

Comparison of Virtual Reality and Mannequin-based Postpartum Hemorrhage Training: A Randomized Controlled Trial



Ide Pustaka Setiawan^{1,*} , Yoyo Suhoyo¹ , Doni Widyandana¹  and Ova Emilia¹ 

¹Department of Medical Education and Bioethics, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia

Abstract:

Introduction: Training that lacks psychological and emotional preparation for high-pressure emergencies can leave healthcare providers unprepared during actual cases. Virtual Reality (VR) offers immersive learning experiences that enhance preparedness and confidence in healthcare providers. To address this, the Gadjah Mada Virtual Reality on Obstetrics and Gynecology - postpartum hemorrhage case (GAMA VROG), a virtual reality-based training application, was developed. Its effectiveness compared to traditional mannequin-based training remains unclear.

Objective: This study evaluates the effectiveness of VR-based training compared to traditional mannequin-based training on the learning experience, knowledge, perceived skills, and readiness level of practicing midwives.

Methods: A non-blinded randomized controlled trial was conducted involving 90 practicing midwives. Participants were allocated to either a control group (mannequin-based training) or an intervention group (VR-based training). Both groups underwent face-to-face training on postpartum hemorrhage, followed by skill practice using their respective methods. Data were collected via pre- and post-training questionnaires, which assessed the learning experience, knowledge, perceived skills, and readiness. Statistical analyses included the Wilcoxon signed-rank test, the Mann-Whitney U test, and independent sample t-tests were conducted using SPSS version 25.00.

Results and Discussion: Both the mannequin and VR groups showed significant improvement in knowledge (mannequin: 55.44 to 78.44, $p = 0.000$; VR: 50.67 to 76.78, $p = 0.000$). However, neither group demonstrated significant improvement in perceived skills (mannequin: 83.37 to 87.87, $p = 0.060$; VR: 86.60 to 89.94, $p = 0.070$). The VR group showed a within-group increase in readiness (83.54 to 88.81, $p = 0.015$), but this did not reach statistical significance after Bonferroni correction ($p < 0.0029$). In learning experience domains, VR significantly outperformed mannequins across all indicators: contextual (58.03 vs. 32.97, $p = 0.000$), enjoyable (54.17 vs. 36.83, $p = 0.000$), focused (53.40 vs. 37.60, $p = 0.001$), interactive (53.28 vs. 37.72, $p = 0.001$), and readiness (50.33 vs. 40.67, $p = 0.044$).

Conclusion and Recommendations: VR-based training demonstrated clear benefits in enhancing knowledge and learner engagement, especially in providing an immersive experience. However, these advantages did not extend to improvements in perceived skills or readiness after statistical adjustment. These findings suggest that while VR can enrich the educational atmosphere, its integration should complement, not substitute for, hands-on simulation in midwifery training.

Keywords: Virtual reality, Postpartum hemorrhage, Midwifery, Clinical competence, Education, Simulation training.

© 2025 The Author(s). Published by Bentham Open.

This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International Public License (CC-BY 4.0), a copy of which is available at: <https://creativecommons.org/licenses/by/4.0/legalcode>. This license permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

*Address correspondence to this author at the Department of Medical Education and Bioethics, Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia; Tel: +6282137481104; E-mail: idepustaka@ugm.ac.id

Cite as: Setiawan I, Suhoyo Y, Widyandana D, Emilia O. Comparison of Virtual Reality and Mannequin-based Postpartum Hemorrhage Training: A Randomized Controlled Trial. Open Nurs J, 2025; 19: e18744346426336. <http://dx.doi.org/10.2174/0118744346426336251001074230>



Received: July 07, 2025
Revised: August 31, 2025
Accepted: September 16, 2025
Published: October 02, 2025



Send Orders for Reprints to
reprints@benthamscience.net

1. INTRODUCTION

Postpartum Hemorrhage (PPH) remains the leading cause of maternal mortality worldwide, responsible for approximately 25% of maternal deaths globally and disproportionately affecting Low- and Middle-Income Countries (LMICs) [1, 2]. According to the World Health Organization, there were an estimated 287,000 maternal deaths in 2020, with most of them preventable [2]. In Indonesia, the maternal mortality ratio remains high and far from the Sustainable Development Goals (SDGs) target of fewer than 70 deaths per 100,000 live births by 2030 [3]. These alarming figures underscore the need for strategic interventions that prioritize improving the competencies of maternal healthcare providers.

In response, the Indonesian Midwives Association (IBI) has established Continuing Professional Development (CPD) training to maintain midwives' clinical competencies. However, these programs are conducted infrequently, typically once every five years, and rely on traditional didactic and mannequin-based simulations [4]. While mannequins provide opportunities for hands-on skill practice, they frequently fail to foster psychological fidelity, contextual realism, and emotional preparedness for rare but high-stakes emergencies such as PPH. Learners may struggle to transfer skills from static mannequins to dynamic clinical settings [5]. This educational shortcoming leaves midwives theoretically competent but less confident and underprepared to act decisively in real clinical crises [6-8]. This shortfall has led to growing interest in immersive training technologies, such as Virtual Reality (VR), which allow healthcare professionals to experience realistic, emotionally engaging environments that mirror clinical emergencies [9].

Over the past decade, VR-based simulation has been increasingly explored as a complement or, in some cases, an alternative to traditional mannequin-based training. Several studies have shown that VR can enhance learner engagement, decision-making, and confidence [10, 11]. VR simultaneously enhances emotional engagement throughout emergency simulations by constructing a real-world-like multisensory clinical learning environment [11-13]. However, findings from comparative trials remain mixed. A systematic review by Rourke (2020) reported that while VR may enhance engagement and knowledge retention, its impact on procedural skill development is often comparable or even inferior to that of mannequin-based practice [5]. This highlights the importance of examining the functional fidelity of each method.

Moreover, most existing comparative studies have been conducted in high-resource settings, where advanced infrastructure, skilled facilitators, and continuous access to high-fidelity simulators are available. In contrast, limited attention has been given to the Indonesian context, where disparities in healthcare resources and geographical challenges make access to repeated, high-fidelity training for midwives far less feasible. This situation creates a significant evidence gap, particularly in understanding how innovative technologies such as VR can be effectively and sustainably integrated into midwifery CPD training. As Indonesia represents one of the largest LMICs, generating

context-specific evidence is critical to ensure that VR interventions are not only pedagogically effective but also scalable and relevant to local needs.

