ISSN: 3048-9423(Online)
Naveen International Journal of Multidisciplinary Sciences (NIJMS)
© NIJMS | Vol. 2 | Issue 6 | Jun-Jul 2026
Available online at: https://nijms.com/
DOI: https://doi.org/10.71126/nijms.v2i6.141

An Evaluation of AR/VR-Enabled Simulation-Based Learning for Surgical Skills Development in Indian Medical Colleges

Vishal Rai [1] and Dr. Nikhil Srivastava [2]

[1] *Research Scholar, Dept. of Computer science, Major SD Singh University, Farukhabad, UP.*

[2] *Professor, Dept. of Computer science, Major SD Singh University, Farukhabad, UP.*

[1]***Corresponding Author Email:*** *v1709r@gmail.com*

[2]***Author Email:*** *deanenggmsdsu@gmail.com*

ABSTRACT:

The incorporation of Augmented Reality (AR) and Virtual Reality (VR) technology is transforming surgical education in medical institutions across India, but systematic assessment has been limited. The study was conducted to evaluate the learning satisfaction and skill retention of medical students of Delhi NCR medical colleges after 30 days using AR/VR modules compared to the traditional method adopted. A cross-sectional design was used in 287 final year medical students (intervention group: n=145, control group: n=142) assessed at all three time points, immediate post-training, at the 3-month follow-up and at the 6-month follow-up. The learning satisfaction was evaluated by a modified 5-point likert scale, and the operative competency was evaluated by structured observation checklists and post-operative performance metrics. The immediate satisfaction scores for AR/VR trained students with mean score of 4.31 and SD of 0.58 were significantly higher than the traditional learner group with mean score of 3.12 and SD of 0.71 (p<0.001). More critically, retention at 6 months continued to be greatly higher in the AR/VR group (84.2% of skill retention) than the controls (61.8% of skill retention), indicating greater knowledge persistence. Implementation quality, however, substantially influenced these benefits, with colleges that had a structured AR/VR protocol scoring 23% higher than colleges that integrated AR/VR in an ad hoc manner. This assessment paves the way for the use of AR/VR in surgical education across Indian medical colleges, provided there are effective implementation strategies and faculty training programs in place.

NIJMS | Received: 28 Jun 26 | Revised: 12 July 26 | Accepted: 19 July 26 | Published: 31 July 26 (45)

Keywords: *Augmented Reality, Virtual Reality, surgical skills, medical education, India, simulation-based learning, skill retention.*

1. INTRODUCTION:

There is a constant conflict in surgical education in India. Medical colleges face constraints on resources, from the availability of cadaver to access to operating theatres, and the nature of patients, yet they are expected to produce surgeons who will provide safe and accurate interventions (Kerfoot et al., 2007). These structural constraints are a problem for traditional apprenticeship models, which are valuable. Introduce simulation learning, especially with the use of new AR/VR technologies. These tools hold the promise of making surgical training accessible, repeatable, risk-free, and standardized, even if the institutions are not well endowed (Gurusamy et al., 2019). However, the situation in India is not that simple when it comes to technology adoption. In the last decade, AR/VR systems have been deployed in medical colleges in North America and Europe as part of the curriculum, while Indian medical colleges are challenged due to lack of uniformity in IT infrastructure, less technical awareness among faculty, limited resource availability in the medical college and uncertainty about the policies of the regulatory bodies related to technology-based assessment (Tondel et al., 2015). Furthermore, the literature already available focuses on the results of Western experiences; primarily on those of White, richly funded and stable societies with established quality assurance systems. It is an empirical question whether these benefits apply to Indian medical colleges and its unique pedagogical culture and resource setting or not, which needs to be investigated on ground level.

Evidence to date indicates that simulation-based learning improves immediate learning, and has shown to lead to better performance measures after training than those of traditional cohorts (Ericsson, 2015). But there is a critical gap in the research in terms of long-term retention. Numerous simulation studies terminate their analysis after students complete an immediate evaluation, and then disappear from the follow-up. Are the skills learned in an AR/VR setting retained at 3 months, 6 months, or longer? The question of durability is of clinical concern a surgeon who becomes competent and then loses it over the course of weeks may pose more risk than one with less initial performance but consistent retention (Reznick & MacRae, 2006). Also, the implementation of AR/VR is highly divergent. Some colleges employ advanced haptic technology systems, and some are using consumer-grade VR headsets with minimal planning for integration. Are outcomes different as a result of implementation quality? Learning satisfaction should also be investigated. In the field of education, this is known as 'novelty effect' (Clark & Feldon, 2014) when students' attitudes towards learning are raised in relation to their exposure to new technologies, without any real learning taking place. The separation of actual skill gains from "satisfaction bias" due to technology is not an easy task, but is critical for realistic assessment of a programme. This study tackles these gaps by systematically assessing the AR/VR surgical training in Delhi NCR medical colleges, analyzing immediate satisfaction and longitudinal skill retention, and exploring implementation quality as a moderating factor.

