Virtual Reality (2026) 30:70
https://doi.org/10.1007/s10055-026-01338-x

ORIGINAL ARTICLE

Towards the target and not beyond: 2D vs 3D visual aids in MR-based neurosurgical simulation

Pasquale Cascarano[1] · Andrea Loretti[1] · Luca Zanuttini[2] · Alessio Di Pasquale[1] · Shirin Hajahmadi[3] · Daniele Giunchi[4] · Riccardo Bovo[5] · Matteo Martinoni[5] · Gustavo Marfia[5]

Received: 26 July 2025 / Accepted: 2 February 2026

Abstract

Neurosurgery increasingly employs Mixed Reality (MR) to support procedures such as External Ventricular Drain (EVD) placement, a task that requires high spatial accuracy, real-time anatomical orientation, and precise trajectory planning. To address these challenges, we present NeuroMix, an MR-based simulator designed to strengthen independent procedural skills without relying on intraoperative visual aids. After an initial pilot study with 10 participants, we conducted a full-scale experiment with 36 medical residents to evaluate three training modalities: CT-only (No Aid), CT with 2D overlay (2D Aid), and CT with 2D overlay plus animated 3D trajectory guidance (2D–3D Aid). All training was performed in a virtual environment using Meta Quest 3, followed by an unaided testing phase in which participants executed EVD placement on a physical phantom using real instruments. An additional control group of 12 residents received no MR training. Our findings show that participants trained with combined 2D–3D visual aids achieved a 44% improvement in targeting precision during unaided testing compared to the control group, significantly outperforming the other training modalities. Crucially, the combination of 2D and 3D visual aids improved retention of spatial knowledge without increasing cognitive workload or reducing usability. These results demonstrate that immersive MR training can effectively foster procedural skill retention in realistic, unaided execution, bridging a critical gap in current surgical education systems.

Keywords Mixed reality · Neurosurgical training · External ventricular drainage · Catherer placement · Visual aids · Skill retention

✉ Pasquale Cascarano
pasquale.cascarano2@unibo.it

Andrea Loretti
a.loretti@unibo.it

Luca Zanuttini
luca.zanuttini@studio.unibo.it

Alessio Di Pasquale
alessio.dipasquale@studio.unibo.it

Shirin Hajahmadi
shirin.hajahmadi2@unibo.it

Daniele Giunchi
d.giunchi@bham.ac.uk

Riccardo Bovo
R.Bovo@greenwich.ac.uk

Matteo Martinoni
m.martinoni@isnb.it

Gustavo Marfia
gustavo.marfia@unibo.it

[1] Department of the Arts, University of Bologna, Via Barberia, 40123 Bologna, Italy

[2] Department of Neurosurgery, IRCCS Institute of Neurological Sciences, Via Altura 3, 40139 Bologna, Italy

[3] Department of Computer Science and Engineering, University of Bologna, via Mura Anteo Zamboni, 40126 Bologna, Italy

[4] Department of Computer Science, University of Birmingham, University Rd W, Edgbaston, Birmingham B15 2TT, UK

[5] Faculty of Engineering and Science, University of Greenwich, Old Royal Naval College, Park Row, London SE10 9LS, UK

Published online: 24 February 2026

1 Introduction

Augmented Reality (AR) and Mixed Reality (MR) technologies are revolutionizing medical education and intraoperative assistance (Giraldo et al. 2025). The application of experiential training for medical students and residents involves the use of immersive and interactive simulators, eliminating the need for cadaver practice and thereby avoiding inherent ethical concerns (James 2020). During surgery, AR/MR systems overlay virtual anatomical information or guidance visual cues onto the real operative field, allowing surgeons to visualize target structures and optimal trajectories in situ (Colombo et al. 2023; Kazemzadeh et al. 2023). These intuitive and real-time visual aids have the potential to enhance intraoperative navigation and reduce the cognitive load of correlating imaging with patient anatomy (Isikay et al. 2024), and to improve surgical accuracy (Lungu et al. 2021).

External Ventricular Drain (EVD) placement (or ventriculostomy) is a critical neurosurgical procedure that stands to benefit from AR/MR technologies. The procedure involves inserting a catheter into the brain's ventricular system for CerebroSpinal Fluid (CSF) drainage, to relieve intracranial pressure (Anatomy 2023). The surgeon identifies the correct entry point on the skull (commonly Kocher's point), performs a burr hole craniotomy, opens the dura mater, and carefully advances the catheter through brain tissue into the lateral ventricle (commonly the Monro's foramen is the target point), ensuring proper depth and trajectory to avoid complications. This procedural task is typically performed freehand at the bedside, using anatomical landmarks to identify the ventricular targets (Anatomy 2023; Flint et al. 2013). However, this blind technique has some limitations, e.g. catheter misplacement occurs in up to 60% of cases (Alizadeh et al. 2024). These misplacements result in inadequate CSF drainage but may also lead to other complications, potentially necessitating multiple repositioning attempts, each carrying added risk. The high error rate of the freehand approach underscores the clinical need for better guidance tools in ventriculostomy. A possible alternative to freehand catheter placement, already employed in clinical practice, is the use of neuronavigation systems (Gumprecht et al. 1999; AlAzri et al. 2017). These systems rely on the real-time alignment of preoperative Computer Tomography (CT) or Magnetic Resonance Imaging (MRI) scans with the patient's physical anatomy via fiducial markers and/or surface registration. The surgical instruments are then tracked (typically via optical or electromagnetic sensors), allowing the clinicians to visualize the catheter's position and trajectory on a monitor. The literature indicates that when residents perform EVD placement under neuronavigation guidance, they achieve significantly higher accuracy compared to the conventional freehand approach (AlAzri et al. 2017). Despite their effectiveness, these systems are often costly, require substantial setup time, are not always available in emergency contexts, and there is a potential risk that overreliance on such technology may hinder the development of skill retention (AlAzri et al. 2017).

This underscores the need for training systems that not only enhance performance but also promote skill retention. To this end, AR/MR technologies have proven to be highly effective in a range of training-intensive domains, including procedural operations, technical maintenance, and complex assembly tasks (Kaplan et al. 2021; Daling and Schlittmeier 2024; Borsci et al. 2015). A key characteristic of procedural tasks in the medical domain lies in the nature of the visual data involved. For example, unlike operators engaged in assembly or maintenance tasks, where 2D representations are typically schematic, symbolic, and standardized (Agrawala et al. 2003; Katsioloudis et al. 2014), medical professionals must interpret CT scans (Kazemzadeh et al. 2023; Durrani et al. 2022) that portray complex, patient-specific internal anatomy. These 2D images are not abstract representations, but real cross-sections of anatomical structures that vary from individual to individual (Krabbe-Hartkamp et al. 1998). Accordingly, clinicians are required to conduct complex spatial reasoning, combine information across various planes, and make precise judgments in conditions of uncertainty and clinical risk (Sidhu et al. 2004; Yohannan et al. 2024).

This complexity makes the design of medical AR/MR training interfaces especially crucial, as they must support intuitive spatial reasoning, reinforce procedural steps, and promote skill retention for accurate performance in unaided scenarios. Within this scope, visual aids such as contextual overlays, annotations, ghosted objects, and animated instructions (Lin et al. 2021; Hajahmadi et al. 2024) play a crucial role in enhancing user understanding and task execution (Krüger et al. 2022; Tang et al. 2020).

The present study evaluates NeuroMix, an MR simulator for EVD placement, involving medical residents with basic neuroimaging knowledge but no prior hands-on experience with the procedure. We compare three distinct learning modalities, namely No Aid, 2D Aid, and 2D-3D Aid, where users perform the EVD procedure virtually. In the No Aid mode, no visual assistance is provided beyond the standard CT scans. The 2D Aid mode leverages design conventions from clinical neuronavigation systems (Gumprecht et al. 1999; AlAzri et al. 2017). Axial, coronal, and sagittal CT slices are displayed to the user, along with the real-time projection of a virtual catheter overlying the imaging planes to support trajectory planning. The 2D-3D Aid modality builds on the previous one by adding a 3D animated guide that shows the correct placement of the virtual catherer in space.

Following the training session, all participants undergo a testing phase in which they are asked to perform the EVD procedure using a real catheter on a phantom skull, without access to MR-based visual guidances. A separate Control group, which undergoes the familiarization process with the technology but does not receive the training, completes only the testing phase, allowing for comparative assessment of performance and skill retention across conditions. A pilot study involving 10 participants was conducted prior to the full-scale experiment to validate the system and address any potential challenges that may arise. The follow-up study involved 48 medical residents (38 experimental and 10 control) to address the following Research Questions (RQs):

RQ1: How do users' perceived usability (measured by SUS), workload (measured by NASA-TLX), and technology acceptance (measured by TAM) differ when comparing the No Aid, 2D Aid, and 2D-3D Aid groups in EVD placement during the training phase?

RQ2: How do users' procedure precision and execution time differ when comparing the No Aid, 2D Aid, and 2D-3D Aid groups in EVD placement during the training phase?

RQ3: What is the impact of the No Aid, 2D Aid, and 2D-3D learning modalities for EVD placement on users' skill retention, measured by procedure accuracy and execution time, once these aids are no longer provided?

This research assesses the usability, cognitive impact, and effectiveness of different training modalities in maximising performance and short-term skill retention without aids. Our work results in the following contributions:

- We introduce NeuroMix, a cost-effective MR training system for EVD using Meta Quest 3.
- We design and evaluate three training modalities with increasing spatial support (baseline No Aid, 2D Aid, 2D-3D Aid).
- We assess skill retention in an unaided testing phase using catheter placement, a gap not addressed by most prior MR studies.
- We find that 2D–3D training aids significantly improve precision even in unaided execution, without raising cognitive workload.

2 Related work

AR/MR systems for EVD placement have been designed primarily to support intraoperative performance (Buwaider et al. 2024). In most experimental studies (Van Gestel et al. 2021; Eom et al. 2024; Li et al. 2019; Grunert et al. 2024; Vychopen et al. 2025; Eom et al. 2024) systems deployed have been evaluated mainly in terms of the accuracy of practitioners when supported by the AR/MR interface via 2D and 3D visual aids. The pioneering works (Yudkowsky et al. 2013; Hooten et al. 2014) presented the first EVD guidance systems, using haptic feedback, varied virtual brains to simulate different realistic clinical scenarios and 3D overlays displayed on an external monitor. Similarly, in Li et al. (2019), the authors introduce one of the first MR systems for EVD placement, which reduces catheter misplacement and the number of insertion attempts in bedside procedures. Later studies (Van Gestel et al. 2021; Schneider et al. 2023) found that AR increased placement accuracy and helps novices reach expert-level performance.

The role played by visual aids in these AR/MR systems is crucial in providing contextual and real-time guidance regarding catheter trajectory, insertion angle, and distance to the target. Similarly, recent MR research in surgical contexts has shown that the design of visual guidance widgets strongly influences spatial accuracy and user workload. For example, in Dastan et al. (2024) the authors demonstrated that dynamically co-designed MR cues can enhance instrument alignment precision, though with higher cognitive demand. In Domínguez-Velasco et al. (2023); Benmahdjoub et al. (2023); Eom et al. (2024); Skyrman et al. (2021) the authors proposed systems visualizing the ideal catheter trajectory on a physical skull model, along with virtual target point markers (Kocher's point and/or Monro's foramen) or segmented anatomical structures such as ventricles. Similarly, some systems (Skyrman et al. 2021; Domínguez-Velasco et al. 2023; Grunert et al. 2024) provided diagnostic images, textual instructions, and feedback panels to support users during EVD placement. For example, Eom et al. (2024) proposed a system that tracks user gaze and triggers visual cues when key interface areas are overlooked. In Benmahdjoub et al. (2023), different AR visualization modes were evaluated for the EVD procedure: 3D aids resulted in being more accurate, intuitive, and cognitively efficient, whereas 2D aids provided useful support for setting the catheter's insertion depth and angles.