The Gadjah Mada Virtual Reality Obstetrics and Gynecology - postpartum hemorrhage case (GAMA VROG) was developed as a VR-based simulation tool for PPH training among midwives as part of their CPD. However, its comparative effectiveness with conventional mannequin-based training remains unclear. This study was therefore designed not only to compare learning outcomes between the two methods but also to examine how midwives experience each type of training and whether VR adds value beyond content delivery.

2. METHODS

2.1. Context

GAMA VROG is a virtual reality-based training application. The application immerses users in a simulated delivery room using a Head-Mounted Display (HMD), where they engage with interactive 3D images, animations, and audio-visual cues that replicate real-life emergency situations. The learning environment supports both practice mode, which offers guided learning with feedback, and assessment mode, which challenges users to manage cases independently based on clinical judgment (Fig. 1).



Fig. (1). Illustration of the GAMA VROG interface.

Within the VR scenario, learners are required to identify the cause of PPH, perform initial interventions such as uterine massage or perineal inspection, and make time-sensitive decisions. The scenario aligns with national clinical guidelines and CPD competency frameworks, targeting three core learning objectives which were designed for PPH management in primary healthcare facilities: (1) early recognition of PPH, (2) implementation of initial clinical management steps, and (3) emergency decision-making under pressure.

Instructional design follows key simulation principles. High psychological fidelity is achieved through scenario branching, real-time feedback, and consequence-driven outcomes (e.g., stabilization or deterioration of the virtual patient). Moderate physical fidelity is built into the

interface through natural user hand gestures to navigate, select tools, and perform simulated clinical tasks. To reduce extraneous cognitive load, interface instructions are concise and intuitive, allowing learners to focus on clinical reasoning rather than technical navigation.

The content and structure of GAMA VROG were validated through iterative feedback from obstetricians as maternal health experts, medical and health profession education experts, and practicing midwives during the previous phase of the study. Besides, preliminary usability and content validation were conducted. However, we acknowledge that a full-scale psychometric validation of the VR platform has not yet been completed.

2.2. Trial Design

This study employed a non-blinded, parallel-group randomized controlled trial design, which is common in educational research involving visible interventions such as VR. Participants were randomly assigned to either the control group (mannequin-based training) or the intervention group (VR-based training) using a 1:1 allocation ratio. Each participant received the assigned intervention once, and there were no deviations or modifications to the protocol after the trial commenced.

To mitigate potential bias despite the non-blinded nature of the study, several safeguards were implemented: participants completed all assessments independently *via* digital forms, ensuring anonymity, and used personal digital devices to avoid group influence; the same standardized instruments were applied for both pre- and post-test evaluations across groups; and training facilitators were not involved in either data collection or analysis. However, it should be noted that no assessor blinding was feasible due to the visible differences between VR and mannequin interventions, and no objective performance metrics, such as OSCEs, were employed.

The null hypothesis (H_0) of this study was that there would be no significant differences between VR-based and mannequin-based training in terms of learning experience, knowledge, perceived skills, and readiness following training. The alternative hypothesis (H_1) posited that VR-based training significantly improves the learning experience compared to mannequin-based training.

2.3. Participants

The study participants were practicing midwives currently providing maternal healthcare services across primary, secondary, and tertiary healthcare facilities in the Special Region of Yogyakarta, Indonesia. Inclusion criteria were: (1) being an active practicing midwife at level I, II, or III health care facilities; (2) holding a valid registration certificate (STR); and (3) providing informed consent to voluntarily participate in the study. Exclusion criteria included a history of vertigo, severe motion sickness, or balance disorders, which are known risk factors for

cybersickness during immersive VR experiences. Cybersickness-characterized by nausea, dizziness, and disorientation-is a well-documented side effect of VR delivered *via* Head-Mounted Displays (HMDs) and may interfere with both safety and learning engagement [14, 15]. Participants who submitted incomplete responses or withdrew before completing the post-test assessment were also excluded. Demographic variables-including age, education level, and work experience-were collected and reported descriptively to characterize the study sample. However, due to the limited sample size, these variables were not included as covariates in the primary inferential analyses to avoid overfitting and preserve statistical power.

2.4. The Training and Interventions

This incidental training was part of the CPD program organized by the Yogyakarta branch of the Indonesian Midwives Association. All participants followed a standardized training agenda, which began with a pre-test to assess their baseline knowledge, perceived skills, and readiness in managing PPH. This was followed by a 120-minute refresher session delivered through lectures and facilitated discussions, ensuring consistent content across all participants.

Following the knowledge session, participants were randomly allocated into two groups-the intervention group (VR-based training) and the control group (mannequin-based training)-using an online randomization platform. Each group underwent a 90-minute practice session followed by a 30-minute structured debriefing. The simulation was based on the same clinical scenario and learning objectives across both groups. Each simulation room was equipped with three GAMA VROG VR units or three birthing mannequins, respectively. With this setup, participants rotated through the stations, receiving 15 minutes of direct hands-on training and 3 minutes of preparation time per person. The complete agenda is presented in Table 1.

Technical facilitators (identical for both groups) were trained to maintain consistency in instruction, timing, and facilitate data collection. Although the VR and mannequin sessions were conducted in separate rooms, instructional materials, facilitator scripts, and task sequences were identical to ensure standardization. Immediately after their simulation session, participants completed a structured questionnaire evaluating their learning experience. The training concluded with a unified debriefing session. One week after the intervention, a post-test-identical to the pre-test-was administered to assess any changes in knowledge, perceived ability, and readiness in handling PPH.

Although training time, task structure, and facilitator's interaction were standardized, the immersive nature of VR may evoke different levels of cognitive load, emotional arousal, and situational presence compared to mannequin training. These differences may influence learners' perception and retention, and were not quantitatively assessed in this study.

Table 1. Rundown agenda of training.