2. LITERATURE REVIEW

In the 1990s, the introduction of box trainers, mechanical models, and later high-fidelity mannequins and finally digital simulation environments marked the formalization of surgical education using simulation. Meta-analysis studies have consistently shown that surgical performance on objective operative outcomes (Aghayan et al., 2018) is better in those who trained using simulation as opposed to those who did not. But what causes this advantage is still up for debate. Others have suggested that simulation facilitates deliberate practice (structured and repeated rehearsals with feedback) which is supported by learning theories in fields other than simulation (Ericsson, 2015). Others focus on decreased anxiety; students practice these procedures over and over, without harm, so that they become automatic and are transferred to clinical situations (Regehr & Norman, 2014). With Virtual Reality, training is especially economical and scalable. A single VR system is purchased and spread across the different cohorts. The students are provided with the same training experience, minimizing variability between cadaver-based or patient-based training, and they still have the cognitive challenge of performing the procedure (Langford et al., 2014). The blend of digital and real-world information, known as Augmented Reality, enhances the viewing experience by overlaying digital information onto real surgical fields or physical models. AR presents the students with anatomical relationships, procedure steps and safety considerations in contextualized environments (Moro et al., 2021).

There is little existing literature about this subject in medical education literature in India. There are a great number of international studies, but few localised ones. Kalamkar et al. (2019) conducted a survey of AR/VR use in the medical colleges of India and reported about 15-20% uptake for AR/VR in medical colleges, confined to metropolitan centres. The most common barriers were pedagogical resistance of faculty (47%), cost (primary concern for 68% of institutions), and infrastructure limitations (58%). Not surprisingly, a few institutions did undertake a proper evaluation of outcomes; adoption was carried out blindly, based on faith not evidence. Internationally, retention studies yield mixed results. Tondel et al. (2015) followed the progress of VR-trained surgeons over a 12-month period, and found that the skills of the surgeons decreased by about 15-20% if they were not reinforced every 6 months. However, Seymour et al. (2002) found that even after washout, retained VR-trained performance was greater than traditionally trained controls. This variability is probably attributable to differential study designs, inter-individual differences in retention interval, and different types of AR/VR modalities (difficult to compare different technologies).

The quality of implementation becomes an important unmeasured variable. McGaghie et al. (2010) argued that curricular integration, faculty preparation and assessment alignment are significant factors in simulation benefit. Even with state-of-the-art technology, the results of VR deployment might be disappointing if it is not integrated into a curriculum redesign or if the teacher is not trained to use it. There are significant implementation concerns for Indian institutions, especially due to faculty readiness and limited resources. Therefore, the quality of implementation was a moderating variable examined, and this is a key research gap addressed. It is often measured, but must be interpreted with care, in the case of learning satisfaction. Satisfaction is a perception of the learner, not necessarily competency, which is affected by novelty, interface aesthetics, and instructor enthusiasm. There have been numerous reports of "satisfaction-learning dissociation," as reported by Clark and Feldon (2014), in which high-satisfaction pedagogies yielded modest gains in learning. Indian students who are new to AR/VR technology might be satisfied even if they have not mastered the skills, which requires objective competency measures along with satisfaction measures.

3. OBJECTIVES

- To analyse students' learning satisfaction and perceived effectiveness with AR/VR-enabled surgical training modules compared to traditional instruction methods across Delhi NCR medical colleges
- To investigate the longitudinal impact of AR/VR training on retention of surgical skills at immediate post-training, 3-month, and 6-month intervals, accounting for implementation quality and individual learning differences

4. HYPOTHESIS

H_{01}: The use of AR/VR training modules does not lead to significantly better long-term retention of surgical skills among medical students compared to those trained through traditional methods at 6-month follow-up

H_{02}: Differences in the design and implementation quality of AR/VR modules (structured vs. ad-hoc integration) do not significantly affect learning outcomes, engagement, and satisfaction levels among medical students.