Most of the AR/MR systems developed for EVD placement rely on head-mounted displays, primarily the Microsoft HoloLens (Eom et al. 2024; Umana et al. 2021). To align virtual content with the patient's anatomy, systems adopted a combination of registration methods and tracking technologies. Registration is typically performed either manually, by visually aligning holograms to anatomical landmarks (Eom et al. 2024, 2024), or via fiducial markers placed on the scalp to enable more precise and repeatable alignment (Li et al. 2019; Skyrman et al. 2021; Schneider et al. 2021).For spatial tracking, systems employ either outside-in tracking with external optical devices (Skyrman et al. 2021; Eom et al. 2024), or inside-out tracking, which uses the headset's built-in sensors for autonomous localization without external references (Van Gestel et al. 2021;

Grunert et al. 2024). A possible alternative for both registration and real-time tracking is also represented by computer vision techniques (Robertson et al. 2021).

Validation of these systems has been conducted on phantom heads (Van Gestel et al. 2021; Eom et al. 2024), 3D-printed skulls (Domínguez-Velasco et al. 2023), cadaver specimens (Grunert et al. 2024), and real patients (Li et al. 2019). Evaluated outcomes included target deviation from the planned trajectory (Li et al. 2019; Skyrman et al. 2021), number of catheter passes (Eom et al. 2024), procedural time (Domínguez-Velasco et al. 2023), and trajectory alignment error (Benmahdjoub et al. 2023; Grunert et al. 2024). Some studies also reported subjective metrics, such as user confidence and perceived spatial awareness (Eom et al. 2024; Benmahdjoub et al. 2023).

This paper offers new contributions to system design and methodological analysis. In contrast to existing literature (Skyrman et al. 2021; Domínguez-Velasco et al. 20230, the system is constructed on the Meta Quest 3, utilizing its increased field of view, improved rendering, and reduced cost compared to the Microsoft HoloLens. The lower cost of the Meta Quest 3 also enhances scalability and practical adoption, enabling institutions to deploy multiple units and incorporate MR-based procedural training into routine educational programs. This reduction in economic barriers makes immersive training more accessible across diverse clinical and educational environments. By leveraging the headset's inside-out tracking, we developed a tracking system using the Quest controllers to offer a precise and portable simulator with no need for external trackers. We developed and compared three different interface modalities, each incorporating varying levels of 2D and 3D visual aids. Notably, we proposed a hybrid visualization, combining both 2D and 3D elements in combination with CT image slices, which are familiar to users and aligned with their formal training. From a methodological standpoint, another key innovation of our study is the assessment of training impact on short-term skill retention in an unaided setting. After completing the training, participants were asked to perform the task again without any visual support from the system. This highlights a gap in the current literature, as many studies demonstrate the effectiveness of AR/MR systems in improving procedural performance (Eom et al. 2024; Benmahdjoub et al. 2023). However, they often do not assess whether those improvements persist once the visual aids are removed.

3 Materials and methods

In this section, we describe the training and testing systems, the initial pilot study for system refinement and the experimental design.

3.1 Training system design and architecture

The architecture shown in Fig. 1 illustrates the NeuroMix training system (left panel), which has been developed using Unity3D (v2022.3.10f1) and the Meta SDK (v62.0.0) (Meta 2023), and deployed on the Meta Quest 3 headset for UI interaction and controller tracking. The system architecture comprises three primary modules: the Application Engine, the Data Logger and the Training Modality Interface (see Fig. 1, left panel).

3.1.1 The application engine and the data logger

The Application Engine is responsible for visualizing and synchronizing data among three virtual components: the Virtual Panel, the Digital Skull, and the Digital Catheter.

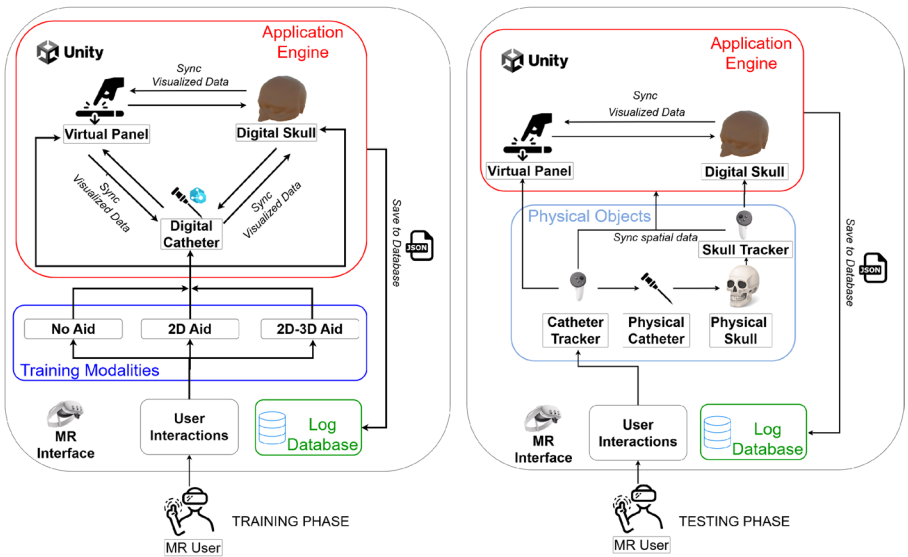


Fig. 1 NeuroMix's system. The architectures used during the Training Phase (left panel) and Testing Phase (right panel)

These components are synchronized through a real-time, bidirectional synchronization system.

The Virtual Panel serves as an in-world interface that provides real-time feedback and data visualization to the user. The content displayed depends on the specific Training Modality Interface selected. In general, the Virtual Panel presents key information such as insertion depth, trajectory angles, and brain CT scans. The design aims at providing users with an intuitive interface (Salomoni et al. 2017). The Digital Skull represents the human skull of an anonymous patient. This 3D digital element allows users to explore key anatomical landmarks, such as Kocher's point, the standard neurosurgical entry site for EVD catheter insertion, and the Monro's foramen, which serves as the anatomical target for precise catheter placement (Anatomy 2023). The Digital Skull was reconstructed using the 3D Slicer software (Pieper et al. 2004) from a real brain CT dataset consisting of 512 sagittal, 512 coronal, and 128 axial images in DICOM format, each with a resolution of 512×512 pixels. The resulting 3D model was optimized in Blender (Mullen 2011) to ensure smooth performance and compatibility with the hardware constraints and calibrated in Unity to accurately align DICOM slices within the volume. The Digital Catheter is the virtual surgical tool. A green sphere is placed at the tip of the Digital Catheter to enhance the user's manipulation (see Fig. 3).

These components are integrated into an interactive workflow that guides the user through the key steps of the EVD placement procedure. Initially, the system enables the user to position the Digital Skull comfortably in the environment, anchor it using a dedicated button, and select one of the three entry points highlighted on the Digital Skull. The user must identify and select the Kocher's point. If the selected point is incorrect, the system provides error feedback on the Virtual Panel and requests a new selection. This encourages users to actively recognize anatomical landmarks, just as they would be required to do in a real surgical setting where the Kocher's point is not pre-marked. Once the correct entry point is chosen, the virtual simulation of the EVD placement procedure begins. The user can define the Digital Catheter trajectory based on three degrees of freedom: Tilt, Rotation, and Depth. The Digital Catheter is anchored at the selected point on the Digital Skull, and the user can manipulate the Tilt and Rotation components, either through direct hand manipulation or by using the dedicated sliders on the Virtual Panel. However, Depth can only be chosen through the Virtual Panel sliders, allowing millimeter-level precision (Mendes et al. 2019). After confirming the choices, the system evaluates the trial and provides the user with some post-trial feedback depending on the Training Modality Interface. To repeat the trial, the user can press the "Retry" button on the Virtual Panel. The system then restarts from the entry point selection phase, allowing iterative practice and progressive refinement of the insertion technique.

Finally, the Data Logger module records user interaction logs during the experiment sessions.

3.1.2 The training modality interface

NeuroMix offers three distinct training modalities, namely No Aid, 2D Aid, and 2D-3D Aid, whose specific features are illustrated in Fig. 3.

No Aid Modality The user is provided only with CT images and the target point (Monro's foramen). All CT images are displayed on the Virtual Panel, and the target point is highlighted with a red dot. By interacting with the Virtual Panel, users can navigate through CT scans across all anatomical planes (sagittal, axial, and coronal). The red dot is visible upon selecting the correct CT slice, requiring the user to actively locate the target. No additional visual aids are provided during the virtual insertion (see Fig. 3A). The user has the option to continue visualizing the CT scans on the Virtual Panel while selecting Tilt, Rotation and Depth values (see Fig. 3B, C). Upon confirming the selection, the Digital Catherer is anchored, and the system provides different forms of feedback to the user. One is a textual feedback appearing on the Virtual Panel, namely the Euclidean distance (in centimeters) between the Digital Catheter's tip (operation point) and the digital anatomical target point (see Fig. 3D). The other is a 3D feedback: the Digital Skull is made transparent, allowing the user to visualize the anchored digital Catheter rendered yellow and the target point within the 3D environment, thus reinforcing spatial understanding and procedural precision (see Fig. 3D).

2D Aid Modality The Virtual Panel displays the anatomical planes simultaneously. To identify the target point highlighted with a red dot, the user can scroll the CT images using the buttons on the Virtual Panel or by dragging green virtual panels that appear on the Digital Skull (see Fig. 3A1). This interface is designed to make the connection between the 2D CT slices on the Virtual Panel and the 3D Digital Skull more intuitive (Mendes et al. 2019). Once the red target point is identified, the user proceeds with the placement, and the selected CT slices are carried over as visual aids to the next steps. Then, the user selects the Rotation, Tilt, and Depth values. However, this interface includes a real-time 2D projection of the Digital Catheter superimposed on the CT scan images. This projection is continuously updated while the user changes the Rotation, Tilt, and Depth values, providing 2D visual aids of the catheter's orientation (see Fig. 3B1-C). After each attempt, the system provides the same feedback as the No Aid modality. In addition, the user

is provided with the final catheter's 2D trajectories across the three anatomical planes on the CT scans (see Fig. 3D).

2D-3D Modality This modality builds upon the 2D Aid modality (see Fig. 3A1). An additional visual aid is introduced in the form of a 3D trajectory guide to support the user in selecting the Tilt and Rotation values. This red guideline visually indicates the optimal insertion path in 3D, including the correct angle. This allows participants to follow a predefined path that would lead to a "perfect" placement if executed accurately (see Fig. 3B2). A dedicated button on the Virtual Panel allows the user to activate an animation. This animation is context-aware: starting from the Digital Catheter's current orientation, it visually shows the movement required for the Digital Catheter to align with the ideal insertion angle. The Depth component is then selected using the sliders of the Virtual Panel (see Fig. 3C). Differently from the previous modalities, the Virtual Panel shows the user the ideal depth component value in order to reach the inner target point. After each virtual EVD placement, the system provides the same feedback as the 2D Aid modality (see Fig. 3D).

