Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: https://www.tandfonline.com/loi/terg20

Does simulation-based training in medical education need additional stressors? An experimental study

Greta Ontrup, Miriam Vogel, Oliver T. Wolf, Peter K. Zahn, Annette Kluge & Vera Hagemann

To cite this article: Greta Ontrup, Miriam Vogel, Oliver T. Wolf, Peter K. Zahn, Annette Kluge & Vera Hagemann (2020) Does simulation-based training in medical education need additional stressors? An experimental study, Ergonomics, 63:1, 80-90, DOI: 10.1080/00140139.2019.1677948

To link to this article: https://doi.org/10.1080/00140139.2019.1677948

Accepted author version posted online: 07 Oct 2019.
Published online: 31 Oct 2019.

ERGONOMICS
2020, VOL. 63, NO. 1, 80–90
https://doi.org/10.1080/00140139.2019.1677948

Taylor & Francis
Taylor & Francis Group

ARTICLE

Does simulation-based training in medical education need additional stressors? An experimental study

Greta Ontrup[a]*, Miriam Vogel[b]*, Oliver T. Wolf[c], Peter K. Zahn[d], Annette Kluge[a] and Vera Hagemann[e]

[a]Department of Work Organizational and Business Psychology, Ruhr University Bochum, Bochum, Germany; [b]Department of Anaesthesiology, Intensive Care Medicine, Emergency Medicine and Pain Management, Klinikum Kassel, Kassel, Germany; [c]Department of Cognitive Psychology, Ruhr University Bochum, Bochum, Germany; [d]Department of Anaesthesiology, Intensive Care Medicine and Pain Management, University Hospital Bergmannsheil, Bochum, Germany; [e]Human Resources, Department of Business Studies and Economics, University of Bremen, Bremen, Germany

ABSTRACT

The increased curricular integration of simulation-based training (SBT) in medical education is accompanied by researchers' calls to examine the effectiveness of SBT. We address conflicting results regarding effects of an added stressor on learning outcomes. In an experimental setting, one group of medical students ($N = 20$) performed cardiopulmonary resuscitation on a patient simulator. For a second group ($N = 21$) the scenario differed in that they encountered a defect defibrillator. We found participants of both groups to show increased biological stress-levels, independent of group allocation. Paradoxically, participants who encountered the equipment failure subjectively reported less stress. We discuss the implications of the comparable high stress levels in both groups with regards to future studies. We further discuss the result regarding subjective stress levels within the framework of attribution theory.

Practitioner summary: The results of our experimental study underline the need for evidence-based choices of additional stressors for the design of simulation scenarios. We describe the choice of stimuli and setting in detail to maximise practical value for the construction of simulation-based medical trainings.

Abbreviations: SBT: simulation-based training interventions; ERC: European Resuscitation Council Guidelines for Resuscitation; ROSC: return of spontaneous circulation; sAA: alpha-amylase; OSCE: Objective Structured Clinical Examination; ECG: electrocardiogram

ARTICLE HISTORY
Received 12 April 2019
Accepted 9 September 2019

KEYWORDS
Simulation-based training; high-fidelity simulator; stress responses; performance; learning

Introduction

Simulation-based training interventions (SBT) are ubiquitous in many high reliability environments (Horn et al. 2015; Saborit et al. 2010) and are increasingly implemented in medical education (Levine 2013). SBT is used for *technical skills* training, e.g. knowledge acquisition and retention (Ackermann 2009; Anderson and Warren 2011; Cotard and Michinov 2018; Rauen 2004; Rezmer et al. 2011), and *non-technical skills* training, e.g. teamwork skills (Hagemann et al. 2017; Merry 2007). In both cases, the interaction of the simulation environment with physiological and psychological characteristics of the learner constitutes a crucial issue regarding the effectiveness of the tool (Bradley 2006; Issenberg et al. 2011). The call for 'further research elucidating what works, for whom, under what circumstances' (Cook et al. 2013) has stimulated valuable investigations on a multitude of factors, relating to characteristics of the learner (Clarke 2018), or to characteristics of the simulation environment (Hamstra et al. 2014) and their impact on the effectiveness of medical SBT. Concerning the latter, one issue that is still controversially discussed concerns the optimal design of the simulation environment regarding (additional) stressors.