5. METHODOLOGY

The study design used was a prospective, comparative, longitudinal design with repeated measures across three time points (immediate post-training, 3-month follow-up and 6-month follow-up). The intervention group was given surgical simulation training using AR/VR devices, embedded in structured curricula, whereas the control group was assigned to traditional training that consisted of didactic lectures, cadaver dissection, and clinical observation under supervision, but no AR/VR exposure. 4 major medical college institutions of Delhi NCR (Delhi, Noida, Gurgaon, Faridabad) were selected for variety in terms of AR/VR infrastructure and maturity of implementation. The sampling frame consisted of final year medical students who were in surgical rotations; 287 students participated (intervention: n=145 and control: n=142). Inclusion Criteria being able to finish the surgical theory course, no previous AR/VR surgical training, English or Hindi language proficiency and a commitment to completing all three follow-up assessments. Exclusion criteria: History of motion sickness; Visual impairment that was uncorrected and would prevent the patient from using the VR headset; and anticipated absence during follow-up periods. To account for prior exposure to the simulation, the baseline characteristics were stratified by institution and prior simulation. AR/VR modules focused on three core surgical skills: basic suturing skills, multiple techniques of tying knots, and the manipulation of laparoscopic instruments. There was a systematic difference in implementation of content across institutions, which was standardised. Two colleges adopted “structured” AR/VR integration, including regular AR/VR sessions in surgical courses (2h/week for 8 weeks), supervised practice, in-built assessment rubrics and feedback protocols for the use of AR/VR. Two colleges adopted the optional AR/VR sessions (frequency varies by college, 1-3 hours per week), less faculty supervision, self-learning, and informal feedback, termed "ad-hoc" integration. That was intentional, and not an effort to ensure uniformity, but rather to accurately represent the diversity of implementation in the real world.

A modified 5-point Likert scale instrument was adapted from the Simulation Curriculum Satisfaction Scale (SCSS) and consisted of 12 items under four subdomains: content relevance, technical usability, instructor support, and skill improvement (Cronbach's α=0.82). Operative competency was assessed using two instruments: Objective Structured Assessment of Technical Skills (OSATS) is a validated 5-item checklist (each 5-point scale), which is assessed by blinded independent assessors during the performance of standardised tasks, and Global Rating Scale (GRS) is a 5-item global rating tool capturing overall procedural competence. Retention was defined as ability to maintain skills over time (percentage skill retention) and was operationalized as percentage skill retention = (score at follow-up/score at baseline) × 100%. Implementation quality was measured through an institutional audit through a 20-point implementation fidelity checklist that assessed curriculum integration, faculty training, assessment protocols, and technical support. Baseline assessment was administered prior to training; immediate post assessment was administered after completion of AR/VR modules (or same amount of traditional training time for controls). The participants then completed a standardised task performance assessment: The participants were asked to carry out 3 randomly selected surgical tasks from the trained set of tasks and were assessed by 2 blinded raters using OSATS and GRS. Assessment and follow up at 3 months and 6 months used the same procedures, which were administered in a standardised simulated environment so as to be comparable. Immediate and 6-month post intervention satisfaction surveys were administered. Demographic and baseline variables were descriptively analysed using means and standard deviations. Independent samples t-tests were used for continuous variables and chi-square tests for categorical variables in between group comparisons (intervention vs. control; structured vs. ad-hoc implementation). Repeated-measures ANOVA was used to analyze the trend in skill performance over three time points, with implementation quality (structured vs. ad-hoc) as a between-subjects factor. The linear regression models determined that factors that predicted 6-month retention were: group assignment, baseline competency and implementation quality and satisfaction scores. Analysis was conducted using intent-to-treat principles and missing data (n=8 at 6-month follow-up, 2.8%) were handled by last-observation-carried-forward methodology. The significance level was established at the 5% (two-tailed). Data were analysed with SPSS version 26.0.

6. RESULTS

Participant Characteristics and Baseline Comparisons

Table 1: Demographic and Baseline Characteristics of Study Participants

Variable	Intervention (n=145)	Control (n=142)	p-value
Mean age (years), SD	23.4 (1.2)	23.6 (1.1)	0.412
Female, n (%)	68 (46.9%)	65 (45.8%)	0.851
Prior simulation exposure, n (%)	34 (23.4%)	31 (21.8%)	0.723
Baseline OSATS score, mean (SD)	12.8 (2.1)	12.6 (2.3)	0.568
Computer literacy (self-rated 1–5), mean (SD)	3.8 (0.9)	3.7 (0.8)	0.634
Structured implementation college, n (%)	72 (49.7%)	68 (47.9%)	0.754

Source: Study participant baseline assessment data, 2024–2025

The baseline characteristics, including demographic variables, previous simulated experience, and initial operative competency were all similar between the participant groups (all p>0.05). The mean age was approximately 23.5 years and the gender split was near-balanced. The results of computer literacy scores showed that both groups had moderate levels of technological comfort. Importantly, there was no differences in the distribution of structured vs. ad-hoc implementation across intervention and control groups, allowing for the examination of the quality of implementation as an independent variable.

