JUTLP
Journal of University Teaching & Learning Practice

Enhancing Simulation Training through Immersive Reality: MESH360

Stephen Aiello[a], Associate Professor Thomas Cochrane[b], and Associate Professor Claudio Aguayo[a]

[a] Auckland University of Technology, New Zealand; [b] The University of Melbourne, Australia.

Abstract

This paper presents a critical evaluation of the third design iteration of the MESH360 project, focusing on immersive reality (XR), enhanced simulation in higher education healthcare. It addresses a significant gap in the current literature: while XR has demonstrated efficacy in developing procedural clinical skills, there is a scarcity of longitudinal studies that are grounded in learning theory to inform the design of these environments. Furthermore, the impact of immersive technologies on learner stress and its relationship to educational outcomes remains underexplored. To address this problem, the MESH360 project uses a Design-Based Research (DBR) framework that utilises an iterative, theory-driven design process since its inception in 2016. The research explores the impact of XR-enhanced simulation on training paramedics through a mixed-methods approach, uniquely triangulating participant subjective feedback, structured observation, and objective biometric data (heart rate) to capture the psycho-physiological dimensions of learning. This study's significance is its contribution to the scholarship of teaching and learning through the generation of provisional design principles and a methodological framework for integrating biometric data into simulation evaluation. The findings suggest potential applications in other high-risk, practice-based disciplines, though further cross-context validation is required.

Editors

Section: Special Issues
Editor in Chief: Dr Joseph Crawford
Senior Editor: Dr Cassandra Colvin

Publication

Submission: 3 October 2025
Revised: 1 April 2026
Accepted: 4 May 2026
Online First: 9 June 2026

Practitioner Notes

1. Clinical simulation is widely used but often lacks contextual authenticity.
2. XR enables immersive, risk-free practice, but its design is under-theorised.
3. XR has predominantly focused on procedural skills, with limited evidence of engagement with critical learning theory or the development of student critical analysis.
4. This paper presents a theory-informed approach to designing XR simulations, triangulated with biometric data, and offers provisional design principles requiring further refinement and validation across contexts.

Keywords

XR, Biometrics, Design-Based Research, Healthcare Education, Clinical Simulation.

Citation:

Aiello, S., Cochrane, T., & Aguayo, C. (2026). Enhancing Simulation Training through Immersive Reality: MESH360. *Journal of University Teaching and Learning Practice,* Advanced Online Publication.

Introduction

Paramedic healthcare education faces a critical challenge: equipping graduates with the skills to assess and manage patients in dynamic, high-risk environments where rapid decision-making is essential (Brooks et al., 2018). Simulation-based training has become a cornerstone of this preparation, providing a safe environment in which to develop skills without compromising patient safety (Shahzeydi et al., 2024). However, a persistent pedagogical problem exists. Traditional manikin-based simulations, while valuable, often fail to capture the environmental and psychological stresses associated with real-world emergencies (sight, sound, feel), creating a significant 'authenticity gap' (Aiello et al., 2023). This gap is particularly problematic, especially in fields such as paramedicine, where clinical reasoning is deeply intertwined with chaotic and uncontrolled contexts.

Immersive reality (XR) technologies, including virtual and augmented reality, present a promising solution by replicating elements of environmental complexity. Yet, their application in healthcare education has often been technologically driven rather than pedagogically grounded. Reviews by Stretton et al. (2018) and Aiello et al. (2023) have highlighted that the design of XR simulations frequently overlooks foundational learning theories, tends to focus on procedural skills over critical thinking, and lacks longitudinal, theory-informed evaluation. This points to a substantive gap in the scholarship of teaching and learning (SoTL): a need for robust, transferable design frameworks that are both theoretically sound and empirically validated to guide the creation of XR-enhanced learning environments.

The Multiple Environments Simulation Hub-360 (MESH360) project was established as a longitudinal Design-Based Research (DBR) project to address this gap. Design-Based Research is a research approach suited to developing usable knowledge and theories for educational practice through iterative design cycles in authentic settings (McKenney & Reeves, 2018). Since 2016, the MESH360 project has investigated cost-effective XR solutions that enhance the authenticity of simulations. A key innovation of this project is its methodological approach, which triangulates subjective feedback, observational analysis, and objective biometric data to explore participants' physiological arousal within these immersive environments.

Therefore, this study, situated within the MESH360 project, is guided by a need for authentic learning environments that offer real-world complexity. To advance this inquiry, this paper evaluates a new iteration of the prototype by providing a rescue helicopter simulation. The significance of this work lies in its contribution to SoTL by generating provisional, theory-informed design principles and demonstrating a mixed-methods evaluation framework that integrates biometric data. This study offers an illustrative model for developing and assessing simulations in similar contexts, though the principles presented require further testing across diverse contexts to establish transferability. Therefore, this study is guided by the following research questions:

1. How can XR-enhanced clinical simulation environments be designed to increase authenticity for paramedic students?
2. What is the relationship between participants' physiological arousal (heart rate) during XR-enhanced simulation and their subjective experiences of stress and learning?
3. How can biometric data be integrated with qualitative feedback to inform iterative DBR cycles?

This paper begins by establishing its theoretical foundations, reviewing the principles of experiential learning, social constructivism, cognitive load theory, and Activity Theory alongside prior work on XR in clinical education and Design-Based Research. It then details the study's methodology, describing the research design, participant sample, measures, data collection procedures, and analytical strategies employed. The findings are subsequently presented in an integrated, thematic manner, weaving together biometric, observational, survey, and interview data to address each research question. The discussion interprets these findings in relation to the international literature, examines their implications for theory and practice, and articulates the study's contributions and limitations. The paper concludes by synthesising its key arguments and outlining directions for future research.

The significance and novelty of this work are threefold. First, it contributes a rare longitudinal example of theory-informed XR simulation design, demonstrating how multiple iterations over eight years have progressively refined both the intervention and the underlying design principles. Second, it presents a methodologically innovative approach to triangulating biometric, observational, and self-report data, including a structured joint display that systematically compares convergent and divergent findings. Third, it advances theoretical understanding by applying Activity Theory retrospectively to analyse contextual systemic tensions in educational technology integration, offering transferable analytical insights for similar projects. Collectively, these contributions provide an illustrative model for researchers and practitioners seeking to develop, evaluate, and theorise XR-enhanced learning in high-stakes professional training domains.

Literature

This review establishes the theoretical and empirical foundation for the study, moving from the overarching learning theories that justify the intervention, to a critique of the current state of XR in healthcare education, and finally to the methodological approaches that informed the design within the MESH360 project.

Theoretical Foundations

Learning theories are conceptual frameworks that attempt to explain how people acquire, process, retain, and apply knowledge; however, there is no single learning theory that can comprehensively capture all aspects of learning. Therefore, this study employs a multi-layered theoretical architecture in which different theories operate at complementary analytical levels. The theories function primarily as interpretative frameworks for analysing the findings rather than as generative design specifications. Experiential learning theory (Kolb, 1984) and social constructivism (Vygotsky, 1978) function as overarching pedagogical frameworks, justifying the use of immersive, context-rich scenarios and highlighting the importance of socially mediated learning. Cognitive load theory (Sweller et al., 2019) operates at the cognitive processing level, providing a lens for interpreting how learners with different levels of expertise might experience the cognitive demands of XR simulations. Activity Theory (Engeström, 2014) functions at the systemic level, offering a framework for analysing the tensions and contradictions that emerge when introducing new technologies into established educational practices within a specific socio-cultural context. These theories are not competing but operate hierarchically: experiential and constructivist theories inform the overall pedagogical approach; cognitive load theory explains individual differences in cognitive processing; and Activity Theory situates the intervention within its socio-cultural context, revealing systemic factors that shape design and implementation.

Experiential learning theory suggests that knowledge is constructed through transformative experience, particularly through the resolution of real-world problems (Kolb, 1984). Kolb's four-stage cycle (concrete experience, reflective observation, abstract conceptualisation, and active experimentation) provides a framework for analysing the XR simulation design. In this iteration, the XR helicopter flight and scene assessment were intended to provide concrete experience; the post-simulation interview was designed to prompt reflective observation; the diagnosis and treatment planning phase engaged participants in abstract conceptualisation; and the hands-on manikin treatment represented active experimentation. However, this mapping is post-hoc rather than embedded in the initial design decisions, and the reflective observation phase was constrained by the immediate transition to practical application, limiting opportunities for structured reflection between stages.

Furthermore, social constructivism argues that learning is a socially mediated process with meaning negotiated through interaction with others and the environment (Vygotsky, 1978). This is particularly relevant to paramedicine, a discipline built on teamwork, communication, and shared decision-making (Mangan et al., 2022). The individual design of the XR simulation, when viewed through a social constructivist lens, reveals a generative tension rather than a limitation. This tension between the individualised technological delivery and the collaborative nature of clinical practice, can be reframed as a design problem that points toward necessary future innovation. Rather than representing a theoretical contradiction, it highlights that the current design occupies one point on a developmental trajectory toward more socially embedded XR environments. The goal of developing collaborative competence can be pursued through preparatory individual experiences that build shared mental models, even in the absence of real-time interaction. This individual-to-collaborative progression itself reflects a Vygotskian trajectory, where individual internalisation of culturally situated tools (here, the XR scenario) precedes and enables more complex socially mediated performance. This gap between aspiration and design is examined further through Activity Theory analysis.

Cognitive load theory (Sweller et al., 2019) is employed in this study as an analytical framework for interpreting heart rate patterns, particularly the observed differences between novice and expert participants. However, it is important to clarify that heart rate is not treated as a direct measure of cognitive load. Rather, heart rate is interpreted as an indicator of autonomic arousal, which may be associated with the cognitive demands of the scenario. Cognitive load theory distinguishes between intrinsic load (inherent complexity of the task), extraneous load (imposed by instructional design), and germane load (devoted to schema construction). In this study, intrinsic load varied with clinical task demands, extraneous load may have been introduced by the XR technology itself for novices, and germane load was not directly measured. The scenario was not explicitly engineered to manage these different load types; instead, cognitive load theory provides a post-hoc interpretive lens for understanding experience-related differences in physiological response.

Activity Theory (Engeström, 2014) is applied retrospectively as an analytical framework for understanding the systemic tensions within the MESH360 project. Rather than functioning as a generative design lens that shaped the intervention from its inception, Activity Theory is used to interpret the contradictions that emerged between project components, such as the tension between individual XR design and collaborative practice, or between psychological safety mandates and deliberate stress induction. This retrospective application provides insights into how systemic factors influenced the design and outcomes, offering transferable analytical insights for similar projects (Greeno, 2016).

