Technology, Knowledge and Learning
https://doi.org/10.1007/s10758-025-09941-8

ORIGINAL RESEARCH

Debriefing in Virtual Reality Simulations for the Development of Counseling Competences: Human-Led or AI-Guided?

Dominik Evangelou[1] · Miriam Mulders[1] · Kristian Heinrich Träg[1]

Received: 24 June 2025 / Revised: 10 December 2025 / Accepted: 12 December 2025

Abstract

Simulation-based learning has established itself as a powerful instructional method in higher education, especially in domains that require complex interpersonal competencies such as counseling. Debriefing plays a pivotal role in transforming simulation experiences into meaningful learning by fostering critical reflection and integration of knowledge. While expert-facilitated debriefings are considered the gold standard, recent advancements in generative Artificial Intelligence (AI) have made chatbot-guided self-debriefings a scalable alternative. This study examines the effectiveness of moderated versus chatbot-supported debriefing formats following a Virtual Reality (VR)-based counseling simulation. A total of 45 undergraduate students in educational science participated in a controlled experiment. All participants engaged in a VR counseling scenario and were subsequently assigned to either a human-moderated or a chatbot-guided debriefing condition. The study investigated changes in counseling competence, self-efficacy, and learner perceptions across the two debriefing formats. Both self-efficacy and counseling competence increased significantly over time. The largest gains, particularly in counseling competence, were recorded after the debriefing, even though these gains were independent of the specific debriefing method employed. These findings underscore the importance of debriefing in educational contexts and indicate that a chatbot-based format could serve as a feasible alternative to traditional, moderator-led debriefings.

Keywords Virtual Reality · Chatbot · AI · GenAI · Counseling · Training · Debriefing

✉ Dominik Evangelou
dominik.evangelou@uni-duisburg-essen.de

Miriam Mulders
miriam.mulders@uni-duisburg-essen.de

Kristian Heinrich Träg
kristian.traeg@uni-duisburg-essen.de

[1] Chair of Educational Technology and Instructional Design, University of Duisburg-Essen, Universitätsstraße 2, 45141 Essen, Germany

Published online: 28 January 2026

1 Introduction

Debriefing is a widely applied technique across various fields, such as trauma processing evaluation of simulated military missions, psychological post-experimental research, and simulation scenario analysis (Dieckmann, 2018). This paper concentrates on post-simulation debriefing following a Virtual Reality (VR) training. Here, debriefing involves reflection on the simulation experience and analysis of psychological and social processes occurring during the simulation, aiming to facilitate the transfer of learning outcomes to real-life contexts (Kriz & Nöbauer, 2015). Integrating debriefing into VR-based educational settings is essential, as it enables learners to critically assesses their decision-making, consequences, and alternative actions (Luctkar-Flude et al., 2021a). Crookall (2023) underscores debriefing as the core-component of simulation learning, where individuals transform their simulation experiences into knowledge. Regardless of a serious game's design, effective learning depends on coupling gameplay with appropriate debriefing. Thatcher (1990) highlighted the pivotal role of debriefing as a mechanism that structures cognitive reflection and synthesizes experiences into a coherent framework, thereby facilitating the transfer of acquired knowledge and skills to new learning environments. Building on Schön´s (1992) theory of reflection, debriefing can be understood as a structured process that bridges reflection-in-action and reflection-on-action. While reflection-in-action occurs during the simulation as participants interpret and respond to dynamic situations in real time, the debriefing phase provides the necessary space for reflection-on-action—the retrospective analysis of experiences, decisions, and emotional responses.

The literature on debriefing generally distinguishes between two primary types: moderated debriefing and self-debriefing. These approaches vary based on factors such as participant numbers, instructional methods, technological requirements, and the role of the instructor (Dufrene & Young, 2014; Favolise, 2024; Luctkar-Flude et al., 2021a). Moderated debriefing is highlighted by the INACSL Board of Directors (2016), the governing body of the International Nursing Association for Clinical Simulation and Learning, as the most effective method for facilitating reflection. While moderated debriefing is widely adopted in practice (e.g., Boet et al., 2011; Tilton, 2013; Tilton et al., 2015; Verkuyl et al., 2018), empirical evidence does not unequivocally support its superiority in terms of effectiveness (Dufrene & Young, 2014). This discrepancy underscores the need for further investigation into the efficacy of different debriefing methods. Moreover, although research on debriefing is expanding, a substantial proportion of studies focus on the healthcare sector (e.g., Cheng et al., 2020). Accordingly, the systematic extension of research to encompass educational settings constitutes a critical area for future research and theory development (Dufrene & Young, 2014; Favolise, 2024; Garden et al., 2015; Luctkar-Flude et al., 2021a).

Based on this background, the present study investigates how different formats of post-simulation debriefing—specifically expert-moderated versus chatbot-guided—affect learning outcomes and student perceptions in a VR-based counseling training. It is assumed that debriefing enhances both students` self-efficacy and counseling competence compared to pre-debriefing levels. Furthermore, based on prior findings emphasizing the importance of human facilitation for deep reflection, it is hypothesized that moderated debriefing will lead to higher gains in self-efficacy and competence, as well as more positive evaluations than chatbot-guided self-debriefing.

This article first reviews the current state of research of various debriefing methods to establish and justify the research question and underlying assumptions. Subsequently, the methods section describes the study procedure in detail, providing in-depth insights into the design of both debriefing approaches, as such descriptions are often underrepresented in the literature. Finally, the results are presented, critically discussed, and contextualized within the fields of educational research and practice.

2 Background

2.1 Simulation-Based Learning in Higher Education

Simulation-based learning has become an increasingly important pedagogical approach in higher education, particularly in fields requiring the development of practical competencies such as counseling, healthcare, and teacher education (Cook et al., 2011; Lateef, 2010; Schmid Mast et al., 2018). By creating authentic, risk-free environments, simulations allow learners to apply theoretical knowledge in realistic scenarios and develop skills such as decision-making, communication, and critical thinking (Neundlinger et al., 2022; Stiefelbauer & Janko, 2023).

With the rapid and continuous advancement of immersive technologies such as VR, simulation experiences have gained new dimensions of realism, interactivity, and emotional engagement (Radianti et al., 2020; Makransky & Petersen, 2021; Caldas & Aviles, 2020). Social VR platforms, a specific form of VR that allows individuals to meet and interact synchronously in a shared virtual space, create environments for role-play and experiential learning that closely mimic real-life situations (Mystakidis et al., 2021).

Recent advances in artificial intelligence (AI), particularly in generative AI and large language models (LLMs), are further transforming the landscape of simulation-based learning. AI-driven avatars and conversational agents can now simulate complex social interactions, such as client interviews, feedback sessions, or emotionally charged counseling scenarios (Benfatah et al., 2024; van As & Cooke, 2024). These intelligent agents allow for adaptive, context-sensitive responses that create more personalized and engaging learning experiences. Moreover, AI can support learners not only during the simulation itself but also in reflective processes afterward, for instance by guiding self-debriefings via chatbots or evaluating performance metrics in real time (Evangelou et al., 2025; Bimpong et al., 2024). The integration of AI in simulation environments thus holds considerable potential to increase scalability, accessibility, and adaptability of educational interventions, especially in higher education contexts where instructional resources are limited. However, these possibilities also raise questions about the pedagogical quality and effectiveness of AI-mediated experiences compared to traditional, human-guided approaches.

While simulations provide rich experiential contexts for learning, their effectiveness depends on the extent which learners engage in systematic reflection on their experiences. This underscores the importance of debriefing as the pedagogical mechanism that transform experience into learning.

2.2 The Role of Debriefing

Within simulation-based learning, debriefing serves as the central reflective component that bridges experience and conceptual understanding. Debriefing is widely regarded as a central component of simulation-based learning, as it facilitates reflective processes that consolidate and deepen learning (Dreifuerst, 2015; Fey & Jenkins, 2015). In educational contexts, debriefing provides an opportunity for learners to analyze their actions, consider alternatives, and connect simulated experiences with theoretical concepts and real-world applications (Sawyer et al., 2016).

2.2.1 Why Debriefing Enhances Learning

From a theoretical standpoint, the effectiveness of debriefing can be explained through Kolb's (2014) experiential learning theory, which identifies reflection as the central mechanism that transforms concrete experience into abstract knowledge. In Kolb's model, learning unfolds as a cyclical process comprising four stages: concrete experience, reflective observation, abstract conceptualization, and active experimentation. Within this cycle, debriefing constitutes the critical transition from experience to abstraction, providing the cognitive and social space for learners to interpret events, uncover underlying principles, and integrate new insights into their mental frameworks.

Kolb's assumptions are further supported by Gibbs (1988), who underscored that in experiential learning, not only the performance of an action but, more importantly, the subsequent reflection on that action and the contextualization of what has been learned within theoretical frameworks are indispensable for meaningful knowledge construction.

Building on these foundations, virtual simulations extend experiential learning by enabling learners to acquire and apply practical skills in a safe and controlled environment while engaging in structured reflection afterward. In this context, experience serves as an essential source of learning, whereas systematic reflection provides the foundation for transforming practice into understanding (Fanning & David, 2007).