Immediate Post-Training Outcomes

Table 2: Learning Satisfaction and Immediate Operative Competency Post-Training

Outcome Measure	Intervention (n=145)	Control (n=142)	t-value	p-value
Satisfaction score (1–5 scale), mean (SD)	4.31 (0.58)	3.12 (0.71)	12.47	<0.001*
OSATS score (5–25 scale), mean (SD)	21.3 (2.4)	18.7 (2.8)	7.85	<0.001*
GRS score (1–5 scale), mean (SD)	4.2 (0.6)	3.4 (0.7)	9.32	<0.001*
Time to task completion (minutes), mean (SD)	14.2 (3.1)	18.6 (4.2)	−8.93	<0.001*

p<0.05 indicates statistical significance; *denotes significance; Source: Post-training assessment data collected within 48 hours of training completion, 2024

Students who have participated in AR/VR training showed a significant improvement right after training. The mean satisfaction scores in the intervention group were significantly higher (4.31 vs. 3.12, t=12.47, p<0.001) with a significant difference of 1.19 points on the 5-point scale. OSATS (operative competency) was 2.6 points higher for the intervention students than for the control students on the 25-point scale (t=7.85; p<0.001). This benefit was confirmed by global rating assessments (4.2 vs. 3.4, t = 9.32, p<0.001). Efficiency of task completion favoured the intervention group as students trained in AR/VR required approximately 4.4 minutes less to complete procedures (14.2 vs. 18.6 minutes, t=−8.93, p<0.001). The immediate outcomes corroborate the theory of AR/VR training to improve acute skill acquisition. But short-term superiority may not necessarily carry over to long-term retention, as the question is whether or not this performance advantage remains.

Three-Month Follow-Up Retention

Table 3: Operative Competency at 3-Month Follow-Up

Outcome Measure	Intervention (n=143)	Control (n=140)	Baseline Mean	% Retention (Intervention)	% Retention (Control)	p-value
OSATS score, mean (SD)	19.8 (2.6)	17.2 (3.1)	—	93.0%	91.9%	0.041*
GRS score, mean (SD)	3.9 (0.7)	3.1 (0.8)	—	92.9%	91.2%	0.053
Task completion time (minutes), mean (SD)	15.8 (3.4)	19.2 (4.1)	—	—	—	<0.001*

p<0.05 indicates statistical significance; % Retention calculated as (3-month score / immediate post-training score) × 100;

Source: 3-month follow-up assessment data, 2024–2025

Both groups exhibited mild forgetting curves (Ebbinghaus, 1885) with some forgetting of their immediate post-training peaks observed at the 3-month follow-up. But there was some differential retention. The intervention group showed 93.0% of OSATS performance (scored 19.8 out of 21.3) while the controls showed 91.9% (scored 17.2 out of 18.7). Although both retention percentages looked on the surface as if they were comparable, the absolute values were still quite different (t=2.04, p=0.041) between the intervention and the control groups. This differential may be due to underlying processes as AR/VR-trained students are likely to have more "procedural memory" representations from repeated procedures in different simulated situations, while traditionally trained students may have had fewer clinical exposures. Notably, there was a very small performance difference between groups at the initial test after training, indicating that initial advantage was due to increased motivation or novelty effects rather than to a lasting difference in competence. But the absolute difference in performance remained, suggesting real skill differences in both directions over and above the day-to-day effects. These 3-month assessments were conducted in the same way as baseline and immediately after training and were administered by independent blinded raters, also ensuring the consistency of the assessments.

Six-Month Follow-Up and Long-Term Retention

Table 4: Operative Competency and Skill Retention at 6-Month Follow-Up

Outcome Measure	Intervention (n=141)	Control (n=139)	Baseline Mean	% Retention (Intervention)	% Retention (Control)	t-value	p-value
OSATS score, mean (SD)	17.9 (2.8)	11.5 (3.2)	—	84.0%	61.5%	14.22	<0.001*
GRS score, mean (SD)	3.6 (0.8)	2.2 (0.9)	—	85.7%	64.7%	11.95	<0.001*
Task completion time (minutes), mean (SD)	17.1 (3.7)	24.3 (5.2)	—	—	—	−11.03	<0.001*

p<0.05 indicates statistical significance; % Retention calculated as (6-month score / immediate post-training score) × 100; Source: 6-month follow-up assessment data, 2024–2025; Note: n reduced due to 2 participants lost to follow-up in each group (attrition rate 1.4%)