3.2 Testing system design and architecture

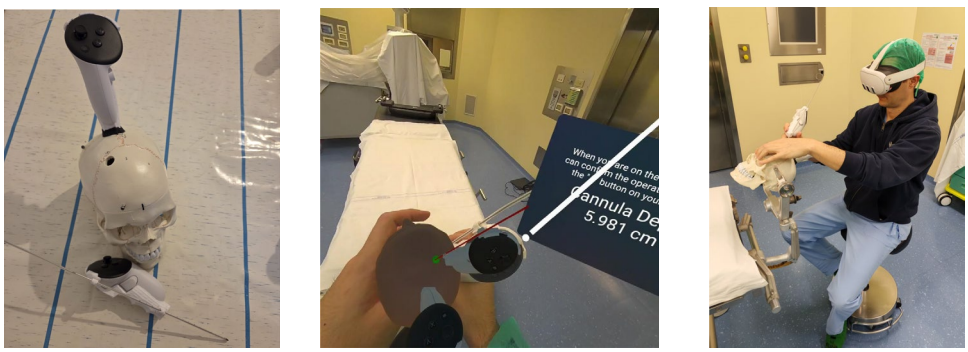
The architecture shown in Fig. 1 illustrates the NeuroMix MR testing system (right panel). The testing system is composed of three main components: the Physical Objects, the Application Engine, and the Data Logger.

Differently from the training system, the testing system considers a tracked Physical Catherer and a tracked Physical Skull (see Fig. 2a). The Physical Catherer is a real graduated catheter used during EVD placement procedures. The markings on the Physical Catheter are intended to indicate the insertion depth to the user. The Physical Skull is a phantom skull. To provide haptic feedback during insertion, an agar-based material (Van Doormaal et al. 2025) was placed inside the cranial cavity, simulating the resistance of brain tissue (gray matter). For the testing phase, the entry point (Kocher's point) is predefined, thus removing the need for manual entry point selection. Hence, the Physical Skull was perforated in advance.

A key technical challenge was achieving accurate tracking of the Physical Skull and the Physical Catheter due to restrictions imposed by Meta on direct camera access within the Meta Quest 3, thus preventing the usage of computer vision techniques as in Robertson et al. (2021). However, camera-based tracking could have introduced latency (40–60 ms), GPU overhead (1–2%), memory usage (45 MB per stream), a data rate capped at 30 Hz, and maximum resolution of 1280 × 960 pixels[1].

To overcome these limitations, we adopted Meta Quest 3's controller-based inside-out tracking, which provides real-time six-degree-of-freedom localization through inertial sensors and infrared LEDs. Meta Quest 3 controllers were selected to ensure portability, simplicity of setup, and full system autonomy. One controller was rigidly attached to the Physical Skull and the other to the Physical Catheter using custom 3D-printed mounts, ensuring stable alignment and minimizing mechanical play. This solution enabled accurate, privacy-preserving tracking at 60 Hz, outperforming the camera-based alternative in responsiveness and computational efficiency.

A rigid-body registration between the Digital Skull and the Physical Skull was performed entirely manually, using three anatomical landmarks (Kocher's point, nasion, and bregma) visually matched in the Quest 3 passthrough view. A fourth independent landmark (the right external auditory meatus) served as validation, yielding a mean registration error of 3.2 ± 1.4 mm across sessions. Since the catheter was tracked via a controller rigidly mounted to its shaft, no additional spatial registration was required. A two-step calibration was performed: (i) tip calibration, obtained by manually aligning the real catheter tip with a virtual reference to compute the fixed offset from the controller frame; and (ii) depth calibration, which mapped the zero-depth contact



(a) (b) (c)

Fig. 2 Testing System setup. **a** Physical Skull and Physical Catheter instrumented with Meta Quest 3 controllers using custom 3D-printed mounts for inside-out tracking. **b** MR view during the catheter manipulation task, showing the virtual catheter and real-time depth feedback on the virtual panel. **c** Participant performing the EVD placement task during the testing phase using the NeuroMix system with the Meta Quest 3 headset

[1] See Meta's documentation.

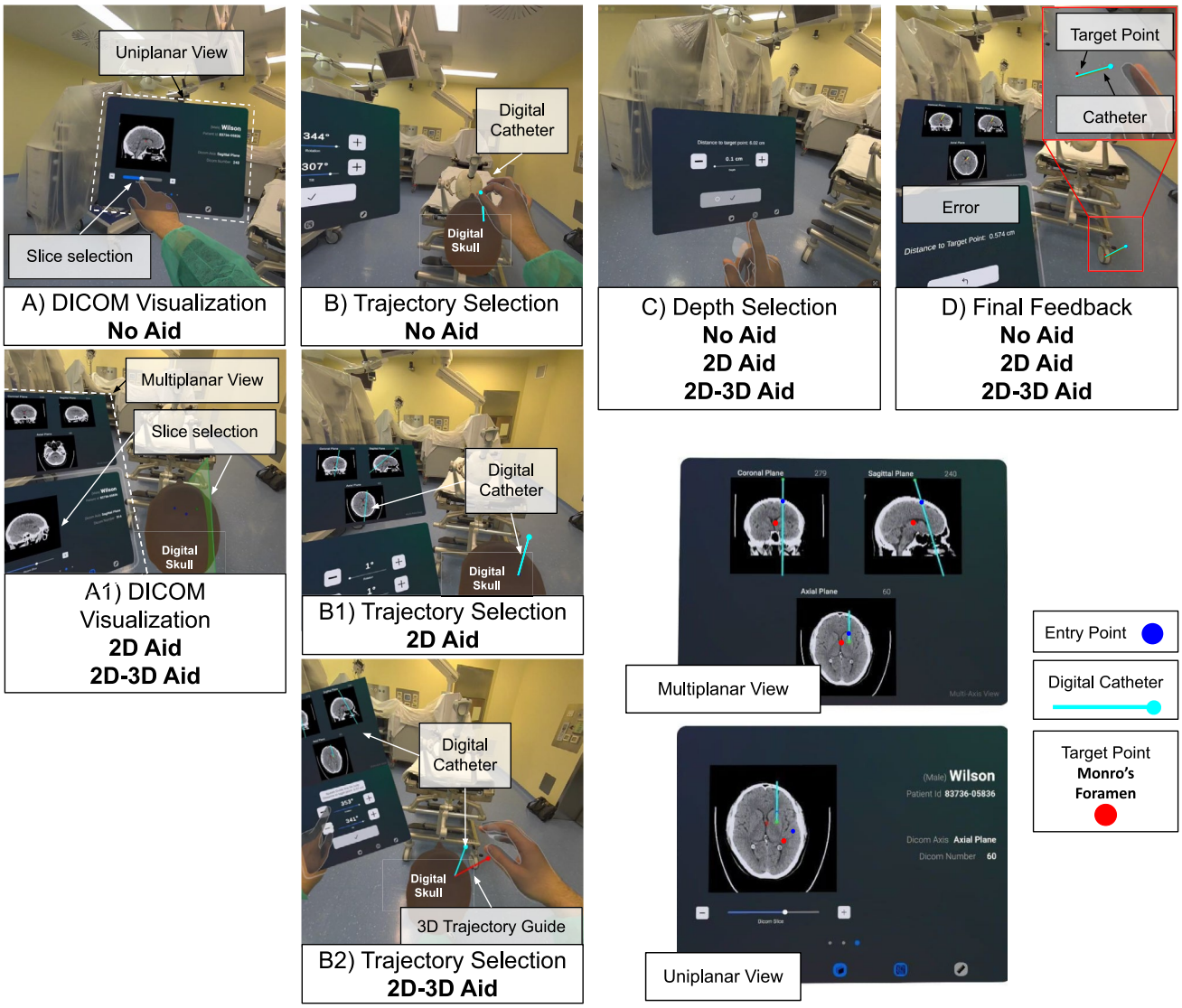


Fig. 3 NeuroMix Training interfaces across the three learning modalities. **A** DICOM Visualization: users inspect CT slices to identify the Monro's foramen target (No Aid, 2D Aid, 2D–3D Aid). On the Right Panel: examples of multi-planar and uni-planar CT views with entry point (blue), catheter (cyan), and anatomical target (red). **B** Trajectory Selection: No Aid provides no visual guidance; 2D Aid adds real-time 2D catheter projections on CT planes; 2D–3D Aid includes an additional animated 3D trajectory guide. **C** Depth Selection: in all modalities, depth is selected via the Virtual Panel sliders. **D** Final Training Feedback: after each virtual insertion, the system provides textual and visual feedback showing the placement error and the catheter–target spatial positioning

point to its virtual counterpart, ensuring millimeter-level consistency between physical and digital insertion depth.

Finally, static and dynamic drift tests were conducted to assess tracking stability. Under static conditions (60 s at 60 Hz), the mean positional drift of the catheter tip was 4.7 mm (SD = 0.9 mm), increasing slightly to 5.3 mm (SD = 1.1 mm) during manual manipulation. This approximately 5 mm systematic uncertainty affects the absolute spatial measurements but does not compromise the relative comparison across training modalities, as all conditions shared the same tracking and calibration setup.

The 3D-printed mount was designed in order to let the user hold the Physical Catheter similarly to how a neurosurgeon would have a catheter in a real scenario. During insertion, the user manipulates the tracked Physical Catheter directly with their hands, with the controller acting as a tracker. This enables simultaneous control of insertion depth and angle, replicating the natural handling of real surgical procedures. By pressing the "A" button on the controller, the user can confirm that the desired target point is reached. To reproduce the tactile experience of catheter insertion, the cranial cavity of the Physical Skull was filled with an

agar-based compound, replicating the resistance of brain parenchyma.

The Application Engine runs within the Unity3D environment and is responsible for real-time visualization and synchronization of the Virtual Panel and the Digital Skull. Additionally, it handles the registration of the Digital Skull to its physical counterpart, the Physical Skull. Concerning the Digital Skull, the same virtual head model used in the training system was rendered in the MR testing environment to provide consistent reference points established during the training phase. Since the graduated catheter is difficult to see clearly while wearing the headset, the insertion depth is also displayed in real-time on the Virtual Panel to support the user throughout the EVD placement (see Fig. 2b). Importantly, we remark that, in this setup, the MR interface was not used to provide any form of assistance (aids or feedback) to the user. This design choice ensures that the testing system can yield meaningful insights into each user's skill retention in an unaided scenario. Differently from the training system, we point out that all performance measurements were computed in a virtual setting, while the procedure is performed with a real catheter on a real phantom skull (see Fig. 2c). For example, the distance between the operation point (from the tip of the tracked Physical Catheter) and the virtual target point was calculated using the MR tracking system. This ensures high-fidelity evaluation of each trial while maintaining full control over data collection and eliminating manual measurement errors.

The Data Logger module collects and stores all interaction data during the EVD procedure.

3.3 A pilot study for system refinement

A pilot study involving 10 participants was conducted to evaluate the initial version of the NeuroMix training and testing systems to identify potential issues in the experimental setup. None of these participants were included in the main experiment to prevent familiarity or learning-related biases. Among them, seven were neurosurgeons with varying levels of experience in EVD placement (one with low experience, two with moderate, and four with high experience) while the remaining three had no prior experience with neurosurgical procedures. All participants tested the No Aid, 2D Aid, and 2D-3D Aid modalities as well as the testing system. Their ages ranged from 24 to 45 years old. Concerning their familiarity with MR technologies, six had never used MR before, and four reported moderate exposure.

