Engagement is not encoding: Outcome alignment determines the effectiveness of mixed reality in medical education

Nadia Karnatz, David Kaffka, Angela Rosa Caso, Michael Schwerter, Shufang Liu, Raymond Wong, Aida Parviz, Majeed Rana

Journal of Cranio-Maxillo-Facial Surgery 54 (2026) 104618

Contents lists available at ScienceDirect

Journal of Cranio-Maxillo-Facial Surgery

journal homepage: www.jcmfs.com

Engagement is not encoding: Outcome alignment determines the effectiveness of mixed reality in medical education

Nadia Karnatz [a,f,*], David Kaffka [b,g], Angela Rosa Caso [c], Michael Schwerter [d,h], Shufang Liu [d,i], Raymond Wong [e], Aida Parviz [a,j], Majeed Rana [c,e,k]

[a] Department of Oral and Plastic Maxillofacial Surgery, Heinrich Heine University Hospital Düsseldorf, Moorenstraße 5, Düsseldorf, 40225, Germany
[b] Medical Faculty, Heinrich Heine University Düsseldorf, Düsseldorf, Germany
[c] Department of Oral and Plastic Maxillofacial Surgery, Albert Einstein University Hospital Ulm, Albert-Einstein-Allee 11, Ulm, 89081, Germany
[d] Brainlab AG, Olof-Palme-Str. 9, München, 81829, Germany
[e] Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National University Singapore, 9, Lower Kent Ridge Road, 119085, Singapore
[f] Department of Otorhinolaryngology, Head and Neck Surgery, Hospital Mutterhaus der Borromäerinnen, Feldstraße 16, Trier, 54290, Germany
[g] CASA DENTAL GbR, Ahrstraße 37, Wesseling, 50389, Germany
[h] KIOTERA GmbH, Speditionsstraße 15a, Düsseldorf, 40221, Germany
[i] Miltenyi Biotec, Friedrich-Ebert-Straße 68, Bergisch Gladbach, 51429, Germany
[j] Department of Oral and Maxillofacial Surgery, Helios St. Josefshospital Uerdingen, Kurfürstenstraße 69, Krefeld, 47829, Germany
[k] Department of Oral and Maxillofacial Surgery, Faculty of Dentistry, National University of Singapore, 9, Lower Kent Ridge Road, 119085, Singapore

A R T I C L E I N F O

Keywords:
Mixed reality
Medical education
Learning outcomes
Perceptions
Randomised controlled trial

A B S T R A C T

Mixed Reality (MR) technologies are increasingly discussed as a resource-efficient option to support high-quality teaching in medical education. Yet the literature is still dominated by feasibility studies and systematic reviews; controlled comparative studies remain less common. The study aimed to assess whether mixed reality provides a learning benefit in a single, time-limited teaching session and to characterise students' perceptions and how these change from pre-to post-session. An exploratory questionnaire was developed in advance based on semi-structured, guide–based interviews ($n = 15$). Subsequently, 46 medical and dental students were randomised to either an intervention group (mixed reality; $n = 23$) or a control group (script-based conventional teaching; $n = 23$). After the respective learning unit (a clinically oriented anatomy session), a knowledge recall test was administered. Subjective ratings in the domains of *prior knowledge*, *educational relevance*, *visualisation*, *understanding*, *interaction*, *technical aspects* and *critical feedback* were collected using pre- and post-questionnaires to capture changes in perception and the extent of opinion change. In the between-group comparison, no consistent differences were observed pre- or post-intervention, nor with regard to the extent of opinion change. Within-group pre–post analyses assessed perception change. In the control group, significant pre–post changes were found in 6 of 7 domains (prior knowledge $p < 0.001$; educational relevance $p < 0.001$; visualisation $p < 0.001$; understanding $p < 0.001$; interaction $p = 0.023$; technical aspects $p = 0.004$). In contrast, in the critical feedback domain, a significant difference was observed only in the intervention group ($p = 0.029$). In the knowledge recall test, the control group performed significantly better ($p = 0.008$).

These findings suggest that immersive technologies do not automatically enhance learning performance. Rather, their effectiveness appears to depend on the alignment between medium characteristics, cognitive demands, and the specific learning outcomes assessed. Engagement and interactivity alone may not translate into improved short-term recall, particularly when outcomes primarily measure factual retrieval.

* Corresponding author. Department of Oral and Plastic Maxillofacial Surgery, Heinrich Heine University Hospital Düsseldorf, Moorenstraße 5, 40225, Düsseldorf, Germany.

E-mail addresses: nadia.karnatz@mutterhaus.de (N. Karnatz), david.kaffka@hhu.de (D. Kaffka), angelarosacaso@gmail.com (A.R. Caso), m.schwerter@kiotera.de (M. Schwerter), shufangl@miltenyi.com (S. Liu), raginwcw@yahoo.com (R. Wong), aida.parviz@helios-gesundheit.de (A. Parviz), majeed.rana@uni-ulm.de (M. Rana).

https://doi.org/10.1016/j.jcms.2026.104618
Received 22 March 2026; Accepted 22 May 2026
Available online 17 June 2026

1. Introduction

Medical education is increasingly influenced by digital technologies. Their use has become an integral component of current curricula (Brann and Sloop 2006; Chen et al. 2019; Barthel et al. 2025). Digital teaching formats may be particularly useful when complex content is being taught in medical and surgical education (Bova et al. 2013). At the same time, there is a growing expectation that new technologies should not only be implementable, but also provide added value in a teaching context (Han et al. 2019; Hernon et al. 2023). In this context, extended reality (XR), such as virtual reality (VR), augmented reality (AR), and mixed reality (MR), has established itself as a promising field of application.