XR in Clinical Education

While the potential of XR in healthcare education is widely acknowledged, systematic analysis revealed a consistent shortfall in its pedagogical implementation. A review by Stretton et al. (2018) identified that most studies were non-longitudinal and focused narrowly on procedural skill development, with limited integration of learning theory into the simulation design. This critique remains relevant, with Aiello et al. (2023) emphasising the continued importance of intentionally pedagogically grounded design when implementing Immersive technologies. They went on to highlight the lack of research examining the effects of XR on learners' stress and anxiety, which are key psychological factors for clinical performance.

Examples of XR in healthcare higher education demonstrate its potential to enhance clinical reasoning and situational awareness (Birt & Cowling, 2018; Cochrane et al., 2018a; Cochrane et al., 2018b; Cabrera & García, 2019; Moro et al., 2017a; Moro et al., 2017b; Khan et al., 2018). However, the literature is still dominated by applications that focus on the individual learner mastering a technical task, often overlooking the socio-collaborative dimensions of practice and the development of higher-order critical thinking under pressure. This gap supports an investigation of how pedagogically grounded learning might be designed when stress is an expected feature of clinical simulation, while acknowledging the distinction between facilitative anxiety (which may enhance performance) and debilitative anxiety (which may impair it) (Nachtigall & Wirth, 2024).

Design-Based Research

To address the real-world complex problem of designing authentic simulations, a Design-Based Research (DBR) approach was adopted. DBR is uniquely suited to this challenge as it moves beyond simply testing an intervention to the development of contextually sensitive and usable educational theories through iterative, collaborative design (McKenney & Reeves, 2019). This aligns with the research questions, which aim to understand how design decisions influence participants' experiences and outcomes. Proponents argue that traditional comparative studies often lead to no significant difference because they substitute technology without redefining the learning design (Reeves, 2005, 2015). DBR therefore aims to avoid this pitfall by aiming for the redefinition level of Puentedura's (2006) SAMR (Substitution, Augmentation, Modification, Redefinition) model (Gogno, 2013). Despite its advantages, DBR is not without its critics.

Proponents of more traditional methodologies highlight inherent limitations in the DBR approach. The deeply contextualised nature of a DBR study, while a strength for ecological validity, challenges the generalisability of its findings, which are often expressed as intermediate-level design principles rather than universal laws (Barab & Squire, 2016). Furthermore, the significant investment of time and resources required for multiple iterative cycles can be a practical barrier (Anderson & Shattuck, 2012). Perhaps most critically, the researcher's dual role as designer and investigator introduces potential biases, necessitating meticulous strategies to ensure methodological rigour and trustworthiness (Anderson & Shattuck, 2012; Collins et al., 2016). In this study, strategies to address potential bias included the involvement of third-party researchers from the university's Centre for Teaching and Learning in data collection, the use of multiple data sources for triangulation, and transparent reporting of the iterative design process.

Despite these challenges, DBR offers advantages that make it a valuable tool within the educational research ecosystem. Its capacity to generate grounded, theory-informed interventions and to explore how and why they work in authentic settings positions DBR as a catalyst for innovation. Rather than replacing traditional experimental approaches, DBR can inform them by providing the conceptual scaffolding and practical insights needed to advance both theory and practice in education. An additional key benefit of DBR in education is its ability to quickly adapt to contextual changes, such as new technology, new data, and social and cultural shifts. This further supports the claim around its usefulness.

Transdisciplinary Collaboration

The capacity for collaborative practice is a foundational epistemic and professional competency in healthcare, exemplified by the clinical handover process; a dialogic exchange that ensures continuity of care across disciplinary boundaries (Shah et al., 2016; Wong et al., 2008). This collaborative paradigm extends into educational design, where transdisciplinary engagement reflects a constructivist paradigm: knowledge is constructed through shared inquiry and contextual negotiation (Laurillard, 2013; Vygotsky, 1978). In this view, curriculum design becomes a reflective practice, mirroring the complexities of the clinical environment and adopting authentic, learner-centred pedagogies (Biggs, 1999; Cronin et al., 2016; Crosby & Morgan, 2017; Laurillard, 2001; Schön, 2017; Weaver et al., 2013). Thus, the philosophical alignment between clinical collaboration and education co-design underscores the ontological and pedagogical significance of teamwork as both a method of practice and a mode of knowing (Amiel & Reeves, 2008; Dewey, 1938).

Transdisciplinary collaboration serves as a foundational principle of Design-Based Research, facilitating the integration of diverse expertise from educational researchers and practitioners to inform, construct and iteratively refine educational interventions (Amiel & Reeves, 2008; Emin-Martinez et al., 2014; Steel, 2012). Khoo et al. (2019) argue that transdisciplinary research embraces 'inclusiveness, tensions, unpredictability and complexity' (p. 181), reflecting the dynamic and multifaceted nature of real-world educational challenges and the need for adaptive, context-sensitive design approaches. For the MESH360 project, this meant integrating the expertise of educational researchers, technologists, paramedic lecturers, and biometric scientists. This approach is not only logistically practical but also philosophically aligned with the collaborative nature of healthcare itself, ensuring that the resulting intervention is both pedagogically sound and clinically relevant.

Biometric Triangulation

Learning is understood to occur through a combination of cognitive processes (Kirschner, 2002), the activation of prior knowledge (Sutherland, 1992), and social interaction (Vygotsky, 1978). This, however, can be accompanied by emotional stress when learners encounter novel challenges (Hase & Kenyon, 2007). Physiological arousal response is indicative of both positive challenge and related stress. Such stress responses can be objectively qualified using biometric indicators such as heart rate, galvanic skin response, and blood pressure (Halbig & Latoschik, 2021). It is important to note that heart rate is a measure of autonomic arousal, not a direct measure of cognitive load or learning. The relationship between arousal and cognitive processing is complex and mediated by numerous factors, including individual differences, context, and the type of task demands. In this study, heart rate is used as one indicator of participants' physiological engagement with the simulation, triangulated with subjective reports to build a more complete picture of the learner experience.

A key methodological component of this research is the triangulation of data, specifically the incorporation of these biometric measures. This approach bridges a disciplinary divide: education's traditional reliance on subjective feedback and healthcare's emphasis on quantitative data (Aguayo et al., 2018; Creswell, 1994; Turner et al., 2012). By aligning physiological markers with clinical simulation and subjective feedback, this study aims to develop a model of educational inquiry that understands not only what learners say but also how they physiologically respond within stressful, authentic learning environments.

Method

Research Design

This study is grounded in a pragmatist paradigm, which justifies the pursuit of usable knowledge through practical intervention (Biesta, 2010). Pragmatism aligns closely with Design-Based Research (DBR), which seeks to develop contextually relevant solutions through iterative design and evaluation, and supports the use of mixed methods, where knowledge is judged by its practical utility (Ryu, 2020).

Methodologically, this study is situated within a DBR framework, employing a convergent mixed-methods design (Creswell & Clark, 2017). DBR provides the overarching, iterative framework for developing interventions in authentic contexts (McKenney & Reeves, 2019). This paper reports on a new evaluation cycle within the broader MESH360 research programme, not previously reported, as shown in Table 1, representing a structured phase of evidence-based refinement. The convergent mixed-methods approach was selected to collect both qualitative and quantitative data simultaneously, enabling direct investigation of the relationship between physiological responses and user experiences.

Table 1

Longitudinal Progression of the MESH360 Project

Phase	Year	Aim	Reference
Phase 1	2016	Analysis and exploration to identify pedagogical issues and initial design principles	Cochrane et al., 2017
Phase 2	2017	Design and construction of the first XR prototype	Cochrane et al., 2018a
Phase 3	2018	Evaluation and reflection of the initial prototype	
Phase 2-3 (Loop)	2018	Iterative redesign and re-evaluation	Aguayo et al., 2018; Cochrane et al., 2018b

The transition from the 2018 ambulance scenario to the current helicopter scenario was informed by participant feedback from the previous iteration, which identified opportunities to enhance authenticity by simulating a context that is logistically impossible to replicate in traditional training due to cost and safety constraints. For example, in the 2018 evaluation, participants noted that while the ambulance environment increased realism, they sought exposure to "extreme environments that we can't normally train in" and scenarios where "the environment itself is the primary stressor." This feedback directly informed the selection of the helicopter rescue context for the current iteration, as helicopter emergency medical services represent an authentic but inaccessible training environment where environmental stressors (noise, vibration, spatial constraints, altitude) are pedagogically valuable. This design decision represents an explicit conjecture refinement: the hypothesis that increasing environmental extremity and inaccessibility would enhance authenticity and stress exposure, thereby better preparing students for real-world practice.

Collaborative Research Team

A foundational principle of this Design-Based Research study is its transdisciplinary, participatory nature. The MESH360 project was not developed by researchers observing practitioners, but through active collaboration among all stakeholders throughout the iterative design cycles (McKenney & Reeves, 2019). Table 2 outlines the composition of the collaborative research team, detailing the specific expertise and role each group contributed to the project. This structure ensured that the intervention was pedagogically sound, technologically feasible, clinically relevant, and rigorously evaluated.

Table 2

Transdisciplinary MESH360 Design-Based Research Study Structure

Team Members	Role in Project
Academic Advisor	Research design facilitator and mentor, and educational technologist
Digital development team	Co-research design and immersive reality application development team
Paramedicine Lecturers	Discipline lecturers and core members of the MESH360 enhanced simulation project development. Lead development of simulation environment, and participatory SOTL researchers.
Biometric researchers	Biometric data researchers and tracking software development facilitating quantitative data gathering and analysis.
Paramedic students and practitioners	Simulation participants: 27 participant volunteers from year 1-3, professional paramedics who were also postgraduate students, and Paramedicine lecturers – providing critical feedback and evaluation of the enhanced simulation environments.

The transdisciplinary team met regularly throughout the design cycle, with paramedicine lecturers leading scenario development, the digital team prototyping based on clinical specifications, and biometric researchers advising on data collection protocols. This put into practice the collaborative ethos central to DBR and ensured the intervention was co-constructed rather than imposed.

Participants

A sampling strategy was used to recruit participants with varying clinical experience. The study aimed not for a single homogeneous group, but for a stratified sample representing five distinct levels of practitioner development, ranging from beginner to expert. The sampling was intentionally designed to facilitate a comparative analysis of how clinical experience influences psychological stress responses and clinical reasoning strategies within the XR environment.

A total of 27 participants were recruited across five representative demographic groups (Table 3) and participated in the XR-enhanced clinical simulation experience over three consecutive days.

Table 3

Participant Level of Experience

Participant Group (Level of experience)	**Number (*n*)**
1st year student (Y1): Beginner	7
2nd year student (Y2): Novice	4
3rd year student (Y3): Competent	6
Postgraduate student (PG): Professional Paramedic ‘proficient’	8
Masters level (MA): Professional Paramedic ‘expert’	2
Total *n* =	27

Note. Five professional participants were also Paramedic lecturers.

The sample size of 27 across five strata, with group sizes ranging from 2 to 8 participants, is appropriate for qualitative analysis and descriptive quantitative purposes but provides limited statistical power for inferential comparisons. The two participants in the Masters-level group represent a particular limitation for between-group statistical analysis. This study is therefore positioned as exploratory, with quantitative findings treated as descriptive trends requiring confirmation in larger samples.

