Comparative Effectiveness of Virtual Reality Simulation versus High-Fidelity Manikins on Critical Care Nurses' Long-Term Retention and Clinical Competency in Complex Arterial Blood Gas Matrix Interpretation

Pranjal Lavin Robinson[1*]
Assistant Professor, SAGE college of Nursing, Bhopal, Madhya Pradesh

Dr Sudha swamy[2]
Vice Principal, College of Nursing, ASMC Shahjahanpur, Uttar Pradesh

Aarushi Saxena[3]
P.G. Tutor, Teerthanker Mahaveer University, Moradabad, Uttar Pradesh

Rooben Lall[4]
Assistant Professor, PIPRAMS Greater Noida, Uttar Pradesh

Neeru Yadav[5]
Assistant Professor, PIPRAMS Greater Noida, Uttar Pradesh

Abstract

Arterial blood gas (ABG) interpretation is among the most cognitively demanding psychomotor-cognitive hybrid competencies required of critical care nurses, requiring rapid, sequential integration of pH, partial pressure of carbon dioxide, bicarbonate, oxygenation status, and compensatory patterns into a single time-sensitive clinical judgment. Simulation-based education has become the dominant strategy for teaching and assessing this competency outside live patient encounters, with high-fidelity manikins (HFM) historically serving as the reference standard and virtual reality (VR) simulation emerging as a scalable, increasingly immersive alternative. This review synthesizes the current comparative evidence on VR and HFM modalities across three outcome domains relevant to complex ABG matrix interpretation: immediate knowledge and skill acquisition, long-term retention, and transfer of clinical competency to authentic practice. Drawing on systematic reviews, meta-analyses, randomized and quasi-experimental trials, and the broader skill-decay literature, the review finds that VR and HFM produce broadly comparable gains in knowledge and self-confidence immediately post-training, that HFM retains an advantage for psychomotor and team-based crisis skills, and that VR offers meaningful advantages in repeatability, standardization, cost, and scalability that are particularly relevant to the deliberate, repeated practice ABG interpretation requires for durable retention. No study identified in this review directly and exclusively evaluated VR versus HFM training on ABG interpretation using a validated long-term retention design, revealing a substantive gap in the evidence base. The review concludes with a proposed blended-modality framework for ABG education and recommendations for future comparative trials incorporating retention intervals of six months or longer and validated competency instruments.

Keywords: virtual reality simulation, high-fidelity manikin, arterial blood gas interpretation, critical care nursing education, skill retention, clinical competency, simulation-based learning

1. Introduction

Critical care nursing practice is defined by the requirement to synthesize large volumes of physiological data under time pressure and translate that synthesis into safe, prioritized action. Few competencies illustrate this demand as sharply as arterial blood gas (ABG) interpretation. A single ABG panel encodes information about ventilation, oxygenation, and acid-base status simultaneously, and the clinical significance of any one value cannot be determined without reference to the others. Nurses who staff intensive care units (ICUs), emergency departments, and post-anesthesia units are routinely expected to recognize primary respiratory or metabolic derangements, distinguish compensated from uncompensated states, identify mixed disorders, and correlate these patterns with the evolving clinical picture, often before a physician or advanced practice provider is at the bedside.

Despite its centrality to safe practice, the literature consistently documents gaps in ABG interpretation competency among both novice and experienced critical care nurses. Structured teaching programs and simulation-based learning have repeatedly been identified as effective strategies for closing this gap, but the optimal simulation modality for teaching a skill that is simultaneously cognitive, pattern-based, and high-stakes remains unresolved.

Two simulation paradigms dominate contemporary critical care nursing education. High-fidelity manikins (HFM) are computerized, physiologically responsive mannequins capable of simulating vital signs, breath sounds, and pharmacologic responses within a physically realistic patient-care environment. Virtual reality (VR) simulation, by contrast, immerses the learner in a computer-generated clinical environment, delivered via head-mounted displays, desktop interfaces, or non-immersive screen-based platforms, and increasingly incorporates adaptive branching scenarios, physiological modeling, and automated performance analytics. Both modalities have an evidence base supporting their use in nursing education broadly, but comparative evidence specific to complex, matrix-style cognitive interpretation tasks such as ABG analysis, and specific to long-term retention rather than immediate post-test performance, remains fragmented.

This review addresses three linked questions. First, how do VR and HFM compare in producing immediate gains in ABG-relevant knowledge, clinical reasoning, and self-confidence among critical care nurses and nursing students preparing for critical care practice? Second, what does the broader simulation literature indicate about the durability of these gains over weeks to months, and does the available evidence suggest differential retention by modality? Third, how do these modalities compare in supporting transfer of ABG interpretation competency into authentic clinical decision-making? The review closes by identifying the specific evidentiary gaps that limit definitive conclusions and proposing a research and implementation agenda.