Mixed reality enables the combination of real learning environments with virtual enhancements. MR appears to be a potentially suitable teaching tool, especially for spatially complex content or the simulation of practical applications (Curran et al. 2023; Leung and Shi 2024). In addition, MR can support communication between teachers and students through interaction in a shared mixed-reality environment. Accordingly, MR has received increasing attention in medical education in recent years, both in student teaching and in skills training and case-based learning (Dhar et al. 2021; Silvero Isidre et al. 2023; Daling and Schlittmeier 2024). From a cognitive perspective, immersive environments may impose additional extraneous load due to navigation, interaction, and device management. According to Cognitive Load Theory, when working memory resources are allocated to interface handling rather than content processing, learning efficiency may decrease. Thus, immersive technologies may enhance certain forms of learning (e.g., spatial integration), while simultaneously impairing short-term factual encoding under time-limited conditions (Dhar et al. 2021; Silvero Isidre et al. 2023; Daling and Schlittmeier 2024).

However, the introduction of new teaching technologies is often accompanied by high expectations in terms of motivation, comprehensibility, interactivity, and degree of innovation. This is offset by the high implementation costs required to enable reliable use in everyday curricular practice (Uymaz and Uymaz 2022; Toni et al. 2024). The technical feasibility and development of MR-supported simulations have been discussed extensively in the literature (Rochlen et al. 2022; Huang et al. 2023; Suresh et al. 2023). Systematic didactic evaluations of teaching conditions using XR have increased in recent years (Dhar et al. 2021; Ryan et al. 2022). Although individual studies report advantages in terms of spatial ability and learning experience, for example in anatomy lessons, it is unclear whether positive perceptions in short-term teaching situations are accompanied by measurable short-term learning gains or whether effects are primarily attributable to interactivity, immersion, and novelty (Bork 2018). From the perspective of students in particular, data on students' perceptions of MR-supported teaching situations are available, but direct comparisons with conventional teaching formats under authentic curricular conditions are rare (Baratz et al. 2022). Against this background, the present randomised study compares an MR-supported learning unit with a paper-based script covering the same content. The influence of the presentation format was evaluated using pre- and post-questionnaires assessing subjective perceptions and a knowledge recall test assessing short-term learning success. The study aimed to investigate whether mixed reality offers a learning advantage beyond positive subjective evaluations in a one-time, time-limited setting, or whether potential effects are primarily reflected in perceptions and interaction.

2. Materials and methods

2.1. Study design

The study included qualitative preliminary work to develop study-specific questionnaires and a randomised main study to investigate perceptions and short-term knowledge recall in the context of mixed reality-supported teaching and to measure learning success (Fig. 1). In the main study, standardised pre- and post-questionnaires were administered before and after a mixed reality-supported or conventional learning unit, respectively. In addition, a knowledge recall test was carried out to assess learning success.

2.2. Questionnaire design

An exploratory questionnaire was developed to record study-specific aspects of the use of mixed reality in a teaching context. The items were generated on the basis of semi-structured, guide-based interviews with 15 students of (dental) medicine in the clinical stage of their studies and with surgeons. The aim of the qualitative preliminary work was to identify relevant content dimensions in connection with the use of mixed reality in teaching. Based on the qualitative analysis, seven main thematic categories were identified (prior knowledge, educational relevance, visualisation, understanding, interaction, technical aspects and critical feedback), which served as the conceptual basis for the development of the questionnaire items. Three to five closed items were formulated for each category. The items were designed as statements with a five-point Likert scale. Lower values corresponded to higher agreement (1 = true; 5 = not true). The critical feedback category contains negatively worded items, so higher values here correspond to lower agreement with critical statements. In addition, the post-questionnaire contained three open-ended questions on perceived barriers to introduction, technical problems that arose during the teaching session, and subjectively perceived advantages over conventional learning. The questionnaire was used specifically for the study and for exploratory purposes and was not formally psychometrically validated.

2.3. Participants

Forty-six dental and medical students at University Hospital Düsseldorf in the clinical study phase (5th to 10th semester) participated in the main study. Participants were randomly assigned to either the mixed reality group (intervention group) or the conventionally taught control group. Randomisation was performed by drawing lots (23 red/23 green slips of paper) from pre-mixed, sealed envelopes; the envelopes were handed out without knowledge of their contents.

2.4. Teaching intervention

During the study, the intervention (MR) and control groups received identical learning content, a clinically oriented anatomy unit including annotations. The only difference was in the form of presentation (mixed reality vs. paper-based script) with comparable implementation of the teaching session. For the intervention group, a mixed reality application was developed for this study that provided an interactive 3D model of the viscerocranium and neurocranium with integrated annotations. The skull bones were colour-coded and could be viewed as a three-dimensional model from different perspectives, scaled and moved, and broken down into individual anatomical components. Annotations with bullet-point-style, anatomically relevant information could be shown and hidden (Fig. 2). The mixed reality application was implemented using a head-mounted display (HMD) from Magic Leap (Plantation, Florida, USA). The system comprised a headset and a hand-held controller input; the computing unit was carried externally (Lightpack). Interaction with the virtual model took place via the controller.

The control group used a conventional, paper-based learning script that was based on the MR application in terms of content and visuals. The script included coloured two-dimensional representations of the human skull (ventral and lateral view; partially disassembled and completely disassembled representation), each with and without annotations. The learning units were self-directed in both groups with a maximum duration of 10 min.