Potential participants were invited via Facebook, Instagram, and an announcement on the learning management system (LMS) to respond to a project email account. The channels were selected to cast a wide net across both the general student body and professional networks, ensuring we could reach individuals of all experience levels. Respondents were then emailed a simulation booking day and time, instructions, a consent form, and a participant information sheet. This was in accordance with an updated amendment to the MESH360 ethics approval from the university's ethics committee, AUTEC 17/29. This included changes to the original scenario details from the 2017 and 2018 iterations. Participation in the XR-enhanced simulation experience was voluntary, non-assessed, and supervised by third-party researchers from the university's Centre for Teaching and Learning to help minimise any potential coercion or assessment bias.

Measures

Three categories of measures were employed to capture the multidimensional nature of the simulation experience.

Physiological Measure

Heart rate (beats per minute) was recorded continuously using an Apple Watch worn on the non-dominant wrist. Heart rate was selected as an indicator of autonomic arousal, acknowledging its limitations as a direct measure of cognitive load or stress. As noted in the literature review, heart rate reflects autonomic arousal influenced by multiple factors including cognitive demand, emotional response, physical exertion, and uncontrolled variables. The interpretation of heart rate in this study is therefore triangulated with other measures and treated as one indicator of physiological engagement rather than a direct measure of psychological constructs.

Self-report Measures

A pre-simulation survey captured demographic information, prior clinical experience, previous exposure to simulation and XR technologies, caffeine consumption within the preceding four hours, and self-reported pre-existing stress levels. A post-simulation survey assessed perceived authenticity, engagement, usability, and desire for future XR use using Likert-scale items. Semi-structured interviews (described below) elicited rich qualitative data on participants' subjective experiences.

Observational Measures

Structured observations were conducted during both the XR and manikin simulation phases. Researchers recorded behavioural indicators of engagement and anxiety, including scene assessment patterns, hand movements (fidgeting), and attention to environmental cues. Video recordings enabled subsequent review and verification of observational notes.

Data Collection Procedures

Data collection followed a structured sequence designed to capture multiple streams of evidence across all phases of the simulation experience.

Pre-simulation Baseline and Profiling

Upon recruitment and consent, participants completed the anonymous pre-survey. A physiological baseline was then established. Participants were fitted with an Apple Watch to measure heart rate while wearing the head-mounted display (HMD: Oculus Go) in a neutral state. This pre-XR experience was important to isolate and account for any physiological response to the hardware itself, ensuring that subsequent HR measurements during the clinical scenario reflected the simulation's psychological load.

The XR Simulation

Participants engaged with the XR clinical scenario, a 360° video designed to create a high-fidelity, psychologically immersive environment. The 11-minute scenario comprised: baseline calibration (calming forest scene); callout (1 minute); pre-flight (2 minutes); flight (5.30 minutes); landing (30 seconds); and scene assessment (2 minutes). Continuous HR monitoring provided an objective, quantitative measure of physiological stress response throughout this phase. Immediately following the scenario, participants verbally presented their initial diagnosis, capturing clinical reasoning based solely on the immersive XR environment experience.

Practical Application

To assess the translation of learning from the virtual to the physical realm, participants then treated a high-fidelity manikin based on their XR-formed diagnosis. To create a seamless transition, the final view from the XR-HMD scene was projected onto the wall of the immersive simulation suite. The manikin was placed in identical clothing and position to those seen in the XR environment. This traditional practical simulation phase allowed for the observation of practical skills application. After treatment, participants stated their final diagnosis, allowing for the measurement of any diagnostic refinement after the combined XR and hands-on experience.

Post-simulation Reflection

A video-recorded post-simulation interview was conducted to gather rich qualitative data on participants' subjective experiences, perceived stress, and learning. The interviews were subsequently transcribed verbatim. The interview questions were:

1. What was the impact of the XR experience on the quality of your clinical simulation learning?
2. What are the key differences between the XR environments and a traditional classroom or simulation suite?
3. Based on your experience during this project, would you like to make use of XR simulation for future practice/training?

A post-simulation survey was also administered to quantify perceptions of usability and educational impact. The simulation cycle concluded with a comprehensive debriefing and feedback on clinical performance.

Data Analysis

Quantitative Analysis

Heart rate data were analysed descriptively using medians and interquartile ranges (IQR) to account for small sample sizes and non-normal distributions. Kruskal-Wallis tests were conducted to examine between-group differences in peak heart rate during XR and simulation phases. Statistical analyses were performed using SPSS Version 28. Given the exploratory nature of the study and limited statistical power, inferential results are interpreted cautiously, with emphasis on descriptive patterns and effect sizes rather than null hypothesis significance testing alone.

Qualitative Analysis

Thematic analysis followed an inductive approach, with codes generated from the data rather than imposed from theoretical frameworks (Braun & Clarke, 2006). To enhance trustworthiness, a multi-stage analytical process was employed. Two researchers independently coded a subset of transcripts to establish initial coding categories. The researchers then met to compare their coding, discuss discrepancies, and reach consensus on a refined coding framework. This framework was subsequently applied to the remaining transcripts by the primary researcher, with regular peer debriefing sessions involving the broader transdisciplinary team to challenge assumptions and interrogate emerging interpretations.

Mixed-methods Integration

Integration of quantitative and qualitative data was achieved through two complementary strategies. First, there was a systematic comparison of convergent and divergent findings across biometric, observational, and self-report data streams for key themes. Second, narrative integration was employed throughout the results section, weaving together different data types within thematic subsections to provide a coherent account of findings. This approach aligns with the convergent mixed-methods design and enables identification of both corroborating evidence and discordant cases that qualify interpretation.

The Redesigned Intervention

The 2018 enhanced simulation utilised a HMD to immerse participants in an XR-simulated ambulance ride. This scenario involved responding to an emergency call, preparing en route, and upon arrival, assessing and viewing (virtually) the scene to formulate a diagnosis and treatment plan before transitioning to a high-fidelity manikin simulated patient (Cochrane et al., 2020). Analysis of participant feedback from this iteration revealed key opportunities to enhance experience authenticity. Consequently, the research team designed a new XR-simulation targeting the specific high-stress environment of a helicopter rescue mission, a scenario logistically prohibitive and unsafe to replicate 'in reality' due to extreme cost and risk. This choice was pedagogically driven to safely expose students to critical stressors endemic to this context, including high ambient noise, equipment alarms, and physical hazards, to replicate the cognitive load associated with a high acuity real-world environment. Link to the redesigned iteration of the XR-experience: https://go.wondavr.com/tpGIvxPbz1

Results

Participant Characteristics and Baseline Data

The participant cohort (*n* = 27) was 51.9% male (*n* = 14) and 48.1% female (*n* = 13). Most participants (59%, *n* = 16) were aged 18 to 25 years, and 55.6% (*n* = 15) were currently working as either a student and an ambulance service volunteer or a paid professional paramedic. Analysis of prior experience revealed that most first-year undergraduate students (22% of the total cohort, *n* = 6) had no prior experience with traditional manikin-based simulation. Among participants with prior manikin-based experience, only 25% agreed it provided sufficient information to treat a patient. The majority (66.7%, *n* = 18) had never experienced XR. Immediately prior to the simulation, 14.8% (*n* = 4) of participants reported feeling stressed, and 40.7% (*n* = 11) reported consuming a stimulant (caffeine) within the previous four hours. These uncontrolled variables (particularly caffeine consumption and pre-existing stress) represent confounds that limit the interpretability of heart rate data as an indicator of simulation-specific arousal.

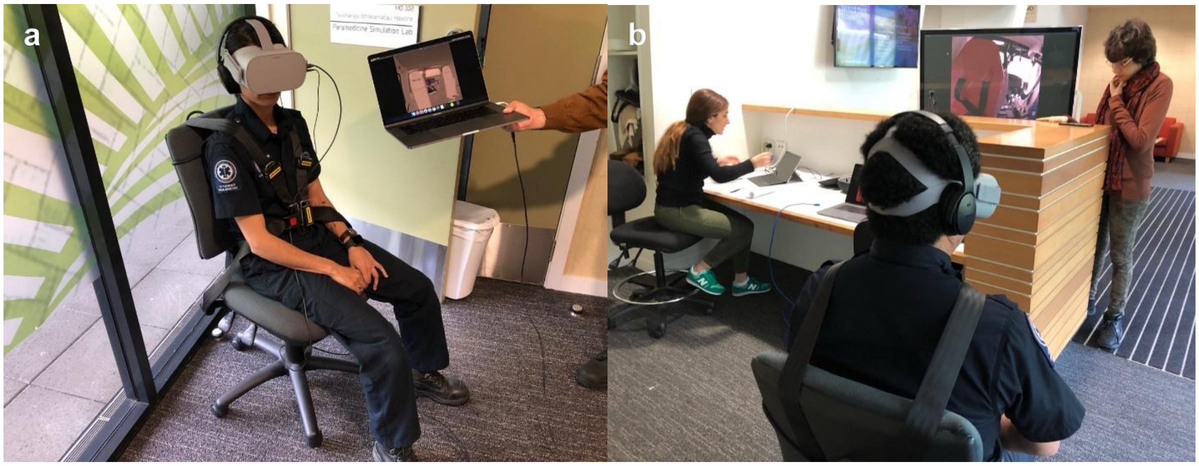
Thematic Integration of Findings

Theme 1: Authenticity and Engagement

The XR environment successfully created an authentic learning experience, as evidenced across all data streams. Biometrically, participants showed elevated heart rate during the helicopter flight and landing phases, with a mean increase of 18 bpm from baseline. Figure 1 (a) illustrates the physical setup with HMD and seat harness designed to enhance presence, while Figure 1 (b) shows the time-stamped biometric data recording during the XR scenario. Together, these figures demonstrate how the combination of physical props and continuous physiological monitoring enabled capture of engagement patterns.