2. Background: The Clinical Significance and Complexity of ABG Interpretation

2.1 Why ABG Interpretation Is a Distinct Educational Challenge

ABG analysis provides critical care teams with real-time information about a patient's oxygenation, ventilation, and acid-base balance, and is used to guide decisions ranging from ventilator adjustment to renal replacement therapy initiation. Correct interpretation requires the learner to hold multiple interacting variables in working memory simultaneously: pH as the net indicator of acid-base status, partial pressure of carbon dioxide ($PaCO_2$) as the respiratory component, bicarbonate (HCO_3^-) as the metabolic component, and oxygenation indices as a partially independent but clinically entangled fourth axis. Determining whether a derangement is respiratory or metabolic in origin, and whether compensation is absent, partial, or complete, requires the learner to apply a decision algorithm consistently under time pressure, precisely the kind of task that cognitive load theory identifies as vulnerable to instructional design failures.

A recent multicenter simulation-based performance evaluation of emergency department and intensive care unit nurses found meaningful variability in nurses' ability to correctly interpret blood gas results and translate that interpretation into appropriate urgency-level judgments and therapy recommendations, underscoring that ABG interpretation is not a competency reliably acquired through exposure alone and that structured, simulation-based training and decision-support integration are needed to close performance gaps. A companion review of the ABG literature spanning 2010 to 2025 concluded that knowledge and interpretation gaps persist among nurses and other health professionals despite the technique's importance for early detection of respiratory failure, metabolic derangement, and other life-threatening conditions, and that structured teaching programs combined with simulation-based learning significantly improve competency.

2.2 The Educational Response: Structured Teaching and Simulation

Several quasi-experimental and pre-post intervention studies conducted with intensive care and staff nurses have demonstrated that structured teaching programs meaningfully improve ABG-related knowledge scores relative to baseline, and reviews of ABG education consistently recommend simulation-based training that replicates high-stakes scenarios encountered in emergency department and ICU settings as the preferred vehicle for this instruction, rather than didactic teaching alone. However, the majority of this literature evaluates a single simulation modality against a no-intervention or lecture-based control, rather than comparing simulation modalities against one another, and few studies extend assessment beyond an immediate post-test.

This asymmetry in the evidence base, extensive documentation that simulation improves ABG competency in general, but limited head-to-head modality comparison and minimal long-term follow-up, motivates the present review's focus on drawing inferences from the broader comparative simulation literature and applying them, with appropriate caution, to the specific demands of ABG matrix interpretation.

3. Theoretical Frameworks: Fidelity, Cognitive Load, and Skill Acquisition

3.1 Simulation Fidelity and Its Limits

Simulation fidelity, the degree to which a simulated environment replicates the physical, semantic, and psychological features of authentic practice, has traditionally been treated as a proxy for educational value: the assumption that more realistic simulation produces better learning. High-fidelity manikins were built explicitly on this premise, offering physical presence, tactile feedback, and integration with real clinical equipment and team-based workflows. Naturalistic and social-constructivist learning theories provide a rationale for this approach, holding that learning is most transferable when it is situated in activity that closely resembles genuine professional practice.

Cognitive load theory complicates this picture. Instructional design research applying cognitive load theory to simulation has found that immersing learners in an environment that fully replicates the complexity of real clinical practice, without regard to the learner's existing capacity, can increase extraneous cognitive load and impair learning rather than enhance it. Working memory is limited, and when a task's intrinsic complexity, the germane load associated with genuine schema-building, and extraneous load generated by environmental noise, time pressure, or unfamiliar interfaces together exceed the learner's processing capacity, performance and retention suffer. For a task like ABG interpretation, which is already intrinsically demanding because it requires simultaneous integration of multiple interacting variables, the instructional environment's capacity to add or remove extraneous load is a first-order design consideration, arguably more consequential than physical realism per se.

3.2 Applying Cognitive Load Theory to VR and HFM

This theoretical lens reframes the VR-versus-HFM comparison. High-fidelity manikins introduce a rich, multisensory, team-based environment that closely mirrors bedside practice but may also introduce extraneous cognitive load through equipment management, communication demands, and environmental noise, load that may be justified when the target competency is procedural or team-based crisis management, but potentially counterproductive when the target competency is a discrete cognitive interpretation task. VR simulation, particularly non-immersive or moderately immersive platforms, can be designed to isolate the ABG interpretation task from many of these extraneous demands, presenting the data set, allowing repeated practice, and providing immediate, structured feedback, features consistent with cognitive load theory's recommendations for scaffolded, staged instruction. At the same time, VR's own interface, particularly head-mounted immersive systems, can introduce its own extraneous load in the form of novel controls, simulator sickness, or visual complexity, especially among learners unfamiliar with the technology.

A further relevant framework is Benner's novice-to-expert model, which describes clinical competence as progressing through stages, from rule-based novice performance toward the intuitive, pattern-recognition-based judgment characteristic of expert practice. ABG interpretation is frequently taught initially as a rule-based algorithm and, with repeated exposure, becomes for experienced clinicians a rapid pattern-recognition task. This developmental trajectory has direct implications for simulation modality selection: the repeatability and low marginal cost of VR-based practice may be particularly well suited to supporting the high-volume, spaced repetition required to move a learner from rule-based to pattern-based ABG interpretation, a claim examined further in Section 7.