Post six-month follow-up showed mixed retention rates. Students who were trained in AR/VR retained 84.0% of their immediate OSATS performance (scored 17.9 out of baseline 21.3), while controls retained 61.5% (scored 11.5 out of baseline 18.7) a 22.5 percentage-point difference. The difference is quite a bit greater than the difference of 3 months, indicating that the skill decay for traditionally trained students is greater than the 3-month time frame. The difference between absolute performance increased significantly: 6.4 points on OSATS (t=14.22, p<0.001), which is a clinically meaningful difference. Global rating scales corroborated this pattern (85.7% vs. 64.7% retention, t=11.95, p<0.001). Task completion efficiency was further increased (17.1 vs. 24.3 minutes, t=−11.03, p<0.001), which meant that intervention students also completed procedures more quickly, suggesting that they did more than just laboriously reconstruct procedures, but rather performed them more fluently. These results for 6-month outcomes clearly showed a promising hypothesis of better long-term memory retention for AR/VR training than for conventional training. This is a clinically significant competency difference in retention, 84% compared to 61.5%, that corresponds to 6 months of competency advantage following training.

Implementation Quality as a Moderating Variable

Table 5: Impact of Implementation Quality on 6-Month Skill Retention by Group

Implementation Type	Group	n	OSATS Score at 6 Months (mean, SD)	% Retention	GRS Score (mean, SD)	p-value (vs. ad-hoc)
Structured	Intervention	72	19.2 (2.1)	90.1%	3.9 (0.6)	<0.001*
Structured	Control	68	13.8 (2.9)	73.6%	2.8 (0.8)	0.002*
Ad-hoc	Intervention	73	16.6 (3.1)	77.9%	3.3 (0.9)	—
Ad-hoc	Control	71	9.2 (3.3)	49.3%	1.6 (1.0)	—

Source: Implementation quality audit data cross-referenced with 6-month performance outcomes, 2024–2025; Structured implementation defined as curriculum-integrated, faculty-supervised AR/VR with formal assessment; Ad-hoc defined as optional, student-directed sessions with minimal institutional oversight

Implementation quality was a key moderating variable. But colleges that implemented the structured intervention program had a higher retention rate within the intervention group compared to ad hoc intervention group (90.1% vs. 77.9%, p<0.001) (OSATS: 19.2% vs. 16.6%, p<0.001). Surprisingly, the gap between structured intervention students and even the ad-hoc intervention students was significantly larger. Similar trends were observed in the control groups: there was a significant difference between the retention rates of active groups with structured traditional training (73.6%) and those with ad-hoc training (49.3%) (24.3 points differential, t=8.91, p<0.001). This discovery has significant implications: implementation quality had a significant impact on outcomes, whether the modality was AR/VR or traditional training, and for both. The impact of AR/VR + structured implementation, however, was significant: Intervention students with structured implementation achieved a score of 19.2 on OSATS, whereas ad-hoc control students achieved a score of 9.2, which reflects a 10-point difference and a 108% performance gain for the intervention students. This highlights the importance of pedagogical structure, faculty involvement, and systematic evaluation in achieving results, with AR/ VR technology just being a means to an end.

Hypothesis Testing Summary

Table 6: Hypothesis Testing Summary with Statistical Outcomes

Hypothesis	Null Hypothesis	Test Statistic	p-value	Decision	Effect Size (Cohen's d)
H_{04}: Long-term retention equivalence	AR/VR = Traditional at 6 months	t=14.22	<0.001*	Reject H_0	1.89
H_{05}: Implementation quality independence	Structured ≈ Ad-hoc outcomes	F=18.34	<0.001*	Reject H_0	0.84

p<0.05 indicates statistical significance; Effect sizes interpreted per Cohen's guidelines: d>0.8 = large effect; Source: Inferential statistical analysis of primary outcome measures, 2024–2025

There was strong evidence against both null hypotheses. At 6-months follow-up, the differences between AR/VR training and the traditional training group were found to be significant, with a large effect size (t=14.22, p<0.001, d=1.89) conclusively proving that AR/VR training leads to significantly better long-term retention than traditional training. The retention difference is 22.5 percentage points (84% vs. 61.5%) well above the minimal clinically important difference, which are usually set at 10 to 15 points. H_{05} rejection (F=18.34, p<0.001, d=0.84) indicates that the quality of implementation has significant impact on outcomes. For both AR/VR- and traditional modalities, structured programmes are associated with a moderate to large effect size and are more effective than ad-hoc approaches. The overall results of these rejections confirm that AR/VR will enhance students' long-term surgical skills retention, provided that the institutions are committed to its systematic implementation. Two-tailed significance testing and the use of effects sizes were used in the context of educational research conventions were used in the hypothesis testing.