Stressors are internal or external demands imposed on or inherent to the learner; widely recognised stimuli in the research domain concern personal factors, characteristics of the simulation environment or the confederate, technical factors, events or distractions (Arora et al. 2010; DeMaria & Levine 2013; Wetzel et al. 2006). For examples, see Table 1. On the one hand, researchers argue that as medical professionals work in high-stress environments, high-stress during

CONTACT Greta Ontrup greta.ontrup@rub.de Department of Work Organizational and Business Psychology, Ruhr University Bochum, Universitätsstraße 150, Bochum 44801, Germany
*These authors contributed equally to this work.

Table 1. Summary of possible additional stressors during SBT

Category	Stressor	Source
Intrinsic to all immersive simulation	public demonstration of performance	A
	managing rare events	A
	managing life-threatening events	A
	public justification of actions and performance shortcomings	A
	time compressed simulations/time pressured	A
	treating the simulator/simulation as if it were real embarrassing	A
	watching and listening to one's self on video	A
Participants/Personal factors	complexity of case out of proportion to participant level of training	A
	learning new techniques	B
	taking a leadership role	A
	taking a team membership role	A
	relying on people during a critical event who you never worked with before	A
	errors and mistakes/failure	A
	fatigue, tiredness	B; C
	personal issues, personal problems	B; C
	hunger	C
	illness	C
	physical discomfort	C
Technical factors	unfamiliar equipment	A
	failing devices	A
	backup devices unavailable	A
Confederates/team factors	dismissive colleagues	A
	less than helpful support staff, flack of support	B; A
	confrontation with patient, family, other practitioners, senior faculty	A
	anxious, nervous, screaming confederates (patient, family, other healthcare providers)	A
	supervising uncooperative, poorly performing, or generally obstructive and unhelpful subordinates	A
	inexperienced, incompetent staff	B; C
	staff paying no attention	C
	interpersonal problems	C
	language problems	C
Events	rapidly deteriorating vital signs	A
	being rushed	A
	bad outcome, patient injury or demise	A
	ineffective therapeutic interventions	A
	unanticipated difficulties	A
	diagnostic dilemmas	A
	bleeding	B
	unexpected anatomy	B
	procedure not going according to plan	B
	complications	B
	immediate decision making	C
	complex procedure	C
	high-risk patients	C
	difficulties finding the source of a problem	C
	no progress	C
Distractions	interruptions	B
	noise/radio, bleeps	B; C
	Performance pressure	B
	multitasking, triaging, and managing multiple patients simultaneously/multiple demands	B; A
	people walking in and out	C
	phone calls	C

Note. Source A = DeMaria & Levine, 2013 ; B = Arora et al., 2009 ; C = Wetzel et al., 2006.
Stressors printed in italics were included in the piloting; stressors printed in bold were included in the present study.

SBT is ideal for skills transfer (Andreatta, Hillard, and Krain 2010). This assumption is supported by research that found stress to increase participants' performance and memory after SBT (DeMaria et al. 2010; DeMaria & Levine 2013). Consequently, multiple ways to enhance stress during SBT through additional stimuli are proposed (Andreatta, Hillard, and Krain 2010; DeMaria et al. 2010; DeMaria & Levine 2013; Wetzel et al. 2006). On the other hand, studies found increased stress to deteriorate participants performance, knowledge retrieval and skills transfer (Hunziker et al. 2011; Prabhu et al. 2010; Valentin et al. 2015). These studies question the educational value of highly stressful simulations and consequently the use and value of additional stressors (Keitel et al. 2011; Piquette et al. 2014). Broadly summarised, studies propagating different stressors, like interruptions or lack of support, to enrich the simulation environment are at odds with controversial evidence on the (enhanced) usefulness of such 'enriched' scenarios.

Our study targeted the question of how to design SBT regarding stressfulness of the learning environment. We address conflicting results regarding the design of SBT by investigating the effects of an added stimuli serving as a stressors on learning outcomes. We will precisely elaborate on the choice of the stimuli below, to maximise practical value regarding the design of SBT within medical education.

However, to delineate substantial hypotheses regarding the effects of stress on meaningful learning and performance outcomes, we find it inevitable to draw on fundamental research concerning stress and memory processes, as stress has been shown to influence multiple learning systems in various ways (Schwabe and Wolf 2013). First, stress was found to have positive effects on memory encoding (the creation of new memories) and consolidation (Wolf 2017). Second, stress is suggested to cause a shift from a flexible, declarative form of learning (i.e. cognitive learning) towards habit based (procedural) learning: the higher the stress, the more rigid the form of learning (DeMaria & Levine 2013; Schwabe and Wolf 2013). Third, stress impairs memory retrieval. Retrieving important information during stressful episodes is hindered by stress (Kluge et al. 2019; Wolf 2017).