During the pilot study sessions, some technical issues emerged, including application crashes due to a memory leak caused by rendering the high-resolution Digital Skull model within both the training and testing systems. To address this, the 3D model was optimized in Blender to reduce complexity while preserving visual fidelity, and GPU-based rendering was implemented to offload image processing from the CPU, leveraging the Snapdragon XR2 Gen 2 architecture. These optimizations enabled the system to run at a full resolution of 2064×2208 pixels per eye with a stable 90 Hz refresh rate, ensuring smooth and responsive performance.

In terms of usability, users were initially required to interact with virtual objects in the training system using controllers, which many found to be cumbersome. Additionally, the only method for selecting the Rotation and Tilt components of the Digital Catheter was through sliders on the Virtual Panel. To improve ease of interaction, hand tracking was introduced, allowing users to manipulate the virtual catheter directly with their hands. However, the slider-based controls for Tilt and Rotation were retained at the participants' request, as they allowed for precise millimetric adjustments, increased accuracy, reduced task frustration, and improved reproducibility across trials. These adjustments align with findings in the literature (Jeffri and Rambli 2021) emphasizing that reducing cognitive load through effective interface and interaction modality design (Mendes et al. 2019) is essential to ensure both improved user performance and reliable experimental results, particularly in tasks requiring high levels of spatial reasoning.

During the testing phase, although the tracked Physical Catheter featured graduated markings, participants had difficulty seeing them clearly in passthrough mode (Guo et al. 2022). To address this, the insertion depth values were displayed in real-time on the Virtual Panel. Additionally, due to camera distortion in passthrough mode (Guo et al. 2022), users experienced some difficulties when inserting the Physical Catheter into the hole at the Kocher's point on the Physical Skull. To mitigate this issue, a digital red stick was registered to the Physical Catheter to provide a clearer visual reference. Finally, participants also confirmed that the agar-based medium provided realistic tactile feedback, leading to its inclusion in the final testing setup.

3.4 Study design

This section outlines the experimental design, procedure, participants, and measurements used in the study.

3.4.1 Study objective

The experiment employed a between-subjects design to assess the effectiveness of the developed training and testing systems, as well as the impact of three different training modalities (No Aid, 2D Aid, and 2D-3D Aid) offering varying levels of visual aids on short-term skill retention for the EVD placement procedure. Participants were divided into

two main groups: the Experimental Group and the Control Group. The Experimental Group completed both the training and testing phases, whereas the Control Group took part only to the testing phase. Specifically, the experiment aims to assess how different learning modalities influence users' perceived usability, cognitive workload, and acceptance of the training system during the training phase. Additionally, the study examines the impact of each modality on procedural precision and execution time, both during training and in a subsequent testing phase where visual aids are no longer available. A comparison with a Control Group, which did not receive training through the developed system, is included to evaluate the retention effects more rigorously.

3.4.2 Participants

A total of 48 participants took part in the main experiment, ranging in age from 23 to 38 years (M = 28.04, SD = 3.13), including 21 females and 28 males. This sample size aligns with prior research in the field (Buwaider et al. 2024; Butaslac et al. 2022). Results from the background questionnaire indicated that all participants had a medical background and were selected among recently hired clinicians with limited experience in EVD placement (M = 2.00, SD = 0), limited familiarity with neuronavigation systems (M = 1.19, SD = 0.57), and moderate experience in interpreting neuroimaging such as CT and MRI scans (M = 3.41, SD = 0.69). In terms of handedness, 43 participants were right-handed, 4 left-handed, and 1 ambidextrous. Regarding familiarity with immersive technologies, participants had limited familiarity with Augmented, Mixed, or Virtual Reality (M = 2.57, SD = 0.68). Specifically, the vast majority (45 out of 48) had no prior experience using the Meta Quest 3 device employed in this study, and limited experience with other head-mounted displays (M = 2.60, SD = 0.63). Of the 48 participants, 36 were assigned to the Experimental Group and evenly distributed across the three training modalities: No Aid (n = 12), 2D Aid (n = 12), and 2D-3D Aid (n = 12). The remaining 12 participants were assigned to the Control Group.

3.4.3 Procedure and task

We now outline the experimental procedure followed by the Experimental Group and the Control Group (see Fig. 4). Participants were randomly assigned to one of the two groups. All participants were asked to complete a demographic questionnaire, which included information such as age, gender, and education level. This was followed by a background questionnaire aimed at assessing their prior experience with EVD procedures, neuronavigation systems, interpretation of brain CT DICOM images, the use of software such as 3D Slicer, and their familiarity with various immersive technologies.

Then, participants in the Experimental Group were randomly assigned to one of the three learning modalities: No Aid, 2D Aid, or 2D-3D Aid. For each participant in the Experimental Group, the experiment consisted of three distinct phases: the Familiarization Phase, the Training Phase, and the Testing Phase.

During the Familiarization Phase all participants took part in an introductory session led by an experienced operator, who explained the theoretical and practical aspects of

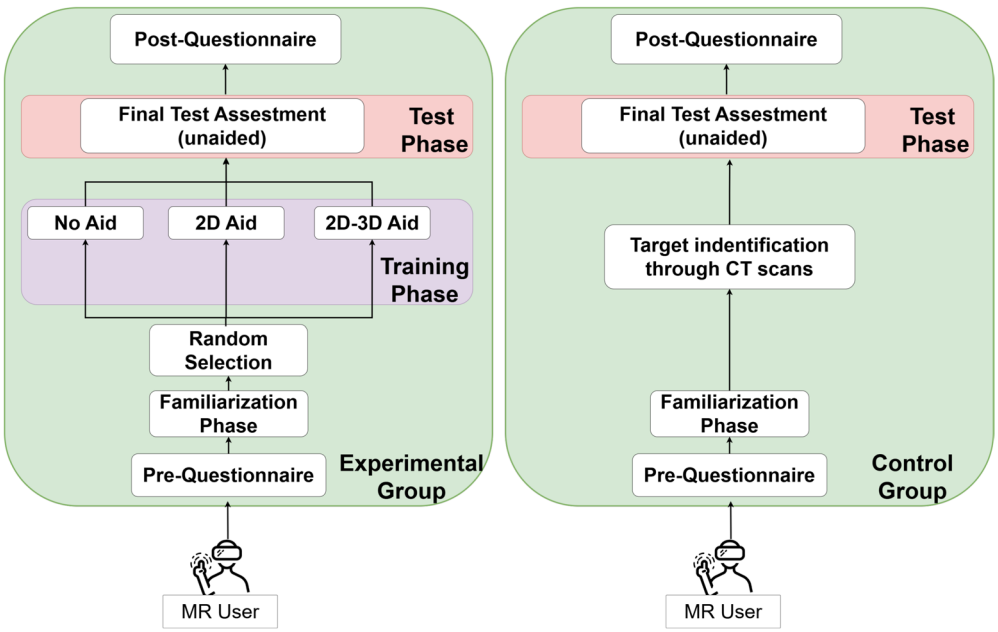


Fig. 4 Experimental Design for the Experimental Group (left panel) and Control Group (right panel)

the EVD procedure according to standard neurosurgical protocols (Flint et al. 2013). This session covered topics such as the clinical indications for EVD placement, common anatomical landmarks (with an emphasis on Kocher's point and Monro's foramen), the ideal catheter trajectory and depth (generally 5 to 7 cm), confirmation cues for correct catheter positioning (e.g., release of cerebrospinal fluid), and potential complications including infections, hemorrhages, device failure, and misplacement. Following this, participants were guided through a familiarization session specific to the learning modality they had been assigned to. During this session, they were introduced to the system's features and instructed on how to interact with the Virtual Panel, Digital Catheter, and Digital Skull provided by their respective setup.

After the Familiarization Phase, the Training Phase begins. During this phase, the participants interact exclusively with virtual tools (Seymour et al. 2002). Each participant is asked to perform five virtual EVD placements. Depending on the assigned learning modality, participants practice inserting the Digital Catherer into the Digital Skull towards the target point (Monro's foramen) highlighted on the CT scans. Depending on the training modalities, the procedure is supported by visual aids, when available, in order to assist participants in selecting the correct entry point and achieving proper catheter placement. After each attempt, participants receive feedback tailored to their assigned modality. This feedback is designed to help them understand how to improve their technique, in line with evidence from the literature supporting the role of feedback in enhancing procedural learning (Eom et al. 2024; Butaslac et al. 2022).

Once the Training Phase is completed, the Testing Phase begins. In this phase, participants interact with the Physical Skull, a phantom model enhanced for realism by filling the cranial cavity with an agar-based material that mimics the consistency of brain tissue. This setup provides users with haptic feedback, simulating the EVD placement procedure in a more lifelike manner. The entry point is indicated to the participant by a pre-existing borehole on the Digital Skull. To ensure visual continuity and consistency, the system registers the same Digital Skull used during the Training Phase on the Physical Skull. All sessions were conducted using a single anonymized CT dataset, ensuring full experimental control and consistent anatomical landmarks across participants (Anatomy 2023; Flint et al. 2013). A Physical Catheter, identical to the one used in real surgical procedures, is also employed. Participants are required to perform five EVD placements during this phase. No visual aids or feedback are provided during this phase. This number was chosen based on prior work in procedural learning and MR-based surgical simulation (Kaplan et al. 2021; Borsci et al. 2015), which shows that performance in immersive task-based training generally stabilizes within 4–6 repetitions.

The Control Group did not undergo any virtual Training Phase but performed five EVD placements as described for the Testing Phase performed by the Experimental Group. However, before the insertions, the Control Group was allowed to visualize the CT scans on a standard computer monitor using Slicer3D for reference.

At the end of the experiment, all participants were asked to complete questionnaires tailored to their respective groups.

3.4.4 Measurements

Various objective and subjective metrics were collected to assess the impact of the NeuroMix training on procedural accuracy and efficiency. Concerning objective measures, during the Training Phase the system recorded: (a) the number of attempts required to identify the correct entry point, (b) the 3D coordinates of the operation points, (c) the Rotation and Tilt components selected, and (d) the Total Time (TT) for each insertion. During the Testing Phase, only metrics (b), (c) and (d) were recorded, as the entry point was predefined and provided to participants. From (b), we derived the Error Score (ES) by computing the Euclidean distance between the operation point and the target point. From (c), we derived the Rotation Error (RE) and the Tilt Error (TE), which are defined as the absolute angular differences between the user-selected Rotation and Tilt and the ideal trajectory toward the target point. For each participant and training modality, we computed the Average metric by averaging the individual error values (e.g., Error Score, Tilt Error, Rotation Error, Total Time) across all trials (5 trials for the Training Phase and 5 trials for the Testing Phase for each participant).

Subjective measures were measured through standardized questionnaires. The Experimental Group completed the System Usability Scale (SUS), a 10-item questionnaire rated on a 5-point Likert scale. The items are grouped into two dimensions: Usability and Learnability (Lewis and Sauro 2009). In addition to analyzing these two subscales, we also calculated the Total score as the sum of all item responses across the entire SUS questionnaire. In addition, the Experimental Group completed the NASA Task Load Index (NASA-TLX) (Hart and Staveland 1988), which is based on a 10-point Likert scale and evaluates six dimensions of perceived workload: Mental Demand, Physical Demand, Temporal Demand, Performance, Effort, and Frustration. We also calculated the Average workload score by averaging across all dimensions. The Experimental Group was asked to complete the Technology Acceptance Model (TAM) questionnaire (Davis 1989), which was also

administered using a 5-point Likert scale. It includes two core dimensions: Perceived Usefulness and Perceived Ease of Use. Finally, for the Experimental Group we developed a custom Final Test Assessment (FTA) questionnaire to evaluate whether participants' perceived complexity of the Testing Phase aligned with their actual performance. The questionnaire consisted of five items, each rated on a 5-point Likert scale (e.g., "How do you rate the difficulty of accomplishing the X insertion of the catheter during the testing phase?").