Figure 1

HMD and Seat Harness, with XR-view mirrored to the laptop | *Recording Time-stamped Biometric Data during XR Scenario.*

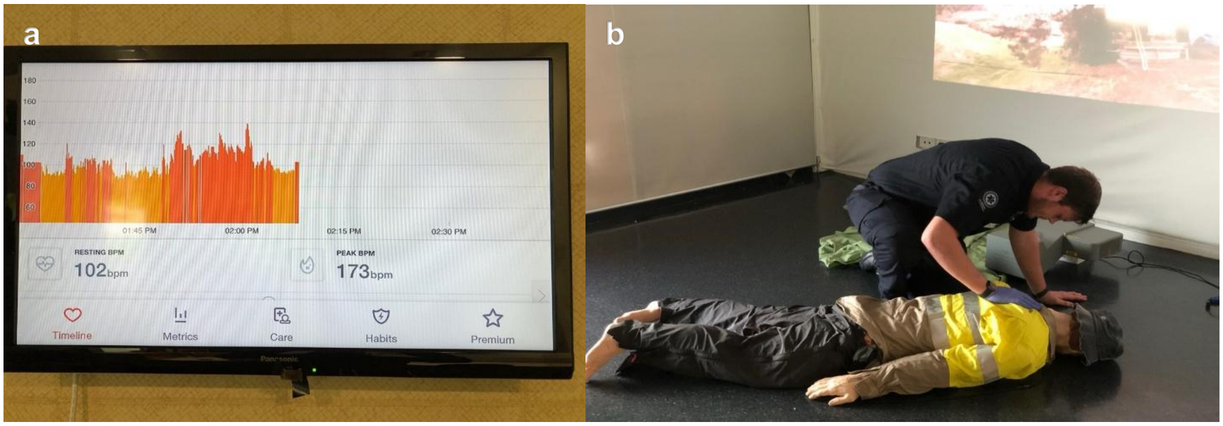


Observational data corroborated physiological engagement, with participants exhibiting focused attention, minimal distraction, and physical engagement such as gripping the harness during flight sequences. Survey data revealed that 96.5% of participants rated the XR flight as authentic, and all participants agreed that XR increased the quality of clinical simulation learning and improved their awareness of the clinical environment.

Interview themes reinforced these findings, with participants describing "enhanced realism" and "the ability to prepare." One participant noted that XR provided "perspective" to form "clinical reasoning" without "the need for [descriptive] questions and imagination" required in traditional simulation. As shown in Figure 2 (a), the timeline view of biometric data mirrored to a large screen allowed researchers to monitor real-time arousal fluctuations in relation to scenario events. Figure 2 (b) demonstrates the critical element of environmental continuity, showing the projection of the XR scene adjacent to the identically dressed manikin, which participants referenced to confirm the mechanism of injury during treatment.

Figure 2

Post XR-Scenario: Timeline View of Participant Biometric Data Mirrored to Large Screen | Transition to High-Fidelity Manikin Treatment (Simulation).



Note. A replication of the XR scene is projected onto the wall, and the manikin patient is dressed and positioned in the same manner for continuity.

Theme 2: Stress as Facilitative vs. Debilitative

A consistent pattern emerged across data streams suggesting that participants experienced stress as a feature of authentic engagement rather than an impediment to learning. Biometric data showed elevated heart rates across all participants, with values ranging from 75 to 173 bpm, with the highest values among novices. To provide a clearer summary of these patterns, Table 4 presents descriptive statistics for peak heart rate during the XR and simulation phases, grouped by participant experience level.

Table 4

Summary of Peak Heart Rate (bpm) by Participant Experience Level

Experience Level	***n***	**XR Phase: Median [IQR]**	**XR Phase: Range**	**Simulation Phase: Median [IQR]**	**Simulation Phase: Range**
1st Year (Y1)	7	118.0 [98 - 125]	88 - 138	143.0 [108 - 165]	88 - 173
2nd Year (Y2)	4	103.0 [100 - 106]	100 - 106	126.0 [125 - 134]	125 - 134
3rd Year (Y3)	6	101.0 [99 - 110]	98 - 127	130.0 [124 - 143]	121 - 143
Postgraduate (PG)	8	100.5 [85 - 115]	82 - 123	132.5 [109 - 142]	106 - 143
Masters (MA)	2	77.5 [75 - 80]	75 - 80	97.5 [94 - 101]	94 - 101

Experience Level	*n*	XR Phase: Median [IQR]	XR Phase: Range	Simulation Phase: Median [IQR]	Simulation Phase: Range
All Participants	27	102.0	75 - 138	128.0 [106 - 143]	88 - 173

Table 4 illustrates a descriptive trend of decreasing heart rate with increasing clinical experience. First-year students showed the highest median peak heart rates in both phases, while Masters-level experts exhibited the lowest. However, the wide ranges and overlapping interquartile ranges indicate substantial individual variation within groups. To support the specific participant references made throughout the text, Table 5 presents the detailed heart rate data for all 27 participants.

Table 5

Detailed Participant Heart Rate Data by Experience Level

Participant ID	Experience Level	XR HR-Hi	XR Range	XR Experience	SIM HR-Hi	SIM Range	ATP	Clinical Experience
01-04 ID4144	MA	80	L1	Yes	94	L2	ICP	>10
01-07 ID234123	MA	75	L2	Yes	101	L3	ICP	>10
03-08 ID25	PG	93	L2	No	106	L3	ILS	>10
03-07 ID26	PG	83	L2	No	129	L4	ILS	8
01-08 ID23031992	PG	111	L2	No	137	L4	ILS	8
03-10 ID5146	PG	88	L2	Yes	109	L3	ILS	5
02-04 ID1082390	PG	82	L2	Yes	143	L4	EMT	5
03-04 ID249	Y2	102	L2	No	125	L4	NA	3
03-03 ID13	Y1	98	L2	Yes	123	L4	NA	0
03-05 ID3882	Y1	97	L2	No	105	L3	NA	0
03-06 ID18015707	Y1	88	L2	No	96	L3	NA	0
01-10 ID2288	PG	118	L3	Yes	142	L4	ILS	>10
03-09 ID12345	PG	123	L3	Yes	136	L4	ILS	>10
01-09 ID1262	PG	110	L3	No	128	L4	ILS	10
02-06 ID29	Y3	102	L3	No	137	L4	EMT	5

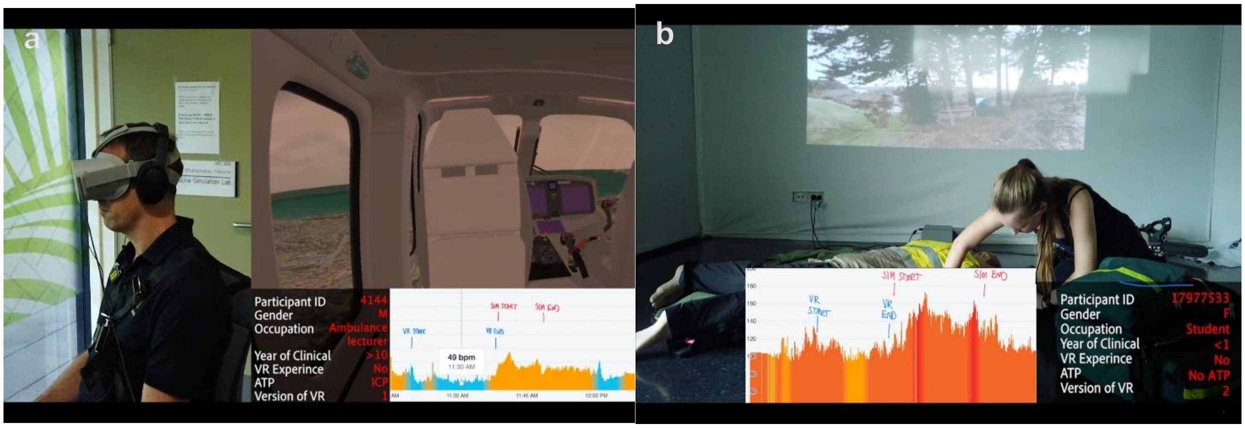
Participant ID	**Experience Level**	**XR HR-Hi**	**XR Range**	**XR Experience**	**SIM HR-Hi**	**SIM Range**	**ATP**	**Clinical Experience**
01-06 ID16939259	Y3	98	L3	No	143	L5	EMT	0
01-01 ID4980	Y1	102	L3	No	88	L3	EMT	1
02-02 ID16940870	Y3	99	L3	Yes	121	L4	FR	3
02-03 ID18	Y3	100	L3	No	124	L4	NA	3
03-01 ID16938624	Y3	110	L3	No	128	L4	NA	0
01-05 ID7415	Y2	100	L3	No	127	L4	NA	3
02-07 ID5371	Y2	106	L3	No	134	L4	NA	0
01-03 ID13111995	Y1	118	L3	No	123	L4	NA	0
02-01 ID132854	Y3	127	L4	No	132	L5	NA	0
01-02 ID2666940	Y1	125	L4	Yes	143	L5	NA	1
02-05 ID17977533	Y1	122	L4	No	173	L6	NA	0
03-02 ID2001	Y1	138	L4	No	165	L5	NA	0

Note. HR values are grouped by range average (L1: <60bpm, L2: 60-80 bpm, L3: 80-100 bpm, L4: 100-120 bpm, L5: >120 bpm). XR HR-Hi and SIM-Hi represent the peak HR value for each participant. Experience levels are: MA-Masters/expert (*n* = 2), PG-professional (*n* = 8), Y3-third year undergraduate student (*n* = 6), Y2-second year undergraduate student (*n* = 4), Y1-first year undergraduate student (*n* = 7). Authority to practice (ATP) levels are defined by New Zealand ambulance standards. Two participants (IDs 4144 and 234123) had additional paramedic helicopter experience.

Figures 3 (a) and 3 (b) contrast the heart rate profiles of an expert and a novice participant, visually addressing RQ2 by illustrating the divergent physiological responses associated with different levels of clinical experience. The expert (ID4144) maintained a stable profile with XR HR 80 bpm and SIM HR 94 bpm, while the novice (ID17977533) peaked at XR HR 122 bpm and SIM HR 173 bpm, demonstrating the magnitude of experience-related differences in physiological arousal.

Figure 3

Expert Paramedic Heart Rate Profile | Novice Paramedic student (Y1) Heart Rate Profile.



Participant ID 4144
Gender M
Occupation Ambulance lecturer
Year of Clinical >10
VR Experince No
ATP ICP
Version of VR 1

Participant ID 17977533
Gender F
Occupation Student
Year of Clinical <1
VR Experince No
ATP No ATP
Version of VR 2

Observational data revealed behavioural indicators of anxiety, including fidgeting, rapid scanning, and hesitation, particularly among novices. Figure 4 captures a typical example of repetitive hand movements observed during the XR experience. As Lang et al. (2015) note, such movements serve as self-regulatory strategies to cope with stress. The presence of these behaviours alongside positive learning outcomes (all participants correctly diagnosed the patient) suggests that observed anxiety did not preclude successful performance.

