A., S.E., L.M., F.C.G.D.Z. and R.D.A.; Formal Analysis, F.M., R.D.A., V.R. and P.A.; Investigation, F.M., V.R., S.B., L.D.D., G.D.G., E.P., P.A., A.T., S.E., L.M., F.C.G.D.Z. and R.D.A.; Resources, L.D.D., G.D.G., E.P. and P.A.; Data Curation, V.R., P.A. and S.E.; Writing—Original Draft preparation, F.M. and R.D.A.; writing—review and editing, F.M., R.D.A. and G.D.G.; Visualization, F.M. and R.D.A.; Supervision, R.D.A. and G.D.G.; Project Administration, G.D.G. and S.B.; funding acquisition, L.D.D., G.D.G., E.P., P.A. and A.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institute for Insurance against Accidents at Work, through the “GURU” project, within the program BRIC, grant number CUP B53C22008970005.

Institutional Review Board Statement: The study was conducted in accordance with the local legislation and institutional requirements. The experimental protocol design was approved by the Sapienza University Transdisciplinary Ethics Committee (ID: 191/2024, approved on 15 May 2024).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The original contributions presented in the study are included in the article/Supplementary Materials, further inquiries can be directed to the corresponding author.

Acknowledgments: This work was conducted within the GURU PROJECT promoted by INAIL within the program BRIC, grant number CUP B53C22008970005. We would like to thank all colleagues working at Department of Technological Innovations and Safety of Plants, Products and Anthropic Settlements, Istituto Nazionale per l’Assicurazione contro gli Infortuni sul Lavoro (INAIL), who contribute every day to ergonomics and safety of workers in Italy. We sincerely appreciate their professionalism, technical contribution, and constant availability during meetings and discussions aimed at improving the application. Among all, we are profoundly grateful to Massimo Spagnuolo, Marco Pirozzi, Lorian Ricciardi, whose guidance and continuous support represented the key enabling element in the collaboration with INAIL. Finally, we would like to thank all the people who participated in the laboratory experiments.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. HMD-VR User Scenario

Table A1. Procedural workflow tasks implemented in the HMD-VR training system.

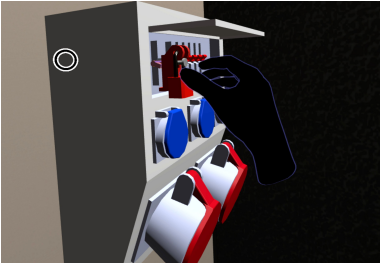
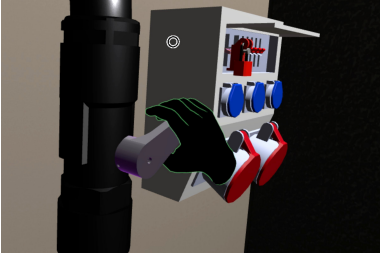
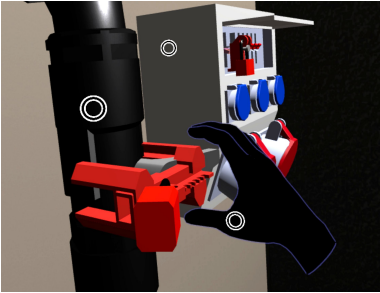
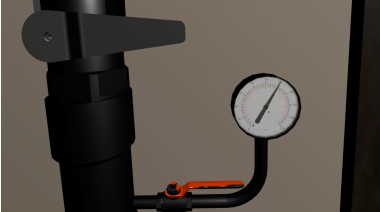
Task	Figure	Title and Description
T1		Electrical Power Interruption : This task requires interruption of the electrical power supply through interaction with the main electrical control panel. The user opens the protective panel cover and engages the main circuit breaker. Correct interaction is confirmed through visual highlighting, vibrotactile feedback, and auditory signals. Successful task completion is indicated by the switch transitioning to the off position and by activation of the corresponding indicator light
T2		Toolbox Access: The user approaches the workbench and opens the toolbox by interacting with its central metal latch. The contents of the toolbox become visible and accessible, enabling retrieval of safety devices and hand tools required for subsequent tasks.
T3		Application of Electrical LockOut Device: The user retrieves a LockOut device from the toolbox and applies it to the electrical circuit breaker to mechanically prevent reactivation. Successful placement is confirmed via visual, auditory, and vibrotactile feedback. This task must be completed before any downstream operations are enabled.
T4		Upstream Steam Valve Closure: The user closes the upstream steam valve located near the electrical control panel. The task requires grasping and rotating the valve by 90 degrees to interrupt steam flow. During the operation, the system provides resistive force feedback and visual cues to simulate real mechanical resistance.
T5		LockOut/TagOut (LOTO) of Upstream Valve: A second LockOut/TagOut device is retrieved from the toolbox and applied to the upstream steam valve to prevent accidental rotation. Task completion is confirmed via multimodal feedback and is mandatory before proceeding.
T6		Visual Verification of Pressure Verification: The user visually inspects the pressure gauge and verifies that the safety valve is correctly aligned with the gauge conduit and that the pressure reading corresponds to the expected operational value. This task requires perceptual assessment without physical interaction.