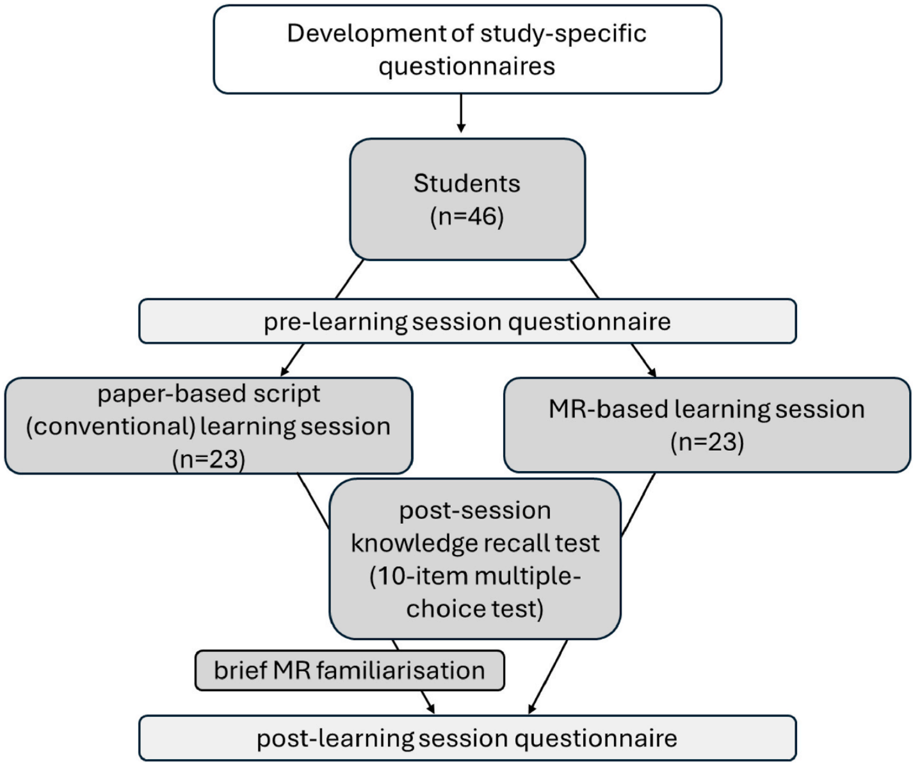

Fig. 1. Study design.

Fig. 2. Mixed reality learning display (head-mounted display).

2.5. Data collection

A pre- and post-questionnaire was used to record students' perceptions of the use of mixed reality in a teaching context. Before the start of each learning unit, all participants completed the pre-questionnaire. An individual anonymised identification number was used to clearly assign the pre- and post-questionnaires. Depending on the group assignment, the mixed reality-supported learning unit or the conventional processing of the learning script was then carried out. After completing the learning unit, a knowledge recall test was carried out after a 10-min break. It comprised ten multiple-choice questions on the ten skull bones covered and the associated annotations. There was exactly one correct answer per question; the items were designed to test only the content of the annotations or the learning script provided, without requiring any further knowledge. Learning success was recorded as the number of questions answered correctly (0–10). Afterwards, the post-questionnaire was completed. To enable the participants in the control group to answer the items in the post-questionnaire on mixed reality use based on their own experience, they were given a brief introduction to the mixed reality application after completing the knowledge recall test and before completing the post-questionnaire. The post-questionnaire also contained three open-ended questions on (i) perceived barriers to introduction, (ii) technical problems encountered during the session, and (iii) subjectively perceived advantages over conventional learning. The total time required per participant was approximately 60 min.

The open-ended responses in the post-questionnaire were coded and categorised based on Mayring's qualitative content analysis. The categories and assignments were checked independently and, in case of discrepancies, clarified by consensus with the aim of providing a structured presentation of recurring themes and additions to the closed items (Mayring 2015).

2.6. Statistics

Statistical analysis was performed using IBM SPSS Statistics (version 29.0.1.0, Armonk, NY, USA). The closed questionnaire items were summarised and averaged according to the predefined categories. Likert responses were treated as ordinal data. Group differences between the intervention and control groups were analysed using the Mann–Whitney U test. Intra-individual changes between pre- and post-questionnaires were evaluated using the Wilcoxon signed-rank test; in addition, change scores were calculated as differences. Knowledge recall test (0–10 correctly answered questions) was compared between groups using the Mann–Whitney U test. The significance level was set at $\alpha = 0.05$. The effect size r was calculated as $r = |Z|/\sqrt{N}$, where N corresponds to the total sample (n_intervention + n_control) in the Mann–Whitney U test and to the number of valid pairs in the Wilcoxon signed-rank test.

2.7. Ethics approval

This study received an ethics approval from the Ethics Committee of the Medical Faculty of Heinrich Heine University Düsseldorf (Study number 2022-1849). This work was funded by the Ministry of Economic Affairs, Innovation, Digitalisation, and Energy of the State of North Rhine-Westphalia (MWIDE NRW) and the 5G.NRW Competence Center under grant number 005-01903-0047. Informed consent was obtained from all participants.

3. Results

3.1. Study population

A total of 46 dental and medical students participated in the study and were randomly assigned to either an intervention (MR) or control group, with $n = 23$ per group. All questionnaires were completed in full. There were no major technical interruptions requiring termination of the MR session.