Schön's (1992) theory of reflection complements this perspective by highlighting that learning emerges from the dynamic interplay between action and thought. While reflection-in-action occurs during the simulation as participants respond to unfolding situations, reflection-on-action takes place afterward, most prominently during debriefing, when learners critically analyze their previous decisions and performance. Debriefing thus operationalizes reflection-on-action, turning practice into insight and fostering the development of professional competence through guided reflection.

In addition to its cognitive role, debriefing encompasses strong affective and metacognitive dimensions. It allows learners to process emotions that arise during simulation, to evaluate their performance, and to articulate personal learning goals. This emotional integration enhances self-efficacy, supports professional identity formation, and strengthens the capacity for critical self-reflection (Cantrell, 2008; Rudolph et al., 2006).

Empirical research substantiates these theoretical principles. Kriz et al. (2007) and Hense and Kriz (2008) showed that even highly realistic simulations fail to produce lasting learning effects when debriefing is absent or poorly structured. Similarly, Ryoo and Ha (2015) found in a pre–post study that students, specifically nursing professionals, exhibited significantly improved clinical competencies and self-reflection skills following a debriefing session. In line with these findings, Shinnick et al. (2011) argued that learning does not occur primarily or exclusively during the simulation itself, but rather that the subsequent debriefing is the decisive factor in achieving measurable learning gains. This insight is particularly relevant for educators employing (virtual) simulations, as it underscores the pivotal role of well-structured and adequately timed debriefing sessions in facilitating meaningful and sustainable learning outcomes.

Taken together, these theoretical and empirical insights illustrate why debriefing is indispensable in simulation-based education: it transforms action into understanding, emotion into insight, and isolated experience into transferable knowledge.

2.3 The Structure and Formats of Debriefing

Having established *why* debriefing facilitates learning, the next question concerns *how* it should be designed and conducted to maximize educational effectiveness.

2.3.1 How Debriefing Should be Structured

Various models describe how post-simulation debriefing should be structured. These typically comprise between three and seven stages that define the process in detail. Generally, debriefing begins with the pre-briefing phase, during which the conditions for structured reflection and the overall procedure are discussed with learners. This includes establishing a psychologically safe environment and clearly defining learning objectives (Crookall, 2023; Jain, 2022).

Regardless of the specific model, debriefing commonly follows three main phases: a *reaction phase*, in which participants share initial emotional responses; an *analysis phase*, focusing on actions and applied skills; and a *summary phase*, where insights are linked to objectives and future learning activities (Jain, 2022).

Thatcher and Robinson (1985) describe the debriefing process in five stages: identifying the impacts of the experience, examining processes in detail, clarifying key concepts and principles, analyzing emotions, and considering the perspectives formed by participants. Their framework represents one of the earliest attempts to systematize reflection, emphasizing the progression from emotional to analytical engagement.

The *Structured Debriefing in Simulation-Based Education* model by Palaganas et al. (2016) comprises three consecutive phases. In the *Reactions* phase, participants share immediate impressions and emotions following the simulation, which helps to reduce tension and ease the transition into reflection. The *Understanding* phase forms the core of the debriefing, focusing on analyzing performance gaps and exploring underlying thought processes. Finally, the *Summary* phase consolidates key learning points and formulates take-home messages. This model provides a concise and learner-centered structure that integrates emotion, analysis, and application.

The *Debriefing with Good Judgment* (DWGJ) model by Rudolph et al. (2006) focuses less on procedural structure and more on the conditions that enable effective reflection. The authors argue that debriefing can never be entirely free of judgment; thus, evaluative feedback must be communicated in ways that maintain psychological safety and motivation to learn. Three elements are central: supporting learners in recognizing and reframing their *frames*—the assumptions or emotions shaping their actions; adopting an attitude of genuine curiosity toward the learner; and applying the *advocacy–inquiry* technique, combining open feedback (*advocacy*) with exploration of the learner's reasoning (*inquiry*).

Zinns et al. (2020) developed the seven-step *REFLECT* framework to provide educators with a standardized yet time-efficient guide for debriefing. The acronym stands for sequential steps, *Review, Encourage, Focused Feedback, Listen, Emphasize, Communicate,* and *Transform*, hat structure the dialogue from collective review to actionable outcomes. The framework offers a clear conversational pathway that integrates emotional expression, feedback, and forward planning.

The *PEARLS for Systems Integration* (PSI) framework by Dubé et al. (2019) represents a higher-order, system-oriented model that extends the focus of debriefing beyond individual learning toward organizational improvement. It comprises five phases: defining simulation objectives with relevant stakeholders; presenting the scenario to create shared understanding; optionally allowing participants to express impressions or emotions; conducting a structured analysis using tools such as targeted questioning to identify systemic weaknesses; and concluding with a summary that defines responsibilities and next steps for improvement.

Despite their conceptual differences, these models share several fundamental design principles. Each emphasizes the creation of psychological safety to foster trust and openness for honest reflection, a structured progression that guides learners from emotional response to analytical understanding and synthesis, and the integration and transfer of insights into future professional practice.

As Crookall (2023) emphasizes, transparent documentation of debriefing design, its structure, timing, and pedagogical rationale, is crucial for advancing both educational practice and research.

2.3.2 Moderated and Self-Guided Debriefing

Debriefing formats can be broadly categorized into moderated (facilitator-led) and self-guided (independent) approaches.

In moderated debriefing, a trained facilitator leads participants through structured reflection, providing feedback, clarifying misconceptions, and fostering collaborative meaning-making (Cheng et al., 2017; Sawyer et al., 2016). This approach is widely considered the gold standard in simulation-based education because facilitators can dynamically adapt discussions to learners' needs and model critical thinking (Fey & Jenkins, 2015). However, moderated debriefing is resource-intensive, relies on facilitator expertise, and may vary in quality (Cheng et al., 2017).

To address scalability challenges, self-guided debriefing formats have been developed, allowing learners to reflect independently using structured prompts, reflection guides, or digital tools (Boet et al., 2014; Tosterud et al., 2013). Such approaches align with constructivist learning principles, emphasizing learner autonomy and self-regulated learning. Nonetheless, evidence is mixed: self-debriefing may yield superficial reflection among novices (Dufrene & Young, 2014), but when supported by adequate scaffolding—such as reflection questions, visual aids, or conversational AI—it can achieve comparable results to facilitated sessions (Koole et al., 2012; Luctkar-Flude et al., 2021b).

A recent development within this domain is chatbot-guided debriefing, which utilizes conversational AI to emulate facilitator behavior through question sequencing, summarization, and empathy cues. Initial research indicates that chatbot-guided reflection can enhance engagement and accessibility (Favolise, 2024; Verkuyl et al., 2018), though empirical validation remains limited. Although debriefing is widely recognized as essential for learning, its implementation remains resource-intensive and context-dependent. Emerging technologies—particularly conversational agents—offer new opportunities to facilitate reflective processes at scale.

2.4 Conversational Agents in Education

Building on this, conversational agents represent a technological extension of debriefing practice, emulating key aspects of human facilitation to guide learners through structured reflection dialogues. Driven by advances in natural language processing and generative AI, such agents, often referred to as chatbots, are increasingly integrated into educational contexts to support learning, feedback, and reflection (Winkler & Söllner, 2018). They can engage learners in context-sensitive dialogues that approximate human tutoring, enhancing self-regulation and metacognitive engagement (Holmes et al., 2019; Kerlyl et al., 2007).

Within simulation-based learning, conversational agents are gaining traction as scalable alternatives to human facilitators. Recent studies suggest that AI-driven chatbots can lead learners through structured reflection protocols, pose adaptive questions, and provide empathic feedback (Favolise, 2024; Kumar et al., 2025; Verkuyl et al., 2018, Evangelou et al., 2025). Their main advantages lie in scalability, consistency, and independence from time or personnel constraints making them particularly attractive in higher education, where trained facilitators are often scarce.

The emergence of generative AI has further expanded this potential. Modern language models can produce coherent, contextually appropriate responses that support metacognitive processes such as reflection and goal setting (Kasneci et al., 2023; Zawacki-Richter et al., 2019). When pedagogically designed, learners perceive AI-guided debriefings as engaging and helpful (Nghi & Anh, 2024; Wang & Akhter, 2025). Nonetheless, current systems still face notable limitations, such as limited emotional sensitivity, difficulty with nuanced follow-up questioning, and a tendency to confirm rather than challenge learner input (Liang & Hwang, 2023; Sharma et al., 2025).

In summary, conversational agents offer promising opportunities to scale reflective learning processes but remain pedagogically underexplored. Empirical evidence directly comparing AI-guided and human-facilitated debriefings is still scarce, particularly in higher education contexts involving immersive VR simulations. Addressing this gap, the present study investigates how chatbot-supported debriefing compares to expert-moderated sessions in promoting learning and reflection.

2.5 Research Gap and Focus of the Present Study

Empirical comparisons between moderated and self-guided debriefing have yielded inconsistent findings. Some studies report equivalent outcomes regarding learner satisfaction and knowledge gains (Verkuyl et al., 2018), while others find that moderated sessions foster deeper reflection and higher engagement (Garden et al., 2015). Given this heterogeneity, meta-analytic conclusions remain premature.