Time	Activity Agenda	
08.30 - 09.00	Re-registration	
09.00 - 09.30	Pre-test	
09.30 - 11.30	Refreshment of knowledge - postpartum hemorrhage management	
11.30 - 12.00	Explanation of research and randomization	
12.00 - 13.00	Breaks	
13.00 - 14.30	Control Group: Exercise using a mannequin	Intervention Group: Exercise using GAMA VROG
14.30 - 15.00	Filling out the questionnaires about their learning experience using a mannequin	Filling out the questionnaires about their learning experience using GAMA VROG
15.00 - 15.30	Debriefing	
15.30 - 16.00	Closing	
	Post-test (1 week after pre-test)	

2.5. Instruments

On the day of each training batch, all participants completed an online pre-test using Google Forms on their respective gadgets prior to receiving a 120-minute refresher session on PPH management. The same instruments were used for the post-test, which was conducted one week later. The instruments assessed participants' knowledge, perceived ability, and perceived readiness in handling PPH cases.

To assess knowledge, participants answered 20 multiple-choice questions covering theoretical and procedural aspects of PPH management. To measure perceived ability and readiness, participants completed a self-assessment of 28 key clinical tasks related to PPH using four-point Likert scales. For perceived ability, the scale was adapted to reflect levels aligned with Miller's Pyramid of clinical competence:

(1) Knows - the participant perceive understands the theoretical concept; (2) Knows - how the participant perceive has observed or demonstrated the procedure; (3) Shows - the participant perceive able to performs the skill under supervision or with team collaboration; and (4) Does - the participant perceive able to perform the skill independently. This study was specifically designed to measure self-perception and does not equate to objective performance outcomes; therefore, this scale was designed to capture self-perceived competence, not actual performance. Similarly, perceived readiness was assessed using a four-point Likert scale: (1) Very unprepared, (2) Unprepared, (3) Ready/prepared, and (4) Very ready/Very prepared. The total scores for perceived ability and readiness were calculated by summing item responses and dividing by the maximum possible score (Table 2).

Table 2. Instruments to assess the perceived skills and readiness in handling PPH cases.

No.	List of Skills related to PPH Management	Perceived Ability				Readiness Level			
		1	2	3	4	1	2	3	4
1	Vital Signs Monitoring								
2	Infection control and prevention in each treatment								
3	Implementation of patient safety in every treatment								
4	Intravenous insertion								
5	Urinary catheter insertion								
6	Physical examination								
7	Monitoring the patient's level of consciousness								
8	Using a speculum for examination								
9	Administering drugs in various ways								
10	Hydration and rehydration management (fluid balance)								
11	Oxygen installation								
12	Patient positioning								
13	Basic life support								
14	Interpersonal communication/counseling								
15	Communication, information, and education								
16	Providing motivation								
17	Referral								
18	Documentation								
19	Examination of the amount of vaginal blood discharge								
20	Examination of birth canal wounds								

(Table 2) contd.....

No.	List of Skills related to PPH Management	Perceived Ability				Readiness Level			
		1	2	3	4	1	2	3	4
21	Suturing of grade 1 and grade 2 perineum rupture								
22	Suturing of the portio rupture								
23	Stage IV of labour monitoring								
24	Manual placenta with bleeding								
25	Bimanual compression (external, internal)								
26	Catheter condom insertion								
27	Initial management of the most frequent emergency cases in labour (postpartum hemorrhage - uterine massage)								
28	Initial management of basic emergencies on the maternity mother (cardio-respiratory arrest, hemorrhage shock, shortness of breath and fainting)								

Additionally, a five-item questionnaire was used to explore participants' learning experience after engaging in either mannequin-based or VR-based practice. The questions focused on whether the training medium provided contextual learning, fun learning, enhanced focus, interactive engagement, and increased confidence in performing procedures. This instrument was subjective by design and intended for reflective evaluation in the CPD context, rather than objective performance assessment.

All instruments underwent content validation through expert review by obstetrics professionals as well as medical and health profession education experts. Validity testing using the Pearson product-moment correlation confirmed that all items were valid ($p < 0.05$). Reliability testing using Cronbach's alpha demonstrated excellent internal consistency, with values of 0.95 and 0.93 for the respective instruments. However, as no factor analysis was conducted, we acknowledge that the psychometric strength of the "learning experience" tool is limited and should be interpreted accordingly.

2.6. Outcome Measures

The present study assessed and compared participants' learning experiences, knowledge, perceived ability, and readiness in managing PPH across control and intervention groups, both before and after training. All participants completed the outcome questionnaires prior to the knowledge refresher session on the training day and again one week later to evaluate the intervention's impact. Baseline comparisons between groups were conducted to detect any initial differences in knowledge, perceived ability, or readiness level. No changes to the outcome measures were made after the study commenced.

Due to logistical constraints and the scale of the training, implementing resource-intensive measures such as instructor ratings or video-based assessments was not feasible at this stage. To partially address this issue, triangulation was applied through the inclusion of both objective knowledge assessments (multiple-choice questions) and subjective measures of perceived ability and readiness. The design of this study aimed to measure self-perception, not to assess actual performance; therefore, reliance on such self-reported data may introduce bias. Furthermore, a five-item instrument was utilized to explore participants' learning experiences post-intervention, providing additional evaluative depth.

In order to minimize potential response bias associated with the repeated use of the same questionnaire, specific procedural safeguards were implemented. These included randomized ordering of questionnaire items and the removal of item numbers in both pre- and post-tests to reduce memorization effects and answer pattern recognition.

2.7. Sample Size

The sample size for this study was calculated using the minimum sample size formula by Lemeshow: $n = (Z^2 \times N \times p \times (1-p)) / (d^2 \times (N-1)) + (Z^2 \times p \times (1-p))$, with a confidence level of 95%, a degree of precision of 0.1 (corresponding to a 90% confidence level), and an estimated population proportion of 0.5. Based on a total target population of 2,976 practicing midwives, this yielded a minimum required sample size of 78 participants, equally divided into the control and intervention groups. This sample size was determined to achieve a statistical accuracy level of approximately 89% [17].

