

The effectiveness of debriefing in simulation-based learning for health professionals: A systematic review

Prof. Tracy Levett-Jones, RN, PhD, MEd & Work, BN, DipAppSc (Nursing)¹

Samuel Lapkin (PhD Candidate), RN, BN (Hons)¹

1. University of Newcastle Evidence Based Health Care Group: a JBI Evidence Synthesis Group

Corresponding Author: Tracy.Levett-jones@newcastle.edu.au

Executive summary

Background

Simulation is defined as a technique used to replace or amplify real experiences with guided experiences that evoke or replace substantial aspects of the real world in a fully interactive manner. The use of simulation for educational purposes began decades ago with the use of low-fidelity simulations and has evolved at an unprecedented pace. Debriefing is considered by many to be an integral and critical part of the simulation process. However, different debriefing approaches have developed with little objective evidence of their effectiveness.

Objectives

The aim of this review was to identify, appraise and synthesise the best available evidence for the effectiveness of debriefing as it relates to simulation-based learning for health professionals.

Inclusion criteria

Types of participants

This review included any health professional participants regardless of gender, age or profession.

Types of intervention(s)/phenomena of interest

Studies that evaluated the use of debriefing for the purpose of simulation-based learning were included.

Types of studies

The review included randomised controlled trials.

Types of outcomes

Outcome measures included objectively measured outcomes such as situational awareness, communication skills, teamwork, knowledge acquisition, and performance of psychomotor skills as assessed by validated instruments such the Anaesthesia Non-Technical Skills Scale, the Emergency Response Performance Tool, and the Clinical Simulation Tool.

Search strategy

A three-stage comprehensive search strategy was utilised to search across ten electronic databases. English language studies published between January 2000 and September 2011 were considered for inclusion.

Methodological quality

Two independent reviewers assessed the methodological quality of each study selected for retrieval using standardised Joanna Briggs Institute critical appraisal tools.

Data collection

Data were extracted from studies using the standardised data extraction tool from the Joanna Briggs Institute

Data synthesis

Meta-analysis was not possible because of the different outcomes, control groups and interventions in the selected studies. Findings are therefore presented in narrative form.

Results

Ten randomised controlled trials involving various debriefing methods were eligible for inclusion in this review. The methods of debriefing included: post simulation debriefing, in-simulation debriefing, instructor facilitated debriefing and video-assisted instructor debriefing. In the included studies there was a statistically significant improvement pre-test to post-test in the performance of technical and nontechnical skills such as: vital signs assessment; psychomotor skills; cardiopulmonary resuscitation; task management; team working; and situational awareness, regardless of the type of debriefing conducted. Additionally, four of the studies reported improvement in these outcomes without the use of video playback during debriefing. In one study the effect of the educational intervention was evident 6–9 months after the initial simulation experience.

Conclusions

The results of this systematic review support the widely held assumption that debriefing is an important component of simulation. It is recommended therefore that debriefing remain an integral component of all simulation-based learning experiences. However, the fact that there were no clinical or practical differences in outcomes when instructor facilitated debriefing was enhanced by video playback is an important finding since this approach is currently considered to be the 'gold standard' for debriefing.

Keywords

debriefing, simulation, systematic review.

Background

Simulation is defined as a technique used to “replace or amplify real experiences with guided experiences that evoke or replace substantial aspects of the real world in a fully interactive manner”.^{1(p126)} The use of simulation for educational purposes began decades ago with the use of low-fidelity simulations² and has evolved at an unprecedented pace. Debriefing is considered by many to be an integral and critical part of the simulation process.³⁻⁵ However, different debriefing approaches have developed with little objective evidence of their effectiveness.

Some suggest that a structured debriefing should occur immediately after simulation⁶⁻⁸; other researchers advocate that debriefing should include reflection on and for practice.^{9,10} Opportunities for formative feedback and self-evaluation are claimed to be essential components of debriefing¹¹; and the use of video recordings of the simulation during are said to enhance debriefing sessions by stimulating learning and discussion based on an accurate account of events.¹² Depending on the simulation objectives, opportunities for discussion of students' non-technical skills such as clinical reasoning, communication, leadership and teamwork skills are also considered important to debriefing.⁹

While simulation-based learning and debriefing have been adopted and used extensively in health care the use of the term 'debriefing' originated in the army.¹³ Colonel Marshall, a United States (US) Army historian from World War II and the Korean and Vietnam wars is attributed with first developing debriefing methods.¹⁴ He developed 'group historical debriefing' as a method of conducting interviews with soldiers. These interviews were conducted on the battlefield soon after the combat had ceased and involved all ranks; with the emphasis being on learning from the experience.¹⁵ Other debriefing techniques currently used in the military such as after-action reviews are based on Marshall's original approaches. After-action reviews are debriefings conducted as part of military training exercises and involve immediate feedback on proficiency.¹⁶ These examples of military debriefings are educational and do not usually involve counselling or therapy.¹⁷ The debriefings are aimed at improving combat performance by reflective learning and developing new military tactics as a result of the experience.¹⁸

Debriefing has also been adopted by the airline industry in response to aviation incidents. Analysis of 35,000 National Aeronautics and Space Administration (NASA) Aviation Safety Reporting System reports between 1976-1983 indicated that nontechnical and communications skills rather than technical flying abilities or aircraft mechanical malfunctions contributed to most aviation incidents.¹⁹ For example, a report by the National Transportation Safety Board concluded that the captain's failure to accept input from junior crew members and the flight engineer's lack of assertiveness contributed to a United Airlines crash in 1978.²⁰ In response to similar findings from other accidents the aviation industry developed training programs called Crew Resource Management (CRM) in the 1970s.²¹ These programs typically focus on a range of knowledge, skills and attitudes including communications, situational awareness, problem solving, decision making, and teamwork.²² In addition, components of the CRM training include simulated flights scenarios. At the end of the simulated flights scenario the flight instructor facilitates a debriefing. The crew critically analyses performance during the simulated flight scenarios to reinforce newly acquired skills.²³

Debriefing has also been used in experimental psychology in research involving deception (the purposeful provision of ambiguous details about the research and procedures when it is thought that truthful disclosure to participants may influence the phenomena under investigation).²⁴ Although commonly used in psychological and neuroscience research the use of deception remains ethically

controversial²⁵ and for this reason debriefing is used to reverse any adverse effects on participants from the experience. During the debriefing participants who have been 'deceived' as a part of the study are informed of the true nature of the experiment. Another common example of the use of debriefing in psychology is the critical incident debriefing developed as a structured therapeutic approach to mitigate acute post-crisis psychological symptoms.²⁶

While the debriefing approaches outlined above vary in terms of process and terminology, they each include structured and purposive discussions about and reflection on prior experiences. However, in health care this discussion is aimed at facilitating learning to enhance future performance and ultimately improve patient outcomes. This is achieved, in part, by providing an opportunity to clarify the learner's knowledge and rationale for actions during the simulation experience.¹⁸ Debriefing is considered critical to experiential learning as it encourages the learner to reflect on their performance and construct meaning from that experience with a view to clinical improvement.²⁷ In a systematic review of high-fidelity simulation literature Issenberg, et al.⁴ reported that 51 studies listed educational feedback during debriefing as the single most important feature of simulation-based medical education, albeit with little empirical evidence to support these assertions. It is noted however, that the effectiveness of the debriefing process may potentially be compromised by behaviours such as the use of ridicule, focusing on errors and non-constructive criticism.⁸

Although debriefing following the simulation experience (post-simulation debriefing) is common practice with effectiveness a taken-for-granted assumption, there is little empirical evidence to support this approach. Additionally, differences in learning outcomes and effectiveness in relation to other types of debriefing are unclear. Some studies have examined pre-briefing during which the facilitator explains the purpose of the simulation and any learning objectives before the simulation experience²⁸ whilst others have investigated the use of debriefing during the simulation experience (in-simulation debriefing).²⁹ Another common approach is the use of reflective journals as an alternative or supplementary method to oral debriefing.³⁰

There are conflicting views regarding the ideal length of debriefing with some proposing it should typically be three times longer than the length of the scenario³; and others limiting it to 10 minutes after a 45 minutes simulation.⁷ There is also uncertainty about the ideal number of participants in debriefing and who should be involved³¹, with one study claiming that four participants per debrief is appropriate.³²

The issues highlighted here, along with the limited number of empirical studies, illustrate the gaps that currently exist in relation to the effectiveness of debriefing in simulation-based learning. A search of Cochrane Database of Systematic Reviews and the Joanna Briggs Institute data base did not identify any systematic reviews focusing on simulation debriefing. This gap in evidence is an important finding given the assumption that the purpose of debriefing is to facilitate learning. This presents an opportunity for systematically searching, synthesising and summarising the best available evidence on the effectiveness of debriefing in simulation-based learning.

Definitions

For the purpose of this review, the following definitions of terms were used:

Anaesthesia resident is a medical graduate undertaking a post-graduate accredited residency program (usually of three to five years duration) in management of patients undergoing surgical procedures. This period is after the supervised medical practice undertaken by recent medical school graduates known as internship.

Intern refers to medical professional who has recently graduated from a medical school and is completing their first year in medical practice in a hospital under supervision and usually rotating through different speciality areas.

Non-technical skills refers to a set of skills not related to knowledge or technical procedures, but instead encompassing cognitive and interpersonal skills such as clinical reasoning, leadership, communication, resource management and situational awareness. Fletcher, McGeorge, Flin, Glavin, & Maran,³³

Simulation debriefing is a planned feedback session during or after the hands on component of simulation and is usually facilitated by an expert instructor, who provides learners with the time to assess their decisions, actions, communication, and ability to deal with the unexpected.^{8, 34} In this systematic review the term 'debriefing' will be used to refer to simulation debriefing.

Simulation is a technique used "to replace or amplify real experiences with guided experiences that evoke or replicate substantial aspects of the real world in a fully interactive manner".^{1(p126)}

Objectives

The objective of this review was to identify the best available evidence on the effectiveness of debriefing as it relates to simulation-based learning for health professionals.