Despite the well-established importance of debriefing in simulation-based learning, existing research is largely concentrated in the healthcare sector (e.g., Cheng et al., 2014, 2020). There is a growing need to explore debriefing practices in other domains, such as teacher education and counseling training, where reflection plays an equally critical role in developing professional competencies.

At the same time, moderated debriefings, although considered the *gold standard*, are often resource-intensive and may not be feasible in larger-scale or technology-enhanced educational contexts. To address these challenges, recent studies have begun to investigate alternative formats such as self-guided or AI-supported debriefings (e.g., Liang & Hwang, 2023). Conversational agents, particularly chatbots, have shown promise in facilitating structured reflection without requiring continuous instructor involvement. However, empirical evidence directly comparing AI-guided and human-facilitated debriefings remains limited, especially within higher education contexts that incorporate immersive VR technologies.

Against this background, the present study investigates how different formats of post-simulation debriefing, specifically expert-moderated versus chatbot-guided, influence learning outcomes and student perceptions in a VR-based counseling training. The study thereby extends existing simulation research into the field of teacher education and explores scalable, digitally supported reflection practices. Based on the reviewed literature, the study seeks to answer the following research question: To what extent does the format of post-simulation debriefing (moderated vs. chatbot-guided) influence learning outcomes and student perceptions in a VR-based counseling training?

From this question, the following hypotheses were derived and empirically tested:

1. Students' self-efficacy is higher after a debriefing than before a debriefing.
2. Students' counseling competence is higher after a debriefing than before a debriefing.
3. A moderated debriefing is rated more positively by students than a chatbot-guided debriefing.
4. Students' self-efficacy is higher after a moderated debriefing than after a chatbot-guided debriefing.
5. Students' counseling competence is higher after a moderated debriefing than after a chatbot-guided debriefing.

3 Methods

The complete pre-registered study design is available on the *Open Science Framework* (*OSF*).[1] The study was conducted in January 2025 at a large university in Germany and involved undergraduate students of educational science who were enrolled in multiple seminars focusing on the development of counseling competencies. The VR training sessions were implemented using the social VR platform *Engage*, which enabled immersive, simulation-based learning experiences. Prior to participating in the VR training, students had been introduced to the communication techniques—summarizing, paraphrasing and mirroring, using I-messages and questioning techniques—during the respective seminars. These techniques were taught through a dedicated in-person session as well as a screencast presentation, ensuring that all participants had received comparable theoretical input before engaging in the practical simulation. The VR training centered on counseling a fictional client seeking a career change. The client was

[1] OSF-Link: https://osf.io/jguh4/?view_only=44b600c21f864e0794bb41f904f973d2

portrayed by an experienced student research assistant. A more detailed description of the VR training can be also found in Sect. 3.4.

3.1 Sample

Participants were undergraduate students from a large university in Germany who were enrolled in a seminar focused on counseling techniques. All participants had prior knowledge of fundamental conversational strategies. Initially, 46 individuals took part in the study. However, one participant withdrew early due to motion sickness, resulting in a final sample of $N=45$ (age $M=22$; $SD=8.7$). Of these, 22 were randomly assigned to the chatbot self-debriefing condition and 23 to the moderated debriefing condition. The final sample included 6 male and 39 female participants.

3.2 Debriefing Design

To ensure replicability and transparency, the debriefing design was systematically planned and documented (Crookall, 2023; Palaganas et al., 2016). Debriefing was integrated as a core instructional element within the VR-based training and implemented immediately after each individual simulation. This sequencing was intended to promote cognitive consolidation and emotional processing while the experience was still fresh in participants´ minds.

3.2.1 Instructional Design Considerations

The debriefing design was guided by pedagogical and structural principles derived from the theoretical frameworks discussed in 2.3.1. During the planning phase, competency-oriented learning objectives were defined as the foundation for both the simulation and the debriefing. The VR training aimed to foster key communication skills—including summarizing, paraphrasing and mirroring, using I-messages, and applying questioning strategies. To ensure alignment between the VR scenario and these learning objectives, the simulation depicted a realistic counseling situation with a virtual client ("Lena") who was considering a career change. Students assumed the role of counselor and were tasked with supporting the client using the targeted communication techniques. In preparation for the debriefing, contextual factors such as group size, participant composition, and prior experience were considered to establish a psychologically safe and supportive environment (Crookall, 2023). The role of the debriefer followed the framework proposed by Palaganas et al. (2016). As a reminder, this framework describes a simple three-step flow for guiding reflection in simulation-based learning. It begins with participants' immediate reactions, moves into a focused exploration of what happened and why, and concludes with a brief recap of the main insights. This structure supports clear and meaningful reflection on the learning experience.

3.2.2 Debriefing Conditions

Two debriefing conditions were implemented to compare the effects of human- versus AI-guided facilitation:

Moderated debriefing (facilitator-led):

In this condition, the facilitator actively guided the reflection process, asking open-ended questions, clarifying misunderstandings, and helping learners articulate key insights. The facilitator also summarized the discussion and explicitly linked observations to theoretical concepts and practical applications.

Chatbot-guided self-debriefing:

In this condition, the facilitator provided only the introductory instructions, while the debriefing itself was conducted by students using an generative chatbot. The chatbot prompted participants with reflection questions modeled on the same structure as the moderated sessions, encouraging emotional articulation, analytical reasoning, and synthesis.

Both conditions followed an identical temporal structure and shared reflection prompts to ensure comparability of content, duration, and cognitive demand.

3.2.3 Sequence and Timing

Given the short, single session nature of the VR training, a simple sequence model (Crookall, 2023) was applied. The debriefing began immediately after the simulation and consisted of a single reflective episode lasting approximately 20–25 min, corresponding to the duration of the simulation itself. The debriefing followed the Structured Debriefing in Simulation-Based Education by Palaganas et al. (2016), encompassing reaction, understanding, and summary phases:

Reaction Phase (~5 min):

Learners were invited to express immediate emotional responses to the simulation. Facilitators (or the Chatbot) used opening prompts such as *How are you feeling right now?* or *What stood out to you most?* to promote emotional processing and transition toward cognitive reflection. This phase concluded with a short facilitator summary of key scenario elements to ensure shared situational understanding.

Understanding Phase (~10 min):

The reflective discussion focused on analyzing communication strategies and decision-making processes during the simulation. Participants reconstructed their actions, identified challenges, and discussed the effectiveness of applied techniques. Guiding questions included, for example: *Which communication technique did you find most difficult to apply?* or *How did you decide when to use a specific strategy?* Learners were encouraged to generalize their insights and relate them to real-world counseling contexts.

Summary Phase (~5–10 min):

Learners summarized their key takeaways in their own words (e.g., *What is the most important insight you gained form today´s session?*). The facilitator or chatbot synthesized the discussion and connected it to broader professional practice, marking the conclusion of the debriefing.

3.3 Instruments

Quantitative data were collected at three points as part of the experimental study: prior to the VR training, immediately after the training, and following the debriefing. Before completing the first questionnaire, participants generated a unique four-digit personal code to ensure pseudonymized data matching across measurement points. Subsequently, demographic information such as age and gender was collected, along with prior experience using VR technologies.

Counseling competence and self-efficacy were assessed using a shortened version of the questionnaire developed by Hertel (2009). Items referring to the knowledge acquired during teacher education, specifically in the context of school-based counseling, were excluded from the present study. Additionally, the term "*parent counseling*" was replaced with "*counseling*" to ensure the questionnaire addressed counseling in a more general sense, independent of the school setting. To measure counseling competence, 22 items from the corresponding subscale were used (e.g., "*I am able to structure a counseling session in a way that is easy for the client to follow."*), rated on an 8-point Likert scale ranging from 1 ("*does not apply"*) to 8 ("*applies"*). Self-efficacy was measured with 9 items (e.g.*, "I am confident that my counseling can make a difference*."), using a 6-point Likert scale ranging from 1 ("*does not apply at all"*) to 6 ("*fully applies"*).

Debriefing experiences and the perceived importance of the debriefing process were assessed using the Debriefing Experience Scale (Reed, 2012). The questionnaire was adapted for the present study by modifying items originally referring to the healthcare context (e.g., "*The facilitator reinforced aspects of the health care team's behavior."*) to reflect the counseling context. In the self-guided debriefing condition, references to a "*facilitator"* were replaced with "*chatbot"*. The 20-item instrument is composed of four subscales: analysis of thoughts and feelings (e.g., "*Debriefing helped me to analyze my thoughts."*), learning and making connections (e.g., "*Debriefing helped me to make connections in my learning."*), facilitator skill in conducting the debriefing (e.g., "*The facilitator allowed me enough time to verbalize my feelings before commenting."*), and appropriate facilitator guidance (e.g., "*The facilitator taught the right amount during the debriefing session."*). All items were rated on a 5-point Likert scale ranging from 1 ("*strongly disagree"*) to 5 ("*strongly agree"*).

3.4 Procedure

The experimental study was conducted in a university laboratory under the supervision of an experimenter. Students registered for individual 90-min VR training sessions via an online scheduling tool. At the beginning of each session, participants received a concise overview of the procedure. During this briefing, they were informed that a debriefing would follow immediately after the simulation to reflect on their learning experience.