3.2. Between-group questionnaire results (pre and post)

To analyse differences between the groups, the category values of the pre- and post-questionnaires were compared between the MR and control groups (Table 1). In the main categories prior knowledge, educational relevance, visualisation, understanding and interaction, there were no statistically significant group differences in either the pre- or post-questionnaires. In contrast, significant differences were found in the main categories technical aspects and critical feedback: In the critical feedback category, a significant group difference was found in the pre-questionnaire ($p = 0.039$*), while no significant difference was found in the post-questionnaire ($p = 0.068$). In the technical aspects category, there was no group difference in the pre-questionnaire ($p = 0.911$), but in the post-questionnaire there was a significant difference between the MR and control groups ($p = 0.018$*). Full results are provided in Table S1.

Table 1
Overview of group comparisons between the intervention and control groups in the main categories of the pre- and post-questionnaires (Mann–Whitney U test; Z, p, r values). Significant p values ($p < 0.05$) are marked with *.

Category	Intervention and control group comparison	
	Pre-questionnaire	Post-questionnaire
Prior knowledge	$Z = 0.73$; $p = 0.466$; $r = 0.11$	$Z = -1.30$; $p = 0.192$; $r = 0.19$
Educational relevance	$Z = 1.61$; $p = 0.108$; $r = 0.24$	$Z = -0.37$; $p = 0.711$; $r = 0.05$
Visualisation	$Z = 1.35$; $p = 0.177$; $r = 0.20$	$Z = -0.19$; $p = 0.847$; $r = 0.03$
Understanding	$Z = 1.49$; $p = 0.136$; $r = 0.22$	$Z = -1.66$; $p = 0.098$; $r = 0.24$
Interaction	$Z = 1.84$; $p = 0.065$; $r = 0.27$	$Z = -0.35$; $p = 0.728$; $r = 0.05$
Technical aspects	$Z = 0.11$; $p = 0.911$; $r = 0.02$	$Z = -2.37$; $\mathbf{p = 0.018}$*; $r = 0.35$
Critical feedback	$Z = 2.07$; $\mathbf{p = 0.039}$*; $r = 0.30$	$Z = 1.83$; $p = 0.068$; $r = 0.27$

3.3. Qualitative evaluation of the open questions in the post-questionnaires

The qualitative content analysis of the open answers in the post-questionnaire identified various topics within the subject areas “Barriers to introduction,” “Technical problems,” and “Advantages over conventional learning” (Table 2). Overall, structural (including inadequate digital infrastructure, high costs, limited range of applications), technical (e.g., high system requirements, software errors/crashes, long loading times, display and tracking problems), and individual factors (including limited technical understanding, perception problems, high implementation costs) were described as aggravating factors. In practical use, complex operation, orientation and interaction problems, physical discomfort, and specific barriers to use for people who wear glasses and those who are less tech-savvy were also highlighted.

3.4. Analysis of changes in opinion within the intervention and control groups

To examine changes in opinion within the groups, the category values between the pre- and post-questionnaires were compared separately for the intervention and control groups (Table 3). In the intervention group, there were no statistically significant changes between the pre- and post-questionnaires in the categories prior knowledge, educational relevance, visualisation, understanding, interaction and technical aspects. In the control group, however, the category values fell significantly in all six categories: prior knowledge ($p = 0.001$), educational relevance ($p < 0.001$), visualisation ($p < 0.001$), understanding ($p < 0.001$), interaction ($p = 0.023$), and technical aspects ($p = 0.004$).

Table 2
Overview of topics and aspects based on the open questions in the post-questionnaire in coded form, following Mayring's qualitative content analysis (Mayring 2015).

Topics	Aspects
Barriers to the introduction of MR technology	Digital infrastructure; High technical requirements; Low wearing comfort of the HMD; High implementation costs; Limited range of applications; Display problems; Perception issues (e.g. Motion sickness)
Technical problems	Difficulties in assembling puzzles; Eye discomfort; Orientation problems in the MR environment; Confusing user guidance; Comfort; Compatibility problems with visual aids; Operating problems; Interaction problems
Advantages compared to conventional learning	Improved visualisation; Better understanding; Authentic presentation; Interaction; Motivation

Table 3
Within-group comparison (pre vs. post) of category values in the intervention and control groups. Lower scores indicate higher agreement; for critical feedback, higher scores indicate lower agreement with critical statements. Wilcoxon signed-rank test results are reported as Z, p, and r values; significant p values ($p < 0.05$) are marked with *.

Category	Intervention group pre vs. post	Control group pre vs. post
Prior knowledge	Z = −1.54; p = 0.128; r = 0.32	Z = −3.23; **p = 0.001***; r = 0.67
Educational relevance	Z = −1.52; p = 0.129; r = 0.32	Z = −3.58; **p < 0.001***; r = 0.75
Visualisation	Z = −1.65; p = 0.100; r = 0.34	Z = −3.55; **p < 0.001***; r = 0.74
Understanding	Z = −0.31; p = 0.754; r = 0.07	Z = −3.46; **p < 0.001***; r = 0.72
Interaction	Z = −0.36; p = 0.716; r = 0.08	Z = −2.28; **p = 0.023***; r = 0.48
Technical aspects	Z = −1.52; p = 0.129; r = 0.32	Z = −2.89; **p = 0.004***; r = 0.60
Critical feedback	Z = 2.19; **p = 0.029***; r = 0.46	Z = 1.72; p = 0.086; r = 0.36

This corresponds to a more positive assessment. In the critical feedback category, the category values increased significantly in the intervention group ($p = 0.029$), which means a decrease in critical assessment, while no significant change was detectable in the control group ($p = 0.086$). Full results are provided in Table S2.