Inclusion criteria

Types of participants

This review included any health professional participants regardless of gender, age or profession.

Types of intervention(s)/phenomena of interest

Studies that evaluated the use of debriefing for the purpose of simulation-based learning were included.

Types of studies

This review considered any experimental or quasi-experimental studies that addressed the effectiveness of debriefing as it relates to simulation-based learning for health professionals.

Types of outcomes

Objectively measured outcomes related to debriefing conducted as part of simulation-based learning included: self-confidence, knowledge acquisition, performance of psychomotor skills, performance of non-technical skills such as situational awareness; communication and teamwork were considered. Skill performance was assessed through validated instruments such as the Anaesthesia Non-Technical Skills (ANTS) scale, performance checklists, and global rating scales. Knowledge acquisition was

assessed through questionnaires and multiple choice tests. Self-confidence was assessed through Likert scales.

Search strategy

The search sought to find published studies and papers using a variety of databases. A three-step search strategy was utilised in this review. An initial limited search of MEDLINE and Proquest databases was undertaken using the search terms “debriefing” AND “simulation” AND “Health professional”. The second step involved developing search strategies using identified keywords and MeSH headings identified during the initial literature scoping. Search terms were amended depending on the database used. Where appropriate, key words were exploded and truncated. Detailed descriptions of the search strategies for each electronic database are given in Appendix I.

The databases accessed and searched included:

1. AMED
2. CINAHL
3. Cochrane Central Register of Controlled Trials (CENTRAL)
4. Dissertation and Theses
5. EMBASE
6. ERIC
7. Journals@Ovid
8. MEDLINE
9. ProQuest Nursing Journals
10. PsycINFO

The following were hand searched to find any additional articles:

- MedNar
- Directory of open access journals
- Conference Proceedings

Lastly, the reference lists and bibliographies of all identified reports, articles and abstracts were examined for additional studies relevant to the systematic review. All searches were limited to studies undertaken between January 2000 and September 2011. The searches were limited to English language papers. Studies identified during the search were initially assessed for their relevance using information found in their title, abstract and descriptors. The bibliographical software package EndnoteTM was utilised to facilitate the importation of references from electronic databases and linkage of references into the Joanna Briggs Institute (JBI) Comprehensive Review Management System (CReMSTM).

Assessment of methodological quality

Two independent reviewers assessed selected studies for methodological validity prior to inclusion in the review. The reviewers used standardised critical appraisal instruments from the Joanna Briggs Institute Meta-Analysis of Statistics Assessment and Review Instrument (JBI-MASARI) to assess methodological quality. For this process, the reviewers used the JBI Critical Appraisal Tools Critical Appraisal of Evidence of Effectiveness (Appendix II). These guidelines have been developed to minimise bias and establish validity of the findings. Each retrieved study was critically appraised and the methodological quality assessed using the ten questions in the instrument. Due to the nature of the studies under review some of the criteria; for example concealing treatment groups and blinding participants; were not practical. Therefore, it was decided to include those studies with at least five of the ten criteria of methodological quality as determined by the JBI critical appraisal instrument (Appendix III).

Method of the review

Data collection

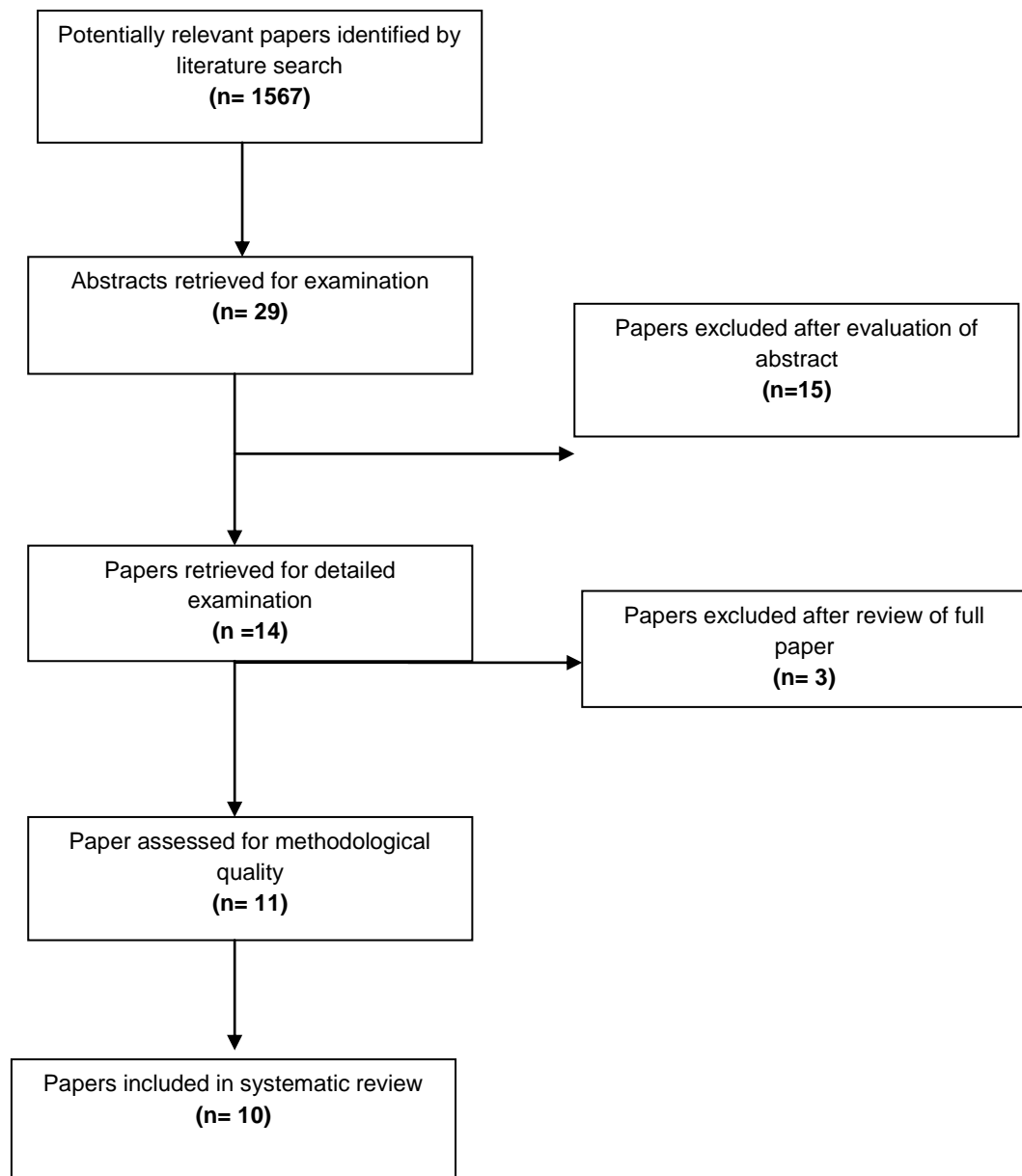
Data were extracted from the papers included in the review using the standardised the JBI Data Extraction of Evidence of Effectiveness tool (Appendix IV). The extracted data included specific details about the interventions, participant demographics, sample size and reasons for withdrawals and dropouts, study methods and any outcomes of significance to the objective of the review.

Data synthesis

It was planned to pool quantitative papers in statistical meta-analysis where appropriate. Odds ratio (for categorical data) and weighted mean differences (for continuous data) and their 95% confidence intervals were to be calculated for each analysis. Where possible, heterogeneity between comparable studies was to be assessed using the standard chi-square analysis. However, as there were no comparable RCTs found for this review and as quantitative data could not be statistically combined for a meta-analysis, extracted data were synthesised into a narrative summary. There was wide heterogeneity, with studies comparing different interventions, with different outcome measures.

Results

The initial search strategy identified 1567 papers, of which 29 were deemed potentially relevant to this review, based on the assessment of title and abstracts. After removal of duplicates and detailed examination, full text papers were retrieved for the remaining 14 studies. After analysis of methodological quality ten papers were identified for data extraction and analysis of results. Details of the study excluded from the review following the methodological assessment with reason for exclusion is provided in Appendix V. The details of the selection process are presented in Figure 1.

Figure 1: Study selection flow diagram

Description of included studies

Appendix VI provides a table of the included studies that addressed the effectiveness of debriefing as it relates to simulation-based learning for health professionals. Ten randomised experimental trials (RCTs)^{5, 12, 29, 35-41} were included in the review. Five of the studies were conducted in the United States of America^{5, 12, 29, 36, 41} a further four in Canada^{35, 37-39}, and one in the United Kingdom.⁴⁰

Anaesthetists and anaesthesia residents accounted for the majority of participants in the included studies^{35, 37-40} followed by nursing students^{5, 12, 41}, medical students²⁹ and qualified nurses.³⁶ The number of participants ranged from 30³⁹ to 162.⁵ All of the studies used convenience sampling of health professionals and health professional students.

In all of the included studies the control and experimental groups received different debriefing methods as part of simulation experience. In four studies outcomes were compared between participants exposed to either instructor facilitated debriefing or video-assisted instructor facilitated debriefing^{12, 35, 38, 41}. Other debriefing methods were: self-debriefing³⁵, multimedia debriefing³⁹; and in-simulation instructor facilitated debriefing²⁹. In two studies^{37, 38} participants in the one of the groups did not receive any debriefing.

In six studies³⁵⁻⁴⁰ participants completed the simulation sessions individually with researchers or actors undertaking scripted roles. In one study¹² participants completed the simulation scenarios in groups of five or six students each taking randomly assigned roles of team leader, airway manager, crash cart manager, recorder and medication nurse. In five studies^{12, 29, 35, 37, 39} the authors reported that the participants undertook pre-simulation orientation to become familiar with the simulation environment and equipment.