Before the training, participants provided written informed consent and were explicitly informed that they could withdraw from the study at any time without facing any negative consequences. They then completed an initial questionnaire and read a case vignette describing the simulated counseling scenario.

Prior to entering the virtual environment, participants received a short technical introduction to the Head-Mounted Display (HMD) and the VR platform *Engage*, which hosted the simulation. During the VR session (M = 13.91 min; SD = 3.53 min), participants engaged in a counseling conversation with the virtual client "Lena", portrayed by one of three trained female research assistants following a standardized script. They were instructed to apply the three communication techniques targeted in the training *summarizing, paraphrasing and mirroring, using I-messages,* and *questioning strategies.*

Following the VR session, participants completed a second questionnaire and were then randomly assigned to either the moderated or chatbot-guided debriefing condition. In both conditions, the objective of the 20-min debriefing, namely to reflect on the VR experience, was explicitly stated by the experimenter.

In the moderated debriefing, the experimenter sat face-to-face with the participant and followed the three-phase model by Palaganas et al. (2016), asking standardized open- and closed-ended reflection questions for each phase. Each phase focused on different aspects of the VR experience, and key insights were summarized at the end of the session. Experimenters encouraged elaboration by asking open-ended questions (e.g., "*What made you decide to respond in that way?*") and by linking students' reflections to theoretical principles or previous experiences. These sessions typically evolved into dynamic dialogues characterized by follow-up questions and paraphrasing, allowing participants to explore multiple perspectives and evaluate their performance collaboratively.

In contrast, in the chatbot-guided self-debriefing, participants reviewed an instruction sheet outlining the debriefing's purpose and five key reflection points (e.g., *"Which communication techniques did you apply successfully? Why?"*). The printed instruction sheet was placed next to a laptop, where participants interacted with an AI-based chatbot. The chatbot introduced itself, explained the procedure, and led participants through the same three-phase structure used in the moderated sessions. Similar to the moderated debriefing, the chatbot asked learners to describe, analyze, and generalize their experiences.

The chatbot's role and dialogue structure were developed iteratively in consultation with four researchers and pre-tested for clarity and usability by four trained student assistants. The chatbot was implemented using *Meta Llama 3.1 8B Instruct*, a compact instruction-based language model by *Meta AI*, chosen for its reliability and consistent conversational behavior. To balance creativity with response stability, both the *temperature* and *nucleus sampling* parameters were set to 0.5.

The system prompt used to configure the chatbot defined its role as a debriefer and instructed it to guide participants through the *reaction*, *understanding*, and *summary* phases. The initial chatbot message was standardized as follows:

Welcome to the debriefing! I am your debriefer and would like to analyze with you your experiences and insights from the VR training on learning counselling techniques. Our debriefing will be conducted in three phases: Reaction Phase, Understanding Phase, and Summary Phase. In the reaction phase, we will clarify your feelings and emotions after the simulation. In the comprehension phase, we will analyze the activities in the simulation and understand the content learned. In the summary phase, we will generalize your findings, discuss their applicability to practice, and summarize what we have discussed. Let's get started! Briefly describe how you feel at this moment after doing the VR training.

After completing the debriefing, participants filled out a final questionnaire, marking the end of the experimental session. The study procedure is illustrated in Fig. 1. All study

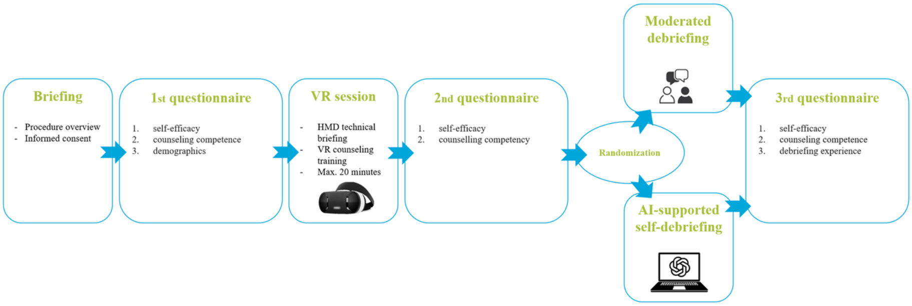

Fig. 1 Experimental procedure

materials—including questionnaires, the moderated debriefing guide, the chatbot instruction sheet, and the vignette versions—are available via OSF.[2]

Following the study, the experiences of the VR simulation and the debriefing were revisited and discussed within the seminar group.

4 Results

The following section presents the quantitative results of our experiment. Prerequisites checks and descriptive statistics are reported first. Subsequently, our findings are presented in relation to the hypotheses outlined in Sect. 2.5. All analyses were conducted using *R* version 4.4.2. To evaluate differences between experimental conditions and test the stated hypotheses, we performed a series of t-tests and analyses of variance (ANOVA).

4.1 Prerequisites

Several Shapiro–Wilk tests were conducted to check whether the different variables follow a normal distribution. Table 1 shows that all variables are normally distributed. In addition, we performed several Levene tests to assess the homogeneity of variance across conditions. Table 2 demonstrates that the variables are homogeneous across the conditions.

Table 1 Normality test—results

	Time	*W*	*p*
Self-efficacy	t1	.98	.667
Self-efficacy	t2	.97	.232
Self-efficacy	t3	.97	.310
Counseling competence	t1	.98	.720
Counseling competence	t2	.96	.177
Counseling competence	t3	.96	.088
Moderated debriefing	t3	.93	.109
Chatbot-debriefing	t3	.98	.914

t measurement time point

[2] https://osf.io/j6c84/?view_only=1ad6da215cdb45dcab42fda76d0b4a3b

Table 2 Homogeneity of variance tests—results

	Time	*df*	*F*	*p*
Self-efficacy	t1	1.43	0.17	.681
Self-efficacy	t2	1.43	0.08	.772
Self-efficacy	t3	1.43	0.10	.748
Counseling competence	t1	1.43	0.19	.661
Counseling competence	t2	1.43	0.18	.673
Counseling competence	t3	1.43	0.22	.641

t measurement time point

Table 3 Descriptive statistics

	Min	*Md*	Max	*M*	*SD*	α
Self-efficacy (t1)	2.38	4.1	5.75	4.1	0.82	.89
Self-efficacy (t2)	1.5	4.13	5.75	4.2	0.91	.92
Self-efficacy (t3)	2	4.5	6	4.4	0.94	.94
Counseling competence (t1)	3.86	5.4	7.29	5.4	0.83	.94
Counseling competence (t2)	2.23	5.59	7.52	5.6	1.14	.96
Counseling competence (t3)	2.82	6	7.68	5.9	1.11	.96
Moderated debriefing	3.74	4.4	5	4.4	0.39	.93
Self-debriefing	3	3.9	4.95	3.9	0.50	.91

Md median, *t* measurement time point

To control for potential confounding effects related to task duration, we examined whether the length of the debriefing sessions differed substantially between conditions. The overall average duration was 14.48 min ($SD = 4.2$). Participants in the moderated debriefing (MB) condition spent an average of 14.6 min ($SD = 2.92$), whereas those in the chatbot debriefing (SB) condition spent 14.4 min ($SD = 5.35$). A Levene's test indicated a significant difference in variance between the two groups $F(1.43) = 6.59$, $p = .014$, suggesting heterogeneity of variance. Accordingly, Welch's t-test was conducted and revealed no significant difference in debriefing duration between the two conditions, $t(32.18) = 0.12$, $p = .905$.

4.2 Descriptive Statistics

Table 3 presents the descriptive statistics for all variables. All scales demonstrate good to excellent internal consistency, as indicated by Cronbach's Alpha (Cronbach, 1951). Statistical significance tests for individual values are described in the hypothesis testing section.

4.3 Hypothesis Testing

To investigate the development of self-efficacy and counseling competence across the intervention, a series of paired-samples t-tests and repeated-measures ANOVAs were conducted, incorporating both debriefing methods and measurement time points.

A paired-samples t-test revealed a significant increase in self-efficacy from measurement time point 1 to measurement time point 3, $t(44) = 3.47$, $p < .001$. A comparison between measurement time point 2 (following the VR simulation) and measurement time point 3 similarly yielded a significant effect, $t(44) = 3.94$, $p < .0001$, supporting the hypothesis that self-efficacy improves following a debriefing session.