Table A1. Cont.

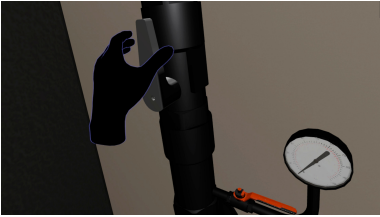
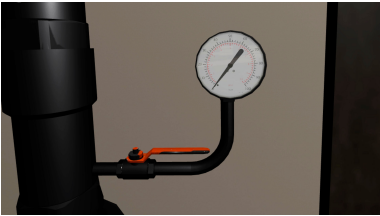
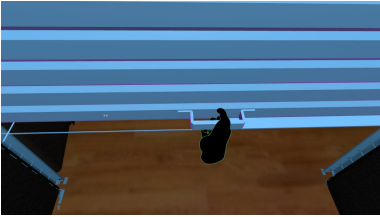
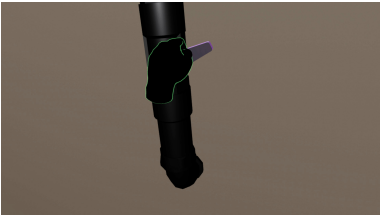
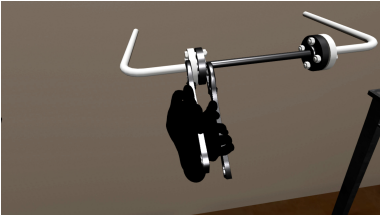
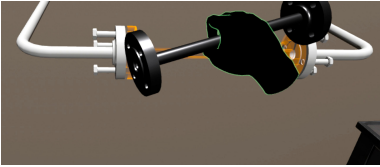
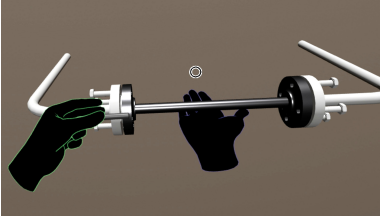
Task	Figure	Title and Description
T7		Vent Valve Opening : The user opens the vent valve located on the main pipeline by rotating it 90 degrees. Visual highlighting, auditory cues, vibrotactile feedback, and resistive force are provided during manipulation to confirm correct execution and simulate physical effort.
T8		Pressure Discharge Verification: The user observes the pressure gauge display and confirms the progressive reduction of pressure until stabilization at 0 bar, indicating successful system depressurization. Upon verification, the toolbox is closed and transported to the next operational area.
T9		Multigas Measurement (Entry Area): The user retrieves the portable multigas detector from the toolbox and activates measurement. Successful acquisition of environmental data is indicated through a clearance message on the display, accompanied by auditory and vibrotactile feedback. The detector is then worn on the user's torso.
T10		Gate Valve Opening: The user opens the vertical gate valve (<i>saracinesca</i>) by grasping and lifting its handle. The task provides force feedback opposing the lifting motion, followed by confirmation of completion via vibrotactile or force-based feedback.
T11		Multigas Measurement (Operational Area): From the entry area, the user performs a second gas measurement by extending the multigas detector into the operational area. Authorization to enter the area is granted only after successful clearance confirmation. The detector is then stored, and the toolbox is repositioned inside the operational space.
T12		Downstream Steam Valve Closure: The user closes the downstream steam valve inside the operational area. Similar to the upstream valve task, operation requires grasping and rotating the valve under simulated mechanical resistance, with multimodal feedback confirming task completion.
T13		Bolt Unfastening Using Wrenches: The user retrieves two open-ended wrenches from the toolbox and performs coordinated bimanual interaction to unfasten bolts on the flanged pipe joint. One wrench stabilizes the bolt head while the second applies rotational torque to loosen the nut. Once the main bolt is removed, remaining bolts are released via procedural animation.

Table A1. Cont.