4. Simulation Modalities Compared: High-Fidelity Manikins and Virtual Reality

4.1 High-Fidelity Manikins

High-fidelity manikins such as those in the Laerdal SimMan and Gaumard Susie families are wireless, computer-controlled simulators capable of producing physiologically responsive vital signs, breath and heart sounds, pupillary responses, and pharmacologic responses to administered medications, embedded within a physical clinical environment complete with real equipment, monitors, and, typically, a multi-person care team. A substantial evidence base, including meta-analytic work, supports HFM as more effective than many traditional teaching methods for building knowledge, technical skill, interprofessional cooperation, and learner engagement among nursing students, and systematic reviews have found simulation-based training broadly useful for strengthening human-factors skills among qualified healthcare teams with downstream implications for patient safety.

4.2 Virtual Reality Simulation

Virtual reality simulation exists on a spectrum from fully immersive, head-mounted-display-based systems to semi-immersive and non-immersive screen-based platforms, sometimes described using virtual-patient software such as Body Interact. VR simulations can incorporate physiological modeling, branching scenario logic, automated scoring, and in some implementations, adaptive difficulty. A scoping review of VR in critical care nursing education found that the modality's immersive and interactive characteristics promote learner engagement, permit unlimited procedural repetition without risk to patients, and support the development of critical thinking and problem-solving skills, while allowing students to progress at an individualized pace. Analyses comparing operating costs have suggested that VR simulations can require substantially less time and lower cost than equivalent high-fidelity simulation to achieve comparable learning outcomes, an efficiency argument increasingly cited in discussions of simulation program scalability.

4.3 Blended and Hybrid Approaches

A growing subset of the literature examines VR as a primer or complement to, rather than a replacement for, manikin-based simulation. One quasi-experimental study comparing virtual patient simulation alone, high-fidelity manikin simulation alone, and virtual patient simulation used as a primer before manikin-based simulation found no statistically significant interaction effect of time by study group on clinical judgment scores, suggesting that a blended sequential approach may produce outcomes comparable to either modality used in isolation, an important finding for programs seeking to combine the scalability of VR with the psychomotor and team-based realism of HFM rather than treating the two as mutually exclusive investments.

5. Methodology of This Review

This review adopts a narrative, comparative-synthesis approach rather than a formal systematic review methodology, reflecting the objective of integrating evidence across three distinct but interlocking literatures: (a) studies directly comparing VR and HFM simulation modalities in nursing and medical education; (b) the ABG interpretation education literature, spanning structured teaching programs, simulation-based ABG training, and nurse performance evaluation studies; and (c) the broader simulation skill-decay and long-term retention literature, which is rarely specific to ABG interpretation but provides the best available evidence on how simulation-acquired competencies behave over time.

Literature was identified through targeted searches of recent systematic reviews, meta-analyses, randomized and quasi-experimental trials, and scoping reviews published predominantly between 2015 and 2026, with priority given to studies published in 2023 through 2026 to capture the rapidly evolving VR evidence base, supplemented by foundational cognitive load theory and skill-decay literature published earlier where no more recent equivalent existed. Because no study meeting strict inclusion criteria, a head-to-head randomized comparison of VR versus HFM specifically on ABG interpretation with a long-term retention arm, was identified, this review draws inferences by triangulating across the three literatures above rather than reporting pooled effect sizes for a single outcome. This approach is explicitly interpretive rather than quantitative and its limitations are discussed in Section 12.

6. Comparative Effectiveness: Immediate Learning Outcomes

6.1 Knowledge and Skill Acquisition

Multiple recent systematic reviews and meta-analyses of randomized controlled trials have examined VR's effect on nursing students' clinical skills, knowledge, and self-confidence relative to traditional teaching methods, including manikin-based simulation. A 2024 meta-analysis of virtual simulation versus mannequins and real persons across medical and nursing education, applying Cochrane methodology with trial sequential analysis, found that virtual simulation produced meaningful gains in knowledge, procedural skill, clinical reasoning, and communication competencies, while also noting that VR simulations required substantially less time and lower cost than high-fidelity simulation to reach comparable learning outcomes. A separate 2024 systematic review and meta-analysis focused specifically on nursing students found that virtual reality training effectively improved knowledge but was not consistently superior to other simulation methods in enhancing all measured competencies, indicating a pattern of approximate equivalence rather than clear VR superiority on cognitive outcomes.

A cross-sectional comparative study of 130 medical students randomized to either virtual high-fidelity simulation (Body Interact) or traditional manikin-based high-fidelity simulation found no significant difference in effectiveness between the two modalities on measures derived from the Simulation Design Scale, Educational Practices Questionnaire, and Student Satisfaction and Self-Confidence in Learning Scale, reinforcing the broader meta-analytic pattern of rough parity on self-reported and cognitive measures. A crossover study directly comparing high-fidelity manikin and VR simulation as assessment tools for acute clinical care skills likewise found both modalities capable of discriminating clinical performance, though sample sizes in this line of research remain small.

6.2 Confidence, Engagement, and Psychological Safety

Several studies report that VR modalities are associated with increased learner confidence, attributed to the psychologically safe, repeatable nature of virtual practice environments that allow errors without real-world consequence and without the social visibility of a shared physical simulation lab. A scoping review of VR in critical care nursing education highlighted consistent findings of increased learner confidence when immersive technology was used, alongside enhanced engagement and motivation. However, self-confidence gains do not always correspond to objectively measured competency gains, and several reviews caution that VR's engagement advantages should not be conflated with superior skill transfer.