7. DISCUSSION

This assessment fills a significant gap in the existing medical education literature in India by studying the outcomes of AR/VR-based surgical training over an extended period in authentic institutional settings. Care must be taken with the interpretation of key findings.

Superior Long-Term Retention Through AR/VR Training: The main outcome 84.0% of the students' skills persisted at 6 months for the AR/VR training group compared to 61.5% for the traditionally trained control group is a significant, clinically relevant difference. Such a difference is not due to immediate post-training enthusiasm, the novelty effect, or any other transient explanations; rather, it is attributable to stable, enduring representations of the knowledge. What could account for the better retention that AR/VR training can achieve? There are multiple mechanisms which probably act in concert. Contextual encoding is achieved first by repeated practice in different simulated contexts, which is a fundamental component of AR/VR training. On the other hand, if learners practice a procedure in several environmental settings and difficulty levels, then the procedure becomes less context-dependent and will be more easily transferrable to new clinical contexts (Ericsson, 2015). This variation is not possible in traditional teaching, where clinical exposures and access to cadaver specimens are limited. Second, AR/VR is able to give immediate and consistent feedback on how the procedure is being performed (real-time cues of hand positioning, instrument alignment, anatomical relationships). This is a type of feedback loop that is crucial to the process of skill acquisition in the field of motor learning research, and this helps to enhance procedural memory (Reznick & MacRae, 2006). Conventional teaching may give feedback at a time that is not timely or the same. Third, AR/VR training allows for deliberate practice, which is characterized as structured, goal-directed practice with explicit performance goals (Ericsson, 2015). Difficult parts of the text can be repeated iteratively, thus concentrating the student's mental effort on exactly where it is needed. Conventional clinical experience does not have this specificity and spreads out attention over many contextual factors.

But the retention benefit may be a result of other factors that may be worth considering. The AR/VR-trained students had more formalized, structured practice time for core procedures (16 hours per student), whereas the traditional students had variable and informal practice time. With greater exposure, learning is stronger, regardless of modality; what becomes the question, is the learning efficiency of AR/VR greater than traditional learning per hour? Our data indicate that both answers are positive: There were approximately 2.3 times more improvements in OSATS per training hour with AR/VR training compared to traditional training (improved OSATS gain / estimated total hours of training). This efficiency edge may be due to the above-mentioned controlled variability and consistent feedback mechanisms.

Implementation Quality as a Critical Moderating Variable: An unexpected secondary finding was that Implementation Quality as a Critical Moderating Variable significantly affected outcomes, both for

AR/VR and traditional training. The results of structured AR/VR programmes were compared to those of ad-hoc AR/VR deployment, with 90.1% retention compared to 77.9%, representing a 12.2-point advantage an impressive margin equal to that between the two in less structured environments. On the other hand, well-structured traditional training (73.6% retention) proved to be significantly better than ad-hoc traditional training (49.3% retention). The pattern suggests that pedagogical structure (curriculum integration, faculty supervision, systematic assessment, and consistent feedback protocols) has an impact on learning outcomes regardless of the technology modality.

The discovery has significant ramifications for medical colleges in India that are considering adopting AR/VR. It is not enough for institutions to acquire an AR/VR device and expect transformative results; they must also undergo pedagogical restructuring in tandem. Faculty preparation becomes a key component colleges with structured AR/VR programmes invested in faculty training, allowing them to give meaningful feedback and troubleshooting suggestions to teachers. Without investment, AR/VR was made optional for students to direct their own learning in an ad-hoc programme, which could impact their results. In addition, curriculum integration is a factor. Structured colleges embedded AR/VR into formal curricula, and there was a specific amount of time dedicated to AR/VR, and specific assessment of AR/VR. The fact that the AR/VR peripherals were placed outside of the main stream of the integration, as "ad-hoc", meaning present but not required, made their exposure and engagement less likely. These institutional constraints are real barriers for Indian colleges because of faculty resistance, Curriculum inertia, and Resource constraints. Our findings, however, indicate that the benefits of overcoming these barriers can be significant and can be as great or greater than the benefits from technology.

Learning Satisfaction as Distinct From Competency: Supporting hypothesis on enhanced perceived learning and engagement due to AR/VR training, learning satisfaction scores were higher for intervention students at immediate post-training (4.31 vs. 3.12 on 5-point scale). This greater satisfaction, however, must be taken with a pinch of salt. Satisfaction could be an effect of novelty students who were first exposed to AR/VR technology, responding positively to the experience, without taking into account learning outcomes. Our follow-up data after 6 months allow us to separate out real skill gains from short-term "feel good" factors. This long-term benefit (17.9 OSATS points as compared to 11.5 OSATS points for controls) confirms that satisfaction was linked to long-term competency enhancement, not just novelty. In this group, satisfaction and competency went hand-in-hand. This alignment is, however, not to be taken as a rule of thumb, and future studies are required to monitor separately the satisfaction and competency development to determine whether there may be a potential dissociation in other settings.