Cardiopulmonary resuscitation has been taught and studied within SBT before (Hunziker et al. 2011; Müller et al. 2009), we thus chose it as the educational case for our scenario. One group of students attended a scenario without manipulation of the stimuli ('functioning equipment'), the second group attended the same scenario but with manipulated stimuli, i.e. equipment (see methods section for choice of stressor) ('equipment failure').

Hypothesis (1): Participants of the two groups show increased biological stress levels directly after the simulation, compared to before the simulation.

Hypothesis (2): Participants of the 'equipment failure' group show higher biological and subjective stress levels after the simulation, compared to participants of the 'functioning equipment' group.

Drawing on fundamental research reported above, we expect the following:

Hypothesis (3). Compared to the 'functioning equipment' group, the 'equipment failure' group should a) perform worse during the simulation (retrieval), b) show less declarative learning achievements on the subject of resuscitation procedure (declarative vs. procedural knowledge acquisition), and at the same time c) score better at an episodic memory test (encoding and consolidation).

Materials and methods

Sampling and study procedure

Medical students (human medicine) in their 7th to 9th semester were recruited using posters and mailing lists of the Ruhr University Bochum (standard period of study in Germany are 10 semesters). Due to their known influences on cortisol levels, exclusion criteria consisted of a history of neurological or psychiatric conditions, prescription of drugs influencing the central nervous system or corticosteroids, regular shift work or substance abuse. To minimise confounding influences, people with a professional background in emergency medicine were excluded. Participation was voluntarily. Participants received 30 Euro remuneration. The ethical committee of the Ruhr University Bochum (medical faculty) approved the study (registration number: 15-5570). We followed a single-blind design and assigned participants randomly and concealed to one of the two groups by drawing lots (see Figure 1 for trial flow).

The study took place on two consecutive days. Before the first testing day, participants received an obligatory electronic invitation to refresh their knowledge on the European Resuscitation Council Guidelines for Resuscitation (ERC) on an online-learning platform (Soar et al. 2015). The first testing day took place at the 'Bergmannsheil' hospital in Bochum. Participants answered demographic questionnaires, the knowledge test on the ERC guidelines (Soar et al. 2015), completed a hands-on training and the actual simulation (see below). On day two, participants again completed the knowledge test on the ERC guidelines and an episodic memory test (see below). They received written debriefing information a week after completing the study. Figure 1 displays the detailed study procedure.

Choice of stressor

We first summarised possible stimuli that have been described as additional stressors, see Table 1. We chose one stimuli per category, not including the categories 'intrinsic to all immersive simulation' and 'participants/personal factors'. As a 'failing technical device' we included a defect defibrillator, as a 'confederate/team'-factor the accompanying nurse was instructed to induce verbal pressure (see below), as 'unanticipated difficulties' the tongue of the mannequin was programmed to swell during the procedure, and as a 'distraction' loud background noises were included.