6.3 Psychomotor and Team-Based Skills

Where the target competency involves physical procedural skill or multi-person team coordination, such as invasive procedures, code-team leadership, or crisis resource management, the evidence more consistently favors

HFM or hybrid approaches. A quasi-experimental study comparing VR and manikin-based practice of invasive nursing procedures, including nasogastric feeding and catheterization, found no significant difference in confidence, task engagement, or learning satisfaction between groups, but the VR group demonstrated significantly higher proficiency on specific procedural sub-components, suggesting that VR's advantages may be procedure-specific rather than uniform. A randomized controlled trial comparing bronchoscope manipulation training using VR simulation versus high-fidelity mannequins found both approaches effective, with efficiency differences favoring VR for certain skill components. These findings collectively suggest that for purely cognitive, matrix-based interpretation tasks such as ABG analysis, which do not require fine psychomotor control or physical team coordination, the modality-parity pattern observed for knowledge outcomes is likely to generalize, while HFM's advantages are more likely to be concentrated in psychomotor and interprofessional dimensions of critical care training that surround, but are distinct from, the interpretive task itself.

7. Comparative Effectiveness: Long-Term Retention

7.1 The General Pattern of Skill and Knowledge Decay

Independent of simulation modality, the broader literature on simulation-based education consistently documents that both knowledge and, especially, psychomotor skill decay over time following training. Systematic reviews of resuscitation training retention describe a tendency toward decay beginning as early as two to three months post-training, with simulated skill performance deteriorating faster than declarative knowledge. A long-term retention study of pediatric procedural skills among emergency physicians found performance already reduced at six months post-training, with near-total loss of specific technical skills by four years, regardless of clinician experience level, underscoring that decay is a near-universal phenomenon rather than a failure specific to any one training modality. A quantitative review and analysis of skill decay and retention identified task characteristics, retention interval, and the availability of overlearning or refresher opportunities as the primary determinants of how much skill is retained, factors that are, notably, modifiable through instructional design rather than fixed properties of the skill itself.

A systematic review of simulation-based learning in nursing programs, covering literature from 2017 to 2023, concluded that while simulation reliably improves immediate knowledge and skill acquisition, substantial evidence gaps remain regarding long-term retention, with the majority of included studies not evaluating outcomes beyond five months post-intervention. The review's authors recommended that nurse educators hold realistic expectations regarding the durability of simulation-acquired competencies and integrate deliberate strategies, rather than one-off training events, to sustain them.

7.2 Does Modality Affect the Rate of Decay?

Direct comparative evidence on whether VR or HFM produces more durable retention is limited, but several considerations from the retention literature bear on the question. First, the skill-decay literature consistently identifies distributed or spaced repetition, sometimes termed low-dose, high-frequency (LDHF) training, as one of the most effective countermeasures to decay. A study of LDHF simulation training among certified registered nurse anesthetists found this approach effective in reducing skill decay and improving retained proficiency for central venous catheter insertion, a finding whose underlying mechanism, brief, frequent, low-stakes practice opportunities, maps naturally onto VR's core operational advantages: low marginal cost per repetition, availability outside scheduled simulation-lab time, and the capacity for automated, standardized scenario delivery without requiring simulation technician staffing for each session.

Second, retention research on obstetric emergency simulation training found that knowledge and skill decay, when it occurred, began within two to nine months of a single training exposure, and that studies varied considerably in whether meaningful decay was detected at all within that window, a heterogeneity the original authors attributed partly to differences in how frequently learners were exposed to the target skill in real clinical practice between training and retention testing. This finding is directly relevant to ABG interpretation: critical care nurses who interpret ABGs frequently in daily practice are likely to experience less decay regardless of training modality, because clinical exposure itself functions as ongoing spaced practice, whereas nurses in lower-acuity settings with infrequent ABG exposure are likely to show more pronounced decay and may derive proportionally greater benefit from a modality that supports low-cost, self-directed refresher practice.

Third, the cognitive versus psychomotor distinction introduced in Section 6 likely extends to retention. Skill-decay reviews consistently find that psychomotor and procedural skills decay faster than declarative or rule-based knowledge. Because ABG interpretation is predominantly a cognitive, algorithmic, pattern-recognition task rather than a psychomotor one, the retention advantage that HFM might hold for procedural competencies is less likely to apply, while VR's capacity to deliver frequent, low-friction refresher practice may confer a comparatively larger retention benefit for this specific competency than it would for, for example, central line insertion or airway management.

7.3 Peer-Learning and Continuing Education Evidence

An explorative study of a simulation-based peer-learning intervention in continuing nursing education found a significant increase in knowledge from baseline to immediate post-test and to a delayed retention test, but also a measurable decline in scores between the immediate post-test and the retention test, attributing the decline to factors including learner interest, prior work experience, and the extent to which the learned content was subsequently reinforced through on-the-job practice. This pattern, meaningful net gain from baseline but partial decay from peak, recurs across the retention literature regardless of simulation modality and should inform realistic expectations for any single-exposure ABG training intervention, VR or HFM alike.