Task	Figure	Title and Description
T14		Removal of Damaged Flanged Joint: The user grasps and removes the damaged flanged joint from the pipeline and places it into the toolbox. Visual and vibrotactile feedback confirms successful removal.
T15		Installation of Replacement Flanged Joint: The user retrieves and positions a new flanged joint between the pipe ends. Bolts and nuts are inserted and fastened using coordinated bimanual wrench interaction. Correct alignment and tightening are required to complete the task. Upon successful completion, the system displays a final confirmation message indicating that the system is ready for commissioning tests according to applicable regulations.

References

- Vignoli, M.; Nielsen, K.; Guglielmi, D.; Mariani, M.G.; Patras, L.; Peirò, J.M. Design of a safety training package for migrant workers in the construction industry. *Saf. Sci.* **2021**, *136*, 105124. [\[CrossRef\]](#)
- Pupazan, D.; Gaman, A.G.; Ilie, C.; Calamar, A.; Irimia, A.; Gireada, A. Training methods for intervention and rescue personnel in confined spaces depending on their physiological parameter changes. *Calitatea* **2017**, *18*, 87.
- Di Donato, L.; Longo, F.; Ferraro, A.; Pirozzi, M. An advanced solutions for operators' training working in confined and/or pollution suspected space. *Procedia Manuf.* **2020**, *42*, 254–258. [\[CrossRef\]](#)
- Anderson, A.; Stankovic, A.; Cowan, D.; Fellows, A.; Buckey Jr, J. Natural scene virtual reality as a behavioral health countermeasure in isolated, confined, and extreme environments: Three isolated, confined, extreme analog case studies. *Hum. Factors* **2023**, *65*, 1266–1278. [\[PubMed\]](#)
- Lovreglio, R.; Duan, X.; Rahouti, A.; Phipps, R.; Nilsson, D. Comparing the effectiveness of fire extinguisher virtual reality and video training. *Virtual Real.* **2021**, *25*, 133–145.
- Bailenson, J.; Leshner, M. Virtual reality and its opportunities and risks. In *OECD Digital Economy Outlook; Embracing the Technology Frontier*; OECD Publishing: Paris, France, 2024; Volume 1.
- Johnson, C.I.; Bailey, S.K.; Schroeder, B.L.; Marraffino, M.D. Procedural learning in virtual reality: The role of immersion, interactivity, and spatial ability. *Technol. Mind Behav.* **2022**, *3*, 551. [\[CrossRef\]](#)
- Evangelista, A.; Manghisi, V.M.; De Giglio, V.; Mariconte, R.; Giliberti, C.; Uva, A.E. From knowledge to action: Assessing the effectiveness of immersive virtual reality training on safety behaviors in confined spaces using the Kirkpatrick model. *Saf. Sci.* **2025**, *181*, 106693. [\[CrossRef\]](#)
- Ronca, V.; Babiloni, F.; Aricò, P. A novel mutual information-based approach for neurophysiological characterization of sense of presence in virtual reality. *IEEE Trans. Biomed. Eng.* **2025**, *72*, 2313–2320. [\[CrossRef\]](#) [\[PubMed\]](#)
- Pulcini, V.; Modoni, G. Hazardous Workplaces and the Role of VR in Stress Mitigation. In *2025 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*; IEEE: New York, NY, USA, 2025; pp. 229–232.
- Ronca, V.; Ricci, A.; Capotorto, R.; Di Donato, L.; Freda, D.; Pirozzi, M.; Palermo, E.; Mattioli, L.; Di Gironimo, G.; Coccoresse, D.; et al. How immersed are you? State of the art of the neurophysiological characterization of embodiment in mixed reality for out-of-the-lab applications. *Appl. Sci.* **2024**, *14*, 8192. [\[CrossRef\]](#)
- Henstrom, J.; De Amicis, R.; Sanchez, C.A.; Turkan, Y. Immersive engineering instruction: Using Virtual Reality to enhance students' experience in the classroom. *Comput. Graph.* **2024**, *121*, 103944. [\[CrossRef\]](#)
- Rangarajan, V.; Shahbaz Badr, A.; De Amicis, R. Evaluating Virtual Reality in Education: An Analysis of VR through the Instructors' Lens. *Multimodal Technol. Interact.* **2024**, *8*, 72. [\[CrossRef\]](#)
- Abbas Al-Khafaji, Z.K.; Keyhanian, M.; Altalbawy, F.M.; Jawad, M.A.; Saeed, S.M.; Aldulaimi, A.; Albadr, R.J.; Alwan, M.; Smerat, A.; Esmaeili, S.V. Virtual reality and augmented reality technology for safety training in high-risk industries: A literature review on applications, perspectives and challenges. *WORK* **2026**. [\[CrossRef\]](#) [\[PubMed\]](#)
- Osti, F.; De Amicis, R.; Sanchez, C.A.; Tilt, A.B.; Prather, E.; Liverani, A. A VR training system for learning and skills development for construction workers. *Virtual Real.* **2021**, *25*, 523–538. [\[CrossRef\]](#)
- Radhakrishnan, U.; Koumaditis, K.; Chinello, F. A systematic review of immersive virtual reality for industrial skills training. *Behav. Inf. Technol.* **2021**, *40*, 1310–1339. [\[CrossRef\]](#)