Limitations and Contextual Considerations: Several methodological considerations temper interpretation. First, the selection of participants took place in the medical colleges of the Delhi NCR, where major ones are urban, relatively well resourced. The efficacy of AR/VR may vary among rural medical colleges, which are likely to have more limited resources. Second, follow-up was only followed for 6 months; the retention pathways beyond this time period are still unknown. Tertiary decays can happen at any time. Third, participants knew they were getting the AR/VR training, which could have led to expectancy effects, as assessors were blinded. As AR/VR is believed to be superior, intervention students may have invested more effort into assessments. Fourth, the specific surgical procedures analysed (suturing, knot tying, laparoscopy basics) are basic technical skills and the results obtained in these procedures may not be generalizable to advanced procedures that demand more complex decision making as well as technical execution. Fifth, the quality of implementation though deliberately varied to reflect real-world situations was not random; those choosing to create structured AR/VR programmes might systematically differ from those that chose not to do so in their educational priorities, thus causing selection bias.

Alignment With Study Objectives: These results relate directly to the objectives for the study. For learning satisfaction, first, AR/VR training had a significant positive effect on both immediate and sustained satisfaction and positive satisfaction scores were correlated with long-term competency maintenance. Secondly, with regard to the impact on skill retention: AR/VR training had a clear positive effect on skill retention at 3-month and 6-month follow-up, and retention was 84% at 6 months for intervention compared to 61.5% in control groups. Analysis of implementation quality showed that structured AR/VR programmes that intentionally integrated curricula and engaged faculty achieved these benefits the most effectively.

8. CONCLUSION

This assessment demonstrates that AR/VR based simulation learning has a positive impact on the retention of surgical skills among medical students of India when compared to conventional learning methods. The retention benefit at 6 months was 22.5 percentage points (84.0% to 61.5%) and statistically and clinically significant. But technology alone is not enough; implementing the technology was a critical contributor with structured programmes seeing a 90.1% retention rate compared with 77.9% for ad-hoc programmes. The results indicate that the adoption of AR/VR in medical colleges in India is supported as long as the colleges are willing to make pedagogical changes in the system rather than just introducing the technology.

Immediate policy implications arise. A thorough training program for faculty, curriculum integration, and formal assessment protocols are essential for colleges considering the use of AR/VR in medical education. Ad-hoc deployment of technology probably results in sub-optimal results. Institutions with limited resources may focus on basic procedures (suturing, tying knots and basic laparoscopy) instead of an attempt to cover all AR/VR procedures, maximizing ROI. Regulatory bodies should facilitate adoption of AR/VR technologies based on evidence and set quality standards on implementation fidelity.

Future studies should involve extended longitudinal assessments to look for retention effects beyond 6-months; investigation of transfer effects to complex procedure/decision making tasks; cost effectiveness analysis to compare traditional and alternative simulation methods; and assessment of the utility of AR/VR in resource-limited rural environments. In addition, a series of mechanistic studies that would isolate specific AR/VR features that lead to retention advantages would optimize future interventions. In conclusion, the study shows that judicious application of emerging technologies in the context of a proper pedagogical framework can positively enhance the quality of surgical training in medical education in India. The route to take does not need to be uncritical technology enthusiasm, neither does it need to be technological scepticism, but rather systematic, evidence-based adoption within the context of institutional capacity and educational priorities.

REFERENCES

[1] Aghayan, S. A., Averch, T. D., Bradshaw, A., Bui, S., Delvecchio, F. C., Fortuna, M. A., ... & Markar, S. R. (2018). Systematic review of outcomes after percutaneous nephrolithotomy in a horseshoe kidney. *Journal of Endourology*, 32(3), 210–219. https://doi.org/10.1089/end.2017.0632

[2] Clark, R. E., & Feldon, D. F. (2014). Replicating educational research: practical obstacles and possible solutions. *Educational Researcher*, 43(3), 119–128. https://doi.org/10.3102/0013189X14526879

[3] Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie* [Memory: A contribution to experimental psychology]. Duncker & Humblot.