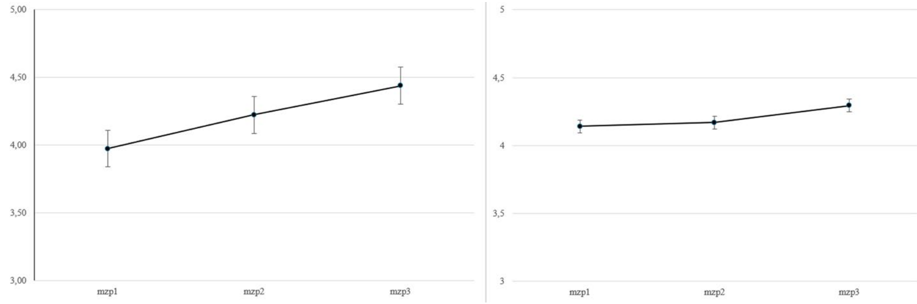

Fig. 2 Average increase in self-efficacy for moderated debriefing (left) and self-debriefing (right). While the original Likert scale ranged from 1 to 6, only the relevant range from 3 to 5 is displayed here (mp = measurement point)

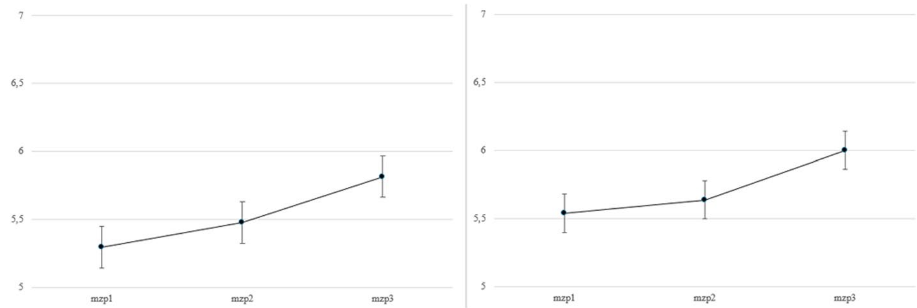

Fig. 3 Average increase in counseling competence for moderated debriefing (left) and self-debriefing (right). While the original Likert scale ranged from 1 to 8, only the relevant range from 5 to 7 is displayed here (mp = measurement point)

In terms of counseling competence, the data revealed even more robust results. A significant increase was observed between measurement time points 1 and 3, $t(44) = 4.27$, $p < .001$ ($\Delta M = -.49$), and between time points 2 and 3, $t(44) = 5.56$, $p < .001$ ($\Delta M = -0.35$). No significant difference was found between time points 1 and 2. These findings suggest a marked improvement in perceived counseling competence over time, particularly after debriefing. Figures 2 and 3 illustrate the average increase in self-efficacy and counseling competence across the three measurement time points for the two debriefing methods.

An ANOVA for repeated measurements examined the effects of debriefing method and measurement time points on self-efficacy and counseling competence. Mauchly´s test indicated that the assumption of sphericity was violated for both outcomes (all $p < .001$). As a result, the effect of time was evaluated using Greenhouse–Geisser adjusted degrees of freedom.

Results showed a significant main effect of measurement time point on self-efficacy, $F(1.38,59.55) = 8.47$, $p = .002$, $\eta^2(g) = .020$, and on counseling competence, $F(1.45,62.40) = 12.20$, $p < .001$, $\eta^2(g) = .039$. These findings confirm that both values improved significantly over the course of the intervention. In contrast, no main effect of the debriefing method was found for self-efficacy ($F(1.43) = 0.001$, $p = .971$) or counseling competence ($F(1.43) = 0.46$, $p = .500$), nor were any significant interaction effects detected. Post

Table 4 Post-hoc comparisons

	SE	*df*	*Estimate*	*p*
S (t1)–S (t2)	.0766	88	.142	.2033
S (t1)–S (t3)	.0766	88	.314	.0003
S (t2)–S (t3)	.0766	88	.172	.0812
C (t1)–C (t2)	.101	88	.143	.4884
C (t1)–C (t3)	.101	88	.492	<.0001
C (t2)–C (t3)	.101	88	.350	.0026

S self-efficacy, *C* counseling competence, *t* measurement time point, *SE* standard error

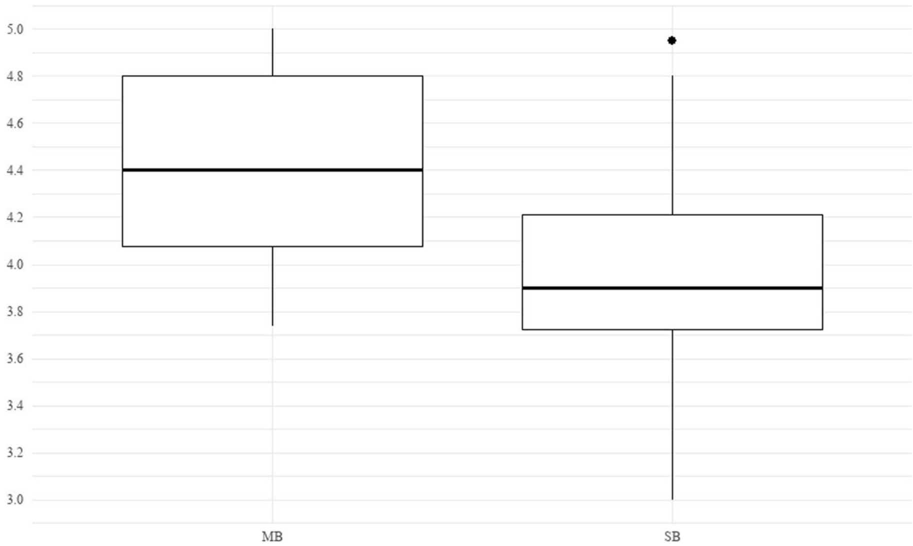

Fig. 4 Boxplot for the evaluation of the moderated debriefing (MB) and the chatbot debriefing (SB). While the original Likert scale ranged from 1 to 5, only the relevant range from 3 to 5 is displayed here

hoc comparisons using Estimated Marginal Means (EMMs) provided further insights into temporal trends, as detailed in Table 4. Self-efficacy showed a significant increase between measurement time points 1 and 3, with a non-significant difference between time points 1 and 2, and a marginal trend between time points 2 and 3. For counseling competence, significant improvements emerged between time points 1 and 3 and between time points 2 and 3, while no significant change was observed from time point 1 to 2.

To compare the perceived effectiveness of the two debriefing formats, an independent samples t-test was conducted. Participants rated the MB significantly higher than the SB, $t(39.806)=3.68$, $p=.0007$, $d=1.10$, indicating greater satisfaction and perceived value in the MB condition (Fig. 4). This remains consistent across all four sub-facets of the debriefing scale: Analyzing thoughts and feelings ($t(42.2)=2.87$, $p=.006$, $d=0.86$), learning and making connections ($t(37.8)=3.57$, $p<.001$, $d=1.07$), facilitator skill ($t(40.2)=2.41$, $p=.021$, $d=0.72$), and appropriate facilitator guidance ($t(37.5)=3.57$, $p<.001$, $d=1.07$) were all rated higher in the moderated group. It may be notable that in the human-moderated condition, the subscales referring to the facilitator do not differ significantly from each other, ($t(22)=0.06$, $p=.953$, $d=0.01$), while they do in the AI-led condition ($t(21)=2.32$, $p=.031$, $d=0.42$), with facilitator skill showing higher values than appropriate facilitator guidance.

In sum, these findings demonstrate that both self-efficacy and counseling competence improved significantly across measurement time points, particularly following the debriefing session. However, the format of the debriefing, moderated versus chatbot-based, did not exert a differential impact on these outcomes. Consequently, the hypothesis that moderated debriefing leads to superior learning outcomes compared to chatbot-based debriefing cannot be supported. These compelling findings on the effects of additional debriefing in its various forms will be analyzed in detail in the following discussion section.

5 Discussion

5.1 Theoretical and Practical Contributions

The findings of our study underscore the pivotal role of debriefing in simulation-based learning, particularly within the domain of professional skill acquisition such as counseling. Both self-efficacy and counseling competence showed significant improvement following the debriefing phase, confirming theoretical models that position reflection as a key mechanism in the transformation of experience into learning (e.g., Kolb, 2014). These results also lend support to instructional design principles that emphasize the importance of post-experiential reflection for cognitive consolidation and transfer (DiStefano et al., 2014; Träg & Mulders, 2025). Consequently, the isolated use of VR is insufficient. Rather, it requires meaningful, contextually integrated accompanying activities to facilitate a sustainable transfer of learning into real-world practice.

An interesting finding of the study concerns the divergent development of self-efficacy and counseling competence over time. Self-efficacy increased most markedly between measurement time points 1 and 3, whereas only a slight upward trend was observed from time point 2 to 3. In contrast, counseling competence exhibited a significant improvement specifically between the second and third measurement. One plausible explanation for this divergence is the nature of the feedback received during the debriefing sessions: while affirming, it also included constructive elements that may have enhanced students' awareness of their actual professional competence, thereby moderating previously elevated perceptions of self-efficacy. Consequently, students who were confronted with more critical feedback may have reassessed their self-efficacy and reported lower values at subsequent measurement points.

Another particularly noteworthy outcome is the absence of significant differences in learning outcomes between the human-moderated and the chatbot-guided debriefing formats. Although the moderated debriefing was rated more positively by participants in terms of perceived value and satisfaction, both conditions yielded comparable gains in perceived counseling competence and self-efficacy. This result invites further reflection on the underlying mechanisms of effective debriefing. It suggests that the cognitive engagement required for structured reflection may play a more decisive role than the social presence of a human facilitator. From a cognitive-psychological perspective, this aligns with the notion that metacognitive self-regulation processes can be activated through well-designed prompts and structured guidance, even in the absence of direct social interaction (Bannert, 2009). On the other hand, the fact that the human debriefing facilitator was rated as more appropriate and skillful may be hinting towards underlying social factors within the debriefing that a chatbot seemingly cannot imitate. This may be due to its lack of human appearance, which in combination with the human-like conversational style portrayed here might be able to further improve the evaluation of the chatbot (Chen et al., 2023).