To recruit participants, the research team collaborated with the Yogyakarta Branch of IBI, disseminating announcements through midwives' WhatsApp groups to maximize outreach and participation. Interested participants registered through a Google Form platform after receiving detailed study information and giving their informed consent. A total of 90 practicing midwives enrolled in the study and were allocated randomly to one of the three available training batch schedules, all conducted in October 2023 at the IBI Yogyakarta branch office. Each training batch was capped at 30 participants. The research team screened all registrants for eligibility, and ineligible individuals were excluded. Recruitment was concluded once the minimum sample size and the maximum capacity for each training session were reached. Although participants were trained in three separate batches, the number of clusters ($n = 3$) was too small to permit reliable multilevel modelling or cluster-robust adjustments, as such methods require a larger number of clusters to yield stable variance estimates. Therefore, batch effects were not statistically modeled.

2.8. Randomization and Blinding

Participants were randomly allocated to either the VR or the mannequin group in a 1:1 ratio. Randomization was performed using an online sequence generator (randomlists.com), chosen for its transparency and repro-

ducibility. To minimize potential allocation bias, participant identification numbers were assigned prior to randomization, and the principal investigator was not involved in the assignment process. While this platform is less robust than specialized research software, all training protocols and participant characteristics were balanced at baseline, reducing the likelihood of systematic bias.

Randomization was performed separately within each training batch, and participants were divided into control or intervention groups accordingly. Participants were informed of their group assignment shortly before the practice session began. The non-blinded design of the study may influence subjective outcomes such as perceived readiness and learning experience. Due to the nature of the intervention (VR vs. mannequin), participant blinding was not feasible. However, allocation concealment was not performed, and we have acknowledged this as a methodological limitation of the study.

2.9. Statistical Method

The normality tests (Shapiro-Wilk) were performed prior to analysis. The normal data (Shapiro-Wilk $p > 0.05$) were analyzed with parametric tests (independent t-test), while non-normal data were analyzed with non-parametric tests (Mann-Whitney U, Wilcoxon signed-rank) [18, 19]. The choice of statistical tests was therefore driven by the characteristics of the data distribution to ensure appropriate and valid analyses.

To address the potential for Type I error inflation resulting from multiple outcome variables and statistical comparisons, a Bonferroni correction was applied to adjust the significance threshold. Given that a total of 17 hypothesis tests were conducted, the alpha level was adjusted from 0.05 to 0.0029 (0.05/17). Accordingly, statistical significance was defined as $p < 0.0029$ for all comparisons. This adjustment was implemented to enhance the rigor of our analysis and reduce the likelihood of false-positive findings. Effect sizes were calculated using Cohen's d for independent t-tests, and the r coefficient was used for non-parametric tests, including the Wilcoxon Signed Ranks Test and the Mann-Whitney U test.

3. RESULT

Ninety-five practicing midwives were registered to participate in this study. After excluding five respondents who did not meet the inclusion criteria, a total of ninety participants were eligible. They were divided into three batches of training conducted in October 2023. Within each batch, participants were randomly assigned to either the control group (mannequin-based training) or the intervention group (VR-based training), with 15 participants in each group per batch. Participant flow is illustrated in Fig. (2). The participants had diverse demographic and professional backgrounds, as detailed in Table 3 and Fig. (3).

Table 3. The characteristics of respondents.

No.	Characteristic	Frequency (n = 90)	Percentage (%)
1	Age	-	-
	20-35 years old	56	62%
	36-50 years old	28	31%
	>50 years old	6	7%
2	Education	-	-
	3-year associate degree	52	58%
	4-year vocational degree	11	12%
	Bachelor	5	6%
	Profession	12	13%
	Master	8	9%
	Doctoral	2	2%
3	Work Place	-	-
	Primary health care centre	41	46%
	Secondary health care centre	39	43%
	Tertiary health care centre	10	11%
4	District Of Work Place	-	-
	Yogyakarta City	15	17%
	Sleman	25	28%
	Bantul	6	7%
	Kulon Progo	13	14%
	Gunung Kidul	18	20%
	Others	13	14%

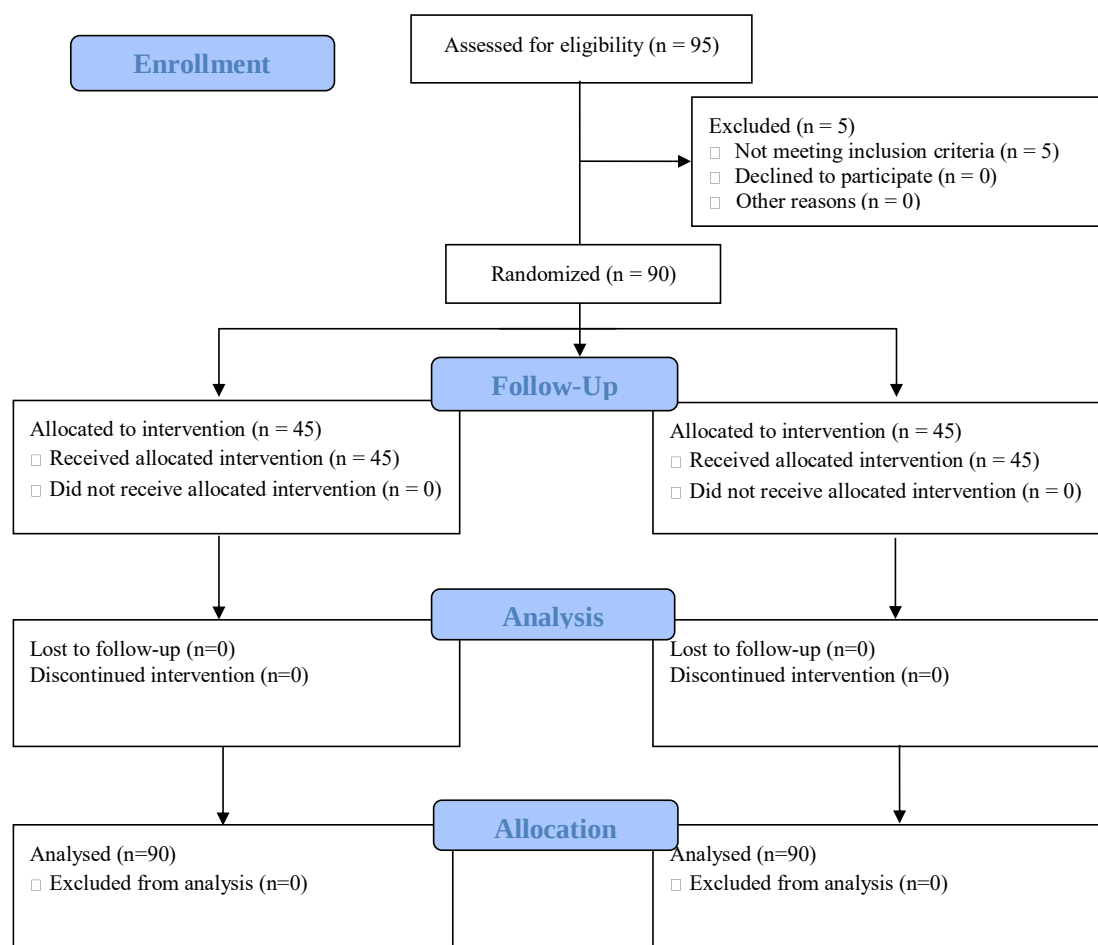


Fig. (2). CONSORT 2010 flow diagram.