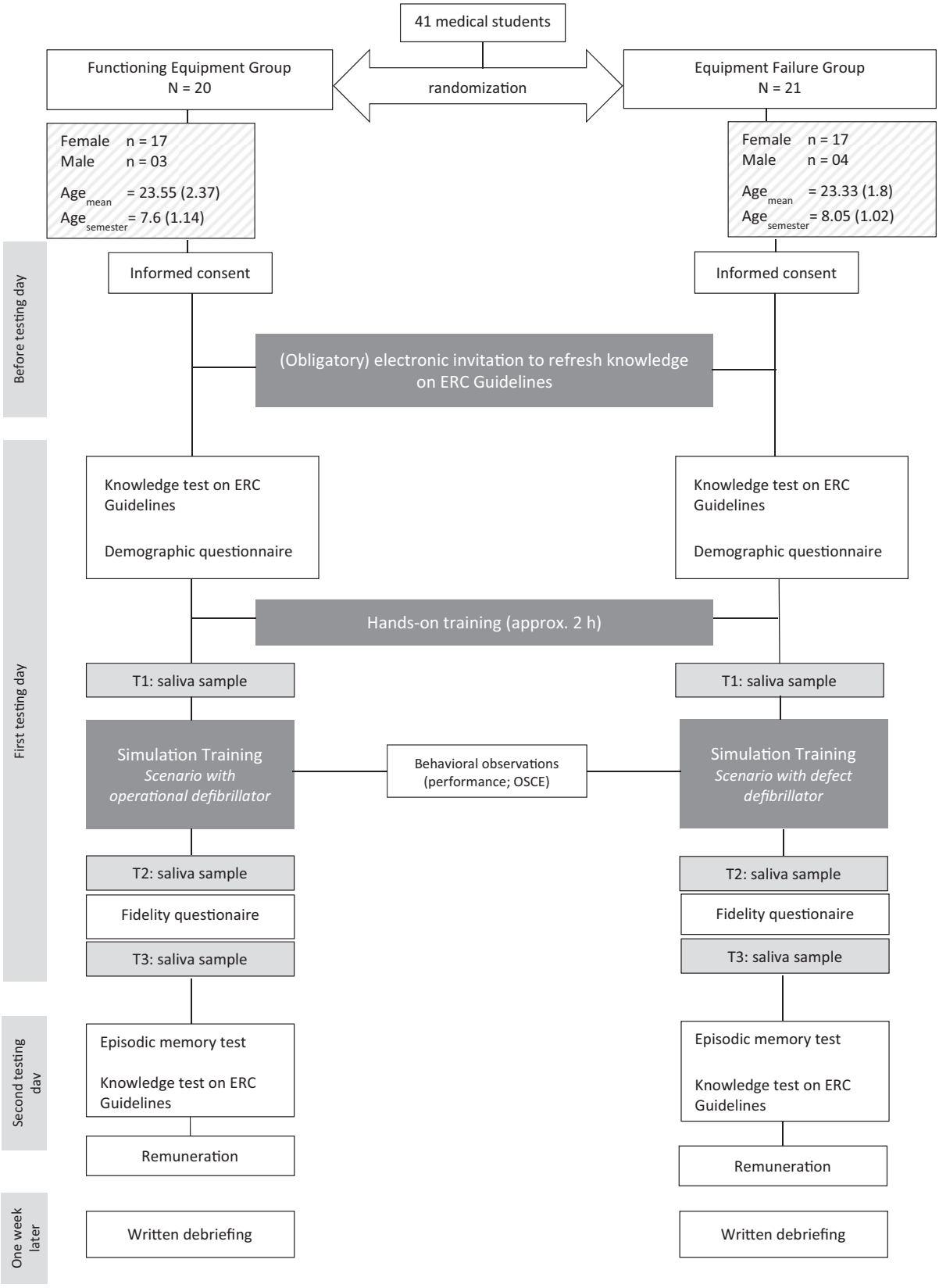

Figure 1. Trial flow and graphical overview of the chronical study procedure.

We conducted five pilot runs including the four stimuli. Ten assistant doctors (two doctors performing one simulation together) completed the pilots. As an evaluation criterion we relied on subjective reports. The assistant doctors reported that they found the scenario very difficult and extremely stressful. As our targeted participants were less experienced (medical students), we decided to reduce the stimuli from four to two.

We chose the defect defibrillator as our main added stimuli due to the following reasons: First, it holds a key position within the algorithm (see below for detailed description of scenario), with a specific time point and its use does not offer alternatives. We could therefore make sure that the scenario was identical for the two groups before and after the stimuli. Second, the defibrillator has incremental importance for the survival of the patient. Third, the stimuli is easily controllable; we were able to intervene by 'finding the cause' of the problem (missing batteries) and thus proceed with the simulation if participants did not recognise the cause after 2 min. This ensured maximal comparability between the groups. Fifth, it was impossible for participants to not recognise the stimuli. We decided to also include verbal pressure by the accompanying nurse as a second stimuli, as it narratively made sense within the scenario (see below).

We completed two pilot runs including these two stimuli. Four medical students within their practical year completed the scenario. They reported that they found the scenario to be appropriate concerning procedure, levels of stress and difficulty.

Simulation scenario

The scenario was designed by a certificated simulation-instructor with experiences in clinical emergency medicine.

Participants received a verbal standardised introduction:

> You are playing an assistant physician in this hospital. Together with the intensive-care nurse you are the appointed emergency team of the hospital for the night. It is 11 pm, the night shift has just taken over when a nurse pages you to the casualty ward that is close to your location. Once you open the door, you arrive at the ward, here is your emergency kit.