The findings also expand upon existing conceptions of debriefing by illustrating that a technology-mediated, self-guided process can achieve similar effects to traditional human-facilitated sessions. However, the effectiveness of the chatbot-based debriefing likely depended on several key factors: the use of a clearly defined theoretical reflection model, the prebriefing instructions, and the conversational flow designed to simulate empathic, learner-centered dialogue. These features align with core principles from established debriefing frameworks such as *PEARLS* (Eppich & Cheng, 2015) or *Debriefing with Good Judgment* (Rudolph et al., 2006), suggesting that structural and procedural quality may be more critical than the specific mode of delivery.

From a practical standpoint, the results point to promising implications for the scalability of simulation-based education. Chatbot-guided debriefing offers a resource-efficient alternative in contexts where access to trained facilitators is limited. It may be particularly relevant in higher education settings characterized by large class sizes, time constraints, or asynchronous learning environments. Furthermore, self-debriefing via chatbot represents a promising addition to the instructional design of Massive Open Online Courses (MOOCs), allowing learners to engage in autonomous reflection and deepen their understanding in the absence of instructor-led feedback. However, while the chatbot was able to guide reflection and foster cognitive engagement, it lacked the nuanced emotional sensitivity and adaptive questioning strategies that human facilitators may provide. This is backed by the fact that the subscale appropriate facilitator guidance was rated significantly lower in the chatbot condition. Additionally, in another paper (Evangelou et al., 2025), we reported that the chatbot occasionally responded with an overly enthusiastic tone, which was perceived as inauthentic by some users. This underscores the importance of calibrating chatbot personality and emotional expressiveness in alignment with the conversational context. Furthermore, the transcript analysis revealed that tutor-led debriefings tended to elicit more elaborated and evaluative reflections, whereas chatbot dialogues remained more structured and confirmatory. This difference likely stems from the tutors' ability to dynamically adapt their questioning, a feature still limited in current conversational agents. These insights are based on ongoing qualitative analyses of the debriefing transcripts, which will be reported in future publications to provide a more in-depth understanding of the conversational mechanisms underlying both facilitation formats. In general, however, chatbot-based formats should be viewed not as a replacement but as a complementary tool within a broader instructional ecosystem, especially suited for preliminary reflection or situations where human moderation is not feasible.

The findings of this study suggest that chatbot-guided debriefing can approximate essential elements of human facilitation, particularly in supporting structured reflection and self-assessment. Comparable to reflective dialogues in counseling training contexts, similar mechanisms may operate in other domains such as medical and teaching education, where debriefing serves as a key vehicle for transforming experience into learning (Cheng et al., 2020; Rudolph et al., 2006). In teacher education, for instance, chatbot-guided debriefings could accompany microteaching sessions or classroom simulations, helping preservice teachers to articulate pedagogical reasoning and critically evaluate their instructional decisions. Similarly, in medical or nursing education, conversational agents could provide standardized post-simulation reflections that promote psychological safety and consistency across cohorts. These scalable formats would enable instructors to review and build upon students' AI-facilitated reflections, reducing the need for continuous human moderation while maintaining pedagogical quality (Kumar et al., 2025; Verkuyl et al., 2018).

However, integrating AI-based debriefings into practice also presents challenges. These include ensuring data protection and ethical transparency (Kasneci et al., 2023), maintaining psychological safety and trust (Cheng et al., 2020), and preventing overly confirmatory or shallow reflection patterns (Sharma et al., 2025). Hybrid models that combine AI and human facilitation may help to overcome these limitations: educators can use chatbot transcripts for meta-reflection, iteratively refine prompts through learner feedback, and embed such systems within institutional quality frameworks (Winkler & Söllner, 2018; Zawacki-Richter et al., 2019). Taken together, these approaches highlight how generative AI can extend reflective learning opportunities while retaining ethical and pedagogical integrity.

5.2 Limitations and Future Directions

While the findings of this experimental study offer valuable insights into the comparative effectiveness of human-moderated and chatbot-guided debriefing formats, several limitations warrant consideration.

First, the relatively small sample size limits the generalizability of our results. Although the observed effects were statistically significant and robust across analyses, larger and more diverse samples are needed to validate the findings and explore potential moderating variables such as prior digital literacy, language proficiency, or interpersonal sensitivity.

Second, the study focused on a single, relatively brief VR simulation embedded in a tightly controlled laboratory setting. While this design ensured experimental control and comparability across conditions, it does not fully capture the complexity and variability of real-world learning environments. In particular, longitudinal designs that embed multiple reflection episodes across extended learning processes, such as semester-long counseling practica, may provide a more comprehensive picture of how debriefing formats contribute to the development of professional competence over time. Future research should explore the cumulative effects of debriefing in iterative learning cycles and investigate how chatbot-based reflection performs in more authentic, situated contexts.

Third, the study focused exclusively on perceived outcomes, namely self-reported self-efficacy and counseling competence, without including independent assessments of performance or behavioral indicators. Although self-perception is closely related to engagement and motivation (Cheng & Tsai, 2020; Makransky et al., 2019), it does not always correspond to actual skill development. More experienced counselors-in-training seem to generally be more accurate in their self-assessments (Lepkowski et al., 2009). For our relatively inexperienced sample, this means that on the one hand, expert ratings might be preferable to accurately assess counseling competence, while being more resource-intensive. Alternatively, creating an objective, reliable, and valid questionnaire to capture a set of complex skills like counseling competence may prove difficult. Still, future studies should incorporate objective performance measures such as objective questionnaire items, behavioral coding of counseling techniques or expert ratings of recorded sessions to triangulate findings.

Fourth, we did not include a control group without any debriefing. While such a comparison could have provided additional insights into the specific contribution of debriefing to learning outcomes, it was deliberately omitted in the current study due to ethical and pedagogical considerations. In the context of higher education, withholding debriefing from one group would have meant depriving students of an essential reflective learning opportunity that is considered integral to simulation-based instruction. Since debriefing serves not only as a feedback mechanism but also as a means of emotional processing and knowledge consolidation, excluding it entirely would have placed participants at a clear disadvantage compared to their peers. Moreover, including a no-debriefing control group would have introduced substantial differences in time on task between conditions, making it difficult to disentangle whether potential learning effects were due to the reflective process itself or simply to the additional time spent engaging with the material (Buchner, 2023). However, future research should address this limitation by systematically comparing conditions with and without debriefing under controlled circumstances, ideally in low-stakes or voluntary learning contexts where the absence of reflection does not negatively affect participants. Such designs could help isolate the incremental impact of debriefing and further clarify the unique benefits of human versus AI-guided facilitation. Including physiological or behavioral indicators of reflection (e.g., think-aloud protocols or eye-tracking measures) might also provide more fine-grained evidence of how debriefing mechanisms influence cognitive and affective learning processes.

From a methodological standpoint, future research should incorporate qualitative data gathered during debriefing sessions to enrich the interpretation of quantitative findings. In particular, it would be worthwhile to examine whether students who received predominantly critical feedback tend to report lower levels of self-efficacy, whereas those exposed to more positive feedback may exhibit higher self-efficacy. A mixed-methods approach could thus yield deeper insight into individual learning processes and the underlying mechanisms shaping self-assessment. Furthermore, the integration of sentiment analysis could help identify emotional patterns within the feedback, thereby offering a more nuanced understanding of how the valence of feedback influences learners' perceptions of their own competence.

Finally, while both debriefing formats were implemented with attention to instructional design, the human-moderated version naturally benefited from dynamic adaptation and personalized follow-up questioning. These are features that current AI systems can only partially emulate. Moreover, generative models often display sycophantic behavior, tending to affirm user input uncritically rather than challenging misconceptions or encouraging deeper reflection (Sharma et al., 2025). This limitation may reduce the potential for meaningful cognitive conflict, which is essential for learning through dialogue. Furthermore, chatbots are not tutoring systems and typically lack an underlying learner model, thereby requiring students to regulate their learning processes through metacognitive strategies (Klar, 2025). As generative AI continues to evolve, future research should investigate how to best balance structure and flexibility in conversational agents, and under which conditions such systems can meaningfully replicate or even enhance the pedagogical value of human facilitation.

6 Conclusion

This study investigated the comparative effectiveness of moderated and chatbot-guided debriefing following a VR-based counseling simulation in higher education. The findings demonstrate that both debriefing formats led to significant improvements in participants' self-efficacy and counseling competence. Notably, the largest gains were observed after the debriefing phase, regardless of whether the session was facilitated by a human moderator or a chatbot. These results highlight the fundamental role of debriefing in consolidating learning outcomes within simulation-based education and suggest that AI-supported, chatbot-guided debriefings can serve as a viable and scalable alternative to traditional, resource-intensive moderator-led sessions.

However, while chatbot-guided debriefing offers logistical advantages and supports learner autonomy, current AI systems only partially replicate the nuanced, adaptive feedback and emotional attunement characteristic of expert human facilitators. As generative AI continues to advance, future research should further explore how conversational agents can be optimized to balance structure with flexibility and to what extent they can meaningfully replicate or even enhance the pedagogical value of human moderation.

Overall, the integration of AI-driven debriefing tools holds considerable promise for increasing the accessibility and scalability of simulation-based learning in higher education, provided that their design remains grounded in evidence-based pedagogical principles.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by all authors. The first draft of the manuscript was written by all authors. All authors read and approved of the final manuscript.