3.5. Analysis of the extent of opinion shift between the groups

To quantify the extent of the change in opinion, the difference in category values between the pre- and post-questionnaires (Δ = pre – post) was calculated and compared for each main category. Positive differences correspond to increased agreement; for the critical feedback category, the opposite applies due to negatively worded items. In the categories prior knowledge, educational relevance, visualisation, interaction, technical aspects and critical feedback, there were no statistically significant differences in the extent of change between the groups. In the category understanding, there was a significant group difference in the extent of change ($p = 0.004$) with larger positive difference values in the control group (Fig. 3). Full results are provided in Table S3.

3.6. Statistical evaluation of the knowledge recall test

In the knowledge recall test, the control group achieved an average

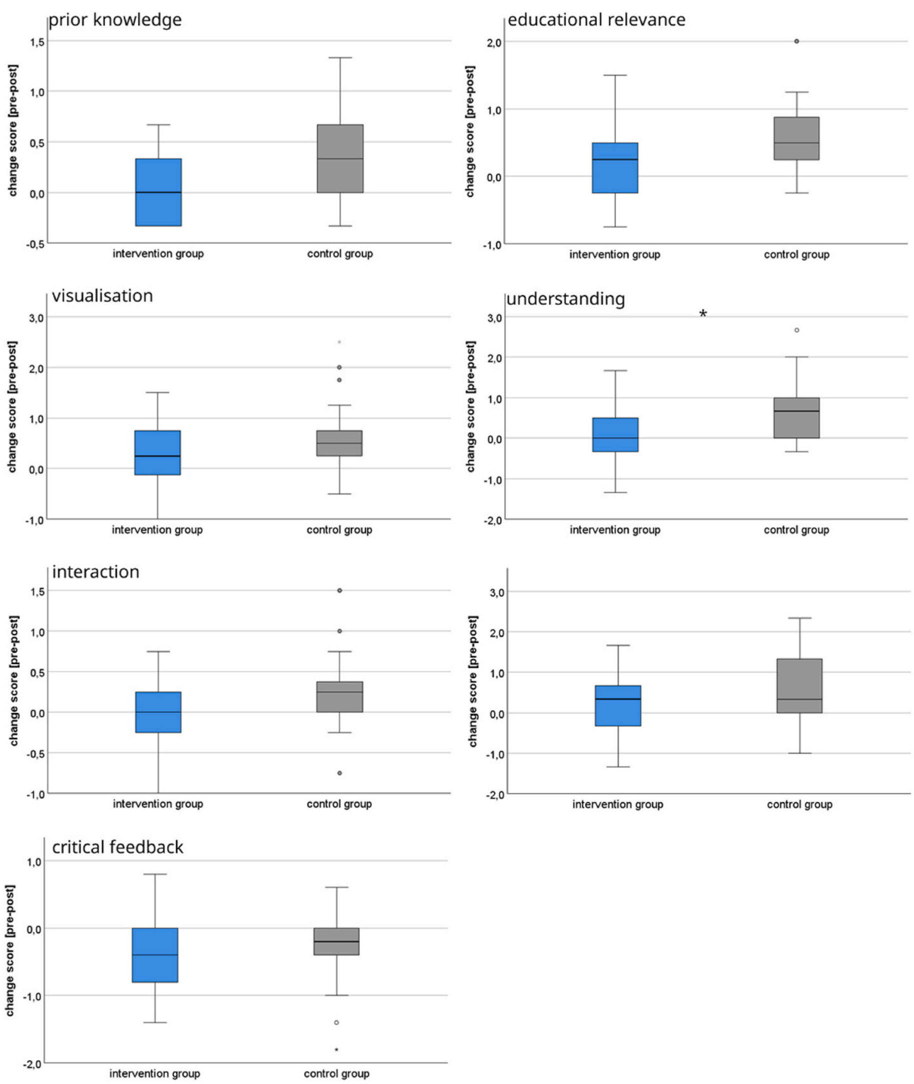

Fig. 3. Box plots of the change in ratings in the main categories (Δ = pre − post) in a comparison between the intervention group and the control group. Group comparison of difference values using the Mann–Whitney U test; * indicates $p < 0.05$. Likert scale (1–5): positive Δ values (pre − post) correspond to increased agreement in the respective category, with the exception of the critical feedback category, where the opposite direction applies due to negatively worded items.

of 8.7 out of 10 correct answers, while the intervention group achieved an average of 7.6 correct answers. The group comparison using the Mann–Whitney U test showed a significant difference between the groups (U = 382; Z = 2.65; p = 0.008; r = 0.39). Descriptively, the results of the control group were higher with a median of 9.0 compared to the intervention group with a median of 8.0 (Fig. 4).

4. Discussion

In this study, learning content was delivered either using mixed reality or a paper-based script. In the pre- and post-questionnaire analyses, no consistent statistically significant group differences between the intervention group (MR) and the control group were found across the main categories examined. In the within-group analysis, a significant change was observed in the intervention group in the category critical feedback, while significant changes were observed in the control group in the categories prior knowledge, educational relevance, visualisation, understanding, interaction and technical aspects. It should be noted that the baseline values in the critical feedback category already differed between the groups prior to the intervention. When comparing the change scores, only the understanding category showed a significant difference with a greater change in the control group. Due to the exploratory design, individual statistically significant findings in the questionnaire analyses are not interpreted as reliable evidence of a specific effect. In addition, the control group achieved significantly higher scores in the knowledge recall test. Overall, the results in this setting do not provide any evidence of a short-term learning advantage of the mixed reality implementation over the conventional script in this one-time, time-limited teaching setting.