8. Comparative Effectiveness: Transfer to Clinical Competency

8.1 Defining Transfer for ABG Interpretation

Transfer, the degree to which a competency demonstrated in a simulated environment predicts performance in authentic clinical practice, is the outcome of greatest practical importance and the hardest to measure. For ABG interpretation specifically, transfer would ideally be operationalized as a nurse's accuracy and speed in interpreting a real patient's ABG panel and appropriately escalating or acting on that interpretation, correlated with prior simulation-based training performance. No study identified in this review directly measured this correlation for either VR or HFM training.

The closest available evidence comes from performance-evaluation studies conducted in live clinical contexts using simulated vignettes based on real patient cases. A multicenter study presenting emergency department and intensive care unit nurses with sixteen clinical vignettes derived from actual patients found that nurses achieved high specificity in excluding patients without time-dependent conditions, but the study also identified meaningful gaps between nurse and physician-assigned urgency judgments, with self-confidence in ABG interpretation identified as one of several personal and professional factors associated with performance. The study's authors explicitly recommended structured, evidence-based training programs integrating simulation-based learning, multidisciplinary collaboration, and decision-support tools, and called for future research evaluating the long-term effectiveness of such interventions, a call this review echoes.

8.2 Clinical Judgment and Reasoning Measures

Where clinical judgment has been measured using validated instruments such as the Lasater Clinical Judgment Rubric, comparative studies of virtual patient simulation and high-fidelity manikin simulation have generally found no statistically significant interaction effect of modality on judgment scores over time, suggesting that neither modality holds a definitive advantage in developing the higher-order reasoning skills that ABG interpretation ultimately depends on, correctly weighing conflicting or ambiguous data, prioritizing differential explanations, and integrating interpretation with action. This null finding is itself informative: it suggests that instructional design quality, feedback structure, and deliberate practice volume may matter more for developing clinical judgment than the specific simulation platform used to deliver them.

8.3 Interprofessional and Contextual Factors in Transfer

HFM's team-based, physically embedded format may offer an advantage specifically for the aspects of ABG-related competency that extend beyond the interpretive task itself, communicating an abnormal result using structured handoff frameworks, escalating appropriately within an interprofessional hierarchy, and coordinating a response with respiratory therapy, nursing, and medical staff simultaneously. VR platforms, particularly those without networked multi-user capability, are less well suited to training this contextual, communicative dimension of competency, even where they may equal or exceed HFM for the discrete interpretive task. A quasi-experimental study of an immersive VR speak-up training program targeting assertive communication among novice nurses found no significant interaction effect on communication clarity between VR and conventional lecture-based training at six weeks, again suggesting that modality alone does not guarantee superior transfer for complex, socially embedded competencies, and that content and feedback design remain central regardless of platform.

9. The Specific Case of Complex ABG Matrix Interpretation

9.1 Task Analysis: What Makes ABG Interpretation a Matrix Problem

The term matrix interpretation reflects the fact that a single ABG result must be evaluated across at least four interacting axes simultaneously, pH, $PaCO_2$, HCO_3^-, and oxygenation indices, with the correct classification of one axis contingent on the values of the others. This structure differentiates ABG interpretation from many other nursing competencies commonly studied in the VR-versus-HFM literature, such as medication administration, catheterization, or basic life support, which are predominantly procedural or single-decision tasks. Task analyses of complex cognitive skills in cognitive load theory literature suggest that matrix-style problems benefit disproportionately from instructional strategies that allow the learner to isolate and practice sub-components before integrating them, worked-example effects, part-task training, and scaffolded complexity, strategies that are comparatively easy to implement in a software-based VR or virtual-patient environment, where sequencing, hint availability, and case difficulty can be programmatically controlled, and comparatively harder to implement consistently within a live, real-time HFM scenario, where scenario pacing is constrained by the physical simulator and the human facilitator running it.

9.2 Evidence from Adjacent Domains

In the absence of ABG-specific comparative trials, the closest analogous evidence comes from training domains that share ABG interpretation's core structure: rapid integration of multiple interacting physiological variables into a single classification judgment. Comparative work on mechanical ventilation training found mannequin-based and computer-based simulation approaches to be similarly effective for building residents' ventilator management competency, a domain that, like ABG interpretation, requires integrating multiple interacting parameters rather than executing a discrete physical procedure. This pattern of approximate modality parity for integrative-cognitive tasks, contrasted with HFM's comparative advantage for procedural and team-coordination tasks documented in Section 6.3, provides indirect but consistent support for the inference that VR platforms are likely to be at least as effective as HFM, and potentially more efficient, for teaching the core interpretive logic of ABG analysis specifically.

9.3 Implications for Long-Term Retention of ABG Competency

Applying the retention principles established in Section 7 to ABG interpretation specifically suggests several testable hypotheses. Because ABG interpretation is a rule-based, pattern-recognition task rather than a psychomotor one, its decay curve is likely to be flatter and slower than that of procedural skills such as central line insertion, consistent with the general finding that declarative and algorithmic knowledge decays more slowly than psychomotor skill. Because VR platforms substantially lower the marginal cost of repeated practice relative to HFM, VR-based or VR-supplemented ABG training programs are structurally better positioned to deliver the low-dose, high-frequency refresher exposures that the skill-decay literature identifies as the most effective countermeasure to decay, independent of any inherent pedagogical superiority of VR over HFM for a single training episode. This suggests that the most consequential comparative advantage of VR for ABG competency may lie not in superior single-session learning, where the evidence suggests rough parity, but in superior support for the spaced repetition regimen required to sustain competency over the months and years of a nursing career.