Figure 4

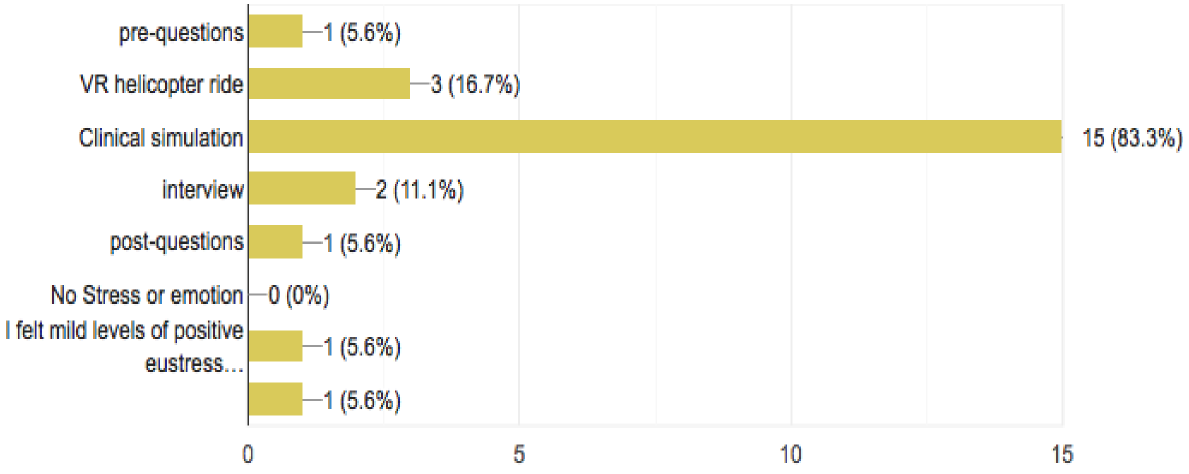
Repetitive Hand Movements Typical of Anxiety

Note. An increase in repetitive hand movement (fidgeting) is a common indicator of anxiety and stress. These movements serve as a self-regulatory strategy to cope and regain control. (Lang et al., 2015).

Despite elevated physiological arousal and observable anxiety, 100% of participants reported stress during the experience yet simultaneously indicated desire for future XR use. Interview data revealed that participants described this stress as "authentic" and "valuable." Survey data showed that participants perceived peak stress during two distinct phases: the helicopter journey and the manikin-based clinical simulation. Figure 5 illustrates this patterned response, showing that stress was not uniform but acutely elevated during specific high-intensity segments, highlighting the role of intentional scenario design in eliciting targeted psychological states.

Figure 5

Subjective points of Highest Stress Intensity During the Experience



Note. More than one point of stress can be reported.

This apparent paradox (measurable physiological arousal co-occurring with positive evaluations) suggests participants distinguished between facilitative stress (associated with authentic challenge) and debilitative anxiety (which would impair performance). However, this study did not directly measure whether stress enhanced learning outcomes, only that participants perceived it as valuable and that learning objectives were achieved.

Theme 3: Experience Related Differences in Response

A descriptive pattern of lower heart rates among expert participants compared to novices emerged across both XR and simulation phases. First-year students (median 118.0 bpm, IQR 98-125 in XR; median 143.0 bpm, IQR 108-165 in simulation) exhibited higher peak rates than Masters-level participants (median 77.5 bpm, IQR 75-80 in XR; median 97.5 bpm, IQR 94-101 in simulation). Kruskal-Wallis tests indicated that between-group differences were not statistically significant for either XR heart rate ($p = 0.175$) or simulation peak HR ($p = 0.303$). Figure 6 compares heart rates across XR and simulation phases for all participants, visually confirming that simulation scenarios consistently elicited higher arousal than XR scenarios across experience levels. Figure 7 presents the comparison of XR peak heart rates across the five experience groups, illustrating the descriptive trend of decreasing physiological arousal with increasing experience, though the overlapping ranges highlight substantial individual variation within groups.

Figure 6

Participant Heart Rate: XR-Experience Versus Manikin-Based Simulation

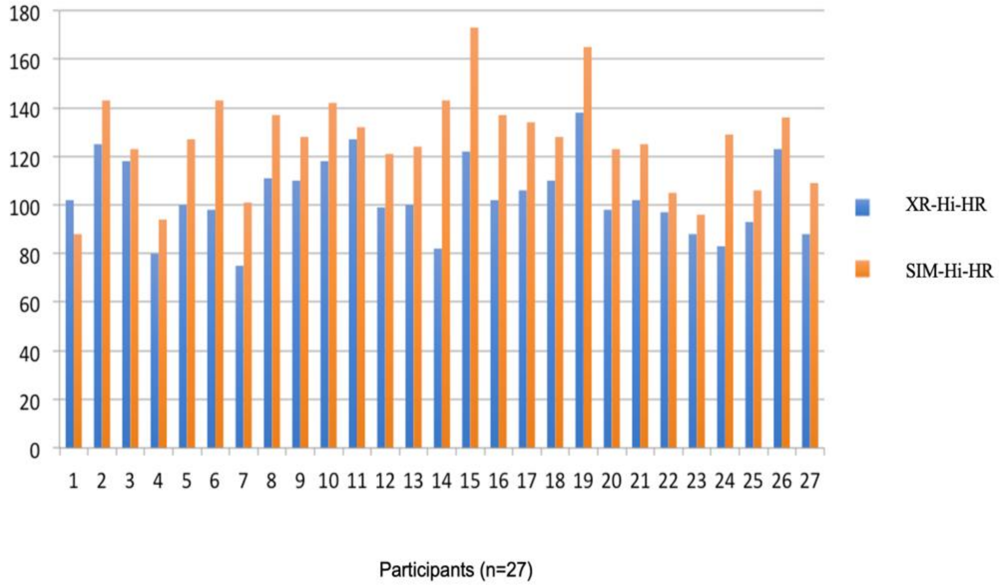
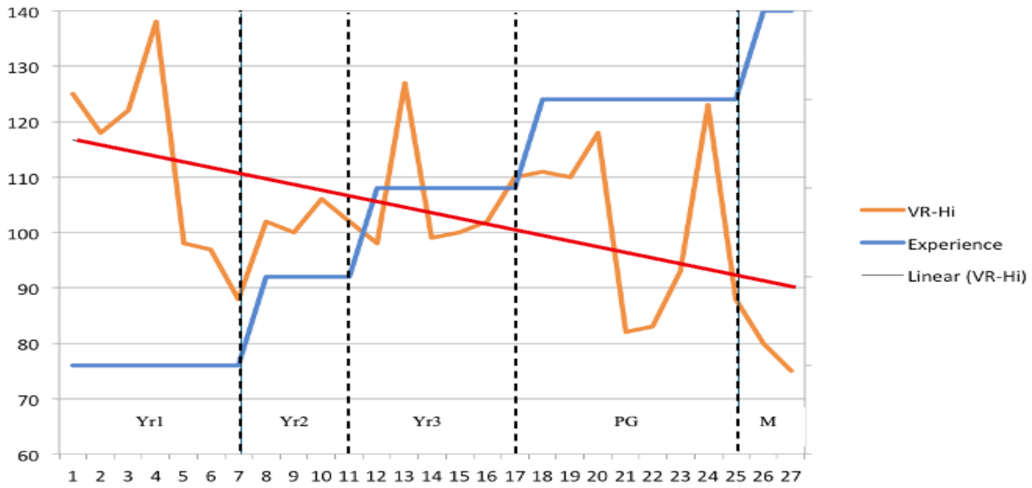


Figure 7

Comparison of XR-Hi (Heart Rate) With Participant Experience



Note. Shows a comparison of impact upon the different levels of experience of the five participant groups: 1st, 2nd, 3rd year undergraduate students, post-graduate (PG), and master's (M).

Observational data aligned with this pattern. Experts conducted thorough scene assessments, scanning for dangers and hazards before approaching the patient, while novices tended to move directly toward the patient. Interview data provided complementary insights, with experts describing the XR experience as a valuable "briefing tool" for mental preparation, whereas novices used terms like "overwhelming" and "intense." Experts appeared to process the scenario through familiar cognitive schemas, while novices simultaneously processed the novelty of both the environment and clinical demands. Individual variation within cohorts indicates that experience alone cannot fully account for stress responses. Factors such as temperament, prior stress exposure, self-efficacy, and broader social or cultural influences may also shape the participant's XR experience.

Theme 4: The Collaboration Gap

A consistent limitation identified across data streams was the absence of collaborative elements in the XR design. Biometric data showed individual engagement patterns only, with no interaction data. Observational data revealed no collaborative behaviour during the XR phase. Most critically, interview and survey data revealed participants explicitly noted the "lack of a crew partner" as a limitation. One participant who experienced a headset error without audio stressed the importance of reliable diegetic sound for immersion, highlighting how technical fidelity supports engagement. However, the collaboration critique was more fundamental: participants reported that paramedic practice is inherently collaborative, and the individual XR experience, while valuable for building situational awareness, did not prepare them for the communication and shared decision-making required in real emergencies. As reframed in the theoretical section, this critique is interpreted not as a fundamental flaw but as an indicator of a generative design tension pointing toward necessary future development. The current individual design can be viewed as a foundational phase for building individual competence and shared mental models, which are prerequisites for effective collaboration.

Integrated Analysis: Joint Display of Convergent and Divergent Findings

Table 6 presents a structured joint display that integrates biometric, observational, and self-report data, explicitly identifying convergent findings (where data streams align) and divergent cases (where data streams conflict). This integration moves beyond narrative triangulation to systematic comparison across data sources.

Table 6

Joint Display of Triangulated Findings: Convergence and Divergence Across Data Streams

Theme/Finding	Biometric Data (Heart Rate)	Observational Data	Self-Report Data	Integration Insight
Authenticity of XR environment	HR elevation during helicopter flight and landing phases (mean increase of 18 bpm from baseline)	Participants exhibited focused attention, minimal distraction, and physical engagement (gripping harness)	96.5% rated flight as authentic; "felt like I was really there"	**Convergent:** All confirm high engagement and perceived authenticity during XR phases
Stress as facilitative vs. debilitative	Elevated HR across all participants (range 75-173 bpm); highest in novices	Fidgeting, rapid scanning, hesitation observed in novices; calm deliberate	100% reported stress but 100% desired future XR use; stress described as	**Convergent:** Stress experienced physiologically and behaviourally but interpreted positively in self-report, suggesting

Theme/Finding	Biometric Data (Heart Rate)	Observational Data	Self-Report Data	Integration Insight
		movement in experts	"authentic" and "valuable"	facilitative rather than debilitative stress
Experience-related differences	Descriptive pattern of lower HR in experts (median 77.5 bpm) vs. novices (median 118.0 bpm) in XR	Experts conducted thorough scene assessment; novices moved directly to patient	Experts described XR as "briefing tool"; novices used terms "overwhelming", "intense"	**Convergent:** All streams suggest experience modulates physiological and psychological response to simulation
Collaboration gap	HR patterns show individual engagement only (no interaction data)	No collaborative behaviour observed during XR phase	Participants explicitly noted "lack of crew partner" as limitation	**Convergent:** All streams confirm individual design limits socio-collaborative learning
Discordant Case: ID4980 (Y1 student)	HR relatively low during XR (102 bpm) compared to Y1 peers (median 118.0)	Observed behaviour: calm, methodical, minimal fidgeting	Self-report: described feeling "stressed but in control"; rated experience highly authentic	**Divergent:** HR lower than peer group despite self-reported stress; suggests individual differences in physiological reactivity or coping strategies; demonstrates that HR alone cannot predict subjective experience
Discordant Case: Audio error participant	HR data erratic during XR phase; unable to establish clear pattern	Visible confusion, looking around, removal of headset briefly	Self-report: noted audio loss significantly reduced immersion; still rated overall experience positively	**Divergent:** Technical failure disrupted physiological and behavioural engagement but did not fully undermine positive evaluation; highlights resilience of overall learning experience
Discordant Finding: Stress-learning relationship	HR elevated during both XR and simulation phases; highest during simulation	Anxious behaviours observed, particularly in novices	100% reported learning was enhanced; all correctly diagnosed patient	**Divergent:** Physiological stress and anxious behaviours co-occurred with successful learning outcomes; suggests stress did not impair performance, though causality cannot be established

The joint display reveals that while most findings converge across data streams, the discordant cases provide important insights. For instance, participant ID4980 demonstrates that low physiological arousal does not necessarily indicate low subjective stress, suggesting individual differences in stress reactivity and coping that are not captured by heart rate alone. The audio error case illustrates that even when technical failures disrupt physiological and behavioural engagement, participants may still evaluate the experience positively, highlighting the complexity of user experience and the importance of multi-modal evaluation. The broader discordance between elevated physiological stress and successful learning outcomes challenges assumptions that stress necessarily impairs performance, supporting the need for nuanced distinctions between facilitative and debilitative anxiety in simulation design.