The Control Group completed only the FTA questionnaire as they did not undergo the virtual training.

3.4.5 Tools for statistical analysis

For the statistical analysis, we primarily relied on a set of Python libraries designed for data manipulation, statistical modeling, and visualization. The goal of the analysis was to identify significant differences or similarities both within the Experimental Group (across the three training modalities) and between the Experimental and Control Groups. We followed the statistical protocol outlined below. First, the normality of the data distributions was assessed using the Shapiro-Wilk test (Razali and Wah 2011). If normality was confirmed, a one-way ANOVA was applied to detect significant differences between groups, followed by pairwise comparisons using independent samples t-tests (Lakens 2013). To evaluate statistical equivalence, we performed a two-sided TOST (Two One-Sided Tests) procedure (Lakens 2017). If normality was not satisfied, non-parametric tests were used: the Kruskal-Wallis H-test (Mckight and Najab 2010) to assess differences across groups and the Mann–Whitney U test for pairwise comparisons (Mcknight and Najab 2010). Finally, we employed the Analysis of Covariance (ANCOVA) (Van Breukelen 2013) test to examine the effect of training modality on procedural performances while controlling for potential covariates.

4 Results

This section reports the outcomes of the statistical analyses, distinguishing between subjective and objective metrics.

4.1 Subjective measures

We present here the results of the subjective evaluations collected through the standardised post-experience questionnaires. For clarity, Table 2 reports a concise summary of the significant pairwise results for all subjective measures.

Table 1 Summary of subjective measures

Variable	Construct	No Aid	2D Aid	2D-3D Aid	Control
		(M, SD)	(M, SD)	(M, SD)	(M, SD)
SUS	Usability	4.24, 0.47	4.16, 0.37	4.02, 0.55	–
	Learnability	3.88, 0.77	3.88, 0.68	3.54, 0.69	–
	Total	91.67, 11.55	90.00, 10.11	85.62, 13.86	–
NASA-TLX	Mental	5.42, 1.88	6.17, 1.34	6.25, 1.48	–
	Physical	3.17, 1.70	3.25, 1.71	4.25, 1.71	–
	Temporal	3.75, 2.05	4.00, 1.28	3.67, 1.44	–
	Performance	7.33, 1.61	6.67, 1.92	7.58, 0.79	–
	Effort	4.92, 2.23	6.17, 1.85	6.08, 1.83	–
	Frustration	2.00, 1.13	3.58, 2.07	3.17, 2.08	–
	Average	4.43, 0.89	4.97, 0.84	5.17, 1.02	–
TAM	Usefulness	4.18, 0.62	4.17, 0.48	4.45, 0.51	–
	Ease of Use	2.85, 0.41	3.07, 0.30	3.09, 0.32	–
FTA	Average	2.98, 0.78	2.95, 0.36	3.07, 0.83	2.75, 0.86

Table 2 Summary of significant statistical results for subjective measures

Measure	NA vs 2D	NA vs 3D	2D vs 3D
SUS – Usability	=	=	=
NASA-TLX – Frustration	<	None	None
TAM – Ease of Use	=	=	=
TAM – Usefulness	=	None	None
FTA – Average	=	None	None

Comparisons use the following abbreviations: NA = No Aid, 2D = 2D Aid, 3D = 2D–3D Aid. "=" indicates significant equivalence; "<" / ">" indicate significant differences; "None" indicates no significant result

4.1.1 SUS questionnaire

In Table 1 we report the mean (M) and standard deviation (SD) by construct of the SUS questionnaire for each of the learning modalities. All modalities received high Total values (>70). Normality was confirmed for each of the constructs across all the learning modalities. The ANOVA test did not show any significant differences among the three modalities for Usability (p=0.521), Learnability (p=0.428) and Overall (p=0.499). A pair-wise TOST analysis revealed significant equivalence in terms of Usability: No Aid vs. 2D

Aid (p = 0.012), 2D Aid vs. 2D-3D Aid (p = 0.035), and No Aid vs. 2D-3D Aid (p = 0.046).

4.1.2 NASA questionnaire

The mean and standard deviation of each NASA dimension are reported in Table 1. Normality was achieved by the Physical, Temporal, and Effort dimensions, with the Mental, Performance, and Frustration dimensions failing to meet the normality assumption. A one-way ANOVA did not reveal significant differences among the three learning modes in the Physical, Temporal, and Effort dimensions. However, the Kruskal–Wallis H test revealed a significant difference in Frustration scores (p = 0.049), with no differences found for the Mental and Performance dimensions. Pair-wise comparisons using the Mann–Whitney U test revealed a significant difference for No Aid vs 2D Aid (p = 0.038), but not for No Aid vs 2D-3D Aid (p = 0.081), or for 2D Aid vs 2D-3D Aid (p = 0.498). Finally, the TOST procedure did not reveal any significant equivalences between the modalities.

4.1.3 TAM questionnaire

Detailed scores for the TAM metric, including means and standard deviations for all dimensions, are reported in Table 1. For all these dimensions of the TAM questionnaire, the normality was achieved. The ANOVA test did not reveal significant differences for Usefulness (p = 0.370) and for Ease of Use (p = 0.183) among the three learning modalities. A pair-wise TOST test showed significant equivalence for the Ease of Use construct for No Aid vs 2D Aid (p = 0.034), No Aid vs 2D-3D Aid (p = 0.046) and 2D Aid vs 2D-3D Aid (p = 0.0001). Moreover, significant equivalence was found when applying the TOST test to No Aid vs 2D Aid (p = 0.021) for the Usefulness construct.

4.1.4 FTA questionnaire

In Table 1, we reported the mean and standard deviation scores for the Overall FTA questionnaire. A Kruskal–Wallis H test conducted across all groups did not reveal any statistically significant differences. However, the TOST analysis indicated a statistically significant equivalence between the No Aid and 2D Aid groups, while no equivalence was found between the other group pairs.

Table 3 Summary of the average error scores [cm]

Phase	Metric	No Aid	2D Aid	2D-3D Aid	Control
		(M, SD)	(M, SD)	(M, SD)	(M, SD)
Training	Average ES	1.05, 0.57	0.67, 0.65	0.42, 0.27	–
Testing	Average ES	2.00, 0.98	2.30, 1.15	1.47, 0.60	2.64, 1.06

4.2 Objective measures

This section presents the statistical analysis results of the objective metrics collected during the Training phase and the Testing phase. For clarity, Table 6 provides a concise summary of the main significant pairwise differences observed across all objective measures.

4.2.1 Accuracy analysis: entry, target and operation points

Regarding Entry Point selection, in 94% of the total instances, the correct option (Kocher's point) was identified on the first attempt during the Training Phase.

In Table 3 we report the mean and standard deviation of the Average Error Scores for the training and testing phases.

The results of the Shapiro-Wilk test indicated that all the variables violated the assumption of normality. The Kruskal–Wallis H-test revealed statistically significant differences between the three learning modalities during the Training phase (p = 0.0018). The Mann–Whitney U test showed the following significant differences: No Aid vs 2D Aid (p = 0.0001), No Aid vs 2D-3D Aid (p = 0.0001), and 2D Aid vs 2D-3D Aid (p = 0.011), suggesting improvements in precision with increased system support.

Concerning the Testing Phase, the four modalities (including Control) showed a significant difference (p =0.0023). In contrast, pairwise comparisons showed that the 2D-3D Aid condition differed significantly from all others: No Aid (p = 0.0053), 2D Aid (p = 0.00021), and Control (p = 0.00014). Significant differences were observed between the Control and the No Aid groups (p = 0.001). No significant differences were found between 2D Aid and No Aid (p = 0.184), and between Control and 2D Aid (p = 0.068).

To perform a more detailed analysis on the Testing Phase data, we computed the component-wise absolute differences between the target point and the operation point along the sagittal X, axial Y, and coronal Z axes. In the sagittal direction X, significant differences were observed only between the Control condition and the other three modalities: No Aid (p = 0.0012), 2D Aid (p = 0.0040), and 2D-3D Aid (p = 0.0188). In the axial direction Y, more widespread differences emerged, particularly involving the 2D-3D Aid condition, which significantly differed from all others: 2D Aid (p = 0.00012), Control (p = 5.02e−7), and No Aid (p = 0.036). Similarly, in the coronal direction Z, the 2D-3D Aid modality showed significantly different results when compared to No Aid (p = 0.00066), 2D Aid (p = 0.00399), and Control (p = 0.0181). For the Testing Phase, Fig. 5 illustrates the spatial deviations from the target point (fixed in the origin) and the

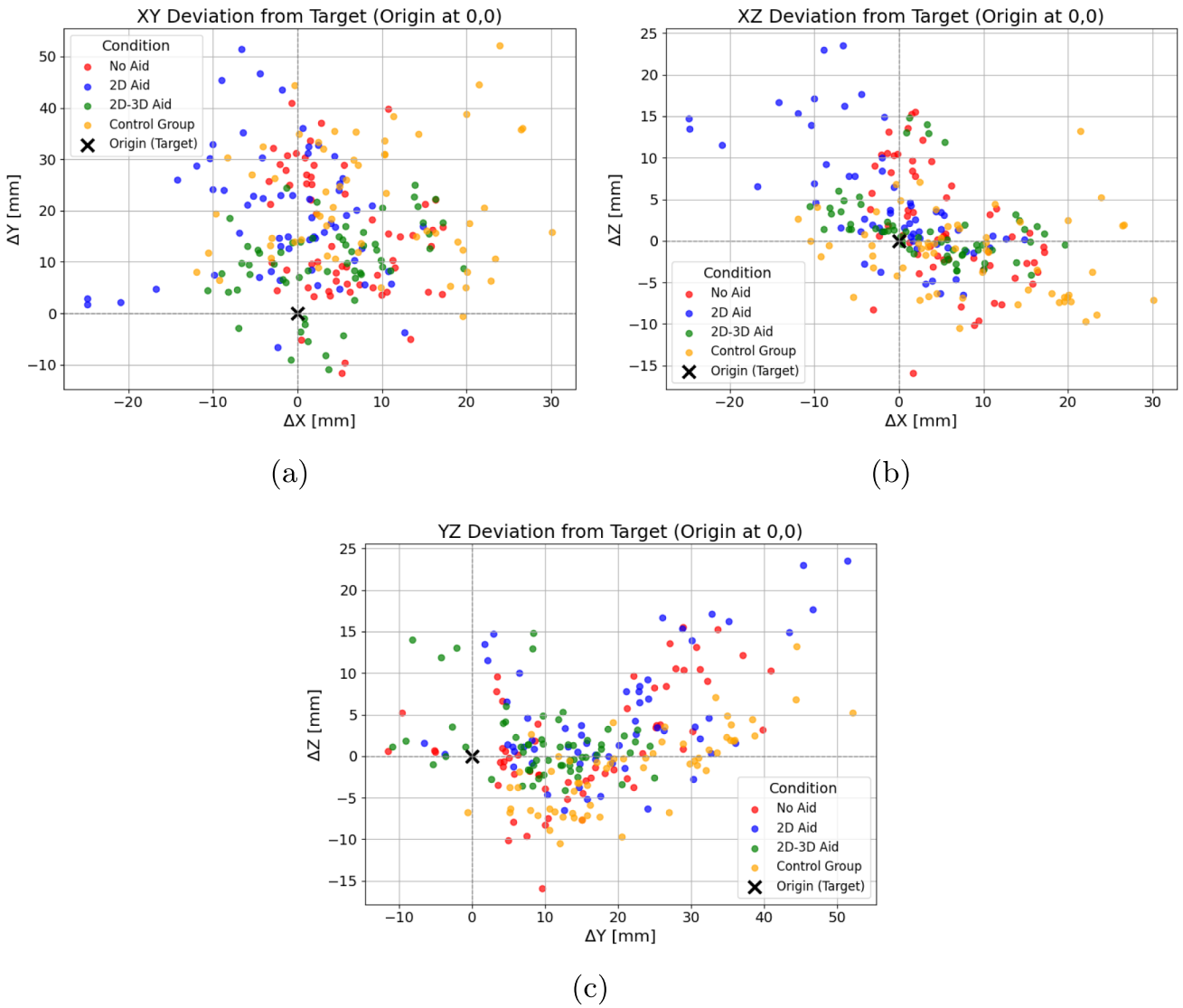


Fig. 5 Spatial deviations from the target across sagittal [XY], coronal [XZ], and axial [YZ] planes

operation points across the three main anatomical planes: sagittal [XY], coronal [XZ], and axial [YZ]. In all the views, the 2D-3D Aid condition (green) shows the highest concentration of data points around the target. In contrast, the Control condition (orange) displays the widest dispersion, particularly along the positive vertical Y axis (see Fig. 5a, C). The No Aid (red) and 2D Aid (blue) conditions show intermediate behavior. From visual inspection, Control, No Aid and 2D Aid groups seem to show a deviation along the positive axial direction Y, differently from the 2D-3D Aid group.