10. Cost, Feasibility, and Implementation Considerations

Implementation considerations are inseparable from any recommendation regarding simulation modality, since even a modality with superior learning outcomes is of limited value if it cannot be feasibly deployed at the scale required for ongoing, spaced-repetition-based ABG training across a critical care nursing workforce. High-fidelity manikins require substantial capital investment, dedicated simulation laboratory space, trained simulation technicians and facilitators, and scheduled protected time for both learners and faculty, constraints that in practice limit the frequency with which any individual nurse can access HFM-based refresher training. VR platforms, by contrast, have been associated in comparative cost analyses with meaningfully lower per-session time and cost requirements relative to high-fidelity simulation for achieving comparable learning outcomes, and can, particularly in non-immersive or desktop-based formats, be deployed asynchronously without requiring a dedicated facility or live facilitator for every session.

These operational advantages carry particular weight for ABG interpretation training specifically, given the retention literature's emphasis on distributed, frequent practice as the primary defense against skill and knowledge decay. A hospital or nursing program seeking to implement quarterly or monthly ABG refresher exercises across an entire critical care nursing staff is likely to find VR-based delivery logistically feasible in a way that equivalent

HFM-based delivery, requiring facility time, technician staffing, and facilitator availability for every learner at every interval, is not. This logistical asymmetry does not resolve the pedagogical comparison documented in Sections 6 through 9, where the evidence suggests approximate parity for single-session cognitive learning outcomes, but it does suggest that VR's practical scalability may translate into superior real-world retention outcomes even where its per-session pedagogical effectiveness is merely equivalent to, rather than superior to, HFM.

Implementation is not cost-free, however. VR platforms require upfront software licensing or development costs, hardware refresh cycles, technical support infrastructure, and faculty development to design scenarios that avoid extraneous cognitive load from unfamiliar interfaces, particularly for immersive head-mounted-display systems, which have documented associations with simulator sickness and orientation difficulty in a subset of learners, especially those with limited prior exposure to the technology. Programs adopting VR for ABG training should budget for iterative scenario design and onboarding support rather than treating VR as a lower-cost substitute requiring no additional pedagogical investment.

11. Gaps in the Evidence Base

- No identified study has directly compared VR and HFM simulation modalities specifically on ABG interpretation using a randomized or quasi-experimental design with a validated long-term retention arm.
- The great majority of VR-versus-HFM comparative studies in nursing and medical education measure outcomes immediately or within days of the intervention; systematic reviews of simulation-based learning consistently identify follow-up beyond five months as rare, leaving the durability question for both modalities, and their relative durability, largely unanswered.
- Existing ABG education studies predominantly evaluate a single modality against a lecture-based or no-intervention control rather than comparing simulation modalities against each other, limiting the ability to draw modality-specific conclusions even for immediate outcomes.
- Clinical transfer, the correlation between simulation performance and subsequent real-patient ABG interpretation accuracy, has not been directly studied for either modality; available performance-evaluation studies describe live clinical competency without linking it back to prior training modality.
- Most VR nursing-education studies to date involve nursing students rather than practicing critical care nurses, limiting generalizability to the postlicensure, continuing-education context in which most ABG competency maintenance actually occurs.

- Heterogeneity in VR immersion level, fully immersive head-mounted-display systems, semi-immersive, and non-immersive desktop platforms, is rarely analyzed as a moderating variable, despite plausible differences in cognitive load and accessibility across these formats.
- Cost-effectiveness analyses comparing VR and HFM remain limited in number and are rarely conducted alongside outcome data, making it difficult to jointly evaluate pedagogical and economic tradeoffs within a single study.

12. Discussion and Recommendations for Practice

12.1 Synthesis

Triangulating across the comparative simulation literature, the ABG education literature, and the skill-decay literature yields a coherent, if evidence-limited, picture. For the specific competency of complex ABG matrix interpretation, a predominantly cognitive, algorithmic, pattern-recognition task, the available evidence suggests that VR and HFM are likely to produce broadly comparable immediate learning gains, consistent with meta-analytic findings of approximate parity for cognitive and knowledge outcomes across nursing education generally. HFM's documented advantages concentrate in psychomotor procedural skill and interprofessional, team-based competencies that surround but are analytically separable from the interpretive task itself. VR's most distinctive and best-supported advantage is not superior single-session pedagogy but superior scalability, standardization, and cost-efficiency, properties that map directly onto what the retention literature identifies as the single most effective countermeasure to skill and knowledge decay: frequent, low-stakes, spaced repetition.