Summary of Integrated Findings

Triangulation across data streams reveals that the immersive, context-rich nature of XR effectively bridged the theory-practice gap for participants across experience levels. Biometric data indicated elevated physiological arousal, particularly in novices, and observation revealed associated anxious behaviours. Qualitative data revealed that participants perceived this arousal as authentic and beneficial for preparing for real-world pressures. A descriptive pattern of lower heart rates among experts emerged, with qualitative data suggesting experts strategically used XR for mental preparation. Novices, who showed higher physiological arousal, successfully integrated environmental clues to inform their clinical reasoning. The joint display analysis confirms overall convergence across data streams while revealing important individual variations that qualify the interpretation of physiological data. The relationship between physiological arousal and learning outcomes remains correlational rather than causal and requires further investigation with direct learning measures.

Discussion

Synthesis of Key Findings

This study triangulated subjective, observational, and biometric data to examine the nature of stress within XR-enhanced simulation and its relationship to authentic learning experiences. Four key findings emerged from the integrated analysis. First, the XR environment successfully created an authentic learning experience, evidenced by convergent physiological engagement, focused observation, and positive participant evaluations. Second, participants experienced stress as a feature of authentic engagement rather than an impediment to learning, with elevated physiological arousal co-occurring with positive learning outcomes and unanimous desire for future XR use. Third, a descriptive pattern of experience-related differences emerged, with experts showing lower physiological arousal and more strategic engagement compared to novices, though individual variation within groups qualified this pattern. Fourth, the absence of collaborative elements in the XR design was identified as a significant limitation, highlighting a generative tension between individual technological delivery and the inherently social nature of paramedic practice.

Revisiting the research Questions

This study set out to address three research questions. First, regarding how XR-enhanced clinical simulation environments can be designed to increase authenticity (RQ1), the findings demonstrate that authenticity emerges from the alignment of multiple design elements rather than technical sophistication alone. The combination of physical props (flight harness), environmental continuity (projected scene matching the manikin), and diegetic audio created a coherent immersive experience that participants rated as highly authentic. The design decisions were explicitly informed by previous DBR cycles, demonstrating how iterative refinement grounded in participant feedback can progressively enhance authenticity. Importantly, authenticity was not merely a technological achievement but a pedagogical one; the XR environment provided contextual cues that eliminated the need for descriptive narration, allowing participants to focus directly on clinical reasoning.

Second, concerning the relationship between physiological arousal and subjective experiences of stress and learning (RQ2), the integrated analysis reveals a complex picture. While elevated heart rates co-occurred with self-reported stress across all participants (see Table 5), qualitative data consistently framed this stress as authentic and valuable rather than debilitating. The discordant case of participant ID4980—who reported subjective stress despite relatively low physiological arousal (102 bpm in XR compared to Y1 peer median of 118.0 bpm highlights that this relationship is mediated by individual differences in stress reactivity and coping. Furthermore, the co-occurrence of elevated arousal with successful learning outcomes (all participants correctly diagnosed the patient) suggests that stress functioned as facilitative challenge rather than debilitative impairment, though causality cannot be established from this correlational data.

Third, regarding how biometric data can be integrated with qualitative feedback to inform iterative DBR cycles (RQ3), Table 6 demonstrates a structured approach to triangulation that moves beyond narrative assertion. By systematically comparing convergent and divergent findings across biometric, observational, and self-report streams, the analysis revealed both corroborating evidence (e.g., all streams confirming authenticity) and critical discordant cases that generated new insights (e.g., the audio error participant). This integration directly informed the identification of the collaboration gap as a priority for future iterations, demonstrating how multi-modal evaluation can guide evidence-based refinement within DBR cycles.

Contributions to Experiential Learning Theory

These findings contribute to ongoing theoretical discussions about the role of authentic experience in professional education. Kolb's (1984) experiential learning cycle provided a post-hoc framework for analysing the simulation design, with the XR helicopter flight and scene assessment representing concrete experience, the post-simulation interview enabling reflective observation, diagnosis and planning engaging abstract conceptualisation, and hands-on manikin treatment constituting active experimentation. However, the constrained reflective observation phase (with immediate transition from XR to practical application) limited full engagement with Kolb's cycle. This observation aligns with Boud et al.'s (2013) emphasis on the necessity of structured reflection for learning from experience. The finding that participants nevertheless reported enhanced learning suggests that immersive concrete experience may partially compensate for constrained reflection, though deliberate integration of reflective pauses would likely strengthen outcomes.

Advancing the Discourse on Stress Simulation

This study contributes to the emerging literature on stress as a pedagogical feature rather than an unwanted byproduct of simulation. Nachtigall and Wirth's (2024) distinction between facilitative and debilitative anxiety provides a valuable lens for interpreting findings. Participants who experienced elevated physiological arousal and exhibited anxious behaviours nevertheless evaluated the experience positively and achieved learning objectives (correct diagnosis), suggesting that stress functioned as facilitative challenge rather than debilitative impairment. This pattern aligns with research in aviation and military training, indicating that controlled stress exposure can enhance performance under pressure (Driskell et al., 2018; Robson & Manacapilli, 2014).

However, the finding also raises important questions about individual differences in stress response. Participant ID4980, who reported subjective stress but showed relatively low physiological arousal, exemplifies the complexity of stress as a multidimensional construct. This discordant case aligns with research on stress reactivity and coping styles (Lazarus & Folkman, 1984), suggesting that effective simulation design must account for individual variation rather than assuming uniform responses. The implication for simulation pedagogy is that stress inoculation training (Meichenbaum, 1988) may need to be personalised, with graduated exposure calibrated to individual stress reactivity.

Implications for Cognitive Load Theory

The descriptive pattern of lower heart rates among experts compared to novices is consistent with cognitive load theory's prediction that experts, possessing automated schemas, experience reduced cognitive burden when processing familiar clinical scenarios (Sweller et al., 2019). This pattern aligns with research by van Merriënboer and Sweller (2010) on expertise reversal effects, where instructional designs beneficial for novices may become redundant or even detrimental for experts. The finding that experts used the XR experience as a "briefing tool" for mental preparation, while novices found it "overwhelming," illustrates this expertise reversal phenomenon.

However, the non-significant statistical results and uncontrolled confounds require cautious interpretation. The experience may account for meaningful variance in physiological response, but adequately powered studies with larger samples are needed to confirm this pattern. Furthermore, the absence of direct measures of intrinsic, extraneous, and germane load limits theoretical precision. Future research should incorporate validated cognitive load instruments (Paas et al., 2016) alongside biometric measures to more precisely map the relationship between experience, cognitive processing, and physiological arousal.

Extending Social Constructivism in XR Design

The collaboration gap identified in this study speaks directly to critiques in the literature regarding XR's individualistic focus (Stretton et al., 2018; Aiello et al., 2023). Social constructivism (Vygotsky, 1978) suggests that learning is fundamentally socially mediated, yet the current XR design positioned learners as isolated individuals. Rather than interpreting this as a theoretical contradiction, the study reframes this gap as a generative tension pointing toward a design trajectory. This reframing draws on Vygotsky's (1978) concept of the zone of proximal development, where individual internalisation of culturally situated tools precedes and enables more complex socially mediated performance. The current individual XR design can be understood as supporting the internalisation phase, building shared mental models and individual competence that serve as foundations for subsequent collaborative performance.

This interpretation aligns with research on collaborative virtual environments (Benford et al., 2001; Steed & Schroeder, 2015), suggesting that effective teamwork in XR requires both individual preparation and multi-user functionality. The participant feedback requesting a "crew partner" indicates that learners recognise this gap and desire more socially authentic experiences. Future iterations should explore multi-user XR environments that enable communication and shared decision-making, while maintaining the individual preparatory phase as a scaffolded introduction.

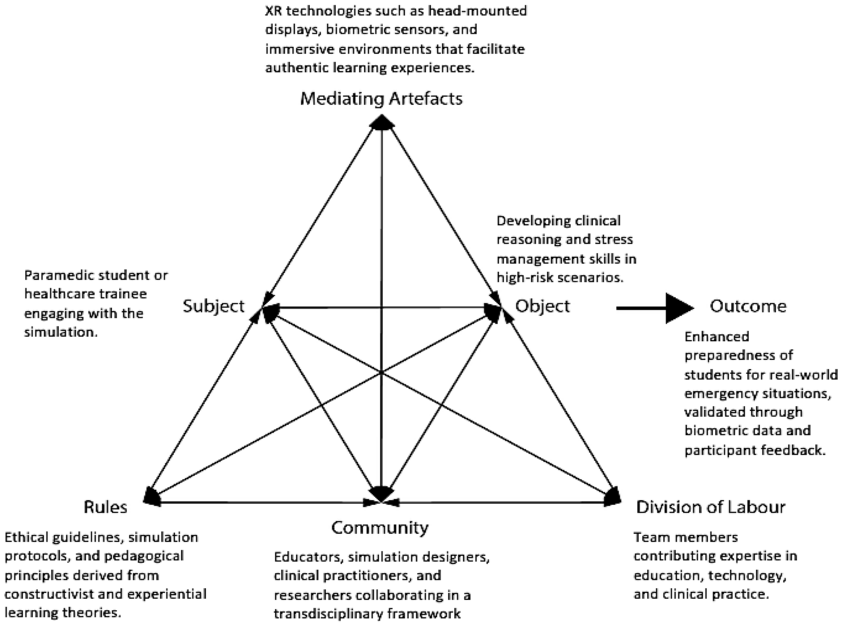
Activity Theory as an Analytical Lens

The retrospective application of Activity Theory (Engeström, 1987) provides a systemic framework for understanding the tensions that shaped the MESH360 project. The contradictions identified between XR tools and traditional manikins, between psychological safety rules and deliberate stress induction, and between individual design and collaborative community are not unique to this project but reflect broader challenges in educational technology integration. Activity Theory's value lies in reframing these tensions as catalysts for innovation rather than obstacles to be eliminated.