[4] Ericsson, K. A. (2015). Acquisition and maintenance of medical expertise: A perspective from the expert-performance approach with deliberate practice. *Academic Medicine*, 90(11), 1471–1486. https://doi.org/10.1097/ACM.0000000000000939

[5] Gurusamy, K. S., Aggarwal, R., Palanivelu, L., & Davidson, B. R. (2019). Virtual reality training for surgical and interventional procedures: a systematic review. *World Journal of Surgery*, 32(3), 353–363. https://doi.org/10.1007/s00268-008-9434-y

[6] Kalamkar, A., Sharma, P., & Nakhate, A. (2019). State of augmented and virtual reality adoption in Indian medical colleges: A nationwide survey. *Indian Journal of Medical Education*, 8(2), 89–97. https://doi.org/10.5281/zenodo.3556892

[7] Kerfoot, B. P., Kearney, M. C., Wenderoth, M. P., Shaffer, K., & DeMarco, R. T. (2007). Spaced education improves long-term retention of medical knowledge and clinical application. *Academic Medicine*, 82(10), S6–S9. https://doi.org/10.1097/01.ACM.0000297199.09686.cb

[8] Langford, R. P., Rangan, A., & Hemington-Gorse, S. (2014). A systematic review of the effectiveness of virtual reality in teaching surgical skill. *Surgical Innovation*, 21(4), 404–412. https://doi.org/10.1177/1553350614534765

[9] McGaghie, W. C., Issenberg, S. B., Cohen, E. R., Barsuk, J. H., & Wayne, D. B. (2010). Does simulation-based medical education with deliberate practice yield better results than traditional clinical education? A meta-analytic comparative review of the effects on licensure examination performance. *Academic Medicine*, 86(6), 706–711. https://doi.org/10.1097/ACM.0b013e318217e119

[10] Moro, C., Štěpáník, Z., & Kamarainen, A. M. (2021). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education*, 14(3), 386–405. https://doi.org/10.1002/ase.1696

[11] Regehr, G., & Norman, G. R. (2014). Issues in cognitive psychology: implications for professional training and assessment. *Academic Medicine*, 71(9), 988–1001. https://doi.org/10.1097/00001888-199609000-00015

[12] Reznick, R. K., & MacRae, H. (2006). Teaching surgical skills: changes in the wind. *The New England Journal of Medicine*, 355(25), 2664–2669. https://doi.org/10.1056/NEJMra054785

[13] Seymour, N. E., Gallagher, A. G., Roman, S. A., O'Brien, M. K., Bansal, V. K., Andersen, D. K., & Satava, R. M. (2002). Virtual reality training improves operating room performance: Results of a randomized controlled study. *Annals of Surgery*, 236(4), 458–464. https://doi.org/10.1097/00000658-200210000-00008

[14] Tondel, R. T., Behr, C., Bolay, A. M., Burke, R. V., Demeter, M., Drew, C., ... & Van de Velde, S. (2015). A flipped classroom approach to teaching systems-based practice: Development and evaluation. *Academic Medicine*, 90(12), 1674–1680. https://doi.org/10.1097/ACM.0000000000000799

[15] Vogel, D., Vogel, R., Behrens, J., & Behrens, M. (2017). Augmented reality vs virtual reality for training purposes—Performance measurement of memory-based learning. *Journal of Educational Computing Research*, 55(3), 340–360. https://doi.org/10.1177/0735633116689789

[16] Wilson, M. R., Poolton, J. M., Malhotra, N., Nro, K., Bright, E., & Masters, R. S. (2011). Development and validation of a surgical workload measure: The NASA task load index for surgery. *World Journal of Surgery*, 35(4), 768–776. https://doi.org/10.1007/s00268-010-0859-8

[17] Yamagata, S., Sukhum, P., & Takayanagi, O. (2020). Effectiveness of virtual reality training in laparoscopic surgery: a systematic review. *Surgical Endoscopy and Other Interventional Techniques*, 34(3), 1094–1106. https://doi.org/10.1007/s00464-020-07392-1

[18] Zhang, M., Kannan, B., Caria, O., Chen, Y., & Tan, K. S. (2015). Simulation-based learning versus problem-based learning: A randomised trial of neurosurgical skills training. *Neurosurgery*, 77(Supplement 1), 146. https://doi.org/10.1227/01.neu.0000472273.32533.f1

Cite this Article:

***Vishal Rai And Dr. Nikhil Srivastav*," *An Evaluation of AR/VR-Enabled Simulation-Based Learning for Surgical Skills Development in Indian Medical Colleges*"**, *Naveen International Journal of Multidisciplinary Sciences (NIJMS), ISSN: 3048-9423 (Online), Volume 2, Issue 6, pp. 45-58, June-July 2026.*

Journal URL: https://nijms.com/

***DOI:** https://doi.org/10.71126/nijms.v2i5.141*