17. Bödding, R.; Schriek, S.A.; Maier, G.W. A systematic review and meta-analysis of mixed reality in vocational education and training: Examining behavioral, cognitive, and affective training outcomes and possible moderators. *Virtual Real.* **2025**, *29*, 44. [CrossRef]
18. Jongbloed, J.; Chaker, R.; Lavoue, E. Immersive procedural training in virtual reality: A systematic literature review. *Comput. Educ.* **2024**, *221*, 105124. [CrossRef]
19. Barve, P.; De Amicis, R. A User-Centered Evaluation of a VR HMD-Based Harvester Training Simulator. *Multimodal Technol. Interact.* **2026**, *10*, 15. [CrossRef]
20. Scorgie, D.; Feng, Z.; Paes, D.; Parisi, F.; Yiu, T.; Lovreglio, R. Virtual reality for safety training: A systematic literature review and meta-analysis. *Saf. Sci.* **2024**, *171*, 106372. [CrossRef]
21. Rostami, M.; Kamoopuri, J.; Pradhan, P.; Chung, J. Development and evaluation of an enhanced virtual reality flight simulation tool for airships. *Aerospace* **2023**, *10*, 457. [CrossRef]
22. Xie, D.; Wang, L. Design of a Sports Training System for Confined Spaces Based on Virtual Reality Technology. *J. Cases Inf. Technol. (JCIT)* **2026**, *28*, 1–16. [CrossRef]
23. Rehman, M.S.U.; Abouelkhier, N.; Shafiq, M. Exploring the Effectiveness of Immersive Virtual Reality for Project Scheduling in Construction Education. *Buildings* **2023**, *13*, 1123. [CrossRef]
24. Dang, H.; Serne, J.; Tafazzoli, M. Virtual reality safety training assessment in construction management and safety and health management programs. In *Computing in Civil Engineering 2023*; American Society of Civil Engineers (ASCE): Reston, VA, USA, 2023. pp. 28–35.
25. Adami, P.; Rodrigues, P.B.; Woods, P.J.; Becerik-Gerber, B.; Soibelman, L.; Copur-Gencturk, Y.; Lucas, G. Effectiveness of VR-based training on improving construction workers' knowledge, skills, and safety behavior in robotic teleoperation. *Adv. Eng. Inform.* **2021**, *50*, 101431. [CrossRef]
26. Harris, D.J.; Arthur, T.; Kears, J.; Olonilua, M.; Hassan, E.K.; De Burgh, T.C.; Wilson, M.R.; Vine, S.J. Exploring the role of virtual reality in military decision training. *Front. Virtual Real.* **2023**, *4*, 1165030. [CrossRef]
27. Xue, S.; Hu, L.; Yongliang, T.; Guanghui, W.; Yuan, G. Team effectiveness evaluation and virtual reality scenario mapping model for helicopter emergency rescue. *Chin. J. Aeronaut.* **2020**, *33*, 3306–3317. [CrossRef]
28. Xu, J.; Tang, Z.; Yuan, X.; Nie, Y.; Ma, Z.; Wei, X.; Zhang, J. A VR-based the emergency rescue training system of railway accident. *Entertain. Comput.* **2018**, *27*, 23–31. [CrossRef]
29. Grabowski, A.; Jach, K. The Use of Virtual Reality in the Training of Professionals: With the Example of Firefighters. *Comput. Animat. Virtual Worlds* **2021**, *32*, e1981.
30. Kim, J.S.; Kim, K.W.; Yang, S.W.; Chung, J.W.; Moon, S.Y. Immersive VR (virtual reality) simulator for vein blood sampling. *Technologies* **2023**, *11*, 158. [CrossRef]
31. Diaz, A.L.; Kantar, R.; Podolsky, D.J.; Flores, R.L. Simulation in Cleft Care: Evolution, Evidence, and Training the Future Surgeon. *Plast. Reconstr. Surg.* **2025**, *156*, 56S–71S. [CrossRef] [PubMed]
32. Lu, S.; Feng, Z.; Lovreglio, R.; Wang, F.; Yuan, X. Comparing the productive failure and directive instruction for declarative safety knowledge training using virtual reality. *J. Comput. Assist. Learn.* **2024**, *40*, 1040–1051.
33. Shen, S.; Chen, H.T.; Leong, T.W. Training transfer of bimanual assembly tasks in cost-differentiated virtual reality systems. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*; IEEE: New York, NY, USA, 2019; pp. 1152–1153.
34. Buckingham, G. Hand tracking for immersive virtual reality: Opportunities and challenges. *Front. Virtual Real.* **2021**, *2*, 728461. [CrossRef]
35. Wei, J.; Siegel, E.; Sundaramoorthy, P.; Gomes, A.; Zhang, S.; Vankipuram, M.; Smathers, K.; Ghosh, S.; Horii, H.; Bailenson, J.; et al. Cognitive load inference using physiological markers in virtual reality. In *2025 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*; IEEE: New York, NY, USA, 2025; pp. 759–769.
36. Evangelista, A.; Manghisi, V.M.; De Giglio, V.; Martellotta, F.; Giliberti, C.; Mariconte, R.; Uva, A.E. Exploiting immersive virtual reality for investigating the effects of industrial noise on cognitive performance and perceived workload. In *Proceedings of the International Conference of the Italian Association of Design Methods and Tools for Industrial Engineering*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 311–318.
37. IngegneriaBiomedica.org. SIDE: Wearable Robot for Simulated Dynamics and Haptic Interface. 2024. Available online: <https://www.ingegneriabiomedica.org/news/biotech-support/side-robot-indossabile-per-dinamica-simulata-interfaccia-aptica/> (accessed on 26 May 2026).
38. Lessiter, J.; Freeman, J.; Keogh, E.; Davidoff, J. A cross-media presence questionnaire: The ITC-Sense of Presence Inventory. *Presence Teleoper. Virtual Environ.* **2001**, *10*, 282–297. [CrossRef]
39. Meta. Meta Quest 3. 2024. Available online: <https://www.meta.com/quest/quest-3/> (accessed on 10 October 2023).
40. Epic Games. Unreal Engine. 2024. Available online: <https://www.unrealengine.com/download> (accessed on 26 May 2026).
41. Meta. Meta OpenXR Mobile SDK. 2024. Available online: <https://developers.meta.com/horizon/downloads/package/oculus-openxr-mobile-sdk/> (accessed on 26 May 2026).