Of the ten studies only half^{5, 35, 36, 38, 41} provided specific details regarding the length of the debriefing sessions. In these studies the length ranged from 2 minutes³⁶ to 30 minutes⁴¹ representing between two and a half and twice the length of the hands on component of the simulation experience respectively. The debriefing sessions were conducted with only the facilitator and one participant present^{35, 37, 39} or as group debriefing sessions with the largest groups having between five to six participants.^{5, 12}

Methodological quality of included studies

All the studies stated that participants were randomly assigned but only seven^{5, 29, 36, 37, 39-41} gave details about the methods of randomisation these included; computer random number generation^{29, 37, 39}, sealed envelope technique^{36, 39, 40}, selecting names from a hat⁴¹, and a combination of coin toss and random number table.⁵

Due to the nature of the intervention and the studies, blinding of the participants or assessors was not practical in most of the studies. Only one study reported that participants were blinded to the topic of the simulation⁵, while outcome assessors and data analysts were blinded to the allocation in five of the ten studies^{12, 35, 37-39}. In five studies^{12, 35, 37, 38, 41} outcome assessors were trained to enhance the quality of outcome measurements.

See Appendix III for critical appraisal results for all included studies.

Baseline comparability of groups

Six studies^{5, 35-37, 39, 41} gave descriptions of baseline comparability of varying details related to participants' demographics, including age, sex and prior experience. In five^{5, 35, 37, 39, 41} of these studies, both the experimental and control groups were relatively homogeneous at baseline. However, in one study³⁶ more members of the experiment group reported having had recent CPR training [$p = 0.018$], whilst more members of the control group had experience working in an intensive care unit setting [$p = 0.0038$].

Outcome assessment

The primary outcomes were technical and non-technical skills such as: situational awareness; communication skills; teamwork; knowledge acquisition; and psychomotor skills. Three of the included studies used the Anaesthesia Non-Technical Skills (ANTS) scale.^{35, 38, 39} The ANTS scale is a validated behavioural marker used to assess nontechnical performance in an anaesthesia context. This scoring system is hierarchical and consists of four main skill categories: situational awareness, team working, decision making, and task management. Each category is further subdivided into a number of elements, and for each element a number of behavioural descriptors for good and poor performance are described. Each category is scored out of 4, with 4 the highest score and 1 the lowest possible score. The categorical scores are summed to provide a total ANTS score of between 4 to 16.⁴²

Two other validated instruments were used, namely the Emergency Response Performance Tool (ERPT)⁴¹, and Clinical Simulation Evaluation Tool (CSET).¹² Morgan, et al³⁷ used a combination of a Dichotomously Scored Checklist (DSC) and a Global Rating Scale of Performance (GRS),⁵ used the Clinical Knowledge Questionnaire (CKQ) and²⁹ used seven-point Likert-scale questionnaire to evaluate outcomes. The remaining two studies^{36, 40} used automatic recordings by the simulation software to evaluate outcomes such as chest compression rate and depth during performance of cardiopulmonary resuscitation. A detailed summary of the characteristics of included studies including all results is given in Appendix VI.

Results of the review

Given the range of debriefing methods used and wide variation in the assessment of the outcomes of interest, it was agreed by the review team that for the purposes of the review and ease of presentation of findings, outcomes should be presented by method of debriefing used.

Video-facilitated instructor debriefing

Six studies^{12, 35, 38-41} compared video- facilitated instructor debriefing with other types of debriefing methods. One study¹² reported a statistically significant increase in outcomes for the groups exposed to video- facilitated instructor debriefing. Chronister and Brown⁴¹ reported mixed findings with differences in some outcomes but not others between the control and experimental groups after the simulation experience. In four of the studies^{35, 38-40} there were no statistically significant difference in improvement between instructor facilitated debriefing and video-facilitated instructor debriefing groups. However, the results from these studies have important clinical and practical implications as they were able to demonstrate improvement in learning outcomes when various debriefing strategies are incorporated as part of simulation experiences.

Study by Grant, et al.,¹²

Grant, et al.,¹² conducted a pilot study with nursing and nurse anaesthetist students to evaluate the effect of instructor facilitated debriefing versus video-facilitated instructor debriefing on clinical performance indicators. Nursing and anaesthetic students in the experimental group (n = 20) participated in simulation scenarios using video-facilitated instructor debriefing. Those in the control group (n= 20) only participated in instructor facilitated debriefing.

Before undertaking the simulation scenarios students participated in a brief pre-simulation orientation to the manikin and the simulation process. After completing the orientation students were then randomly assigned roles and completed two 60-minute practice simulation sessions in groups of five or six nursing students and one nurse anaesthetist student. Participants were instructed to care for the simulated patient and to call the nurse anaesthetist student if intubation was required during the simulation.

The practice simulation scenarios were based on the course objectives and involved providing care for complex patients with pulmonary and cardiac problems. The pulmonary practice simulation sessions related to caring for patients with acute respiratory distress who experienced a pulmonary embolus or pneumothorax. Cardiac practice simulation sessions related to caring for patients with chest pain who experienced sinus bradycardia, ventricular fibrillation, torsades de pointes, or asystole. Participants in both groups received cueing from facilitators using a standardised simulation protocol. Students rotated through the roles of team leader, airway manager, crash cart manager, recorder, and medication nurse.

Immediately after the practice simulation sessions, students in the experimental group underwent video-facilitated instructor debriefing while those in the control group had instructor facilitated debriefing. Participants were debriefed in groups of five to six students. In the discussion of roles and behaviours the control group relied on the recollection of the simulation by students and instructors. The debriefing questions in for both groups of students were based on the simulation protocol used to provide cues to the participants during the simulations.

At the end of the semester both groups participated in a third 60-minute simulation session consisting of two scenarios. One of the scenarios was based on managing a myocardial infarction and the other a patient with a stab wound to the chest. Nursing students participants in each group were randomly assigned to one of the roles used in the practice simulations (i.e., team leader, airway manager, crash cart manager, recorder, medication nurse), and nurse anaesthetists students were assigned to their role. Clinical performance during the scenarios was scored on an adapted version of the Clinical Simulation Evaluation Tool (CSET). The CSET was developed by Radhakrishnan, Roche and Cunningham⁴³ to measure behaviours related to patient safety such as assessment, intervention, delegation and communication skills during congestive heart and pelvic fracture simulation scenarios. The researchers adapted this tool to record the occurrence of expected behaviours related to the scenarios. Examples of behaviours were "checks patient identification," "assesses ABCs," "initiates interventions," and "communicates using situation-background-assessment-recommendation format." The total scores ranged from 0 to 31 for one scenario and from 0 to 34 for the other scenario, with higher scores indicating demonstration of appropriate behaviours. No additional points were allocated if team members initiated target behaviour more than once for each scenario. Each scenario was scored separately, and scores were assigned to individual students, not team performance.

The students in the experimental group were more likely to exhibit desired simulation behaviours as demonstrated by higher mean scores in the following: patient identification [$p < 0.01$], team communication [$p = 0.013$], and assessment of vital signs [$p = 0.047$]. These results were statistically

significant. In addition, the mean scores for the experimental were higher than the control group scores in 9 of the 14 categories of desired behaviours. However, there were no statistically significant difference between the two groups with total performance scores of 9.09 and 8.44 for the experimental group and the control group respectively.

Study by Chronister and Brown⁴¹

Similar results were reported by Chronister and Brown⁴¹ who conducted a comparative crossover pilot study to evaluate the effects of two debriefing styles with 37 undergraduate nursing students. Prior to participation in the study, the students' completed their usual course activities consisting of five 90-minute electrocardiogram classes with both didactic and clinical simulations, 20 hours of classroom lectures, and a critical care clinical practicum of approximately 80 hours. All participants then completed the pre-test knowledge test and were randomly divided to receive either instructor facilitated debriefing (control group) or video-assisted instructor facilitated debriefing (experimental group).

Participants engaged in one of two simulation scenarios involving ventricular fibrillation cardiopulmonary arrest. Each scenario lasted 15 minutes and progressed to successful conversion to a sinus rhythm if participants performed the correct assessment and clinical skills, including oxygen administration, medication administration, defibrillation, and cardiopulmonary resuscitation (CPR).

The participants undertook a 30 minute instructor facilitated debriefing (control group) or video-assisted instructor facilitated debriefing (experimental group) immediately after the first simulation experience. The content of the debriefing for both groups included discussion about: (a) students' feelings about the simulation; (b) review of their initial assessment steps; (c) psychomotor skills demonstrated; (d) communication between team members; and (e) open discussion on points of interest. Participants then completed a second simulation scenario a week after the first.

The three outcomes of interest were: skill response time; quality of skills and knowledge retention. Skill response time was evaluated by time measured in seconds from the onset of cardiopulmonary arrest until the implementation of four skills: verbal recognition of ventricular fibrillation rhythm, assessment of breathing with initiation of rescue breathing, assessment of pulses with initiation of CPR, and delivery of first defibrillation shock. Knowledge retention was assessed by a 10-item multiple-choice exam conducted at baseline and 1 week after the first simulation experience. The Emergency Response Performance Tool (ERPT) was used to evaluate students' assessment and psychomotor skill performance during the scenarios. The ERPT consists of 19-item checklist of skills that are scored by assignment of a whole number ranging from 0 to 2 (with the higher scores indicating better performance) with a total sum ranging from 0 to 32. The developers of the original tool reported test-retest reliability (r) of 0.87 and internal consistency (Cronbach's α) of 0.92.⁴⁴

Analysis of the change in response times showed that both groups had faster times on all skills when responding during the second simulation ($p = 0.025$). However, the response times for the experimental group were faster in three skills: pulse assessment with initiation of CPR [$p = 0.094$]; initial defibrillation shock ($p = 0.042$); and total time to resuscitation [$p = 0.028$]. Emergency Response Performance Tool scores in the second simulation scenario improved significantly for all participants after participation in the simulation and debriefing [$p = 0.025$]. Although the improvement in scores from baseline was higher for the experimental group than the control group, the difference was not statistically significant [$p = 0.71$]. Improvements in mean knowledge retention scores from pre to post-test were higher in the control group (5.14 to 5.57) than in the experimental group (6.3 to 4.95); this

difference was statistically significant [$p = 0.008$]. Therefore, video-facilitated instructor debriefing was more effective for nursing skills and response times, whilst knowledge retention was more positively affected by instructor facilitated debriefing.