Funding Open Access funding enabled and organized by Projekt DEAL. Stiftung Innovation in der Hochschullehre, FR-124/2023, Miriam Mulders.

Declarations

Conflict of interests The authors have no relevant financial or non-financial interests to disclose.

Data Availability Statement The data that support the findings of this study are not publicly available due to privacy and confidentiality restrictions. However, the data are available from the corresponding author upon reasonable request.

References

Bannert, M. (2009). Promoting self-regulated learning through prompts. *Zeitschrift Für Pädagogische Psychologie, 23*(2), 139–145. https://doi.org/10.1024/1010-0652.23.2.139

Benfatah, M., Youlyouz-Marfak, I., Saad, E., Hilali, A., Nejjari, C., & Marfak, A. (2024). Impact of artificial intelligence-enhanced debriefing on clinical skills development in nursing students: A comparative study. *Teaching and Learning in Nursing, 19*(3), e574–e579.

Bimpong, B. W., Atsise, P., & Owusu, F. (2024). Exploring the implementation of artificial intelligence (AI) writing tools in teaching and learning: Faculty and students' perspectives in higher education. *East African Journal of Information Technology, 7*(1), 380–393.

Boet, S., Bould, M. D., Bruppacher, H. R., Desjardins, F., Chandra, D. B., & Naik, V. N. (2011). Looking in the mirror: Self-debriefing versus instructor debriefing for simulated crises*. *Critical Care Medicine, 39*(6), 1377–1381. https://doi.org/10.1097/CCM.0b013e31820eb8be

Boet, S., Bould, M. D., Fung, L., Qosa, H., Perrier, L., Tavares, W., Reeves, S., & Tricco, A. C. (2014). Transfer of learning and patient outcome in simulated crisis resource management: A systematic review. *Canadian Journal of Anesthesia = Journal Canadien D'anesthésie, 61*(6), 571.

Buchner, J. (2023). Beyond media comparison: Investigating when and how learning with augmented and virtual reality works. *Digital Psychology, 4*(1S), 1–3.

Caldas, O. I., Aviles, O. F., & Rodriguez-Guerrero, C. (2020). Effects of presence and challenge variations on emotional engagement in immersive virtual environments. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 28*(5), 1109–1116.

Cantrell, M. A. (2008). The importance of debriefing in clinical simulations. *Clinical Simulation in Nursing, 4*(2), e19–e23.

Chen, J., Guo, F., Ren, Z., Li, M., & Ham, J. (2023). Effects of anthropomorphic design cues of chatbots on users' perception and visual behaviors. *International Journal of Human-Computer Interaction, 40*(14), 3636–3654. https://doi.org/10.1080/10447318.2023.2193514

Cheng, A., Eppich, W., Grant, V., Sherbino, J., Zendejas, B., & Cook, D. A. (2014). Debriefing for technology-enhanced simulation: A systematic review and meta-analysis. *Medical Education, 48*(7), 657–666. https://doi.org/10.1111/medu.12432

Cheng, A., Grant, V., Huffman, J., Burgess, G., Szyld, D., Robinson, T., & Eppich, W. (2017). Coaching the debriefer: Peer coaching to improve debriefing quality in simulation programs. *Simulation in Healthcare, 12*(5), 319–325.

Cheng, A., Kolbe, M., Grant, V., Eller, S., Hales, R., Symon, B., Griswold, S., & Eppich, W. (2020). A practical guide to virtual debriefings: Communities of inquiry perspective. *Advances in Simulation, 5*(1), 18. https://doi.org/10.1186/s41077-020-00141-1

Cheng, K., & Tsai, C. (2020). Students' motivational beliefs and strategies, perceived immersion and attitudes towards science learning with immersive virtual reality: A partial least squares analysis. *British Journal of Educational Technology, 51*(6), 2140–2159. https://doi.org/10.1111/bjet.12956

Cook, D. A., Hatala, R., Brydges, R., Zendejas, B., Szostek, J. H., Wang, A. T., Erwin, P. J., & Hamstra, S. J. (2011). Technology-enhanced simulation for health professions education: A systematic review and meta-analysis. *JAMA, 306*(9), 978–988.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Crookall, D. (2023). Debriefing: A Practical Guide. In: Angelini, M.L., Muñiz, R. (eds) Simulation for Participatory Education. Springer Texts in Education. Springer, Cham. https://doi.org/10.1007/978-3-031-21011-2_6

Dieckmann, P. (2018). *Gute Nachrede – Debriefing*.

DiStefano, G., Gino, F., Pisano, G. P., Staats, B., & Di-Stefano, G. (2014). *Learning by thinking: How reflection aids performance*. Harvard Business School Boston, MA. https://paulpedrazzi.com/s/feedback.pdf

Dreifuerst, K. T. (2015). Getting started with debriefing for meaningful learning. *Clinical Simulation in Nursing, 11*(5), 268–275.

Dubé, M. M., Reid, J., Kaba, A., Cheng, A., Eppich, W., Grant, V., & Stone, K. (2019). PEARLS for systems integration: A modified PEARLS framework for debriefing systems-focused simulations. *Simulation in Healthcare: THe Journal of the Society for Simulation in Healthcare, 14*(5), 333–342. https://doi.org/10.1097/SIH.0000000000000381

Dufrene, C., & Young, A. (2014). Successful debriefing — Best methods to achieve positive learning outcomes: A literature review. *Nurse Education Today, 34*(3), 372–376. https://doi.org/10.1016/j.nedt.2013.06.026

Eppich, W., & Cheng, A. (2015). Promoting excellence and reflective learning in simulation (PEARLS): Development and rationale for a blended approach to health care simulation debriefing. *Simulation in Healthcare, 10*(2), 106–115.

Evangelou, D., Klar, M., Träg K., Mulders M., Marnitz M., & Rahner L. (2025): GenAI-Chatbots as Debriefers: Investigating the Role Conformity and Learner Interaction in Counseling Training. In 23. *Fachtagung Bildungstechnologien (DELFI 2025)* pp. 41-55. Bonn: Gesellschaft für Informatik e.V.. https://doi.org/10.18420/delfi2025_05. PISSN: 2944-7682. Best-Paper-Kandidaten-Langbeiträge. Freiberg. 8.–11. September 2025

Fanning, R. M., & Gaba, D. M. (2007). The role of debriefing in simulation-based learning. *Simulation in Healthcare: THe Journal of the Society for Simulation in Healthcare, 2*(2), 115–125. https://doi.org/10.1097/SIH.0b013e3180315539

Favolise, M. (2024). Post-simulation debriefing methods: A systematic review. *Archives of Physical Medicine and Rehabilitation, 105*(4), Article e146. https://doi.org/10.1016/j.apmr.2024.02.680

Fey, M. K., & Jenkins, L. S. (2015). Debriefing practices in nursing education programs: Results from a national study. *Nursing Education Perspectives, 36*(6), 361–366.

Garden, A. L., Le Fevre, D. M., Waddington, H. L., & Weller, J. M. (2015). Debriefing after simulation-based non-technical skill training in healthcare: A systematic review of effective practice. *Anaesthesia and Intensive Care, 43*(3), 300–308. https://doi.org/10.1177/0310057X1504300303

Gibbs, G. (1988). Learning by doing: A guide to teaching and learning methods. *Further Education Unit*.

Hense, J., & Kriz, W. C. (2008). Making simulation games an even more powerful tool. *Introducing the theory-based evaluation approach. In de Caluwé L., Hofstede GJ, Peters V. Eds., Why do games work*, 211–217.

Hertel, S. (2009). *Beratungskompetenz von Lehrern. Kompetenzdiagnostik, Kompetenzförderung, Kompetenzmodellierung.* Waxmann

Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education promises and implications for teaching and learning*. Center for Curriculum Redesign. https://discovery.ucl.ac.uk/id/eprint/10139722/

INACSL Standards of Best Practice. (2016). SimulationSM Debriefing. *Clinical Simulation in Nursing, 12*, S21–S25. https://doi.org/10.1016/j.ecns.2016.09.008

Jain, S. (2022). Feedback and Debriefing. In: Biyani, C.S., Van Cleynenbreugel, B., Mottrie, A. (eds) Practical Simulation in Urology. Springer, Cham. https://doi.org/10.1007/978-3-030-88789-6_22

Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., & Hüllermeier, E. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, Article 102274.

Kerlyl, A., Hall, P., & Bull, S. (2007). Bringing Chatbots into education: Towards Natural Language Negotiation of Open Learner Models. In R. Ellis, T. Allen, & A. Tuson (Hrsg.), *Applications and Innovations in Intelligent Systems XIV* (S. 179–192). Springer London. https://doi.org/10.1007/978-1-84628-666-7_14

Klar, M. (2025). Using ChatGPT is easy, using it effectively is tough? A mixed methods study on K-12 students' perceptions, interaction patterns, and support for learning with generative AI chatbots. *Smart Learning Environments, 12*(1), 32. https://doi.org/10.1186/s40561-025-00385-2

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Koole, S., Dornan, T., Aper, L., De Wever, B., Scherpbier, A., Valcke, M., Cohen-Schotanus, J., & Derese, A. (2012). Using video-cases to assess student reflection: Development and validation of an instrument. *BMC Medical Education, 12*(1), 22. https://doi.org/10.1186/1472-6920-12-22

Kriz, W. C., Saam, N., Pichlbauer, M., & Fröhlich, W. (2007). Intervention mit Planspielenals Großgruppenmethode–Ergebnisse einer Interviewstudie. *Planspiele für die Organisationsentwicklung. Schriftenreihe: Wandel und Kontinuität in Organisationen (Bd. 8, S. 103–122). Berlin: Wissenschaftlicher Verlag*.