Available online under the terms of the Creative Commons Attribution Non-Commercial License 4.0. [16] Creative Commons Attribution-Non Commercial-ShareAlike License (CC BY-NC-SA).

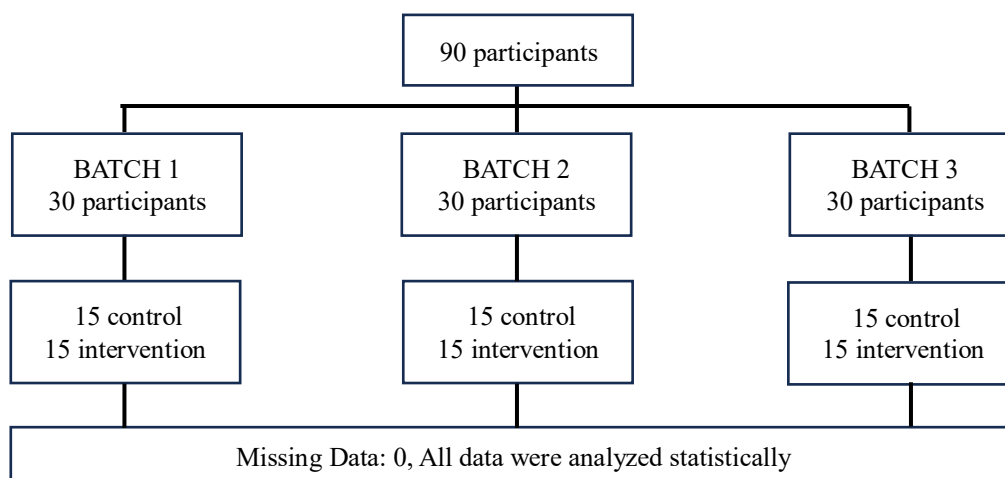


Fig. (3). Participant's flow.

Table 4. Learning experiences using mannequins and virtual reality.

No.	Item	Mean Score		Sig ^a
		Mannequin	GAMA VROG	(Effect Size)
1	The learning media (VR/mannequin) provide a learning experience that closely mirrors a real clinical situation.	32.97	58.03	0.000 (-0.508)
2	The learning media (VR/mannequin) provides a “fun” learning experience.	36.83	54.17	0.000 (-0.495)
3	The learning media (VR/mannequin) enhances focus in learning.	37.60	53.40	0.001 (-0.348)
4	The learning media (VR/mannequin) provides interactive learning to enhance engagement in the learning process.	37.72	53.28	0.001 (-0.348)
5	The learning media (VR/mannequin) is capable of increasing confidence in performing actions.	40.67	50.33	0.044 (-0.211)

Note: ^aMann-Whitney test.

Table 5. Comparison between the control and intervention groups.

No.	Item	Mannequin		Sig ^a (Effect Size)	Virtual Reality		Sig ^a (Effect Size)	Pre-test	Post-test
		Pre-test M ± SD	Post-test M ± SD		Pre-test M ± SD	Post-test M ± SD		Mannequin vs. VR Sig (Effect Size)	Mannequin vs. VR Sig ^b (Effect Size)
1	Knowledge	55.44 ± 14.09	78.44 ± 13.13	0.000 (-0.94)	50.67 ± 12.09	76.78 ± 14.62	0.000 (-1.12)	0.09 ^c (-0.18)	0.64 (-0.45)
2	Perceived ability	83.37 ± 13.11	87.87 ± 12.64	0.060 (-0.28)	86.6 ± 10.02	89.94 ± 12.44	0.070 (-0.32)	0.26 ^b (-0.12)	0.21 (-0.19)
3	Readiness	84.06 ± 11.03	86.54 ± 8.65	0.257 (-0.16)	83.54 ± 10.9	88.81 ± 8.87	0.015 (-0.37)	0.68 ^b (-0.25)	0.16 (-0.13)

Note: ^aWilcoxon Signed Ranks Test, ^b Mann-Whitney test, ^c Independent t-test.

Table 4 summarizes participants' learning experiences using mannequins and virtual reality (GAMA VROG) across five aspects. The VR group consistently reported significantly higher scores than the mannequin group. Specifically, VR was rated higher in providing contextual learning experiences ($M = 58.03$ vs. 32.97 ; $p = 0.000$; $r = -0.508$), fun learning experiences ($M = 54.17$ vs. 36.83 ; $p = 0.000$; $r = -0.495$), enhanced focus ($M = 53.40$ vs. 37.60 ; $p = 0.001$; $r = -0.348$), interactive engagement ($M = 53.28$ vs. 37.72 ; $p = 0.001$; $r = -0.348$), and confidence building ($M = 50.33$ vs. 40.67 ; $p = 0.044$; $r = -0.211$). However, when applying the Bonferroni-corrected significance threshold ($p < 0.0029$), only the first four indicators remained statistically significant. The difference in confidence enhancement ($p = 0.044$) did not reach the adjusted level, suggesting it may reflect a small or variable effect.