*Study by Boet et al*³⁵

The first of the four studies that showed no statistically significant difference in outcomes was a RCT with 50 resident anaesthetists.³⁵ All participants undertook a 30-minute pre-simulation orientation session that covered principles of patient simulation, anaesthesia crisis resource management, discussion of key elements of the Anaesthetists' Non-Technical Skills (ANTS), and familiarisation with the high-fidelity human patient simulator. After the orientation all participants undertook one of the two scenarios as the pre-test followed by debriefing. The order of presentation of the two scenarios was randomised for each participant and equally distributed among the study groups to control for sequencing effect.

Although no detailed descriptions of the simulation scenarios were provided, they lasted 5 minutes and involved an intraoperative cardiac arrest. The authors reported that one of the scenarios was based on ventricular tachycardia secondary to myocardial infarction and the other consisted of a ventricular fibrillation due to hyperkalaemia. During the scenarios two simulation centre staff acted in scripted roles of a surgeon and a circulating operating room nurse. They assisted with performance of tasks but did not offer crisis management advice or differential diagnoses. Participant's performance was video recorded and the simulated patient's vital sign were also obtained. Both simulation scenarios were determined to be of equal difficulty.

After the pre-test simulation scenarios participants were randomised to either self-debriefing (control, $n=25$) or video- assisted instructor facilitated debriefing (experimental, $n=25$). The debriefing sessions were limited to 20 minutes for both groups. The participants in the control group used video recordings of the simulations and the ANTS framework to reflect on their performance. They reviewed the video of the simulation by themselves and were free to fast-forward or rewind segments of the video according to their own preferences. Those in the experimental group had video-assisted instructor debriefing. An expert instructor facilitated a review of the simulation using video playback to encourage participants to comment and reflect on their performance and offered critical feedback based on the ANTS scale as a framework. Immediately following the respective debriefings, all subjects undertook a second simulation.

Performance in the second simulation was evaluated using the ANTS by assessors who were by blinded to allocation. Results indicated that the performance of nontechnical skills such as situational awareness, team working, decision making, and task management improved from simulation one to two [$p < 0.01$] for both groups. However, there was no statistically significant difference in improvement between the video- facilitated instructor debriefing group and the video assisted self-reflection groups [$p = 0.58$]. Although there were no statistically significant differences in outcomes between the two groups, the results highlight that effective teaching of non-technical skills can be achieved through simulation experiences even when instructor facilitated debriefing is not available.

Study by Byrne et al.,⁴⁰

Byrne et al.,⁴⁰ undertook a quasi-experimental study involving 34 anaesthetists from four hospitals to compare performance during simulation scenarios between participants exposed to two types of debriefing. Individuals in the control group (n= 16) completed five scenarios with instructor facilitated debriefing between each simulation. Those in the experimental group (n=16) completed the same scenarios but received video-facilitated instructor debriefing between each of the simulations.

The authors did not give detailed description of the simulation scenarios and debriefing methods but noted that participants completed the five scenarios on the same days in the same order, each followed by a debrief. Participants were required to manage a simulated routine general anaesthetic scenario complicated by either hypotension, ventricular tachycardia, bradycardia, anaphylaxis or oxygen supply failure.

Although adherence to protocols was not required each critical incident had specific actions required for a successful outcome. Participants were expected to undertake actions such as administering intravenous fluids, decreasing the inspired concentration of volatile agent, and increasing the inspired concentration of oxygen. They were also required to accurately document the following five vital signs: pulse rate, systolic and diastolic blood pressure, oxygen saturation, and end-tidal carbon dioxide concentration at 2.5-minute intervals during each simulation. If participants failed to record any of the vital signs a value of 100% error was entered for any missing value.

Data from the automatic recordings by the simulation software and from the anaesthetic charts were entered manually into a spread sheet for analysis and for comparison with the manual recordings completed by the participants. The main outcome measures were time taken to solve the problem and accuracy of participants recording of the vital signs measurements. Time to solve the problem was considered to be the period between the onset of the clinical problem and the last appropriate significant intervention by the participant. Chart error was considered to be any incorrectly documented or missed recordings of the vital signs. The median chart error ratio for each participant was obtained by dividing the median chart errors between the first and fifth scenarios.

The researchers reported that participants in the experimental group had a shorter median time ratio to solve the scenarios of 0.68 minutes compared to a median ratio of 1.18 minutes for those in the control group. The decrease in median chart error ratio between the first and last simulation scenarios was 0.83 and 0.63 between the participants in the experimental and control group respectively. Although the differences between the two groups were not statistically significant [$p > 0.05$] the results indicate that psychomotor skills, such as recoding of vital signs, can be improved when debriefing is included as part of the simulation experience.

Study by Savoldelli et al.³⁸

Savoldelli et al.³⁸ conducted a RCT with 42 anaesthesia residents to compare the educational effectiveness of instructor debriefing, videotape-assisted instructor debriefing, and no debriefing following involvement in a simulation scenarios. Prior to the simulation all participants took part in a group pre-simulation orientation session covering crisis management, patient simulation, and familiarisation with the simulation equipment. Participants then took part in two 8 minute scenarios involving an intraoperative cardiac arrest. The use of two different scenarios was done to minimise the influence of the case or content specificity on participant's performance. One of the scenarios was

based on cardiac arrest due to hyperkalaemia and the other secondary to a massive fat embolism. The simulations and graphical display of patients' vital signs were videotaped. Each participant was supported by clinicians who undertook scripted roles of a surgeon, nurse and a second anaesthesiologist.

After completing the first simulation participants were randomly assigned to receive video-assisted instructor facilitated debriefing (group A, n= 12), instructor facilitated debriefing (group B, n= 15), or no debriefing (group C, n= 15). Participants in groups A and B undertook the second simulation after debriefing. The debriefing focused on nontechnical skills performance guided by crisis resource management principles. Individuals in group A had debriefing with the facilitator using selected video segments to encourage participants to reflect on cognitive and behavioural aspects of their performance. Participants in group B took part in instructor facilitated debriefing where they undertook a self-assessment and reflected on their performance of nontechnical skills. The facilitators also provided constructive comments regarding performance of nontechnical skills. The mean debriefing time was 20 minutes for both groups but each session ended when the instructor's comments and participants questions were exhausted with no set time limit. Individuals allocated to group C receiving no debrief or feedback regarding their performance in the simulation.

Video recording of participants' nontechnical skills performance during the scenarios was rated by two blinded independent assessors using the Anaesthesia Non-Technical Skills (ANTS) scoring system. Analysis of the change in total ANTS scores simulation one to two revealed statistically significant differences between the three groups with improvement being greater among participants who received debriefing, either instructor facilitated (15%) or video-assisted (11%) [$p < 0.005$]. However, the addition of a video review did not offer any advantage as there was no statistically significant difference in the improvement of total ANTS scores between the video-facilitated instructor (group A) and instructor facilitated (group B) debriefing groups. In fact, improvement tended to be lower in the video-assisted instructor facilitated debriefing group than in the instructor facilitated debriefing group.

Study by Welke et al.,³⁹

Welke et al.,³⁹ conducted a study of 30 anaesthesia residents to compare multimedia debriefing and video-facilitated instructor debriefing. All participants were required to manage three advanced cardiac life support (ACLS) resuscitation scenarios of 10 minutes duration each. One scenario consisted of bradycardia and hypotension secondary to pacemaker failure. A second scenario consisted of pulseless electrical activity secondary to hypovolemia; and a third scenario consisted of ventricular fibrillation secondary to cardiac ischemia. Each participant played the role of the primary anaesthesiologist during the simulation experience. They were supported by researchers who had scripted roles of surgeon, nurse and anaesthesiologist but did not offer crisis management advice or differential diagnoses.

Before the first simulation all participants completed a demographic questionnaire and questions related to previous simulation experience. They were also given an orientation to the simulated operating room and asked to review ACLS algorithms before their simulation sessions. All participants managed simulation 1 as a pre-test followed by debriefing. Individuals in the experimental group (n=15) received video-facilitated instructor debriefing after the simulation. Appropriate sections of video recordings were used to provide constructive critique of their crisis management and nontechnical skills during simulation. There was no strict time limit on the duration of the debriefing and it was concluded when all the participant's questions and instructor's observations were addressed.

Participants in the multimedia group (control, $n = 15$) received instruction through the use of standardised multimedia resources after completion of the first simulation. The resources were developed by the researchers and covered the four main principles of crisis management relevant to the simulation objectives: 1) planning and preparing, 2) calling for help early and announcing a crisis clearly, 3) distributing tasks to appropriate individuals, and 4) re-evaluation. The multimedia resources consisted of a combination of text, audio voice-over, and a set of digital videos demonstrating poor performance and ideal management of cardiac arrest. Each demonstration of poor performance was followed by instruction, highlighting how improvements could be made by applying the four principles of crisis management. The multimedia presentation was followed by a demonstration of the same scenario with ideal management through the application of the crisis management principles.

Immediately after the debriefing sessions participants completed a second simulation followed by a third 5 weeks later to test retention of learning. The simulation experience was video recorded and the heart rate, arterial blood pressure, oxygen saturation, electrocardiogram, and end-tidal carbon dioxide were monitored in real-time. Performances of nontechnical skills during the simulation experience were independently rated by two blinded expert assessors using the Anaesthesia Non-Technical Skills (ANTS) scoring system. Analysis of the results indicated that participants in both groups had similar pre-test total ANTS scores [$p = 0.14$]. The improvements in total ANTS score from simulation 1 to 2 [$p = 0.97$], simulation 1 to simulation 3 [$p = 0.94$], and or simulation 2 to 3 [$p = 0.84$] were similar for both group. These results demonstrate only a small effect size of 0.002 standard deviations for comparison of multimedia debriefing and video-facilitated instructor debriefing and that the developed anaesthesia nontechnical skills were retained for a period of at least 5 weeks after the simulation experience. These results suggest that computer-based multimedia may be as effective as video-facilitated instructor debriefing.