Finally, we conducted an ANCOVA test using the number of training attempts as a covariate, the training modality as a between-subject factor and precision during Testing Phase as dependent variable. The analysis revealed a significant main effect of training modalities on operation precision ($p = 0.0021$), indicating that the type of training used had an impact on precision, independent of the number of attempts.

A Spearman correlation analysis (De Winter et al. 2016) was conducted to investigate the relationship between the FTA metric (Table 1) and the operation precision (Table 3) during the Testing Phase. The results revealed a weak but statistically significant ($p = 0.012$) negative correlation, indicating that higher levels of perceived complexity were associated with lower precision scores.

4.2.2 Accuracy analysis: tilt and rotation components

The mean and standard deviation values for the Average Tilt Error and the Average Rotation Error are reported in Table 4.

Table 4 Summary of avarage tilt and rotation errors [°]

Phase	Metric	No Aid	2D Aid	2D–3D Aid	Control
		(M, SD)	(M, SD)	(M, SD)	(M, SD)
Training	Average TE	4.94, 3.87	4.08, 5.33	1.74, 1.58	–
	Average RE	5.05, 3.92	2.78, 2.21	2.20, 1.86	–
Testing	Average TE	14.60, 9.97	16.94, 10.58	10.04, 5.08	17.97, 10.12
	Average RE	6.02, 4.40	7.27, 6.41	6.14, 4.21	9.44, 7.25

Table 5 Summary of the average total time [s]

Phase	Metric	No Aid	2D Aid	2D-3D Aid	Control
		(M, SD)	(M, SD)	(M, SD)	(M, SD)
Training	Average TT	62.21, 46.36	137.84, 78.90	132.09, 93.25	–
Testing	Average TT	42.87, 23.36	47.25, 20.64	49.74, 37.04	33.63, 14.53

A Kruskal–Wallis H test performed on the Training Phase data revealed significant differences in angular accuracy among the three learning modalities, both for Tilt ($p = 0.0034$) and Rotation ($p = 0.012$). Pairwise comparisons using the Mann–Whitney U test confirmed significant differences between all pairs of modalities.

The Kruskal–Wallis H test performed on the Testing Phase data revealed significant differences among the No Aid, 2D Aid, 2D-3D Aid and Control groups for the Tilt component (p=0.0042). No significant differences were found for the Rotation component (p=0.092). For the Tilt-related data, the Mann–Whitney U test confirmed that the 2D–3D Aid condition was significantly different if compared to the 2D Aid (p=0.007) and the Control (p= 0.005), while it was not significant different from the No Aid ($p = 0.071$). Additionally, the No Aid condition was significantly different from Control ($p = 0.036$), whereas it did not show significative difference with respect to 2D Aid (p=0.067).

Finally, we conducted an ANCOVA using the number of training attempts as a covariate, the training modality as a between-subject factor, and Rotation and Tilt Errors during the Testing Phase as the dependent variable. For the Rotation errors, the analysis revealed no significant effect of the training modalities ($p = 0.338$) while the number of attempts had a significant impact (p=0.0004). In contrast, for the Tilt component, both the training modalities ($p = 0.0056$) and the number of attempts (p=0.0002) had a significant main effect, indicating that the type of training aid influenced tilt accuracy beyond mere repetition.

4.2.3 Time analysis

We report in Table 5 the mean and standard deviation values of the Average Total Time.

A Kruskal-Wallis H test was conducted to compare the average total time during the training phase across the three learning modes (No Aid, 2D Aid, and 2D-3D Aid). The test was statistically significant ($p= 0.0001$), indicating that the total time required to complete the operation differed across modes. Pairwise analyses showed that the No Aid group completed the task significantly faster than both aided groups, however, no significant difference was observed between the two aided modalities.

A Kruskal-Wallis H test was conducted to evaluate differences of the Average Total Time for the Testing Phase across No Aid, 2D Aid, 2D-3D Aid and Control groups. The analysis revealed a statistically significant difference within the data ($p = 0.003$), suggesting that the type of aid used had a measurable effect on the operation time. Pairwise comparisons showed that the Control group performed significantly faster than all trained groups, while no significant differences were observed among the No Aid, 2D Aid, and 2D–3D Aid conditions.

Finally, we conducted an ANCOVA using the number of training attempts as a covariate, the training modality as a between-subject factor, and the Total Time during the Testing Phase as the dependent variable. The training modality

Table 6 Summary of significant statistical results for objective measures

Phase	Measure	NA vs 2D	NA vs 3D	2D vs 3D	NA vs C	2D vs C	3D vs C
TR	ES	>	>	>	–	–	–
	TE	>	>	>	–	–	–
	RE	>	>	>	–	–	–
	TT	>	>	None	–	–	–
TS	ES	None	>	>	<	None	<
	TE	None	None	>	<	None	<
	RE	None	None	None	None	None	None
	TT	<	<	None	>	>	>

Abbreviations: TR = Training, TS = Testing; NA = No Aid, 2D = 2D Aid, 3D = 2D–3D Aid, C = Control; ES = Average Error Score (precision), TE = Average Tilt Error, RE = Average Rotation Error, TT = Average Total Time. "<"/">" = significant differences; "None" = no significant result; "–" = not applicable

did not show a significant main effect ($p = 0.396$) on operation time while the number of training attempts did ($p = 0.0045$).

5 Discussion

In this section, we seek answers to the RQs. *RQ1: How do users' perceived usability (measured by SUS), workload (measured by NASA-TLX), and technology acceptance (measured by TAM) differ when comparing the No Aid, 2D Aid, and 2D-3D Aid groups in EVD placement during the training phase?*.

All the three training modes (No Aid, 2D Aid, and 2D–3D Aid) were found to be highly acceptable and usable to the participants. SUS scores were consistently above the usual acceptance level for usability of 70 (Bangor et al. 2009), with no significant differences between the groups and showing significant equivalence in all paired comparisons. These findings are presented in Table 1 and show that even highly advanced visualizations, such as animated 3D trajectory guidance, do not negatively impact perceived usability. Similarly, the TAM measures revealed high scores for Perceived Usefulness and Perceived Ease of Use for all modalities (see Table 1), and statistical equivalence between all pairs was determined based on Ease of Use dimension. This finding is consistent with prior research, which shows that immersive training systems can maintain high usability when interface design is intuitive and familiar (Linte et al. 2013; Mendes et al. 2019; Wells et al. 2024).

Regarding workload, measured through the NASA-TLX, most dimensions showed no significant differences. However, the Frustration dimension was significantly higher for the 2D Aid condition compared to No Aid. This may reflect the higher cognitive effort required to interpret 2D projections without spatial context. In contrast, the 2D–3D Aid group, despite receiving more information, did not show increased frustration or significantly higher workload, suggesting that well-integrated 3D aids may help reduce cognitive effort through intuitive spatial cues (Lin et al. 2021; Wenk et al. 2023). Prior work indicates that fragmented or planar information imposes higher processing demands, whereas adding volumetric visual cues makes spatial relationships more readily interpretable (Sweller 2011; Krüger et al. 2022; Wenk et al. 2023). This may help explain why the 2D–3D Aid interface did not lead to increased workload despite offering richer visual content.

RQ2: How do users' procedure precision and execution time differ when comparing the No Aid, 2D Aid, and 2D-3D Aid groups in EVD placement during the training phase? During the training phase, both catheter placement precision and execution time varied significantly depending on the visual support modality (see Tables 3, 4, 5). Participants trained with 2D and 2D–3D visual aids achieved significant higher accuracy compared to the No Aid group (see Table 3). The highest accuracy was reached by the 2D-3D Aid group. The ANCOVA test controlling for the number of attempts indicated that the improvements were attributable to the training modalities, rather than practice effects. We point out that although the 2D Aid group achieved low error rates during training, it reported significant high levels of Frustration on the NASA-TLX scale, unlike the 2D–3D Aid group. These results are consistent with prior studies demonstrating that visual overlays, particularly when combining 2D imaging with 3D spatial cues, can improve performance in image-guided neurosurgical procedures (Alizadeh et al. 2024; Buwaider et al. 2024). Angular accuracy also improved, with the 2D–3D Aid group participants having significantly lower Tilt and Rotation errors, suggesting that the 3D trajectory guidance facilitated mental rotation and alignment, a benefit of immersive training systems for procedural tasks (Butaslac et al. 2022; Linte et al. 2013). However, the execution time more than doubled in the aided groups. This reflects the well-known trade-off between performance accuracy and task efficiency observed in high-fidelity simulations, and not only for procedural tasks (Buwaider et al. 2024; Daling and Schlittmeier 2024).

RQ3: What is the impact of the No Aid, 2D Aid, and 2D-3D learning modalities for EVD placement on users' skill retention, measured by procedure accuracy and execution time, once these aids are no longer provided? The testing phase was specifically designed to evaluate short-term skill retention by removing all MR-based visual aids and requiring participants to perform EVD placement using only physical instruments. This unaided execution represents a key novelty of our evaluation protocol and distinguishes our study from the majority of prior AR/MR surgical training literature, which typically assesses performance only under guided or partially supported conditions. The results show that prior exposure to training, particularly when enhanced with combined 2D and 3D guidance, led to a substantial improvement in unaided performance (see Table 3). Participants trained with 2D–3D Aid retained the lowest error in terms of Euclidean Distance during the Testing Phase, outperforming also the Control group. A significant difference was observed along the axial (Y) axis between the 2D–3D Aid group and the No Aid, 2D Aid, and Control groups. Interestingly, from visual inspection we observe a noticeable shift towards positive Y values in the No Aid, 2D Aid, and especially the Control groups (see , Fig. 5). In contrast, the 2D–3D Aid group demonstrated more precise performance, with operation points more evenly distributed around the target.