Table 5 presents a comparison of learning outcomes (knowledge, perceived ability, and readiness) between the control and intervention groups before and after the training. Both groups exhibited significant within-group improvements in knowledge, with the mannequin group improving from 55.44 to 78.44 ($p = 0.000$; $r = -0.94$) and the VR group from 50.67 to 76.78 ($p = 0.000$; $r = -1.12$), indicating large effect sizes. These knowledge improvements remained highly significant after Bonferroni correction. In contrast, perceived ability did not significantly improve in either group based on the corrected threshold (mannequin: $p = 0.060$; $r = -0.28$; VR: $p = 0.070$; $r = -0.32$). Readiness showed a statistically significant

improvement in the VR group (83.54 to 88.81 ; $p = 0.015$; $r = -0.37$) but did not show a significant post-correction due to not meeting the Bonferroni-corrected criterion ($p < 0.0029$). Between-group comparisons for post-test scores showed no statistically significant differences in knowledge ($p = 0.64$; $r = -0.45$), perceived ability ($p = 0.21$; $r = -0.19$), or readiness ($p = 0.16$; $r = -0.13$), further confirming the absence of strong evidence for VR superiority when controlling for multiple comparisons. Taken together, these results suggest that both VR and mannequin-based training were effective in improving knowledge, but no clear superiority was observed in enhancing perceived ability or readiness. The study did not control for covariates such as participants' age, years of clinical experience, or prior exposure to digital tools. Multivariate analyses (*e.g.*, regression, ANCOVA) were not performed due to sample size limitations. Therefore, interpretations regarding the comparative effectiveness of each modality should be made with caution.

4. DISCUSSION

This study evaluated the effectiveness of VR-based training compared to mannequin-based training in enhancing midwives' knowledge, perceived skills, and readiness in managing PPH. While both training methods significantly improved knowledge, neither led to statistically significant gains in perceived skills. Notably, although the VR group showed a statistically significant within-group increase in readiness ($p = 0.015$), this did not reach the

Bonferroni-adjusted alpha level ($p < 0.0029$); hence, it should be interpreted with caution. Rather than interpreting this as evidence that VR is superior, we view these findings as evidence that VR and mannequin training each have distinct strengths.

Rather than viewing VR as definitively superior, our findings highlight its unique contributions to learning experiences, particularly in terms of engagement and enjoyment. VR consistently outperformed mannequins in subjective measures of learning experience across all five indicators, including contextual realism, fun, focus, interactivity, and confidence. This means that VR's primary advantage lies in enhancing learner engagement, not necessarily in producing better outcomes. These findings align with prior literature suggesting that immersive environments can boost learner engagement [20], but our results underscore that this engagement may not directly translate into improved procedural or psychomotor performance.

This discrepancy may be illuminated by educational theory. Kolb's experiential learning theory [21] emphasizes the importance of active experimentation and reflective observation for skill acquisition. While VR offers immersive observation and conceptual engagement, it lacks the tactile fidelity essential for practicing psychomotor tasks. Norman *et al.* (2012) similarly argue that functional fidelity-how well a simulation supports the desired learning outcomes-is more important than its technological realism, suggesting that VR may be limited in achieving certain clinical competencies [22]. In other words, the engagement that VR creates does not always guarantee improved hands-on performance, especially for complex psychomotor skills. In addition, VR carries practical limitations, including reduced skill fidelity compared to high-fidelity mannequins, higher implementation costs, and potential digital fatigue or cybersickness, all of which may restrict its scalability in real-world training programs [23].

Cognitive load theory also provides insights into the observed outcomes. VR, although engaging, may impose an extraneous cognitive load on novice learners due to its sensory complexity, potentially distracting from skill acquisition [20]. Learners might allocate cognitive resources to navigating the environment rather than mastering the task. This theory supports the finding that while VR increased readiness perceptions, it failed to enhance perceived skills. This also suggests that overconfidence may develop if increased readiness is not balanced with actual practice of skills. Future research should analyze the differential cognitive load of VR *versus* mannequin training, using Cognitive Load Theory as a guiding framework.

The increase in perceived readiness within the VR group could also reflect an overconfidence bias, a phenomenon in which learners' self-assessment exceeds their actual capabilities. Kovacs *et al.* (2020) warned that this cognitive bias can arise in simulation-based education, especially when learners are exposed to advanced visual environments without corresponding psychomotor challenges [12]. In our study, confidence was enhanced by VR's immersive elements, yet this was not paralleled by demonstrable skill improvement.

Another explanation may lie in the limited duration of exposure. Al-Saud *et al.* (2017) emphasized the importance of repeated practice and feedback in achieving skill mastery [24]. Our single-session training lacked follow-up, formative assessment, or structured reflection, which are critical components of sustained skill development. Future VR-based modules should consider longitudinal delivery, incorporating spaced repetition and immediate feedback mechanisms.

The absence of statistically significant differences in skill acquisition between groups also prompts a discussion on the training design and assessment tools. Our study relied on self-perceived measures, which may not accurately capture procedural competence. Incorporating Objective Structured Clinical Examinations (OSCEs) or direct observation assessments could yield more valid insights into actual performance changes attributable to training modalities.

Moreover, despite randomization, we did not perform regression or ANCOVA to adjust for potential confounders such as age, prior training, or clinical experience. This limits our ability to attribute outcomes solely to the intervention. Although randomization was applied to allocate participants into intervention and control groups, no formal statistical test was conducted to confirm baseline equivalence (Table 3). Therefore, potential baseline differences should be interpreted cautiously.

While VR is often celebrated for its scalability and potential cost-efficiency, our study did not evaluate infrastructure or economic feasibility. Implementing VR in resource-limited settings involves costs for hardware, software, training, and maintenance. Without a cost-effectiveness analysis, it is premature to advocate large-scale adoption. Furthermore, implementation barriers such as digital literacy, institutional readiness, and maintenance logistics must be considered.

Lastly, potential harms associated with VR remain underexplored. Simulation fatigue, cybersickness, and visual strain have been reported in the literature [14]. Moreover, overreliance on VR may inadvertently erode learners' interest in tactile, high-fidelity practice. Future research should systematically assess adverse effects and explore mitigation strategies, including ergonomic design and optimal exposure time.

Given these nuanced findings, we propose that VR-based training be considered as a complementary tool within the broader framework of CPD for midwives. It excels in cognitive and affective engagement and can simulate complex, rare clinical scenarios with consistency. A hybrid model combining VR and traditional mannequin simulations could provide a balanced, context-rich training environment. To support sustainable integration into CPD, future studies should assess long-term outcomes, stakeholder acceptability, and return on investment while incorporating robust educational frameworks and implementation science approaches.