Comparison of other debriefing methods

Four studies^{5, 29, 36, 37} illustrate the variability of debriefing methods used in simulation experiences. These studies involved techniques such as a combination of instructor facilitated debriefing and real-time audio-visual feedback from a defibrillator³⁶; and in-simulation instructor debriefing.²⁹ One study compared the effectiveness of instructor facilitated debriefing to no debriefing.³⁷ Despite differences in the debriefing methods among the four studies, there were significant improvements in outcomes post interventions in all of the studies.

Study by Dine, et al.,³⁶

Dine, et al.,³⁶ undertook a RCT to compare resuscitation skill performance of 80 nurses exposed to two different debriefing methods after completing simulated cardiac arrest scenarios. Participants were randomly assigned into either a control group ($n=34$) or an experimental group ($n= 31$). The control group received instructor facilitated debriefing and the experimental group received a combination of instructor facilitated debriefing and real-time audio-visual feedback from a defibrillator.

All participants were required to complete three simulated cardiac arrest scenarios. During the simulation each participant performed single-person cardiopulmonary resuscitation (CPR) for 2 minutes. The depth and rate of the chest compression measured during the first simulation was considered to be baseline CPR performance for each participant. The participants were not provided with feedback or debriefing before undertaking the second simulated resuscitation.

During the second simulation the experimental group received real-time audio-visual feedback from a defibrillator, whereas the control group performed cardiopulmonary resuscitation without any feedback. After completion of the second scenario participants from both groups received a 5-minute instructor facilitated debriefing. The facilitators used the transcript of each participants CPR performance to provide feedback on how to improve their performance to comply with resuscitation guidelines. All participants then completed the third simulation scenarios.

The main outcome of interest was the CPR performance as indicated by the chest compression rate and depth. An adequate compression rate was defined as 90-110 compressions/minute and an adequate depth as 38-51 mm. The readings were obtained from the cardiopulmonary resuscitation-sensing defibrillator which had the capacity to record the measures of chest compression rate and depth to a memory card. This data was then analysed and used for debriefing the participants and for making comparisons between the two groups. However, data collection was incomplete for 15 participants and these were excluded from data analysis.

Analysis of the results found no significant differences the CPR performance during simulations 1 and 2 for the control group. Providing audio-visual feedback in the second simulation increased the number of participants in the experimental group providing adequate compression depth (58% compared to the control group 38%) but this was not statistically significant [$p=0.110$]. However, after the debriefing the percentage of participants in the control group that performed adequate depth during simulation 3 increased from 19% to 47% [$p = 0.018$]. Comparisons between the two groups during the final simulation demonstrated that the participants in the experimental group improved compression rate compliance from 45% to 84% [$p = 0.001$] and both adequate rate and depth of compressions from 29% to 65% [$p = 0.005$]. Comparison of mean compression rates indicated no significant differences between the two groups over the three simulations. These results imply that debriefing or real-time audio-visual feedback alone when included as part of the simulation experience improved cardiopulmonary resuscitation quality, but the combination led to marked performance improvements.

*Study by Shinnick, et al.*⁵

Shinnick, et al.⁵ conducted a two group repeated measures experimental study with 162 undergraduate nursing students from three nursing schools to determine the effect of simulation debriefing on clinical heart failure (HF) knowledge. All participants had a lecture on HF and a clinical placement experience as part of the usual advanced medical surgical course. Prior to undertaking the simulation all participants completed a baseline online knowledge tests (pre-test).

Participants were randomly assigned by blocks to the control group ($n = 72$) or the experimental ($n= 90$) group. The researchers used the Clinical Knowledge Questionnaires (CKQ) to evaluate participant's HF knowledge outcomes. The CKQ consisted of 12 multiple-choice items focused on a patient showing signs of pulmonary decompensation and nursing interventions for common issues associated with pulmonary problems. The maximum score was 100. The test was administered before the simulation (pre-test), immediately after the simulation (post-test 1) and after the debriefing sessions (post-test 2).

Students in the control group completed post-test 1 within an hour of completing the pre-test but before the simulation. The participants did not have access to textbooks during the period between the pre-test and post-test 1 and they remained in a room that was monitored by the research team. After completing the post-test 1 they then participated in the simulation followed by debriefing, post-test 2 and the demographic questionnaire. Participants in the experimental group completed the pre-test CKQ prior to

simulation. They then completed post-test-1 immediately following the simulation experience but before the debriefing. The third questionnaire (post-test 2) and a demographic questionnaire were administered after the debriefing.

Participants completed one of the three simulation scenarios of a clinical case of acute decompensated HF during the hands on component of the simulation experience. The three scenarios were identical with the exception of the patient history and gender. During the simulation students were expected to respond appropriately initiating nursing actions such as elevating the head of the bed for a dyspnoeic patient, applying oxygen as appropriate, choosing the priority medication from medical orders, and monitoring appropriate electrolytes in a patient receiving diuretic. After the simulation sessions the students then participated in 30 minute instructor facilitated debriefing. The facilitator who was blinded to the contents of the knowledge tests used guided reflection to stimulate conversation among the participants. Each debriefing session involved groups of five participants and were facilitated by same research team member and was taped and reviewed for consistency.

There were no statistically significant differences in pre-test knowledge scores between groups [$p < 0.001$]. The experimental group had significantly higher mean post-test scores than the control group [Experimental = 61; Control = 55, $p < 0.001$]. These results reflect the changes in knowledge scores after the experimental group has been exposed to simulation experience and the control, group had not. Combined analysis of the results indicated that mean knowledge scores for both groups decreased from the pre-test to the first post-test (after the simulation [$M = 5.63$, $SD = 3.89$, $p < 0.001$] but significantly improved after the combination of simulation experience and debriefing sessions [$M = 6.75$, $SD = 4.32$, $p < 0.001$]. There were no significant differences between the groups at post-test 2, after both groups were exposed to the simulations and debriefing. These results suggest that clinical heart failure knowledge decreased after the hands-on component of the simulation and increased only after both the hands-on component and debriefing. Since the knowledge gains were achieved only after debriefing, the authors concluded that debriefing should be included in any simulation learning experience.

Study by Van Heukelom, et al.²⁹

Van Heukelom, et al.²⁹ conducted a RCT with 161 third year medical students to compare post-simulation debriefing versus in-simulation debriefing. Participants were randomised into either the in-simulation debriefing (control, $n = 84$) or post simulation instructor facilitated debriefing (experimental, $n = 77$). Prior to the simulation all participants were given an orientation session that was specific to the type of debriefing they would receive. The facilitators also discussed learning objectives and critical actions required during the simulation. .

One of the simulation scenarios included a patient with ST-elevation myocardial infarction that deteriorated into ventricular fibrillation (VF) and the other a patient with third degree atrioventricular block who required cardiac pacing. The simulations were conducted with two teams of three participants from each group. The first team completed the simulation while the second observed the simulation in the same room. Those observing were provided with an outline of the simulation scenario with critical action steps to follow along. The two teams then swapped roles and the observation team completed the second scenario with the first team observing.

Those in the control group were provided with immediate feedback during the simulation experience. When the participants made an error or failed to perform an expected critical action the scenario was

paused and the facilitator informed the participant about the consequence of the error or inaction and directed them toward the correct action. The instructions given to participants in such cases would include identifying of VF, differentiating it from other rhythms. Expected intervention such as defibrillation and vasoactive medication was also explained and the simulation was recommenced. No feedback on the performance was provided after the simulation but any specific student questions were addressed.

Participants in the experimental group were given a comprehensive instructor facilitated debriefing session after the simulation experience. However, if an error or inaction led to significant deviation from appropriate management while the control group was undertaking the simulation the facilitator would provide limited redirection and the simulation would be recommenced. No further teaching or feedback was given during the simulation experience. Following the simulation participants were debriefed using facilitated reflection, clarification of facts, content and principles involved in the simulation. The feedback during the debriefing was based on the expected critical action simulation objectives determined by the researchers prior to the simulation. The maximum time allowed for the simulation experience and debriefing session 20 minutes.

Outcomes were evaluated using seven-point Likert-scale questionnaire assessing students' self-reported confidence and knowledge levels. It was administered two days after the completion of the simulation experience. Results from both groups showed significantly higher mean post-simulation scores compared with pre-simulation scores for both individual items and overall measures related to students self-reported confidence and knowledge [$p \leq 0.001$]. Participants in the experimental group reported that post-simulation debriefing helped them learn more effectively, better understand the correct and incorrect actions, and overall was more effective than the debriefing that occurred during the simulation [$p = 0.001$]. These results suggest that post-simulation debriefing is more effective than in-simulation debriefing.

Study by Morgan et al.,³⁷

Morgan et al.,³⁷ reported mixed results in performance during high-fidelity simulation scenarios after different debriefing methods. A total of 71 anaesthetists were randomly allocated to three groups: instructor facilitated debriefing (group A, $n=29$); home study (group B, $n=21$); and no intervention (group C, $n=21$). Prior to the scenario all participants were orientated to the purpose and conduct of the study and given time to become familiar with the simulated operating theatre, anaesthetic gas machine, and drug cart. Participants were also given an opportunity to ask questions. The scenario commenced when they had no further questions.

Participants were further randomised to undertake one of two scenarios for their first simulation, with the second conducted six to nine months later. The two scenarios were based on two critical events, one of which required the use of an Advanced Cardiac Life Support (ACLS) protocol. During the scenarios scripted actors filled the roles of circulating nurse, scrub nurse, surgeon, respiratory therapist and anaesthetist. Both scenarios had situations that enhanced the likelihood of human error and all sessions were video recorded.