These results suggest that integrating 2D and 3D spatial guidance during training enhances procedural memory, enabling users to mentally reconstruct accurate trajectories even without external aids. Prior work shows that combining planar and volumetric cues supports the formation of more robust spatial representations, reducing the mental effort required to reproduce complex movements (Castro-Alonso et al. 2019; Krüger et al. 2022). This may explain why the 2D–3D Aid group retained higher accuracy and showed less directional bias during unaided testing, consistent with evidence that immersive spatial cues can strengthen motor skill transfer (Butaslac et al. 2022; Daling and Schlittmeier 2024; Lin et al. 2021).

Moreover, an ANCOVA controlling for the number of training attempts confirmed a significant effect of training modality on retention performance, emphasizing that the quality, not just the quantity, of training matters (Sels 2002). Interestingly, although the 2D–3D Aid group had the best accuracy, it did not translate to faster execution. All trained groups were slower than the Control group during unaided testing. This may reflect a more cautious, deliberate execution strategy, commonly observed after high-fidelity training, where accuracy is prioritized over speed (Batmaz et al. 2016). Finally, the ANCOVA test suggested that while different training modalities significantly impacted skill retention in terms of accuracy, especially when using advanced combined 2D and 3D visual aids, speed relies more on repeated practice. We assume this is due to the fact that participants were not encouraged to prioritize speed during the training phase, but rather to focus on accuracy. Indeed, all the feedback provided during training related solely to placement precision, with no emphasis on execution time. Even when using the 2D–3D Aid, which offered 3D visualization of the trajectory, participants still required a relatively long time on average (Vékony et al. 2022).

6 Limitations and future works

This study demonstrated the effectiveness of the NeuroMix system for EVD placement training, however we acknowledge some limitations. The evaluation focused exclusively on short-term skill retention, as the Testing Phase was conducted immediately following the Training Phase. Future work will aim to investigate long-term retention effects to better understand how procedural skills evolve.

Although the Testing Phase involved a physical catheter and skull phantom, it lacked essential aspects of clinical realism, such as dynamic tissue deformation, bleeding, and the pressures of time-sensitive decision-making. These factors are known to influence user behavior, stress levels, and performance in real-world settings (Zhou et al. 2024).

Moreover, the Training Phase was conducted using a single digital skull and a single patient-specific CT dataset, which limited anatomical variability and may have reduced the diversity of training scenarios. The similarity between the virtual model and the physical skull used in testing could also have introduced a degree of visual familiarity, potentially contributing to the observed accuracy. To mitigate this limitation, future work will incorporate multiple clinical cases and CT datasets, as well as 3D-printed anatomical models, to increase variability, enhance realism, and better assess generalization and skill retention.

Additionally, the tracking system used in this study was based on controller-mounted sensors, which, while being practical and efficient, may not reach the precision levels of optical navigation systems used in operating rooms. Similar trade-offs have been reported in other medical XR applications (Dastan et al. 2024). The choice of employing Meta Quest 3 controllers for spatial tracking, rather than marker-based or fiducial-based optical systems, was primarily driven by the need for portability, low setup complexity, and full system autonomy. As Meta has recently enabled access to its headset cameras, we plan to upgrade the system to incorporate advanced computer vision algorithms for markerless tracking. Regarding tracking accuracy, the system relied on controller-mounted sensors available on the Meta Quest 3. While this approach does not reach sub-millimeter precision typical of optical tracking systems used in intraoperative settings, it was sufficient for our study objectives. Our goal is not to provide intraoperative guidance but to support early-stage procedural training. The same tracking setup was used across all experimental conditions, ensuring consistent measurement of relative performance.

Preliminary testing indicated that the average positional drift remained below approximately 5 mm during stable placement, which was acceptable for the purposes of training and group comparison. Future versions will integrate markerless tracking through headset cameras, which Meta has recently made available, enabling computer vision-based improvements in precision and usability.

7 Conclusions

This study introduced NeuroMix, a Mixed Reality simulator for EVD placement, designed to assess how different visual aids affect procedural performance and short-term skill retention. In a controlled study with 48 participants, those trained with combined 2D–3D visual aids showed the most significant improvement in precision, including reduced tilt and rotation errors, during unaided testing compared to the Control group. The statistical analysis confirmed that training modality significantly influenced retention, highlighting

that training quality matters more than quantity. All three modalities were rated highly in terms of usability, usefulness, and ease of use. Notably, the 2D–3D Aid did not increase cognitive workload, suggesting that rich spatial cues can be well-tolerated when aligned with user familiarity. The execution time was longer with visual aids compared to the Control group. However, the statistical analysis indicated that execution speed improved mainly through repetition, not modality, showing that operational proficiency benefits from practice regardless of the aid used. Beyond these experimental findings, the study also highlights practical considerations for real-world deployment. NeuroMix demonstrates how modern consumer-grade MR devices can reduce the cost and complexity typically associated with immersive medical training. Its standalone architecture allows flexible deployment across educational and clinical settings without requiring external tracking systems or dedicated computing hardware, supporting scalability and facilitating broader adoption of MR-based procedural training.

Author Contributions P.C. and A.L. conceived the study design and coordinated the overall project. P.C., A.L., and A.D.P. developed the NeuroMix system architecture and implemented the MR interface. L.Z. and M.M. contributed to the clinical modeling and selection of neurosurgical procedures. S.H., D.G. and R.B. supported the technical design and experimental validation of the hardware setup. P.C. and A.L. conducted the statistical analysis. A.L. and P.C. wrote the main manuscript text. S.H., D.G., and R.B. contributed to writing and reviewing the manuscript. G.M. supervised the study and contributed to writing and critical revision of the manuscript. All authors reviewed and approved the final version of the manuscript.

Funding Open access funding provided by Alma Mater Studiorum - Università di Bologna within the CRUI-CARE Agreement. This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Data Availability The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no Conflict of interest.

Ethical approval This research was approved by the University of Bologna Ethics Committee (Approval number 0159749) on June 23, 2022. The study adhered to all ethical guidelines, and informed consent was obtained from all participants prior to their involvement.

References

Giraldo JP, Cho SS, Eghrari NB, Dholaria N, Farber SH, Ehredt RB, Michaels C, Fotias DJ, Godzik J, Sonntag VK (2025) Advances in neurosurgical education: literature review of mixed-reality simulation models and novel mixed-reality spine prototype. J Neurosurg Spine 1(AOP), 1–14

James H (2020) Use of cadavers to train surgeons: what are the ethical issues? J Med Ethics 46(7):470–471

Colombo E, Regli L, Esposito G, Germans MR, Fierstra J, Serra C, Sebök M, Van Doormaal T (2023) Mixed reality for cranial neurosurgical planning: a single-center applicability study with the first 107 subsequent holograms. Oper Neurosurg. https://doi.org/10.1227/ons.0000000000001033

Kazemzadeh K, Akhlaghdoust M, Zali A (2023) Advances in artificial intelligence, robotics, augmented and virtual reality in neurosurgery. Front Surg 10:1241923. https://doi.org/10.3389/fsurg.2023.1241923

Isikay I, Cekic E, Baylarov B, Tunc O, Hanalioglu S (2024) Narrative review of patient-specific 3D visualization and reality technologies in skull base neurosurgery: enhancements in surgical training, planning, and navigation. Front Surg 11:1427844. https://doi.org/10.3389/fsurg.2024.1427844

Lungu AJ, Swinkels W, Claesen L, Tu P, Egger J, Chen X (2021) A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. Expert Rev Med Devices 18(1):47–62

Anatomy and Physiology (2023) In: Greenberg MS (ed) Greenberg's Handbook of Neurosurgery, 10th edn. Thieme https://doi.org/10.1055/b000000751https://medone-neurosurgery.thieme.com/ebooks/cs_20449427#/ebook_cs_20449427_cs28259 Accessed 2025-04-03

Flint AC, Rao VA, Renda NC, Faigeles BS, Lasman TE, Sheridan W (2013) A simple protocol to prevent external ventricular drain infections. Neurosurgery 72(6):993–999

Alizadeh M, Xiao Y, Kersten-Oertel M (2024) Virtual and augmented reality in ventriculostomy: a systematic review. World Neurosurg 189:90–107. https://doi.org/10.1016/j.wneu.2024.05.151

Gumprecht HK, Widenka DC, Lumenta CB (1999) Brainlab vectorvision neuronavigation system: technology and clinical experiences in 131 cases. Neurosurgery 44(1):97–104

AlAzri A, Mok K, Chankowsky J, Mullah M, Marcoux J (2017) Placement accuracy of external ventricular drain when comparing freehand insertion to neuronavigation guidance in severe traumatic brain injury. Acta Neurochir 159:1399–1411

Kaplan AD, Cruit J, Endsley M, Beers SM, Sawyer BD, Hancock PA (2021) The effects of virtual reality, augmented reality, and mixed reality as training enhancement methods: a meta-analysis. Hum Factors 63(4):706–726

Daling LM, Schlittmeier SJ (2024) Effects of augmented reality-, virtual reality-, and mixed reality-based training on objective performance measures and subjective evaluations in manual assembly tasks: a scoping review. Hum Factors 66(2):589–626

Borsci S, Lawson G, Broome S (2015) Empirical evidence, evaluation criteria and challenges for the effectiveness of virtual and mixed reality tools for training operators of car service maintenance. Comput Ind 67:17–26

Agrawala M, Phan D, Heiser J, Haymaker J, Klingner J, Hanrahan P, Tversky B (2003) Designing effective step-by-step assembly instructions. ACM Trans Graph (TOG) 22(3):828–837

Katsioloudis P, Jovanovic V, Jones M (2014) A comparative analysis of spatial visualization ability and drafting models for industrial and technology education students. J Technol Educ 26(1)

Durrani S, Onyedimma C, Jarrah R, Bhatti A, Nathani KR, Bhandarkar AR, Mualem W, Ghaith AK, Zamanian C, Michalopoulos GD, Alexander AY, Jean W, Bydon M (2022) The virtual vision of neurosurgery: how augmented reality and virtual reality are transforming the neurosurgical operating room. World Neurosurg 168:190–201. https://doi.org/10.1016/j.wneu.2022.10.002

Krabbe-Hartkamp MJ, Grond J, De Leeuw FE, De Groot JC, Algra A, Hillen B, Breteler M, Mali W (1998) Circle of Willis: morphologic variation on three-dimensional time-of-flight MR angiograms. Radiology 207(1):103–111

Sidhu R, Tompa D, Jang R, Grober E, Johnston K, Reznick R, Hamstra S (2004) Interpretation of three-dimensional structure from two-dimensional endovascular images: implications for educators in vascular surgery. J Vasc Surg 39(6):1305–1311

Yohannan DG, Oommen AM, Kumar AS, Devanand S, Ut MR, Sajan N, Thomas NE, Anzer N, Raju NK, Thomas B (2024) "visualization matters"-stereoscopic visualization of 3d graphic neuroanatomic models through ANAVU enhances basic recall and radiologic anatomy learning when compared with monoscopy. BMC Med Educ 24(1):932

Lin T, Yang Y, Beyer J, Pfister H (2021) Labeling out-of-view objects in immersive analytics to support situated visual searching. IEEE Trans Visual Comput Graphics 29(3):1831–1844

Hajahmadi S, Stacchio L, Giacché A, Cascarano P, Marfia G (2024) Investigating extended reality-powered digital twins for sequential instruction learning: the case of the Rubik's cube. In: 2024 IEEE international symposium on mixed and augmented reality (ISMAR). IEEE, pp 259–268

Krüger JM, Palzer K, Bodemer D (2022) Learning with augmented reality: impact of dimensionality and spatial abilities. Comput Educ Open 3:100065