In summary, our findings do not demonstrate outcome superiority of VR over mannequin training. Instead, VR should be recognized for its ability to enrich learner

engagement and perceived readiness. A blended training model that combines VR for immersive cognitive and emotional preparation with mannequin-based practice for psychomotor skill rehearsal is likely to provide the most balanced and effective CPD experience for midwives.

5. STRENGTHS AND LIMITATIONS

This study demonstrates several important strengths. Foremost, its randomized controlled design enhances internal validity and provides a robust comparison between VR- and mannequin-based training. The clinical focus on postpartum hemorrhage, a leading cause of maternal mortality, ensures that the findings are directly relevant to maternal health practice and professional training priorities. Unlike many prior studies that assess only knowledge or skills, this research integrates both cognitive and affective learning outcomes, offering a more holistic evaluation of training impact. The study also captures learners' subjective experiences, providing novel insights into how VR influences engagement, confidence, and perceived readiness—dimensions often overlooked in traditional simulation research. In addition, rigorous statistical procedures were applied, including the use of non-parametric analyses with the Bonferroni correction, to minimize Type I error and enhance the reliability of results. Finally, the findings are grounded in established educational theories, such as experiential learning and cognitive load theory, strengthening the theoretical relevance and transferability of the conclusions.

However, this study also has several limitations. This study has several limitations. The GAMA VROG platform has not undergone peer-reviewed psychometric validation, and randomization was performed with an online tool without allocation concealment. Objective assessments such as OSCEs or assessor ratings were not conducted, and the reliance on self-reported measures of perceived skills and readiness may introduce bias, as these reflect self-perception rather than actual performance. Clustering by training batches was not modeled statistically due to the small number of clusters, and stratified analyses by age, experience, or prior VR exposure were not feasible with the limited sample size. Multivariate analyses (*e.g.*, regression, ANCOVA) were also not performed, restricting adjustment for confounders. In addition, the short duration of VR exposure and the absence of long-term follow-up limit the conclusions regarding retention. Finally, higher implementation costs, digital fatigue, and lack of cost-effectiveness evaluation may constrain generalizability. By acknowledging both the strengths and limitations of VR, the study offers a balanced foundation for its proposed use as a complementary tool in CPD for midwives.

CONCLUSION

This study demonstrates that both VR-based and mannequin-based training significantly improves midwives' knowledge in managing PPH. While VR training enhances learners' perceptions of readiness, it does not produce statistically significant improvements in perceived skills compared to traditional training. These findings underscore

that VR should be regarded as a complementary educational tool in CPD, rather than a replacement for hands-on simulation. A hybrid training model that integrates VR for immersive cognitive and emotional preparation with traditional mannequin-based practice for psychomotor rehearsal is likely to offer the most balanced benefit. This blended approach provides a sustainable pathway for CPD in maternal emergency care, maximizing the strengths of both modalities while mitigating their individual limitations. Future programs should consider longitudinal designs, incorporate objective performance assessments, and evaluate logistical feasibility and cost-effectiveness to ensure sustainable and impactful integration of VR into maternal emergency training.

DECLARATION OF ARTIFICIAL INTELLIGENCE USE

Artificial Intelligence (AI) assistance was used in this study solely for language refinement and grammar editing purposes, using ChatGPT (OpenAI, GPT-4). The authors affirm that all content related to study design, data interpretation, and scientific reasoning was developed, validated, and approved by the authors themselves. No original scientific content was generated by AI, and final responsibility for the integrity and interpretation of the findings rests with the authors.

AUTHORS' CONTRIBUTIONS

The authors confirm contribution to the paper as follows: I.P.S., Y.S.: Responsible for the study and manuscript's conceptualization, data collection, data analysis, data interpretation, and writing the paper; D.W.: Writing and refining the paper; O.E.: Study concept and refining the paper. All authors are acknowledged to have taken full responsibility for the content of the manuscript and agreed to its submission. They have thoroughly examined the findings and collectively endorsed the final version for publication.

LIST OF ABBREVIATIONS

VR	= Virtual Reality
GAMA	= Gadjah Mada Virtual Reality on
VROG	Obstetrics and Gynecology
PPH	= Postpartum Hemorrhage
LMICs	= Low and Middle Income Countries
SDGs	= Sustainable Development Goals
CPD	= Continuing Professional Development
HMD	= Head-Mounted Display
OSCE	= Objective Structured Clinical Examination
AI	= Artificial Intelligence

ETHICS APPROVAL AND CONSENT TO PARTICIPATE

Ethical approval was obtained from the Ethics Committee of the Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Yogyakarta, Indonesia with approval number KE/FK/1463/EC/2022.

HUMAN AND ANIMAL RIGHTS

All procedures involving human participants were conducted in compliance with the ethical guidelines established by both institutional and national research ethics committees, and conformed to the principles outlined in the Declaration of Helsinki (1975), including its 2013 revision.

CONSENT FOR PUBLICATION

Informed consent was obtained from all participants of this study.

STANDARDS OF REPORTING

CONSORT guidelines were followed.

AVAILABILITY OF DATA AND MATERIALS

The data supporting the findings of the article is available in the Zenodo repository at <https://doi.org/10.5281/zenodo.17266267>, reference number (between 19-20 in result section) with reference: Setiawan, I. P. (2025). Dataset of Research about Comparison VR and manequin for learning [Data set].

FUNDING

This research was supported by a grant from the Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Indonesia (Award Grant Number: 1045/UN1/FK-KMK.2/AK.1/PJ/2022).

CONFLICT OF INTEREST

The authors declare no conflict of interest, financial or otherwise.

ACKNOWLEDGEMENTS

The author expresses gratitude to the Indonesian Midwife Association (IBI) Yogyakarta branch for granting permission and providing support in organizing the training that contributed to this research. Appreciation is also extended to Mayriyana Kartikasari, MKM, for her invaluable assistance with the technical and administrative aspects of this study.