Following the simulation 1, participants in group A received an individual instructor facilitated debriefing with only the participant and instructor present. The debriefing consisted of a combination of a standardised PowerPoint presentation on medical errors and a review of video-recorded sections of the simulation where technical and non-technical errors were most evident. The debriefing was focused on

discussing only two or three learning points selected by the facilitator based on the participant's performance. The instructor then assisted the participants to reflect on their performance with the aim of mitigating the identified errors. Participants in groups B and C did not receive any debriefing. Individuals in group B received a home study program consisting of peer-reviewed articles outlining the causes of human error in healthcare. Those in group C were dismissed after completion of their pre-test scenario without any intervention.

After 6-9 months from the simulation 1 participants undertook a second simulation. On completion of the second simulation individuals in groups B and C were offered the type of debriefing provided to group A previously. Trained expert raters who were blinded to participant's group allocation evaluated anaesthetists' technical and non-technical skills using two performance checklists through by observing through a one-way window. They also had access to video recordings after the session to ensure completion of the checklists. Two scenario-specific performance assessment tools were used to evaluate the participants' management of the scenarios. The dichotomously scored checklist used for simulation 1 consisted of 104-items and 99-items for simulation 2. The inter-rater reliability coefficients for the scales were ($r = 0.71$ GRS, $r = 0.91$ DSC). In addition, a one-item, five-point response global rating scale of performance was used by assessors to provide an overall rating of performance.

Interim analysis indicated no difference between groups B and C and their results were subsequently combined for analysis and considered as the control group. There was a statistically significant difference in scores from simulation 1 (66.8%) to simulation 2 (70.3%) for group A [$p = 0.03$]. All participants showed significant improvement in the global rating scale over time [$p = 0.018$], although the difference between the groups was not statistically significant [$p = 0.57$]. These results suggest a modest improvement in participants performance technical and non-technical skills on the dichotomously scored checklist in the debriefed group and overall improvement in both control and debriefed groups using the global rating scale post-simulation.

Discussion

This systematic review was undertaken to investigate the effectiveness of debriefing as it relates to simulation-based learning for health professionals. A systematic search of the literature resulted in 10 published studies that were eligible for inclusion in the review. The studies included both male and female participants. The majority of the studies were reported according to the guidelines set out in the Consolidated Standards of Reporting Trials statement, which lists the essential criteria that need to be reported so as to enable readers to determine the validity and reliability of the results. We found that the studies were heterogeneous and this emphasises the gap in high-quality research studies in this area. Meta-analysis was limited by the lack of replication studies, thereby reducing the ability to extract definitive conclusions from the studies detailed in this review. The lack of measures of dispersion (e.g. standard deviations) also prevented a meta-analysis from being conducted. As a result, this report is written in a narrative form. The results in this review should be interpreted cautiously, given the heterogeneity in terms of the follow-up period and the potential for the results of small studies like those reported here, to over or underestimate differences.

Use of video facilitation

It has been suggested that video replay enhances debriefing as it offers the precise portrayal of events.

³ However, others have argued that the use of video may distract participants from learning objectives.

⁴¹ Of the papers that examined the use of video playback to facilitate debriefing, authors reported variable findings, making it difficult to establish with certainty the strength of specific findings.

One study ¹² found that the experimental group exposed to video-facilitated instructor debriefing was significantly more likely to demonstrate desirable behaviours concerning patient identification, team communication, and vital signs. A second study ⁴¹ found mixed results with the experimental group having higher improvement in performance of cardiopulmonary resuscitation than the control group. However, higher levels of knowledge retention were reported for the control group.

In four studies ^{35, 38-40} the addition of video playback did not offer any significant differences in improvements in outcomes when compared to instructor facilitated debriefing. In fact in one study ³⁸ improvement tended to be lower in the video-assisted instructor facilitated debriefing group than in the instructor facilitated debriefing group. Since duration of the debriefing session in both the control and experimental groups was the same in this study it is possible that participants who received video-facilitated debriefing received less verbal feedback as some portion of the time was spent watching the appropriate sections of the video. None the less, these are important finding considering the significant extra cost required for implementing video-facilitated debriefing such as video recording equipment and the cost of training academic and technical staff to competently use the technology for debriefing.

The results of this systematic review indicate that other debriefing methods such as video assisted self-debriefing using multimedia resources may be as effective as video-assisted oral debriefing. Approaches such as these that do not require the presence of a facilitator would be more cost effective novel alternative to video facilitated instructor debriefing. Other advantages of self-guided debriefing approaches include offering participants control over the pace of debriefing and the opportunity to review self-perceived weaknesses. Therefore, valuable outcomes can be achieved even when video technology is not available. This is an important finding given that video facilitated debriefing is currently regarded as the “gold standard”.

Structure of debriefing

Although most studies included in this review were conducted to compare different debriefing methods the majority of the studies did not provide sufficient details about the debriefing making it difficult to establish with certainty the strength of specific findings. In the papers that offered details on the debriefing methods used the format of debriefing interventions were varied. Therefore it is not possible to produce strong conclusions regarding a best practice model or debriefing framework. For example there is no evidence to support the optimum time frame for the debriefing phase of the simulation experience and whether debriefing requires less or more time than the hand on simulation component. These issues require further research to develop best practice debriefing framework.

In one study ³⁷ the effect of the debriefing interventions was still evident after 6–9 months whilst in another ³⁹ the developed nontechnical skills were retained for a period of at least 5 weeks after the debriefing sessions. These results suggest that improvements in outcomes last for a significant amount of time although it remains unclear whether the skills learned through a simulation experience are transferable to clinical settings.

Limitations of the review

All of studies considered for the review had methodological shortcomings. Firstly, the studies used convenience sampling, which limits the generalisability of findings. In addition, the use of small sample sizes in some of the studies may have resulted in insufficient power to detect effects of the various interventions on the outcomes. Four of the included did not establish robust baseline or comparative data in order that changes in outcomes could be accurately identified. Interpreting the findings from these studies was somewhat difficult.

Conclusion

Although the studies in this review reported some positive outcomes, the small number of studies, combined with heterogeneity of interventions, means that it is not possible to draw generalisable inferences and to identify the best available evidence on the effectiveness of debriefing as it relates to simulation-based learning for health professionals. However, the findings can inform the design and delivery of simulation-based learning experiences. First, effective teaching of technical and nontechnical skills such as situational awareness, team working, decision making, task management, performance of cardiopulmonary resuscitation, and other psychomotor skills can be achieved when debriefing is included as part of the simulation experience. Second, four of the ten included studies provide evidence that statistically significant increases in these outcomes can be achieved without the inclusion of video playback during debriefing. Taken together results from this systematic review suggest that simulation experiences that include any form of debriefing are an effective pedagogical strategy

Implications for practice

Based on the findings from the small number of studies included in this review a number of recommendations can be made for practice (the specified levels of evidence are the JBI levels of evidence Appendix VII).

- Any form of debriefing regardless of the methods used contribute to effective learning and should be included as an integral component of all simulation learning experiences (Level II)
- Video assisted debriefing does not offer any statistically significant educational advantages over instructor debriefing and therefore other factors must be taken into account when considering this approach (Level II)

Implications for research

The evidence for the optimum elements of simulation such as the duration of debriefing is inconclusive and also requires further investigation. In addition, there is need for studies examining the role of debriefing alone on performance improvement as opposed to the overall effect of simulation-based training including orientation, hands on simulation and debriefing. Therefore, randomised controlled studies explicitly focused on the debriefing component of the simulation experience with rigorous

randomisation procedures and allocation concealment, larger sample sizes, and, validated outcome assessment tools would improve the evidence base of simulation.

Conflict of Interest

The authors declare no conflict of interest.

References

- 1 Gaba D. The future vision of simulation in healthcare. *Simulation in Healthcare*. 2007;2(2):126 - 35.
- 2 Nehring WM, Lashley FR. Nursing simulation: A review of the past 40 years. *Simulation & Gaming*. 2009;40(4):528-52.
- 3 Arafeh J, Hansen S, Nichols A. Debriefing in simulated-based learning: facilitating a reflective discussion. *Journal of Perinatal & Neonatal Nursing*. 2010;24(4):302-9.
- 4 Issenberg S, McGaghie W, Petrusa E, Gordon D, Scalese R. Features and uses of high-fidelity medical simulations that lead to effective learning: a BEME systematic review. *Med Teach*. 2005;27(1):10 - 28.
- 5 Shinnick M, Woo M, Horwich T, Steadman R. Debriefing: The most important component in simulation? *Clinical Simulation in Nursing*. 2011;7(3):e105-e11.
- 6 Flanagan B. Debriefing: Theory and techniques. In: Riley R, ed. *Manual of Simulation in Healthcare*. New York: Oxford University Press 2008:155–70.
- 7 Cantrell M. The importance of debriefing in clinical simulations. *Clinical Simulation in Nursing*. 2008;4(2).
- 8 Decker S. Integrating guided reflection into simulated learning experiences. In: Jeffries P, ed. *Simulation in nursing education: From conceptualization to evaluation*. New York: National League for Nursing 2007:73–85.
- 9 Dreifuerst K. The essentials of debriefing in simulation learning: a concept analysis. *Nursing Education Perspectives*. 2009;30(2):109.
- 10 Cato M, Lasater K, Peeples A. Student nurses' self-assessment of their simulation experiences. *Nursing Education Perspectives*. 2009;30(2):105-8.
- 11 Lasater K. High-fidelity simulation and the development of clinical judgment: Students' experiences. *Journal of Nursing Education*. 2007;46(6):269-76.
- 12 Grant JS, Moss J, Epps C, Watts P. Using video-facilitated feedback to improve student performance following high-fidelity simulation. *Clinical Simulation in Nursing*. 2010;6(5):e177-e84.
- 13 Pearson M, Smith D. Debriefing in experience-based learning. *Simulation/Games for Learning*. 1986;16:155–72.
- 14 Koshes R, Young S, Stokes H. Debriefing following combat. In: Jones F, Sparacino L, Wilcox V, Rothberg J, Stokes J, eds. *War Psychiatry*. Washington, DC: Office of The Surgeon General United States of America 1994:271-90.