Figure 11 presents the Activity Theory framework applied to the MESH360 project, mapping the key components of the activity system and their interrelationships. The diagram illustrates how the Subject (paramedic students) engaged with the Object (clinical reasoning and stress management skills) through mediating Tools (XR technology, manikins, biometrics), shaped by Rules (psychological safety, assessment protocols), Community (transdisciplinary team, educators, clinicians), and Division of Labour (role differentiation in the project), leading to the Outcome (enhanced learner preparedness).

Figure 11

Activity Theory Framework for the MESH360 Project



Note. The diagram illustrates the activity system adapted from Engeström (2001) showing the mediating relationships between Subject, Tools, Object, Rules, Community, and Division of Labour within the MESH360 project context.

This contributes to the educational technology literature by demonstrating how Activity Theory can be applied retrospectively to generate transferable insights. Similar analyses in other contexts (Castro, 2016; Morrison, 2003) have shown that identifying systemic contradictions can guide iterative design toward more coherent outcomes. For researchers engaged in design-based research, Activity Theory offers a structured vocabulary for analysing the complex interplay of tools, rules, community, and division of labour that shapes technology-enhanced learning interventions.

Methodological Contributions to Mixed-Methods Research

The joint display analysis (Table 6) represents a methodological contribution, responding to calls in the mixed-methods literature for more rigorous integration strategies (Clark, 2019; Zhou et al., 2024). By systematically comparing convergent and divergent findings across biometric, observational, and self-report data streams, the study demonstrates an approach to integration that moves beyond narrative assertion to analytical demonstration. The identification of discordant cases, the audio error participant, and the broader stress-learning discordance, illustrates the value of attending to divergence rather than seeking only confirmation. As Costa and Remedios (2014) argue, divergent findings often generate the most theoretically interesting insights, revealing the complexity of phenomena that single-method studies would miss.

For researchers seeking to integrate biometric data into educational research, this study offers an illustrative model. The triangulation framework demonstrates how physiological measures can complement rather than replace traditional educational research methods, with each data stream contributing unique information while compensating for limitations of others. The cautious interpretation of heart rate as autonomic arousal rather than direct cognitive load measure models the conceptual precision required when importing methods from other disciplines.

Limitations and Directions for Future Research

Several limitations qualify the findings and suggest directions for future research. The sample size of 27 across five strata provides limited statistical power, and the non-significant results indicate that between-group differences, while descriptively interesting, cannot be reliably attributed to experience level. Future research should recruit larger, more balanced samples to enable robust inferential analysis.

The absence of direct measures of learning outcomes beyond correct diagnosis limits claims about the relationship between stress and learning. While participants achieved the immediate learning objective and reported enhanced learning, future studies should incorporate validated measures of knowledge retention, transfer to practice, and clinical performance. Longitudinal designs tracking participants from simulation to clinical placement would provide stronger evidence of transfer.

The uncontrolled confounds (caffeine consumption, pre-existing stress, individual differences in stress reactivity) limit the interpretability of physiological findings. Future research should implement pre-simulation questionnaires to capture these variables for statistical control and should consider within-subjects designs with baseline-adjusted change scores to isolate simulation-specific effects.

The qualitative analysis, while strengthened by independent coding, consensus meetings, peer debriefing, and member checking, lacked a formal inter-coder reliability assessment and a comprehensive reflexive audit trail. Future research should incorporate reliability coefficients and maintain systematic reflexivity documentation to further enhance trustworthiness.

Implications for Practice

Notwithstanding these limitations, the findings offer provisional guidance for educators designing XR-enhanced simulations. First, authenticity emerges from alignment between scenario design and learners' psychological reality rather than technical sophistication alone. Simple elements (harnesses for physical presence, environmental continuity between virtual and physical phases, diegetic audio) contribute meaningfully to perceived authenticity. Second, controlled stress exposure can be intentionally designed as a pedagogical feature, but individual differences in stress reactivity require graduated approaches. Third, the collaboration gap identified suggests that XR design should be conceptualised as part of a broader learning ecosystem that includes collaborative elements, rather than as a standalone solution. Finally, the transdisciplinary collaborative model demonstrated in this project (integrating educators, technologists, clinicians, and biometric researchers) provides a replicable blueprint for innovation.

Conclusion

This study contributes to the international scholarship of technology-enhanced learning by presenting a detailed case of XR-enhanced simulation development within a DBR framework, generating provisional design principles and demonstrating a methodological approach for integrating biometric data. The explicit theoretical architecture, positioning experiential learning and social constructivism as pedagogical foundations, cognitive load theory as an interpretive lens for cognitive processing, and Activity Theory as a systemic analytical framework, provides a coherent structure for understanding the multiple levels at which XR simulation design operates. The correlation between theory-informed design and positive user experience suggests that continuous, iterative feedback loops are valuable for developing simulations that are both pedagogically sound and psychologically authentic. However, the design principles presented require further testing and refinement across diverse contexts to establish their transferability.

The joint display analysis represents a significant methodological contribution, demonstrating how structured integration of biometric, observational, and self-report data can reveal both convergent patterns and critical discordant cases that qualify interpretation. This approach moves beyond narrative triangulation toward systematic mixed-methods integration, offering a replicable illustrative model for future research.

The key theoretical contribution of this work is its demonstration that integrating multiple data sources (biometric, observational, and self-report) can provide a more complete picture of the learner experience in immersive environments. While high fidelity is often associated with technical sophistication, this study suggests that authenticity also emerges from the alignment of scenario design with learners' psychological reality. The success of the transdisciplinary collaborative model provides a replicable blueprint for innovation, demonstrating that bridging disciplinary silos is crucial for developing interventions that are educationally coherent and clinically relevant.

In summary, this research establishes a methodologically innovative approach for designing enhanced simulation to model high-stress environments. Grounded in experiential and constructivist learning theory, and interpreted through cognitive load theory and Activity Theory, the MESH360 framework provides a foundation that extends beyond paramedicine. It offers the wider Learning & Teaching community an illustrative model for designing safe, cost-effective, and pedagogically transformative simulations in other high-risk professional domains, while acknowledging that further validation across contexts is necessary to establish generalisability. This project demonstrates that such interventions require a balance between authenticity, immersion, and psychological safety, achieved through iterative design-based research.

Acknowledgements

The authors disclose that they have no actual or perceived conflicts of interest. The authors disclose that they have not received any funding for this manuscript beyond resourcing for academic time at their respective universities. The authors have not used artificial intelligence in the ideation, design, or write-up of this research as per Crawford et al. (2023). The authors confirm that they have met the ethical standards expected as per Purvis and Crawford (2024).The authors list the following CRediT contributions: Aiello: Investigation, Conceptualisation, Methodology, Review & Editing, Data curation, Writing – original draft. Cochrane: Investigation, Conceptualisation, Methodology, Review & Editing, Supervision. Aguayo: Investigation, Conceptualisation, Methodology, Review & Editing, Data curation. The AppLab development team, Te Ara Poutama, Auckland University of Technology, New Zealand: Resources. Westpac Helicopter Trust, Auckland, New Zealand: Resources. Embodied reports, Santiago, Chile: Investigation.

References

Aguayo, C., Danobeitia, C., Cochrane, T., Aiello, S., Cook, S., & Cuevas, A. (2018). Embodied reports in paramedicine mixed reality learning. *Research in Learning Technology*, *26*, Article 2150. https://doi.org/10.25304/rlt.v26.2150

Aiello, S., Cochrane, T., & Sevigny, C. (2023). The affordances of clinical simulation immersive technology within healthcare education: a scoping review. *Virtual Reality, 27*(4), 3485-3503. https://doi.org/10.1007/s10055-022-00745-0

Amiel, T., & Reeves, T. C. (2008). Design-based research and educational technology: Rethinking technology and the research agenda. *Journal of Educational technology & society, 11*(4), 29-40.

Anderson, T., & Shattuck, J. (2012). Design-based research: A decade of progress in education research? *Educational researcher, 41*(1), 16-25. https://doi.org/10.3102/0013189X11428813

Barab, S., & Squire, K. (2016). Design-based research: Putting a stake in the ground. In *Design-based research* (pp. 1-14). Psychology Press. DOI: 10.1207/s15327809jls1301_1

Benford, S., Greenhalgh, C., Rodden, T., & Pycock, J. (2001). Collaborative virtual environments. *Communications of the ACM, 44*(7), 79-85.

Biesta, G. (2010). Pragmatism and the philosophical foundations of mixed methods research. In *Sage handbook of mixed methods in social and behavioral research, 2*, 95-118.

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher education research & development, 18*(1), 57-75. https://doi.org/10.1080/0729436990180105

Birt, J., Stromberga, Z., Cowling, M., & Moro, C. (2018). Mobile mixed reality for experiential learning and simulation in medical and health sciences education. *Information, 9*(2), 31. https://doi.org/10.3390/info9020031

Boud, D., Keogh, R., & Walker, D. (2013). *Reflection: Turning experience into learning*. Routledge.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101. DOI: https://doi.org/10.1191/1478088706qp063oa

Brooks, I. A., Grantham, H., Spencer, C., & Archer, F. (2018). A review of the literature: the transition of entry-level paramedic education in Australia from vocational to higher education (1961–2017). *Australasian Journal of Paramedicine, 15*, 1-11. https://doi.org/10.33151/ajp.15.2.58

Cabrera, M. V. L., & García, J. E. V. (2019). Virtual reality in residents training. *Annals of Eye Science, 4*, 27-27. DOI: 10.21037/aes.2019.08.01

Castro, W. (2016). An activity theory approach to study barriers of Faculty regarding technology integration in higher education. *INTED2016 Proceedings*. DOI: 10.21125/inted.2016.0709

Clark, V. L. P. (2019). Meaningful integration within mixed methods studies: Identifying why, what, when, and how. *Contemporary educational psychology, 57*, 106-111. https://doi.org/10.1016/j.cedpsych.2019.01.007

Cochrane, T., Cook, S., Aiello, S., Aguayo, C., Danobeitia, C., & Boncompte, G. (2018a). Designing immersive Mobile mixed reality for paramedic education. *2018 IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE),* DOI: 10.1109/TALE.2018.8615124

Cochrane, T., Marriott, P., Mak, V., Molloy, E., Sevigny, C., Aiello, S., Sweeney, R., Vawser, T., Linegar, D., & Allan, T. (2024). Enhancing Clinical Education Training with Immersive Reality (XR). *ASCILITE Publications*, 546-550.