Since both groups were able to try out the mixed reality application themselves before completing the post-questionnaire, the post-evaluations in both groups are based on their own MR experience. This is consistent with the predominantly parallel changes in ratings and may explain why there were no significant differences in the extent of change between the groups in most categories. This design choice was intended to allow experience-based responses but may have reduced between-group differences in post-questionnaire ratings. Differences in the technical aspects category are plausible, as the intervention group used the application in a learning setting, while the control group primarily got to know it as part of a trial. In addition, the ratings in several categories were already at a high baseline level, so that additional improvements in the questionnaire measures could only be reflected to a limited extent (see Table S1). It should also be noted that the questionnaire used was study-specific and exploratory and had not been formally psychometrically validated, which may limit its sensitivity to detect small effects (Terwee et al. 2007; Parai et al. 2020). Furthermore, potential memory effects must be taken into account, as participants could adapt their answers in the post-questionnaire to those in the pre-questionnaire (Schwarz et al. 2020).

The intergroup comparisons of the change scores showed no statistically significant differences in most categories; however, descriptively, shifts toward more positive ratings were predominantly evident in both groups. In contrast, the change in the “understanding” category was more pronounced in the control group. Against this background, mixed reality can be considered a potentially useful addition to established teaching methods, provided that implementation conditions such as technical equipment, structured onboarding and training, and appropriate didactic integration are taken into account (Moro et al. 2017; Ryan et al. 2022; Johnston et al. 2024). This is supplemented by the qualitative evaluation of the open-ended questions, which describes both implementation hurdles and perceived advantages of MR application.

To classify the results of the knowledge recall test, increased extraneous cognitive load due to the use of mixed reality technology should be considered as a plausible explanatory approach. Interaction via head-mounted displays and controllers requires additional operating and navigation processes that can tie up cognitive resources without directly contributing to content processing. In a one-time, time-limited setting, this can reduce the availability of capacity for memorizing and recognizing anatomical facts; when used for the first time, novelty effects are also possible, which divert attention from content-related learning processes to operation (Dontre 2021; Miguel-Alonso et al. 2023, 2024). In contrast, a paper-based script often supports linear and exam-oriented learning strategies, for example through repeated reading and targeted reference. This is particularly relevant in light of the fact that the knowledge recall test, as an immediate multiple-choice test, primarily assessed short-term fact retrieval.

In long-term learning settings, however, a study by Ruthberg et al. (2020) found no evidence of any impairment of test results when using MR (Ruthberg et al. 2020). Mixed reality can also encourage more exploratory interactions, which can be beneficial for certain learning objectives (e.g., spatial relations, transfer), but not necessarily for short-term fact retrieval. The observed differences can therefore be interpreted as an indication of a mismatch between the form of presentation, learning strategy, and tested outcome, since the chosen outcome (multiple-choice knowledge recall test) primarily measures recall and less spatial understanding/transfer. The learning object was taught using a model of the skull. Model-based representations only

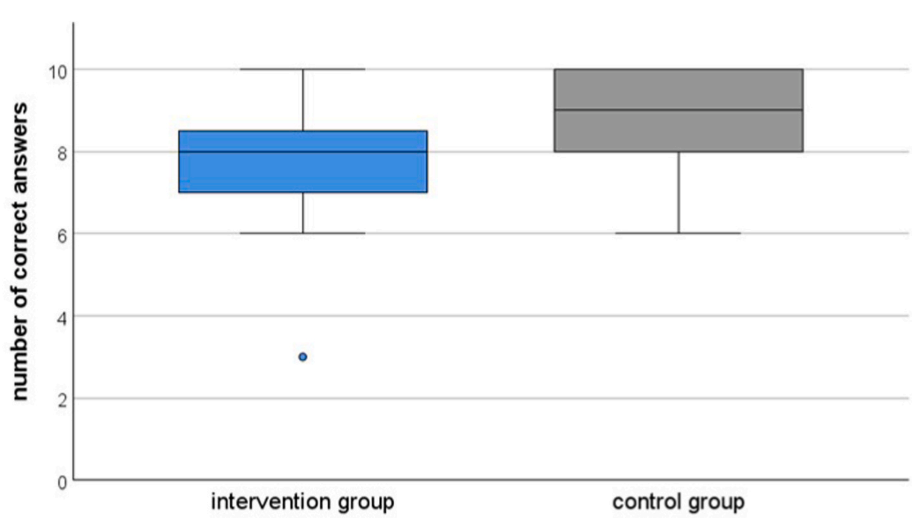

Fig. 4. Box plot of the knowledge recall test (0–10 correct answers) in the intervention- and control groups. Intervention group: Q1 = 7.0; median = 8.0; Q3 = 8.5. Control group: Q1 = 8.0; median = 9.0; Q3 = 10.0. * indicates p < 0.05; Mann–Whitney U test.

depict clinical situations to a limited extent, as variability, soft tissues, and pathological changes are not taken into account. It is therefore possible that the potential advantages of mixed reality are more evident in transfer-oriented outcomes (spatial orientation, application scenarios) than in a short-term, knowledge-oriented multiple-choice recall test.

The positive perception of mixed reality and high awareness among learners could be partly influenced by characteristics such as interactivity, immersion, and novelty, which are discussed in gamification literature as motivational factors (Bölek et al. 2022). However, motivational effects are not necessarily synonymous with sustainable learning gains and offer a possible explanation for a discrepancy between subjective assessments and short-term measurable learning success (Makransky et al. 2019, 2021).