- 15 Marshall S. Bringing Up the Rear: A Memoir. 1st ed. California: Presidio Press 1979.
- 16 United States Army. Leader actions to offset battle fatigue risk factors [Field Manual 22–51]. 1994 [cited 2011 9 September]; Available from:
<http://www.globalsecurity.org/military/library/policy/army/fm/22-51/>
- 17 United States Department of the Army. Debriefers' Guide to Critical Event Debriefing. Texas: Combat Stress Actions Office, Department of Preventive Health Services, Army Medical Department Center and School; 1993.
- 18 Lederman L. Debriefing: Toward a systematic assessment of theory and practice. *Simulation & Gaming*. 1992;23(2):145-60.
- 19 Billings C, Reynard W. Human factors in aircraft incidents: results of a 7-year study. *Aviation, Space, and Environmental Medicine*. 1984;55(10):960-5.
- 20 Eisen L, Savel R. What went right: Lessons for the intensivist from the crew of US airways flight 1549. *Chest*. 2009;136(3):910-7.
- 21 Byrnes R, Black R. Developing and implementing CRM programs: The Delta experience. In: Wiener E, Kanki B, Helmreich R, eds. *Cockpit resource management*. San Diego, CA: Academic Press 1993:421-46.
- 22 Weiner E, Kanki B, Helmreich R, eds. *Cockpit resource management*. San Diego: Academic Press 1993.
- 23 Civil Aviation Authority. Crew Resource Management (CRM) Training. Guidance for Flight Crew, CRM Instructors (CRMIs) and CRM Instructor-Examiners (CRMIEs). CAP 737. Version 2. Gatwick: Civil Aviation Authority; 2006.
- 24 Miller F, Kaptchuk T. Deception of subjects in neuroscience: An ethical analysis. *The Journal of Neuroscience*. 2008 28(19):4841– 3.
- 25 Wendler D, Miller F. Deception in the pursuit of science. *Archives of Internal Medicine*. 2004;164(6):597-600.
- 26 Mitchell J, Everly G. Critical incident stress debriefing: An operations manual for the prevention of traumatic stress among emergency services and disaster workers. 2nd ed. Ellicott City, MD: Chevron Publishing Corp 1996.
- 27 Hattie J, Timperley H. The power of feedback. *Review of Educational Research*. 2007;77(1):81-112.
- 28 Zigmont J, Kappus L, Sudikoff S. The 3D model of debriefing: Defusing, Discovering, and Deepening. *Seminars in Perinatology*. 2011;35(2):52-8.

- 29 Van Heukelom J, Begaz T, Treat R. Comparison of postsimulation debriefing versus in-simulation debriefing in medical simulation. *Simulation in Healthcare*. 2010;5(2):91-7
- 30 Petranek C, Corey S, Black R. Three levels of learning in simulations: Participating, debriefing, and journal writing. *Simulation & Gaming*. 1992;23(2):174-85.
- 31 Mayville M. Debriefing: The essential step in simulation. *Newborn and Infant Nursing Reviews*. 2011;11(1):35-9.
- 32 Wagner D, Bear M, Sander J. Turning simulation into reality: increasing student competence and confidence. *Journal of Nursing Education*. 2009;48(8):465.
- 33 Fletcher G, McGeorge P, Flin RH, Glavin RJ, Maran NJ. The role of non-technical skills in anaesthesia: a review of current literature. *British Journal of Anaesthesia*. 2002;88(3):418-29.
- 34 Fanning R, Gaba D. The role of debriefing in simulation-based Learning. *Simulation in Healthcare*. 2007;2(2):115-25
- 35 Boet S, Bould MD, Bruppacher HR, Desjardins F, Chandra DB, Naik VN. Looking in the mirror: Self-debriefing versus instructor debriefing for simulated crises. *Critical Care Medicine*. 2011;39(6):1377-81
- 36 Dine C, Gersh R, Leary M, Riegel B, Bellini L, Abella B. Improving cardiopulmonary resuscitation quality and resuscitation training by combining audiovisual feedback and debriefing *Critical Care Medicine*. 2008;36(10):2817-22
- 37 Morgan P, Tarshis J, LeBlanc V et al. Efficacy of high-fidelity simulation debriefing on the performance of practicing anaesthetists in simulated scenarios. *British Journal of Anaesthesia*. 2009;103(4):531-7.
- 38 Savoldelli G, Naik V, Park J, Joo H, Chow R, Hamstra S. Value of debriefing during simulated crisis management: Oral versus video-assisted oral feedback. *Anesthesiology*. 2006;105(2):279-85.
- 39 Welke T, LeBlanc V, Savoldelli G et al. Personalized oral debriefing versus standardized multimedia instruction after patient crisis simulation. *Anesthesia & Analgesia*. 2009;109(1):183-9.
- 40 Byrne A, Sellen A, Jones J et al. Effect of videotape feedback on anaesthetists' performance while managing simulated anaesthetic crises: a multicentre study. *Anaesthesia*. 2002;57(2):176-9.
- 41 Chronister C, Brown D. Comparison of simulation debriefing methods. *Clinical Simulation in Nursing*. 2011;In press DOI: 10.1016/j.ecns.2010.12.00.
- 42 Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. *British Journal of Anaesthesia*. 2003;90(5):580-8.

- 43 Radhakrishnan K, Roche J, Cunningham H. Measuring clinical practice parameters with human patient simulation: a pilot study. *International Journal of Nursing Education Scholarship*. 2007;4(1):1 - 11.
- 44 Arnold J, Johnson L, Tucker S, Malec J, Henrickson S, Dunn W. Evaluation tools in simulation learning: Performance and self-efficacy in emergency response. *Clinical Simulation in Nursing*. 2009;5:e35-e43.

Appendix I: Search strategy

AMED- 22 January 2012

1. Manikins.mp.	8
2. Patient Simulation.mp.	10
3. Mannequin.mp.	5
4. Debriefing.mp.	51
5. 1 OR 2 OR 3	23
6. #4 AND # 5	0
6. Number retrieved based on abstract further analysis	0

The Cochrane Central Register of Controlled Trials (CENTRAL)- 22 January 2012

1. Debriefing	161
2. Simulation	4301
3. Mannequin	77
4. Patient Simulation	2628
5. # 2 OR # 3 OR # 4	17850
6. #1 AND # 5	125
7. Number retrieved based on abstract further analysis	6

Directory of open access journals -24 January 2012

1. All Fields=Debriefing AND All Fields=Simulation	4
2. Number retrieved based on abstract further analysis	0

EMBASE-24 January 2012

1. Manikins.mp.	438
2. Patient Simulation.mp.	455
3. Mannequin.mp.	853
4. Debriefing.mp.	1736
5. (#1 OR #2 OR #3)	1695
6. # 4 AND #5	52
7. Number retrieved based on abstract further analysis	2

Medline-24 January 2012

1. exp Manikins/	2567
2. exp Patient Simulation/	2496
3. Mannequin.mp.	610
4. Debriefing.mp.	1180
5. (#1 OR #2 OR #3)	5139
6. # 4 AND # 5	122
7. Number retrieved based on abstract further analysis	5

ProQuest Nursing Journals-24 January 2012

1. Manikins	1129
2. Patient Simulation	586
3. Mannequin	4147
4. Debriefing	3091
5. (#1 OR #2 OR #3)	5731
6. #4 AND # 5	44
7. Number retrieved based on abstract further analysis	0

CINAHL- 26 January 2012

1. Manikins	107
2. Patient Simulation	1744
3. Mannequin	186
4. Debriefing	742
5. (#1 OR #2 OR #3)	1998
6. #4 AND # 5	57
7. Number retrieved based on abstract further analysis	7

Dissertation and Theses-24 January 2012

1. Simulation	102250
2. Debriefing	741
3. #1 AND #2	49
4. Number retrieved based on abstract further analysis	0

ERIC-28 January 2012

1. Manikin	3
2. Patient Simulation	20
3. Mannequin	15
4. Debriefing	719
5. (#1 OR #2 OR #3)	45
6. # 4 AND # 5	2
6. Number retrieved based on abstract further analysis	0

Journals@Ovid-28 January 2012

1. Manikins.mp. [mp=title, abstract, full text, caption text]	1049
2. Patient Simulation.mp. [mp=title, abstract, full text, caption text]	862
3. Mannequin.mp. [mp=title, abstract, full text, caption text]	1713
4. Debriefing.mp. [mp=title, abstract, full text, caption text]	9266

5. (#1 OR #2 OR #3)	3371
6. # 4 AND # 5	479
7. Number retrieved based on abstract further analysis	4

Mednar-28 January 2012

1. Debriefing	1,346
2. Patient Simulation	1,967
3. Healthcare	3,997
4. (#1 AND #2 AND #3)	390
5. Number retrieved based on abstract further analysis	5

PsycINFO-28 January 2012

1. Manikin\$.mp.	514
2. exp Simulation/	29956
3. Patient Simulation.mp.	61
4. Mannequin\$.mp.	98
5. Debrief\$.mp.	1846
6. (#1 OR #2 OR #3 OR #4)	30547
7. # 5 AND # 6	104
7. Number retrieved based on abstract further analysis	0

Appendix II: Joanna Briggs Institute Critical Appraisal Tools**Critical Appraisal of Evidence of Effectiveness**

Reviewer _____ Date _____

Author _____ Year _____ Record Number _____

1) Was the assignment to treatment groups truly random?

yes ☐ no ☐ not clear ☐ NA ☐

2) Were participants blinded to treatment allocation?

yes ☐ no ☐ not clear ☐ NA ☐

3) Was allocation to treatment group concealed from the allocator?

yes ☐ no ☐ not clear ☐ NA ☐

4) Were the outcomes of people who withdrew described and included in the analysis?

yes ☐ no ☐ not clear ☐ NA ☐

5) Were those assessing the outcomes blind to the treatment allocation

yes ☐ no ☐ not clear ☐ NA ☐

6) Were control and treatment groups comparable at entry?