42. Unity Technologies. Hand Tracking in Unity XR Hands Package. 2024. Hand Tracking in Unity XR Hands Package. XR Hands Manual. Available online: <https://docs.unity3d.com/Packages/com.unity.xr.hands@1.1/manual/index.html> (accessed on 10 July 2025).
43. Meta. Meta XR Plugin for Unreal Engine. 2024. Available online: <https://developers.meta.com/horizon/documentation/unreal/unreal-quick-start-install-metaxr-plugin/> (accessed on 26 May 2026).
44. Meta. Meta Quest Casting. 2024. Available online: <https://www.oculus.com/casting/> (accessed on 26 May 2026).
45. bHaptics. TactGlove DK2. 2024. Available online: <https://www.bhaptics.com/tactsuit/tactglove-dk2/> (accessed on 26 May 2026).
46. Mindtooth. Mindtooth EEG System. 2024. Available online: <https://www.mindtooth.com/> (accessed on 26 May 2026).
47. Ronca, V.; Capotorto, R.; Buonocore, S.; Massa, F.; Di Gironimo, G.; De Amicis, R.; Papa, S.; Mattioli, L.; Palermo, E.; Di Donato, L.; et al. A novel neurophysiological approach to evaluate the impact of virtual training on skills acquisition. *J. Neural Eng.* **2026**, *23*, 026032. [[CrossRef](#)]
48. Vincenzo, R.; Marianna, C.; Rossella, C.; Gianluca, D.F.; Andrea, G.; Daniele, G.; Gianluca, B.; Fabio, B.; Pietro, A. Beyond the lab: Real-world benchmarking of wearable EEGs for passive brain-computer interfaces. *Brain Inform.* **2026**, *13*, 3.
49. Ronca, V.; Flumeri, G.D.; Giorgi, A.; Vozzi, A.; Capotorto, R.; Germano, D.; Sciaraffa, N.; Borghini, G.; Babiloni, F.; Aricò, P. o-CLEAN: A novel multi-stage algorithm for the ocular artifacts' correction from EEG data in out-of-the-lab applications. *J. Neural Eng.* **2024**, *21*, 056023. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.