The results suggest that mixed reality should not be viewed primarily as a medium that is more effective for learning per se, but rather as a tool whose effectiveness depends on usability, didactic embedding, task structure, and outcome fit. New technologies should be tested didactically, as their use does not automatically lead to added value in learning.

5. Conclusion

In this one-time, time-limited teaching session, the mixed reality application did not confer a short-term advantage in the knowledge recall test compared with a paper-based script covering the same content. At the same time, participants' questionnaire ratings were generally positive, though not uncritical, and did not show consistent between-group differences. Together, these findings suggest that mixed reality may enrich teaching primarily through subjectively valued features such as interaction and visualisation, rather than by improving immediate fact retrieval. The qualitative feedback underscores that any didactic added value hinges on pragmatic implementation factors—above all technical stability, usability, and thoughtful integration into the teaching concept. Mixed reality therefore seems best understood not as a superior medium per se, but as a complementary tool whose benefit depends on the learning objectives and the way it is embedded. The findings suggest that immersive technologies should not be evaluated in terms of technological superiority, but in terms of cognitive and didactic alignment. Mixed reality may enhance spatial understanding, interaction, and engagement; however, when learning objectives emphasize short-term factual retrieval under time constraints, conventional formats may be equally or more efficient. The educational value of immersive tools therefore depends less on immersion itself than on the congruence between medium affordances, learner strategies, and assessment design.

Informed consent statement

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

Author contributions

Conceptualization, Nadia Karnatz, Majeed Rana; Methodology, Nadia Karnatz, Majeed Rana, David Kaffka; Software, Michael Schwerter and Shufang Liu; Validation, Nadia Karnatz, Angela Rosa Caso, Raymond Wong, Majeed Rana, David Kaffka; Formal analysis, Nadia Karnatz and David Kaffka; Investigation, Nadia Karnatz and David Kaffka; Resources, Michael Schwerter and Majeed Rana; Data curation, Nadia Karnatz; Writing – original draft, Nadia Karnatz, Majeed Rana; Writing – review & editing, Majeed Rana and Aida Parviz; Visualisation, Nadia Karnatz and Majeed Rana; Supervision, Nadia Karnatz, Majeed Rana; Project administration, Nadia Karnatz and Majeed Rana; Funding acquisition, Majeed Rana.

Institutional review board statement

The study was conducted in accordance with the Declaration of Helsinki, and approved by the Ethics Comittee of the University Düsseldorf (protocol code 2022-1849).

Data availability statement

Data are contained within the article and Supplements.

Funding

This work has been supported by the Ministry of Economic Affairs, Innovation, Digitalisation and Energy of the State of North Rhine-Westphalia (MWIDE NRW) along with the Competence Center 5G. NRW under grant number 005-01903-0047.

Conflicts of interest

The authors affiliated with Brainlab (M.S. and S.L.) had financial interests in the subject matter, materials, and equipment, in the sense that they were employees of Brainlab. The extent of influence on the data, manuscript structure and manuscript conclusions by these authors and/or Brainlab was limited to technical knowledge and support for the manuscript preparation. Some features of the Brainlab software described in this article are currently a research prototype and are not for commercial use. Apart from that, the authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jcms.2026.104618.

References

Baratz, G., Sridharan, P.S., Yong, V., Tatsuoka, C., Griswold, M.A., Wish-Baratz, S., 2022. Comparing learning retention in medical students using mixed-reality to supplement dissection: a preliminary study. Int. J. Med. Educ. 13, 107–114.

Barthel, C., Montigel, F., Jobst, S., Büchner, V.A., Kugler, C., Strube-Lahmann, S., 2025. Digital technology and teaching methods in nursing education: a scoping review protocol. BMJ Open 15 (11), e108610.

Bölek, K.A., Jong, G. de, van der Zee, C.E.E.M., van Cappellen Walsum, A.-M., Henssen, D.J.H.A., 2022. Mixed-methods exploration of students' motivation in using augmented reality in neuroanatomy education with prosected specimens anatomical sciences education, 15 (5), 839–849.

Bork, F., 2018. „Interactive augmented reality systems“: hilfsmittel zur personalisierten Patientenaufklärung und Rehabilitation (Interactive augmented reality systems : aid for personalized patient education and rehabilitation. Der Unfallchirurg 121 (4), 286–292.

Bova, F.J., Rajon, D.A., Friedman, W.A., Murad, G.J., Hoh, D.J., Jacob, R.P., Lampotang, S., Lizdas, D.E., Lombard, G., Lister, J.R., 2013. Mixed-reality simulation for neurosurgical procedures. Neurosurgery 73 (Suppl. 1), 138–145.

Brann, D.W., Sloop, S., 2006. Curriculum development and technology incorporation in teaching neuroscience to graduate students in a medical school environment Advances in physiology education, 30 (1), 38–45.

Chen, H.-J., Liao, L.-L., Chang, Y.-C., Hung, C.-C., Chang, L.-C., 2019. Factors influencing technology integration in the curriculum for Taiwanese health profession educators: a Mixed-. Method. Stud. Int.J. Environ.Res. Pub.Health 16 (14).

Curran, V.R., Xu, X., Aydin, M.Y., Meruvia-Pastor, O., 2023. Use of extended reality in medical education: an integrative review medical science educator, 33 (1), 275–286.