12.2 A Proposed Blended Framework

Rather than framing VR and HFM as competing choices, the evidence supports a blended, sequenced framework for ABG competency development and maintenance. Initial conceptual instruction and rule-based algorithm training may be efficiently delivered through VR or virtual-patient platforms, capitalizing on their capacity for scaffolded, part-task practice and immediate feedback while minimizing extraneous cognitive load. Periodic HFM-based scenarios can then be used to integrate ABG interpretation into the full complexity of an authentic critical care encounter, embedding the cognitive task within the communicative, team-based, and psychomotor context in which it must ultimately be applied. Ongoing competency maintenance, the low-dose, high-frequency refresher exposures that the retention literature identifies as necessary to counteract decay, is then delivered primarily through VR, given its logistical feasibility for frequent deployment across an entire nursing workforce without the facility and staffing burden that equivalent HFM-based refresher programs would require.

12.3 Recommendations

- Critical care education programs should not treat VR and HFM as mutually exclusive investments; the strongest available evidence supports a blended, sequenced approach that allocates each modality to the training phase it is best suited for.
- ABG-specific competency programs should incorporate structured refresher intervals of no more than three to six months, consistent with the general finding that decay in simulation-based training frequently begins within this window.
- Institutions adopting VR for ABG training should invest in scenario design that explicitly manages cognitive load, staged case complexity, minimized interface friction, and structured feedback, rather than assuming immersion alone confers educational benefit.
- Simulation programs should incorporate validated clinical judgment and competency instruments, rather than knowledge tests alone, when evaluating ABG training outcomes, to better capture the reasoning and prioritization components of the skill.
- Nurse educators should track and report retention outcomes at intervals of six months or longer as a standard component of ABG training evaluation, addressing the field-wide gap in long-term follow-up data documented across the broader simulation literature.

13. Limitations of This Review

This review is narrative and interpretive rather than a formal systematic review or meta-analysis, and it does not report pooled quantitative effect sizes; its conclusions regarding ABG interpretation specifically are therefore extrapolated from comparative simulation research conducted on related but not identical competencies and populations, and should be interpreted as evidence-informed hypotheses rather than established findings. The core comparative question motivating this review, VR versus HFM effectiveness specifically for ABG interpretation with long-term retention measurement, has not, to the best of this review's search, been directly studied, meaning the central synthesis offered here rests on triangulation across adjacent literatures rather than direct evidence. Additionally, most of the VR comparative literature involves nursing students rather than practicing critical care nurses, limiting confidence in generalizing findings to the postlicensure population that is this review's primary population of interest. Finally, rapid technological change in VR platforms, including advances in haptic feedback, physiological modeling, and multi-user networked scenarios, means that findings drawn from earlier VR implementations may understate the capabilities of current and emerging systems.

14. Conclusion

Complex ABG matrix interpretation sits at the intersection of high clinical stakes and high cognitive demand, making the choice of simulation modality for teaching and maintaining this competency a matter of genuine practical consequence for critical care nursing education. The available comparative evidence, while not specific to ABG interpretation, converges on a consistent pattern: virtual reality and high-fidelity manikin simulation produce broadly comparable gains in knowledge, clinical reasoning, and learner confidence, high-fidelity manikins retain advantages for psychomotor and team-based competencies, and virtual reality offers substantial, well-documented advantages in scalability, repeatability, and cost that are directly relevant to the spaced-repetition strategies known to counteract the skill and knowledge decay documented across the simulation literature. Because no existing study has directly tested VR against HFM for ABG interpretation with a long-term retention design, these conclusions should be treated as a evidence-informed synthesis rather than a settled finding, and this review recommends that future research prioritize exactly this comparison, using validated competency instruments and retention intervals of six months or longer, to place ABG-specific simulation modality decisions on firmer empirical footing.

References

A comparative analysis of the impact of virtual patients versus high-fidelity manikin-based blended simulation on nursing students' self-evaluation of clinical judgment skills. Clinical Simulation in Nursing.

A comparison of high-fidelity and virtual reality simulation as assessment tools in undergraduate medical education. (2025). [Preprint/PMC record].

Ansquer, R., Mesnier, T., Farampour, F., Oriot, D., & Ghazali, D. A. (2019). Long-term retention assessment after simulation-based training of pediatric procedural skills among adult emergency physicians: A multicenter observational study. BMC Medical Education, 19(1), 348.

Arterial blood gas analysis and clinical decision-making in emergency and intensive care unit nurses: A performance evaluation. (2025). Healthcare, 13(3), 261.

Arthur, W., Bennett, W., Stanush, P. L., & McNelly, T. L. (1998). Factors that influence skill decay and retention: A quantitative review and analysis. Human Performance, 11(1), 57–101.

BMC Medical Education. (2024). The effectiveness of simulation-based learning (SBL) on students' knowledge and skills in nursing programs: A systematic review. BMC Medical Education, 24, Article 06080. https://doi.org/10.1186/s12909-024-06080-z

Beckett, A., & Kipnis, E. (2019). Improving clinical reasoning in nurses through structured education: Implications for ABG interpretation. Journal of Nursing Scholarship, 51(3), 309–317.

Brown, K. M., Swoboda, S. M., Gilbert, G. E., Horvath, C., & Sullivan, N. (2023). Curricular integration of virtual reality in nursing education. Journal of Nursing Education, 62(6), 364–373.

Chacko, B., & Peter, J. V. (2018). Role of arterial blood gas in the management of critically ill patients. Indian Journal of Critical Care Medicine, 22(9), 696–704.