REFERENCES

- [1] Pertiwi TS, Temesvari NA, Nurmallasari M. Spatial patterns of maternal mortality causes in West Kalimantan, Indonesia. *Public Health Indones* 2021; 7(3): 101-10. <http://dx.doi.org/10.36685/phi.v7i3.441>
- [2] Maternal mortality. 2025. Available from: <https://www.who.int/news-room/fact-sheets/detail/maternal-mortality>
- [3] Rencana Aksi Kegiatan Direktorat Kesehatan Keluarga. 2020. Available from: https://ditjen-sdmk.kemkes.go.id/be/storage/upload/reports/42501_reports.pdf
- [4] Webber J, Moran K, Cumin D. Paediatric cardiopulmonary resuscitation: Knowledge and perceptions of surf lifeguards. *J Paediatr Child Health* 2019; 55(2): 156-61. <http://dx.doi.org/10.1111/jpc.14097> PMID: 29943876
- [5] Rourke S. How does virtual reality simulation compare to simulated practice in the acquisition of clinical psychomotor skills for pre-registration student nurses? A systematic review. *Int J Nurs Stud* 2020; 102103466 <http://dx.doi.org/10.1016/j.ijnurstu.2019.103466> PMID: 31783192
- [6] Hijazi H, Baniissa W, Al Abdi R, *et al.* Experiences of work-related stress among female healthcare workers during the COVID-19 public health emergency: A qualitative study in the United Arab of Emirates. *Psychol Res Behav Manag* 2022; 15: 2701-15. <http://dx.doi.org/10.2147/PRBM.S381177> PMID: 36172543
- [7] Tefera M, Mezmur H, Jemal M, Assefa N. Midwives' experiences of performing obstetric ultrasounds in antenatal care in eastern Ethiopia: Qualitative exploratory study. *Womens Health* 2024; 2017455057241228135 <http://dx.doi.org/10.1177/17455057241228135> PMID: 38366806
- [8] Carr KC. Using the unfolding case study in midwifery education. *J Midwifery Womens Health* 2015; 60(3): 283-90. <http://dx.doi.org/10.1111/jmwh.12293> PMID: 25959399
- [9] Wenli Lian N. Application of virtual reality technology and its impact on digital health in healthcare industry. *J Commer Biotechnol* 2023; 27(4) <http://dx.doi.org/10.5912/jcb1320>
- [10] Ruthenbeck GS, Reynolds KJ. Virtual reality for medical training: The state-of-the-art. *Journal of Simulation* 2015; 9(1): 16-26. <http://dx.doi.org/10.1057/jos.2014.14>
- [11] Chiang DH, Huang CC, Cheng SC, *et al.* Immersive virtual reality (VR) training increases the self-efficacy of in-hospital healthcare providers and patient families regarding tracheostomy-related knowledge and care skills: A prospective pre-post study. *Medicine* 2022; 101(2): e28570. <http://dx.doi.org/10.1097/MD.00000000000028570> PMID: 35029229
- [12] Kovacs R, Lagarde M, Cairns J. Overconfident health workers provide lower quality healthcare. *J Econ Psychol* 2020; 76: 102213. <http://dx.doi.org/10.1016/j.joep.2019.102213>
- [13] Dunlop K, Dillon G, McEvoy A, *et al.* The virtual reality classroom: A randomized control trial of medical student knowledge of postpartum hemorrhage emergency management. *Front Med* 2024; 11: 1371075. <http://dx.doi.org/10.3389/fmed.2024.1371075> PMID: 38566920
- [14] Rebenitsch L, Owen C. Review on cybersickness in applications and visual displays. *Virtual Reality* 2016; 20: 101-25. <http://dx.doi.org/10.1007/s10055-016-0285-9>
- [15] Lawson BD, Stanney KM. Editorial: Cybersickness in virtual reality and augmented reality. *Front Virtual Real* 2 <http://dx.doi.org/10.3389/frvir.2021.759682>
- [16] Sarriidou DG, Chalmouki G, Braoudaki M, Siafaka I, Asmatzi C, Vadalouka A. Parecoxib possesses anxiolytic properties in patients undergoing total knee arthroplasty: A prospective, randomized, double-blind, placebo-controlled, clinical study. *Pain Ther* 2016; 5(1): 55-62. <http://dx.doi.org/10.1007/s40122-016-0046-1> PMID: 26861666
- [17] Ath-Thahirah AS, Anak Agung Gede Eka Septian Utama N, Agung Wiwiek Indrayani N, Gede Parta Kinandana N. The usage and weight of backpacks are associated with shoulder pain complaints among elementary students. *Phys Ther J Indon* 2024; 5(1): 61-5. <http://dx.doi.org/10.51559/ptji.v5i1.190>
- [18] Aliberti S, D'Elia F, Cherubini D. Tips for statistical tools for research methods in exercise and sport sciences. *Phys Educ Theory Methodol* 2023; 23(3): 470-7. <http://dx.doi.org/10.17309/tmfv.2023.3.20>
- [19] Maharjan A, Wang E, Peng M, Cakmak YO. Improvement of olfactory function with high frequency non-invasive auricular electrostimulation in healthy humans. *Front Neurosci* 2018; 12(225): 1-14. <http://dx.doi.org/10.3389/fnins.2018.00225> PMID: 29740266
- [20] Mayer RE. The Cambridge handbook of multimedia learning. Cambridge handbooks in psychology. New York: Cambridge University Press 2014. <http://dx.doi.org/10.1017/CBO9781139547369>
- [21] Kolb DA. *Experiential Learning: Experience As The Source Of Learning And Development*. Prentice-Hall 2015.
- [22] Norman G, Dore K, Grierson L. The minimal relationship between simulation fidelity and transfer of learning. *Med Educ* 2012; 46(7):

- 636-47.
<http://dx.doi.org/10.1111/j.1365-2923.2012.04243.x> PMID: 22616789
- [23] Yu J, Wu J, Lu J, *et al.* Efficacy of virtual reality training on motor performance, activity of daily living, and quality of life in patients with Parkinson's disease: An umbrella review comprising meta-analyses of randomized controlled trials. *J Neuroeng Rehabil* 2023; 20(1): 133.
<http://dx.doi.org/10.1186/s12984-023-01256-y> PMID: 37777748
- [24] Al-Saud LM, Mushtaq F, Allsop MJ, *et al.* Feedback and motor skill acquisition using a haptic dental simulator. *Eur J Dent Educ* 2017; 21(4): 240-7.
<http://dx.doi.org/10.1111/eje.12214> PMID: 27324833