Tang Y-M, Au KM, Lau HC, Ho GT, Wu C-H (2020) Evaluating the effectiveness of learning design with mixed reality (MR) in higher education. Virtual Reality 24(4):797–807

Buwaider A, El-Hajj VG, Iop A, Romero M, Jean CW, Edström E, Elmi-Terander A (2024) Augmented reality navigation in external ventricular drain insertion-a systematic review and meta-analysis. Virtual Reality 28(3):141. https://doi.org/10.1007/s10055-024-01033-9

Van Gestel F, Frantz T, Vannerom C, Verhellen A, Gallagher AG, Elprama SA, Jacobs A, Buyl R, Bruneau M, Jansen B, Vandemeulebroucke J, Scheerlinck T, Duerinck J (2021) The effect of augmented reality on the accuracy and learning curve of external ventricular drain placement. Neurosurg Focus 51(2):8. https://doi.org/10.3171/2021.5.FOCUS21215

Eom S, Ma TS, Vutakuri N, Hu T, Haskell-Mendoza AP, Sykes DAW, Gorlatova M, Jackson J (2024) Accuracy of routine external ventricular drain placement following a mixed reality-guided twist-drill craniostomy. Neurosurg Focus 56(1):11. https://doi.org/10.3171/2023.10.FOCUS23615

Li Y, Chen X, Wang N, Zhang W, Li D, Zhang L, Qu X, Cheng W, Xu Y, Chen W, Yang Q (2019) A wearable mixed-reality holographic computer for guiding external ventricular drain insertion at the bedside. J Neurosurg 131(5):1599–1606. https://doi.org/10.3171/2018.4.JNS18124

Grunert R, Winkler D, Wach J, Kropla F, Scholz S, Vychopen M, Güresir E (2024) IMAGINER: improving accuracy with a mixed reality navigation system during placement of external ventricular drains. A feasibility study Neurosurg Focus 56(1):8. https://doi.org/10.3171/2023.10.FOCUS23554

Vychopen M, Kropla F, Winkler D, Güresir E, Grunert R, Wach J (2025) IMAGINER 2-improving accuracy with augmented reality navigation system during placement of external ventricular drains over Kaufman's, Keen's, Kocher's and Frazier's point. Front Surg 11:1513899. https://doi.org/10.3389/fsurg.2024.1513899

Eom S, Kim S, Jackson J, Sykes D, Rahimpour S, Gorlatova M (2024) Augmented reality-based contextual guidance through surgical tool tracking in neurosurgery. IEEE Trans Visual Comput Graph. https://doi.org/10.1109/TVCG.2024.3390680

Yudkowsky R, Luciano C, Banerjee P, Schwartz A, Alaraj A, Lemole GM, Charbel F, Smith K, Rizzi S, Byrne R, Bendok B, Frim D (2013) Practice on an augmented reality/haptic simulator and library of virtual brains improves residents' ability to perform a ventriculostomy. Simul Healthc J Soc Simul Healthc 8(1):25–31. https://doi.org/10.1097/SIH.0b013e3182662c69

Hooten KG, Lister JR, Lombard G, Lizdas DE, Lampotang S, Rajon DA, Bova F, Murad GJA (2014) Mixed reality ventriculostomy simulation: experience in neurosurgical residency. Oper Neurosurg 10(4):576–581. https://doi.org/10.1227/NEU.0000000000000503

Schneider M, Kunz C, Wirtz CR, Mathis-Ullrich F, Pala A, Hlavac M (2023) Augmented reality-assisted versus freehand ventriculostomy in a head model. J Neurol Surg Part A Cent Eur Neurosurg 84(06):562–569

Dastan M, Fiorentino M, Walter ED, Diegritz C, Uva AE, Eck U, Navab N (2024) Co-designing dynamic mixed reality drill positioning widgets: a collaborative approach with dentists in a realistic setup. IEEE Trans Vis Comput Graph

Domínguez-Velasco CF, Tello-Mata IE, Guinto-Nishimura G, Martínez-Hernández A, Alcocer-Barradas V, Pérez-Lomelí JS, Padilla-Castañeda MA (2023) Augmented reality simulation as training model of ventricular puncture: Evidence in the improvement of the quality of punctures. Int J Med Robot Comput Assisted Surg 19(5):2529. https://doi.org/10.1002/rcs.2529

Benmahdjoub M, Thabit A, Van Veelen M-LC, Niessen WJ, Wolvius EB, Walsum TV (2023) Evaluation of AR visualization approaches for catheter insertion into the ventricle cavity. IEEE Trans Visual Comput Graphics 29(5):2434–2445. https://doi.org/10.1109/TVCG.2023.3247042

Skyrman S, Lai M, Edström E, Burström G, Förander P, Homan R, Kor F, Holthuizen R, Hendriks BHW, Persson O, Elmi-Terander A (2021) Augmented reality navigation for cranial biopsy and external ventricular drain insertion. Neurosurg Focus 51(2):7. https://doi.org/10.3171/2021.5.FOCUS20813

Eom S, Ma T, Vutakuri N, Du A, Qu Z, Jackson J, Gorlatova M (2024) Did you do well? real-time personalized feedback on catheter placement in augmented reality-assisted neurosurgical training. In: 2024 IEEE conference on virtual reality and 3D user interfaces abstracts and workshops (VRW). IEEE, pp 1208–1209

Umana GE, Scalia G, Yagmurlu K, Mineo R, Di Bella S, Giunta M, Spitaleri A, Maugeri R, Graziano F, Fricia M (2021) Multimodal simulation of a novel device for a safe and effective external ventricular drain placement. Front Neurosci 15:690705

Schneider M, Kunz C, Pal'a A, Wirtz CR, Mathis-Ullrich F, Hlaváč M (2021) Augmented reality-assisted ventriculostomy. Neurosurg Focus 50(1):16

Robertson FC, Raahil MS, Amich JM, Essayed WI, Lal A, Lee BH, Prieto PC, Tokuda J, Weaver JC, Kirollos RW (2021) Frameless neuronavigation with computer vision and real-time tracking for bedside external ventricular drain placement: a cadaveric study. J Neurosurg 136(5):1475–1484

Meta: Meta SDK for Unity: Hand Tracking Documentation. (2023). Retrieved from https://developer.meta.com

Salomoni P, Prandi C, Roccetti M, Casanova L, Marchetti L, Marfia G (2017) Diegetic user interfaces for virtual environments with

HMDS: a user experience study with oculus rift. J Multimodal User Interfaces 11:173–184
Pieper S, Halle M, Kikinis R (2004) 3d slicer. In: 2004 2nd IEEE international symposium on biomedical imaging: nano to macro (IEEE Cat No. 04EX821), IEEE, pp 632–635
Mullen T (2011) Mastering blender. John Wiley & Sons, New Jersey
Mendes D, Caputo FM, Giachetti A, Ferreira A, Jorge J (2019) A survey on 3d virtual object manipulation: from the desktop to immersive virtual environments. In: Computer graphics forum, vol 38, Wiley Online Library, pp 21–45
Van Doormaal JAM, Fick T, Boskovic E, Hoving EW, Robe PAJT, Van Doormaal TPC (2025) Development and validation of a neurosurgical phantom for simulating external ventricular drain placement. J Med Syst 49(1):1. https://doi.org/10.1007/s10916-024-02133-4
Jeffri NFS, Rambli DRA (2021) A review of augmented reality systems and their effects on mental workload and task performance. Heliyon 7(3)
Guo H-J, Bakdash JZ, Marusich LR, Prabhakaran B (2022) Augmented reality and mixed reality measurement under different environments: a survey on head-mounted devices. IEEE Trans Instrum Meas 71:1–15
Butaslac IIM, Fujimoto Y, Sawabe T, Kanbara M, Kato H (2022) Systematic review of augmented reality training systems. IEEE Trans Visual Comput Graphics 29(12):5062–5082
Seymour NE, Gallagher AG, Roman SA, O'Brien MK, Bansal VK, Andersen DK, Satava RM (2002) Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Ann Surg 236(4):458–464
Lewis JR, Sauro J (2009) The factor structure of the system usability scale. In: Human centered design: first international conference, HCD 2009, Held as Part of HCI International 2009, San Diego, CA, USA, July 19–24, 2009 Proceedings 1, Springer, pp 94–103
Hart SG, Staveland LE (1988) Development of NASA-TLX (task load index): results of empirical and theoretical research. In: Hancock PA, Meshkati N (eds) Human mental workload. Advances in psychology, vol 52. North-Holland, pp. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9https://www.sciencedirect.com/science/article/pii/S0166411508623869
Davis FD (1989) Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS quarterly, pp 319–340
Razali NM, Wah YB (2011) Power comparisons of shapiro-wilk, kolmogorov-smirnov, lilliefors and anderson-darling tests. J Stat Model Anal 2(1):21–33
Lakens D (2013) Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and anovas. Front Psychol 4:863
Lakens D (2017) Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. Soc Psychol Personal Sci 8(4):355–362
McKight PE, Najab J (2010) Kruskal-Wallis test. Corsini Encycl Psychol 1–1
McKnight PE, Najab J (2010) Mann-Whitney u test. Corsini Encycl Psychol 1–1
Van Breukelen GJ (2013) ANCOVA versus change from baseline in nonrandomized studies: the difference. Multivar Behav Res 48(6):895–922
De Winter JC, Gosling SD, Potter J (2016) Comparing the Pearson and spearman correlation coefficients across distributions and sample sizes: a tutorial using simulations and empirical data. Psychol Methods 21(3):273
Bangor A, Kortum P, Miller J (2009) Determining what individual SUS scores mean: adding an adjective rating scale. J Usability Stud 4(3):114–123
Linte CA, Davenport KP, Cleary K, Peters C, Vosburgh KG, Navab N, Jannin P, Peters TM, Holmes DR III, Robb RA (2013) On mixed reality environments for minimally invasive therapy guidance: systems architecture, successes and challenges in their implementation from laboratory to clinic. Comput Med Imaging Graph 37(2):83–97
Wells B, Wang N, Maurer F (2024) Technical challenges in building cross reality applications for analyzing 3d medical images. In: 2024 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct). IEEE, pp 152–155
Wenk N, Penalver-Andres J, Buetler KA, Nef T, Müri RM, Marchal-Crespo L (2023) Effect of immersive visualization technologies on cognitive load, motivation, usability, and embodiment. Virtual Reality 27(1):307–331
Sweller J (2011) Cognitive load theory. Psychology of learning and motivation, vol 55. Elsevier, New Jersey, pp 37–76
Castro-Alonso JC, Ayres P, Sweller J (2019) Instructional visualizations, cognitive load theory, and visuospatial processing. Visuospatial processing for education in health and natural sciences. Springer, Berlin, pp 111–143
Sels L (2002) "More is not necessarily better": the relationship between the quantity and quality of training efforts. Int J Hum Resour Manag 13(8):1279–1298
Batmaz AU, Mathelin M, Dresp-Langley B (2016) Getting nowhere fast: Trade-off between speed and precision in training to execute image-guided hand-tool movements. BMC Psychol 4:1–19
Vékony T, Pleche C, Pesthy O, Janacsek K, Nemeth D (2022) Speed and accuracy instructions affect two aspects of skill learning differently. npj Sci Learn 7(1):27
Zhou Y-B, Ruan S-J, Zhang K, Bao Q, Liu H-Z (2024) Time pressure effects on decision-making in intertemporal loss scenarios: an eye-tracking study. Front Psychol 15:1451674