Cho, M. K., & Kim, M. Y. (2024). Enhancing nursing competency through virtual reality simulation among nursing students: A systematic review and meta-analysis. Frontiers in Medicine, 11, 1351300.

Dehghani, K., & Nayeri, N. D. (2021). The effect of simulation-based training on clinical skills of critical care nurses: A systematic review. Nurse Education in Practice, 54, 103128.

Dev, T. A., & Rasool, A. (2025). Arterial blood gas analysis: A review on clinical significance and strategies to enhance interpretation skills among nurses. International Journal of Advance Research in Nursing, 8(2C).

Divya, U., & Mishra, R. (2022). Effectiveness of a structured teaching programme on knowledge regarding arterial blood gas interpretation among staff nurses. International Journal of Nursing Research, 8(2), 112–118.

Effect of virtual reality training on clinical skills of nursing students: A systematic review and meta-analysis of randomized controlled trials. (2024). Nurse Education in Practice, 81, 104182.

Effectiveness of Simulation-Based, Peer Learning Intervention in Continuing Nursing Education: An Explorative Study. (2024). [PMC record].

Effectiveness of immersive virtual reality in nursing education for nursing students and nursing staff: A systematic review and meta-analysis. (2025). Nurse Education Today.

Effectiveness of virtual simulations versus mannequins and real persons in medical and nursing education: Meta-analysis and trial sequential analysis of randomized controlled trials. (2024). Journal of Medical Internet Research, 26, e56195.

El-Sherbiny, N. A., & Ebrahim, R. (2022). Effectiveness of educational programs on nurses' knowledge and practice regarding blood gas interpretation. Journal of Education and Practice, 13(6), 22–30.

Evaluating whether nonimmersion virtual reality simulation training improves nursing competency in isolation wards: Randomized controlled trial. (2025). Journal of Medical Internet Research.

Fraser, K., Ma, I., Teteris, E., Baxter, H., Wright, B., & McLaughlin, K. (2012). Emotion, cognitive load and learning outcomes during simulation training. Medical Education, 46(11), 1055–1062.

Gore, T. (2025). Healthcare simulation research comparison: VR vs in-person. HealthySimulation.com.

Helping Mothers Survive Bleeding After Birth: Retention of knowledge, skills, and confidence nine months after obstetric simulation-based training. (2015). [PMC record].

Huai, P., Li, Y., Wang, X., Zhang, L., Liu, N., & Yang, H. (2024). The effectiveness of virtual reality technology in student nurse education: A systematic review and meta-analysis. Nurse Education Today.

Hur, H.-K., Choi, J.-H., & Jung, J.-S. (2025). Effectiveness of an immersive virtual reality simulation speak-up training program for patient safety in novice nurses: A quasi-experimental study. Healthcare, 13(19), 2435.

Implementation and evaluation of a virtual reality simulation: Intravenous injection training system. (2022). [PMC record].

Jiang, B., Ju, H., Zhao, Y., Yao, L., & Feng, Y. (2018). Comparison of the efficacy and efficiency of the use of virtual reality simulation with high-fidelity mannequins for simulation-based training of fiberoptic bronchoscope manipulation. Simulation in Healthcare, 13(2), 83–87.

Josephsen, J. (2015). Cognitive load theory and nursing simulation: An integrative review. Clinical Simulation in Nursing, 11(5), 259–267.

Kulvi, R., Sharma, P., & Thomas, L. (2023). Impact of a structured teaching program on nurses' knowledge of arterial blood gas analysis: A quasi-experimental study. Indian Journal of Critical Care Nursing, 18(3), 155–162.

Olson, J., Ko, A., Nowak, K. A., Latack, K., & Bozimowsky, G. (2024). Using simulation training to reduce skill decay among certified registered nurse anesthetists. Journal of Continuing Education in Nursing, 55(4), 187–194.

Pruitt, B. (2024). Strategies for interpreting arterial blood gases. Nursing, 54(1), 16–21.

Reedy, G. B. (2015). Using cognitive load theory to inform simulation design and practice. Clinical Simulation in Nursing, 11(8), 355–360.

Sałacińska, I., Trojnar, P., Éles Gebriné, K., Törő, V., Sárváry, A., & Więch, P. (2025). A comparative study of traditional high-fidelity (manikin-based) simulation and virtual high-fidelity simulations concerning their effectiveness and perception. Frontiers in Medicine, 12, 1523768.

Spadaro, S., Karbing, D. S., Fogagnolo, A., Ragazzi, R., Mojoli, F., Astolfi, L., Gioia, A., Marangoni, E., Rees, S. E., & Volta, C. A. (2017). Simulation training for residents focused on mechanical ventilation: A randomized trial using mannequin-based versus computer-based simulation. Simulation in Healthcare, 12(6), 349–355.

The efficacy of virtual-reality teaching and learning method in enhancing interprofessional knowledge, clinical skill performance, and self-confidence in nursing education: A systematic review and meta-analysis. (2025). Nurse Education in Practice.

Virtual Reality Simulation-Based Clinical Procedure Skills Training for Nursing College Students: A Quasi-Experimental Study. (2024). [PMC record].

Virtual Reality in Critical Care Nursing Education: A Scoping Review. (2025). Education Sciences, 15(9), 1258.