Cochrane, T., Stretton, T., Aiello, S., Britnell, S., Cook, S., & Naryan, V. (2018b). Authentic interprofessional health education scenarios using mobile VR. *Research in Learning Technology, 26*. http://dx.doi.org/10.25304/rlt.v26.2130

Cochrane, T. D., Cook, S., Aiello, S., Christie, D., Sinfield, D., Steagall, M., & Aguayo, C. (2017). A DBR framework for designing mobile virtual reality learning environments. *Australasian Journal of Educational Technology*, *33*(6). DOI: https://doi.org/10.14742/ajet.3613

Collins, A., Joseph, D., & Bielaczyc, K. (2016). Design research: Theoretical and methodological issues. In *Design-based research* (pp. 15-42). Psychology Press. https://doi.org/10.4324/9780203764565

Costa, L. d., & Remedios, R. (2014). Different methods, different results: Examining the implications of methodological divergence and implicit processes for achievement goal research. *Journal of Mixed Methods Research, 8*(2), 162-179. https://doi.org/10.1177/155868981349597

Crawford, J., Cowling, M., Ashton-Hay, S., Kelder, J.-A., & Middleton, R. (2023). Artificial intelligence and authorship editor policy: ChatGPT, Bard Bing AI, and beyond. *Journal of University Teaching and Learning Practice, 20*(5), 1-11. DOI: https://doi.org/10.53761/1.20.5.01

Creswell, J. W. (1994). Research Design: *Qualitative & Quantitative Approaches*. SAGE Publications. Inc. USA.

Cronin, C., Cochrane, T., & Gordon, A. (2016). Nurturing global collaboration and networked learning in higher education. *Research in Learning Technology, 24*. https://doi.org/10.3402/rlt.v24.26497

Crosby, A. L., & Morgan, A. C. (2017). Levering critical collaboration: The First Year Interdisciplinary Design Experience. In Collaboration and Student Engagement in Design Education (pp. 169-187). *IGI Global Scientific Publishing*. DOI: 10.4018/978-1-5225-0726-0.ch008

Dewey, J. (1938). *Experience and education*. Macmillan.

Driskell, J. E., Salas, E., Johnston, J. H., & Wollert, T. N. (2018). Stress exposure training: An event-based approach. In *Performance under stress* (pp. 287-302). CRC Press. https://doi.org/10.1201/9781315599946

Emin-Martínez, V., Hansen, C., Rodriguez Triana, M. J., Wasson, B., Mor, Y., Dascalu, M., Ferguson, R., & Pernin, J.-P. (2014). *Towards teacher-led design inquiry of learning. eLearning Papers. 39*, 1-12.

Engeström, Y. (2014). Activity theory and learning at work. In *Tätigkeit-Aneignung-Bildung: Positionierungen zwischen Virtualität und Gegenständlichkeit* (pp. 67-96). Springer. DOI:https://doi.org/10.1007/978-3-658-02120-7_3

Gogno, N. (2013). Educational technology reaches a point of redefinition. In: Wordpress Blog posted to http://wp. lasalle. edu/blog/educational-technology ….

Greeno, J. G. (2016). Cultural-historical activity theory/Design-based research in Pasteur's quadrant. *Journal of the Learning Sciences*, *25*(4), 634-639. https://doi.org/10.1080/10508406.2016.1221718

Halbig, A., & Latoschik, M. E. (2021). A systematic review of physiological measurements, factors, methods, and applications in virtual reality. *Frontiers in Virtual Reality, 2*, 694567. https://doi.org/10.3389/frvir.2021.694567

Hase, S., & Kenyon, C. (2007). Heutagogy: A child of complexity theory. Complicity: *An international journal of complexity and education, 4*(1). DOI: https://doi.org/10.3389/frvir.2021.694567

Khan, R., Plahouras, J., Johnston, B., Scaffidi, M. A., Grover, S. C., & Walsh, C. M. (2018). Virtual reality simulation training in gastrointestinal endoscopy: a Cochrane systematic review. *Gastrointestinal Endoscopy*, *87*(6), AB614.

Khoo, S.-M., Haapakoski, J., Hellstén, M., & Malone, J. (2019). Moving from interdisciplinary research to transdisciplinary educational ethics: Bridging epistemological differences in researching higher education internationalization (s). *European Educational Research Journal, 18*(2), 181-199. https://doi.org/10.1177/14749041187812

Kirschner, P. A. (2002). Cognitive load theory: Implications of cognitive load theory on the design of learning. *Learning and Instruction, 12*(1), 1-10. Elsevier. https://doi.org/10.1016/S0959-4752(01)00014-7

Kolb, D. A. (1984). *Experiential Learning: Experience as the source of learning and development. Upper Sadle River: Prentice Hall*.

Lang, M., Krátký, J., Shaver, J. H., Jerotijević, D., & Xygalatas, D. (2015). Effects of anxiety on spontaneous ritualized behavior. *Current biology, 25*(14), 1892-1897. DOI: 10.1016/j.cub.2015.05.049

Laurillard, D. (2001). *Rethinking University Teaching: What students need from learning technologies; teaching as mediated learning; what students bring to learning; the complexity of coming to know; generating a teaching strategy. Part II Analysing the media for learning and teaching; narrative media; interactive media; adaptive media; communicative media; productive media. Part III The design methodology; designing teaching materials; setting up the learning context; designing an effective organizational infrastructure*. Taylor & Francis Books Limited.

Laurillard, D. (2013). *Teaching as a design science: Building pedagogical patterns for learning and technology*. Routledge.

Lazarus, R. S., & Folkman, S. (1984). *Stress, appraisal, and coping*. Springer.

Mangan, J., Rae, J., Anderson, J., & Jones, D. (2022). Undergraduate paramedic students and interpersonal communication development: a scoping review. Advances in *Health Sciences Education, 27*(4), 1113. doi: 10.1007/s10459-022-10134-6

McKenney, S., & Reeves, T. (2018). *Conducting educational design research*. Routledge.

Meichenbaum, D. H., & Deffenbacher, J. L. (1988). Stress inoculation training. *The counseling psychologist, 16*(1), 69-90. https://doi.org/10.1177/00110000881610

Moro, C., Štromberga, Z., Raikos, A., & Stirling, A. (2017a). The effectiveness of virtual and augmented reality in health sciences and medical anatomy. *Anatomical Sciences Education, 10*(6), 549-559. https://doi.org/10.1002/ase.1696

Moro, C., Štromberga, Z., & Stirling, A. (2017b). Virtualisation devices for student learning: Comparison between desktop-based (Oculus Rift) and mobile-based (Gear VR) virtual reality in medical and health science education. *Australasian Journal of Educational Technology, 33*(6). DOI: https://doi.org/10.14742/ajet.3840

Morrison, D. (2003). Using activity theory to design constructivist online learning environments for higher order thinking: A retrospective analysis. *Canadian Journal of Learning and Technology/La revue canadienne de l'apprentissage et de la technologie*, *29*(3).

Nachtigall, V., & Wirth, J. (2024). Perspectives on authentic learning. *European Journal of Psychology of Education, 39*(4), 3213-3225. DOI: https://doi.org/10.1007/s10212-024-00897-4

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. (2016). Cognitive load measurement as a means to advance cognitive load theory. In *Cognitive load theory* (pp. 63-71). Routledge.

Puentedura, R. (2006). *Transformation, Technology, and Education, Hippasus*. Retrieved from http://hippasus.com/resources/tte/

Purvis, A., & Crawford, J. (2024). Ethical standards in Social Science Publications. *Journal of University Teaching and Learning Practice, 21*(9). DOI: https://doi.org/10.53761/hqnqr710

Reeves, T. (2005). No significant differences revisited: A historical perspective on the research informing contemporary online learning. *Online learning: Personal reflections on the transformation of education*, *2005*, 299-308.

Robson, S., & Manacapilli, T. (2014). *Enhancing performance under stress: Stress inoculation training for battlefield airmen*. RAND corporation

Ryu, S. (2020). The role of mixed methods in conducting design-based research. *Educational Psychologist, 55*(4), 232-243. https://doi.org/10.1080/00461520.2020.1794871

Schön, D. A. (2017). *The reflective practitioner: How professionals think in action*. Routledge.

Shah, Y., Alinier, G., & Pillay, Y. (2016). Clinical handover between paramedics and emergency department staff: SBAR and IMIST-AMBO acronyms. *International Paramedic Practice, 6*(2), 37-44. https://doi.org/10.12968/ippr.2016.6.2.37

Shahzeydi, A., Dianati, M., & Kalhor, F. (2024). Clinical Simulation in Nursing Students' Safe Medication Administration: A Systematic Review. *Iranian Journal of Nursing and Midwifery Research, 29*(5), 522-529. DOI: 10.4103/ijnmr.ijnmr_323_23

Steed, A., & Schroeder, R. (2015). Collaboration in immersive and non-immersive virtual environments. In *Immersed in media: Telepresence theory, measurement & technology* (pp. 263-282). Springer. DOI: https://doi.org/10.1007/978-3-319-10190-3_11

Steel, C. (2012). Linking theory to practice in learning technology research. *Research in Learning Technology*, *20*. DOI: 10.3402/rlt.v20i0.16148

Stretton, T., Cochrane, T., & Narayan, V. (2018). Exploring mobile mixed reality in healthcare higher education: A systematic review. *Research in Learning Technology*, *26*(0). https://doi.org/10.25304/rlt.v26.2131

Sutherland, P. A. (1992). *Cognitive development today: Piaget and his critics*. SAGE.

Sweller, J., Van Merrienboer, J. J., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational psychology review, 31*, 261-292. DOI: https://doi.org/10.1007/s10648-019-09465-5

Turner, M. J., Jones, M. V., Sheffield, D., & Cross, S. L. (2012). Cardiovascular indices of challenge and threat states predict competitive performance. *International Journal of Psychophysiology, 86*(1), 48-57. https://doi.org/10.1016/j.ijpsycho.2012.08.004

Van Merriënboer, J. J., & Sweller, J. (2010). Cognitive load theory in health professional education: design principles and strategies. *Medical Education, 44*(1), 85-93. https://doi.org/10.1111/j.1365-2923.2009.03498.x

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. *Harvard UP*.

Weaver, D., Robbie, D., Kokonis, S., & Miceli, L. (2013). Collaborative scholarship as a means of improving both university teaching practice and research capability. *International journal for academic development, 18*(3), 237-250. https://doi.org/10.1080/1360144X.2012.718993

Wong, M., Yee, K., & Turner, P. (2008). Clinical handover literature review.. eHealth Services Research Group, University of Tasmania. *Tasmania, Australia: Australian Commission on Safety and Quality in Health Care*. 111-114.

Zhou, Y., Zhou, Y., & Machtmes, K. (2024). Mixed methods integration strategies used in education: A systematic review. *Methodological Innovations, 17*(1), 41-49. https://doi.org/10.1177/2059799123121793