Daling, L.M., Schlittmeier, S.J., 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. Rev. Human Factors 66 (2), 589–626.

Dhar, P., Rocks, T., Samarasinghe, R.M., Stephenson, G., Smith, C., 2021. Augmented reality in medical education: students' experiences and learning outcomes. Med. Educ. Online 26 (1), 1953953.

Dontre, A.J., 2021. The influence of technology on academic distraction: a review Human Behav and Emerg Tech, 3 (3), 379–390.

Han, E.-R., Yeo, S., Kim, M.-J., Lee, Y.-H., Park, K.-H., Roh, H., 2019. Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. BMC Med. Educ. 19 (1), 460.

Hernon, O., McSharry, E., MacLaren, I., Carr, P.J., 2023. The use of educational technology in teaching and assessing clinical psychomotor skills in nursing and

midwifery education: a state-of-the-art literature review Journal of professional nursing. Off. J. Am Assoc Colleges Nurs. 45, 35–50.
Huang, X., Yan, Z., Gong, C., Zhou, Z., Xu, H., Qin, C., Wang, Z., 2023. A mixed-reality stimulator for lumbar puncture training. Pilot Study BMC Med. Edu. 23 (1), 178.
Johnston, M., O'Mahony, M., O'Brien, N., Connolly, M., Iohom, G., Kamal, M., Shehata, A., Shorten, G., 2024. The feasibility and usability of mixed reality teaching in a hospital setting based on self-reported perceptions of medical students. BMC Med. Educ. 24 (1), 701.
Leung, R., Shi, G., 2024. Building your future holographic mentor: can we use mixed reality holograms for visual spatial motor skills acquisition in surgical education? Surg. Innov. 31 (1), 82–91.
Makransky, G., Andreasen, N.K., Baceviciute, S., Mayer, R.E., 2021. Immersive virtual reality increases liking but not learning with a science simulation and generative learning strategies promote learning in immersive virtual reality. J. Educ. Psychol. 113 (4), 719–735.
Makransky, G., Terkildsen, T.S., Mayer, R.E., 2019. Adding immersive virtual reality to a science lab simulation causes more presence but less learning. Learn. InStruct. 60 (11), 225–236.
Mayring, P., 2015. Qualitative Inhaltsanalyse: Grundlagen Und Techniken, twelfth ed. Beltz Pädagogik. Beltz, J, Weinheim, Bergstr.
Miguel-Alonso, I., Checa, D., Guillen-Sanz, H., Bustillo, A., 2024. Evaluation of the novelty effect in immersive Virtual Reality learning experiences Virtual Reality, 28 (1), 5501.
Miguel-Alonso, I., Rodriguez-Garcia, B., Checa, D., Bustillo, A., 2023. Countering the novelty effect: a tutorial for immersive virtual reality learning. Environ. Appl. Sci. 13 (1), 593.
Moro, C., Štromberga, Z., Raikos, A., Stirling, A., 2017. The effectiveness of virtual and augmented reality in health sciences and medical anatomy Anatomical sciences education, 10 (6), 549–559.
Parai, C., Hägg, O., Lind, B., Brisby, H., 2020. ISSLS prize in clinical science 2020: the reliability and interpretability of score change in lumbar spine research European spine journal. In: Official publication of the European Spine Society, the European Spinal Deformity Society, and the European Section of the Cervical Spine Research Society, 29, pp. 663–669, 4.
Rochlen, L.R., Putnam, E., Levine, R., Tait, A.R., 2022. Mixed reality simulation for peripheral intravenous catheter placement training. BMC Med. Educ. 22 (1), 876.
Ruthberg, J.S., Tingle, G., Tan, L., Ulrey, L., Simonson-Shick, S., Enterline, R., Eastman, H., Mlakar, J., Gotschall, R., Henninger, E., Griswold, M.A., Wish-Baratz, S., 2020. Mixed reality as a time-efficient alternative to cadaveric dissection Medical teacher, 42 (8), 896–901.
Ryan, G.V., Callaghan, S., Rafferty, A., Higgins, M.F., Mangina, E., McAuliffe, F., 2022. Learning outcomes of immersive technologies in health care student education: systematic review of the literature. J. Med. Internet Res. 24 (2), e30082.
Schwarz, H., Revilla, M., Weber, W., 2020. Memory Effects in Repeated Survey questions: reviving the empirical investigation of the independent measurements assumption SRM, 14 (3), 325–344.
Silvero Isidre, A., Friederichs, H., Müther, M., Gallus, M., Stummer, W., Holling, M., 2023. Mixed reality as a teaching tool for medical students in neurosurgery medicina. Kaunas, Lithuania) 59 (10).
Suresh, D., Aydin, A., James, S., Ahmed, K., Dasgupta, P., 2023. The role of augmented reality in surgical training. Systematic Rev. Surg. Innov. 30 (3), 366–382.
Terwee, C.B., Bot, S.D.M., Boer, M.R. de, van der Windt, D.A.W.M., Knol, D.L., Dekker, J., Bouter, L.M., Vet, H.C.W. de, 2007. Quality criteria were proposed for measurement properties of health status questionnaires. J. Clin. Epidemiol. 60 (1), 34–42.
Toni, E., Toni, E., Fereidooni, M., Ayatollahi, H., 2024. Acceptance and use of extended reality in surgical training. Umbrella Rev.Systematic Rev. 13 (1), 299.
Uymaz, P., Uymaz, A.O., 2022. Assessing acceptance of augmented reality in nursing education. PLoS One 17 (2), e0263937.