Kriz, W. C., & Nöbauer, B. (2015). Den Lernerfolg mit Debriefing von Planspielen sichern. *Planspiele und Serious Games in der beruflichen Bildung. Auswahl, Konzepte, Lernarrangements, Erfahrungen–aktueller Katalog für Planspiele und Serious Games*.

Kumar, P., Harrison, N. M., McAleer, K., Khan, I., & Somerville, S. G. (2025). Exploring the role of self-led debriefings within simulation-based education: Time to challenge the status quo? *Advances in Simulation, 10*(1), 9. https://doi.org/10.1186/s41077-025-00342-6

Lateef, F. (2010). Simulation-based learning: Just like the real thing. *Journal of Emergencies, Trauma, and Shock, 3*(4), 348–352.

Lepkowski, W. J., Packman, J., Smaby, M. H., & Maddux, C. (2009). Comparing self and expert assessments of counseling skills before and after skills training, and upon graduation. *Education, 129*(3), 363+.

Liang, J.-C., & Hwang, G.-J. (2023). A robot-based digital storytelling approach to enhancing EFL learners' multimodal storytelling ability and narrative engagement. *Computers & Education, 201*, Article 104827.

Luctkar-Flude, M., Tyerman, J., Tregunno, D., Bell, C., Lalonde, M., McParland, T., Peachey, L., Verkuyl, M., & Mastrilli, P. (2021b). Designing a virtual simulation game as presimulation preparation for a respiratory distress simulation for senior nursing students: Usability, feasibility, and perceived impact on learning. *Clinical Simulation in Nursing, 52*, 35–42.

Luctkar-Flude, M., Tyerman, J., Verkuyl, M., Goldsworthy, S., Harder, N., Wilson-Keates, B., Kruizinga, J., & Gumapac, N. (2021a). Effectiveness of debriefing methods for virtual simulation: A systematic review. *Clinical Simulation in Nursing, 57*, 18–30. https://doi.org/10.1016/j.ecns.2021.04.009

Makransky, G., Borre-Gude, S., & Mayer, R. E. (2019). Motivational and cognitive benefits of training in immersive virtual reality based on multiple assessments. *Journal of Computer Assisted Learning, 35*(6), 691–707. https://doi.org/10.1111/jcal.12375

Makransky, G., & Petersen, G. B. (2021). The cognitive affective model of immersive learning (CAMIL): A theoretical research-based model of learning in immersive virtual reality. *Educational Psychology Review, 33*(3), 937–958. https://doi.org/10.1007/s10648-020-09586-2

Mystakidis, S., Berki, E., & Valtanen, J.-P. (2021). Deep and meaningful e-learning with social virtual reality environments in higher education: A systematic literature review. *Applied Sciences, 11*(5), 2412.

Neundlinger, K., Mühlegger, M., Kriglstein, S., Layer-Wagner, T., & Regal, G. (2022). Training Social Skills in Virtual Reality Machine Learning as a Process of Co-Creation. In A. Dingli, A. Pfeiffer, A. Serada, M. Bugeja, & S. Bezzina (Hrsg.), *Disruptive Technologies in Media, Arts and Design* (Bd. 382, S. 139–156). Springer International Publishing. https://doi.org/10.1007/978-3-030-93780-5_10

Nghi, T. T., & Anh, L. T. Q. (2024). Promoting student-centered learning strategies via AI chatbot feedback and support: A case study at a public university in Vietnam. *International Journal of Teacher Education and Professional Development (IJTEPD), 7*(1), 1–25.

Palaganas, J. C., Fey, M., & Simon, R. (2016). Structured debriefing in simulation-based education. *AACN Advanced Critical Care, 27*(1), 78–85. https://doi.org/10.4037/aacnacc2016328

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, Article 103778.

Reed, S. J. (2012). Debriefing experience scale: Development of a tool to evaluate the student learning experience in debriefing. *Clinical Simulation in Nursing, 8*(6), e211–e217. https://doi.org/10.1016/j.ecns.2011.11.002

Rudolph, J. W., Simon, R., Dufresne, R. L., & Raemer, D. B. (2006). There's no such thing as "nonjudgmental" debriefing: A theory and method for debriefing with good judgment. *Simulation in Healthcare, 1*(1), 49–55.

Ryoo, E. N., & Ha, E. H. (2015). The importance of debriefing in simulation-based learning: Comparison between debriefing and no debriefing. *CIN: Computers, Informatics, Nursing, 33*(12), 538–545.

Sawyer, T., Eppich, W., Brett-Fleegler, M., Grant, V., & Cheng, A. (2016). More than one way to debrief: A critical review of healthcare simulation debriefing methods. *Simulation in Healthcare, 11*(3), 209–217.

Schmid Mast, M., Kleinlogel, E. P., Tur, B., & Bachmann, M. (2018). The future of interpersonal skills development: Immersive virtual reality training with virtual humans. *Human Resource Development Quarterly, 29*(2), 125–141. https://doi.org/10.1002/hrdq.21307

Schön, D.A. (1992). The Reflective Practitioner: How Professionals Think in Action (1st ed.). Routledge. https://doi.org/10.4324/9781315237473

Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S. R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S. R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M., & Perez, E. (2025). *Towards Understanding Sycophancy in Language Models* (arXiv:2310.13548). arXiv. https://doi.org/10.48550/arXiv.2310.13548

Shinnick, M. A., Woo, M., Horwich, T. B., & Steadman, R. (2011). Debriefing: The most important component in simulation? *Clinical Simulation in Nursing, 7*(3), e105–e111. https://doi.org/10.1016/j.ecns.2010.11.005

Stiefelbauer, C., & Janko, M. (2023). Innovating teaching and learning with immersive technologies at WU Vienna. *EdMedia+ Innovate Learning*, 1051–1054. https://www.learntechlib.org/primary/d/222613/

Thatcher, C., & Robinson, J. (1985). An introduction to games and simulations in education. Simulations.

Thatcher, D. C. (1990). Promoting learning through games and simulations. *Simulation & Gaming, 21*(3), 262–273.

Tilton, K. J. (2013). *Non-acute Care Clinical for BSN Students Chronic Illness Care and Diabetes Self-management Support*.

Tilton, K. J., Tiffany, J., & Hoglund, B. A. (2015). Non-acute-care virtual simulation: Preparing students to provide chronic illness care. *Nursing Education Perspectives, 36*(6), 394–395. https://doi.org/10.5480/14-1532

Tosterud, R., Hedelin, B., & Hall-Lord, M. L. (2013). Nursing students' perceptions of high-and low-fidelity simulation used as learning methods. *Nurse Education in Practice, 13*(4), 262–270.

Träg, K. H., & Mulders, M. (2025). What the Amazon Can't Deliver: Lessons Learned from Virtual Reality-Based Sustainability Education. In J. M. Krüger, D. Pedrosa, D. Beck, M.-L. Bourguet, A. Dengel, R. Ghannam, A. Miller, A. Peña-Rios, & J. Richter (Hrsg.), *Immersive Learning Research Network* (Bd. 2272, S. 68–80). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-80472-4_5

van As, J., & Cooke, R. (2024). Immersive virtual community engagement: Unleashing the potential of AI avatars and virtual reality in education to enhance learning excellence. In *Integrating cutting-edge technology into the classroom* (S. 162–183). IGI Global. https://www.igi-global.com/chapter/immersive-virtual-community-engagement/344306

Verkuyl, M., Atack, L., McCulloch, T., Liu, L., Betts, L., Lapum, J. L., Hughes, M., Mastrilli, P., & Romaniuk, D. (2018). Comparison of debriefing methods after a virtual simulation: An experiment. *Clinical Simulation in Nursing, 19*, 1–7. https://doi.org/10.1016/j.ecns.2018.03.002

Wang, M., & Akhter, S. (2025). Tracing interpersonal emotion regulation, behavioural emotion regulation strategies. hopelessness and vocabulary retention within Bing vs. ChatGPT environments. *British Educational Research Journal*. https://doi.org/10.1002/berj.4137

Winkler, R., & Soellner, M. (2018). Unleashing the potential of chatbots in education: A state-of-the-art analysis. *Academy of Management Proceedings, 2018*(1), 15903. https://doi.org/10.5465/AMBPP.2018.15903abstract

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – Where are the educators? *International Journal of Educational Technology in Higher Education, 16*(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zinns, L. E., Mullan, P. C., O'Connell, K. J., Ryan, L. M., & Wratney, A. T. (2020). An evaluation of a new debriefing framework: REFLECT. *Pediatric Emergency Care, 36*(3), 147–152. https://doi.org/10.1097/PEC.0000000000001111