yes ☐ no ☐ not clear ☐ NA ☐

7) Were groups treated identically other than for the named interventions?

yes ☐ no ☐ not clear ☐ NA ☐

8) Were outcomes measured in the same way for all groups?

yes ☐ no ☐ not clear ☐ NA ☐

9) Were outcomes measured in a reliable way?

yes ☐ no ☐ not clear ☐ NA ☐

10) Was appropriate statistical analysis used?

yes ☐ no ☐ not clear ☐ NA ☐

Overall appraisal: Include ☐ Exclude ☐ Seek further info ☐

Comments (including reasons for Exclusion)

Appendix III: Critical appraisal results for included studies

Author	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
Boet, et al. ³⁵ , 2011	Y	N	U	NA	Y	Y	Y	Y	Y	Y
Byrne, et al. ⁴⁰ , 2002	Y	N	N	NA	N	Y	Y	Y	N	Y
Chronister & Brown ⁴¹ , 2011	Y	N	N	NA	N	Y	Y	Y	Y	Y
Dine, et al. ³⁶ , 2008	Y	N	N	N	N	Y	Y	Y	Y	Y
Grant, et al. ¹² , 2010	U	N	N	NA	Y	Y	Y	Y	Y	Y
Morgan, et al. ³⁷ , 2009	Y	N	N	Y	Y	Y	Y	Y	Y	Y
Savoldelli, et al. ³⁸ , 2006	U	N	N	NA	Y	Y	Y	Y	Y	Y
Shinnick, et al. ⁵ , 2011	Y	N	N	NA	N	Y	Y	Y	Y	Y
Van Heukelom, et al. ²⁹ , 2010	Y	N	N	NA	N	Y	Y	Y	Y	Y
Welke, et al. ³⁹ , 2009	Y	N	N	Y	Y	Y	Y	Y	Y	Y

Key Y = Yes; N = No; U = Unclear; NA = not applicable

Appendix IV: Data Extraction of Evidence of Effectiveness

Data Extraction Form (Quantitative Data)

Author **Record Number**

Journal

Year

Reviewer

Method _____

Setting _____

Participants _____

Number of Participants

Group A **Group B**

Interventions

Intervention A _____

Intervention B

Appendix V: Excluded paper and reason for exclusion

1. Zausig, Y., Grube, C., Boeker-Blum, T., Busch, C., Bayer, Y., Sinner, B., et al. (2009). Inefficacy of simulator-based training on anaesthesiologists' non-technical skills. *Acta Anaesthesiologica Scandinavica*, 53(5), 611-619. **Reason for exclusion:** groups not treated identically other than for the named interventions

Appendix VI: Characteristics of included studies

Author(s) Country, Year	Methodology	Participants Groups (sample size)	Details of simulation scenario	Outcomes	Results
Boet, et al. ³⁵ Canada, 2011	Randomised Controlled Trial	Anaesthesiology residents (N= 50) Control group Self-debriefing (n= 25) Experimental group video- assisted instructor facilitated debriefing (n= 25)	30 minutes orientation followed by 20 minutes debriefing.	Nontechnical skills such as situational awareness, team working, decision making, and task management evaluated using Anaesthesia Non-Technical Skills (ANTS) scale.	No statistically difference in improvement of nontechnical skills performance between control and experimental groups ($p = 0.58$). Effective teaching of nontechnical skills can be achieved through formative video-assisted self-assessment even when facilitators are not available.
Byrne, et al. ⁴⁰ UK, 2002	Randomised Controlled Trial	Anaesthesiology residents (N= 32) Control group instructor facilitated debriefing between each simulation (n= 16) Experimental group video-facilitated instructor debriefing between each of the simulations (n =16)	Five simulations scenarios in a single session with debriefing between each simulation. No details on length of debriefing or scenarios.	Time taken to solve the simulation scenario and accuracy of recording vital signs measurements as measured by automatic recordings by the simulation software.	The experimental group took a shorter median time to solve simulation scenario however, the difference was not statistically significant ($p > 0.05$).
Chronister & Brown ⁴¹ USA, 2011	Randomised Controlled Trial	Undergraduate nursing students (N=37) Control group instructor facilitated debriefing (n= not given) Experimental group video-assisted instructor facilitated debriefing (n =not given)	15-minute simulation scenarios followed by a 30 minutes debriefing.	Quality of students assessment and psychomotor skills, response time and knowledge retention as measured by the Emergency Response Performance Tool (ERPT).	Greater knowledge retention occurred in the control group ($p < 0.008$). Experimental group participants demonstrated significantly faster times for three skills: pulse assessment with initiation of CPR ($p = 0.094$), initial defibrillation shock ($p = 0.042$), and total time to resuscitation ($p = 0.028$).
Dine, et al. ³⁶ USA, 2008	Randomised Controlled Trial	Nurses (N= 80) Control group instructor facilitated debriefing (n=34) Experimental group a combination of instructor facilitated debriefing and real-time audio-visual feedback from a defibrillator (n=31)	Three simulated cardiac arrest scenarios of 2 minutes each and 5 minutes debriefing.	Cardiopulmonary resuscitation quality as measured by chest compression rate/depth as measured by automatic recordings by the simulation software.	Control: Depth increased after debriefing, from 38% to 68% ($p = 0.015$). Experimental group: Depth compliance improved from 19% to 58% ($p = 0.002$). The combination of feedback and debriefing improved participants' compression rate compliance from 45% to 84% ($p = 0.001$) and adequate rate and from 29% to 64% ($p = 0.005$)
Grant, et al. ¹² USA, 2010	Randomised Controlled Trial	Nursing and nurse anaesthetist students (N=40) Control group	Three -60-minute simulation session (two were practice).	Clinical Simulation Evaluation Tool (CSET) scores for target behaviour demonstrating	Experimental group: Higher mean scores in the following: patient identification [$p < 0.01$]; team

Author(s) Country, Year	Methodology	Participants Groups (sample size)	Details of simulation scenario	Outcomes	Results
		instructor facilitated debriefing Verbal debriefing (n= 20) Experimental group simulation scenarios using video-facilitated instructor debriefing (n= 20)		patient identification, team communication, and vital signs	communication [$p = 0.013$]; and assessment of vital signs [$p = 0.047$]. There were no statistically significant difference between the control and experimental groups in their total performance score (8.44 and 9.09, respectively).
Morgan, et al. ³⁷ Canada, 2009	Randomised Controlled Trial	Anaesthetists (n= 74) Control group Home study (n=21) No intervention (n=21) Experimental group Instructor facilitated Debriefing (n=29)	Two simulation scenarios. No details on length of debriefing or scenarios.	Participants' management of the scenarios assessed by Dichotomously scored checklist (DSC) and Global Rating Scale of Performance (GRS).	There was a statistically significant improvement in the scores on the DSC in the debriefing group when compared with the control. Although the two groups improved equally from pre- to post-test there was no statistical difference between overall performance ($p = 0.61$).
Savoldelli, et al. ³⁸ Canada, 2006	Randomised Controlled Trial	Anesthesia residents (N=42) Control group B Instructor facilitated debriefing (n= 15) Control group C No debriefing (n= 15). Experimental group (group A) Video-assisted instructor facilitated debriefing (n= 12),	8 minutes scenarios and debriefing of about 20 minutes.	Nontechnical skills such as situational awareness, team working, decision making, and task management evaluated using ANTS scale.	No statistically significant differences in performance of nontechnical skills such as situational awareness, team working, decision making, and task management between the groups ($p < 0.005$).
Shinnick, et al. ⁵ USA, 2011	Randomised Controlled Trial	Undergraduate nursing students (N =162) Control group (n= 72) Experimental group (n=90)	Simulation scenarios followed by 30-minute group debriefing sessions.	Heart failure (HF) clinical knowledge as measured by the Clinical Knowledge Questionnaire (CKQ).	Mean HF knowledge scores for both groups decreased from the pre-test to the first post-test ($p < 0.001$) but improved after the combination of simulation experience and debriefing sessions ($p < 0.001$)
Van Heukelom, et al. ²⁹ USA, 2010	Randomised Controlled Trial	Undergraduate medical students (N=161) Control group immediate feedback during the simulation experience/in-simulation debriefing (n = 84) Experimental group Instructor facilitated debriefing session after the simulation	Total time for simulation scenario and debriefing limited to 20 minutes.	Self-reported confidence and knowledge in ability to perform medical resuscitation skills as measured by a seven-point Likert-scale questionnaire.	Statistically significant improvement evidenced by higher mean post-simulation scores compared with pre-simulation scores for both individual items and overall measures related to students self-reported confidence and knowledge for both groups ($p \leq 0.001$).

Author(s) Country, Year	Methodology	Participants Groups (sample size)	Details of simulation scenario	Outcomes	Results
		experience/post-simulation debriefing (n= 77).			
Welke, et al. ³⁹ Canada, 2009	Randomised Controlled Trial	Anesthesia residents (N=30) Control group Multimedia resources (n= 15) Experimental group video-facilitated instructor debriefing after the simulation (n=15)	Three simulation scenarios of 10 minutes each. No details on length of debriefing.	Nontechnical skills such as situational awareness, team working, decision making, and task management evaluated using ANTS scale.	The improvements in total ANTS score from simulation 1 to 2 ($p= 0.97$), simulation 1 to simulation 3 ($p = 0.94$), and or simulation 2 to 3 ($p = 0.84$) were similar for both groups.

Appendix VII: JBI levels of evidence

JBI Levels of Evidence	Effectiveness
I	Meta-analysis (with homogeneity) of experimental studies (e.g. Randomized Controlled Trials (RCT) with concealed randomization) OR One or more large experimental studies with narrow confidence intervals
II	One or more smaller RCTs with wider confidence intervals OR Quasi-experimental studies (without randomization)
III	a. Cohort studies (with control group) b. Case-controlled c. Observational studies (without control group)
IV	Expert opinion, or physiology bench research, OR consensus