Participants entered the scenario with the nurse, to meet the nurse who paged them. Two of three student assistants took turn in playing the nurses, they were instructed beforehand which script to follow (from which group condition). The nurses were instructed to be helpful, to not initiate any actions until they got instructions from the participant and to perform tasks conscientiously. The patient simulator (Resusci Anne Simulator®, 'Laerdal Medical') was placed in a supine position on the floor; an emergency cart was standing nearby. No monitoring was installed; the patient had vascular access. The simulator was programmed to show neither respiration, nor cardio-circulation. Participants were expected to start mechanical cardiopulmonary resuscitation. The simulator was programmed to show ventricular fibrillation, prompting the participants to follow the Algorithm of Advanced Life Support (ERC) and thus use the defibrillator (Soar et al. 2015). The simulator was programmed to indicate asystole after the first shock, prompting participants to continue with medicinal treatment. The scenario ended with the return of spontaneous circulation (ROSC) after approximately 10 min.

The scenarios only differed in that the 'equipment failure' group additionally encountered a defective defibrillator, whereas in the 'functioning equipment' group the defibrillator was operational. Alongside, the script for the nurses only differed in that they were instructed to induce verbal pressure on the participants of the 'equipment failure' group when the failing defibrillator became apparent (e.g. 'what do we do now? We have to do something!'). Furthermore, they were instructed to find the cause for the failing defibrillator (missing batteries) after 2 min. if the participant had not offered any solutions up to that point. The nurse then went to get batteries so that participants were able to perform the next steps.

Material

Demographic questions included age, gender, weight, and height. Questions on the medical history of the participant and their experiences with SBT and emergency medicine were asked with regards to the exclusion criteria mentioned above. Internal consistencies (α) are reported for this study in the following.

As biological markers of stress, we used salivary cortisol and alpha-amylase (sAA). The two markers relate to the two primary stress-response systems. Cortisol is a widely used indicator of the 'slow' stress response via the hypothalamus-pituitary-adrenal axis (DeMaria & Levine 2013; Schwabe and Wolf 2013); sAA is assumed to be a reliable indicator of the 'rapid' stress response via the autonomic nervous system (Nater and Rohleder 2009). The assay of the salivary samples followed the procedure described in Herten et al. (2016).

Five items previously designed and validated by one of the authors were used to measure subjective stress, e.g. 'Looking back at the simulation scenario, I felt overwhelmed' ($\alpha = .72$) (Hagemann et al. 2017). Participants responded on a four-point Likert-scale, anchored by 'disagree' to 'agree'.

The authors applied the six-item presence scale to assess fidelity of the scenario (Frank and Kluge 2014). Participants answered on a Likert-Scale, ranging from 'not true at all' to 'very true' (e.g. 'I felt like I was part of the simulation scenario'; $\alpha = .72$).

To assess participants' performance during the simulation, we designed a checklist following the 'Objective Structured Clinical Examination' (Harden and Gleeson 1979). OSCE has been shown to have fewer psychometric problems compared to conventional performance-based assessments (Scalese and Hatala 2013). Participants' performance was rated by a research assistant during the simulation on key actions concerning basic and advanced life support. Deviations and delays in time by the participant during the scenario were counted as errors. Afterwards, a total OSCE score was computed; the maximum were 21 points. To avoid confounding the performance measure with the effect of the defect defibrillator, performance was deliberately not evaluated for the 2 min. of troubleshooting for the 'equipment failure' group. Please see supplemental material 1 for the detailed OSCE.

On the first and second testing day declarative knowledge of the ERC Guidelines was assessed (see Figure 1). The contents of the knowledge test related directly to the processes and procedures that needed to be used and applied – und thus held the chance to be acquainted – during the scenario. The test followed the design of medical exams at the Ruhr University Bochum, so that participants would be familiar with the outline. It consisted of 24 multiple-choice questions concerning basic life support, advanced life support, defibrillation and ECG, ventilation and securing breathing, return of spontaneous circulation and reversible causes. The answers were aggregated into a knowledge index for each participant; the maximum score was 120. Higher scores signified greater knowledge. For the pre simulation test, item difficulties fell between $p_i = 33.2$–99; post simulation, difficulties ranged from $p_i = 41.4$–99. Generally, items showed lower difficulties post simulation, yet the range can be considered balanced for the two measurements. The supplemental material 2 displays the item analysis for the items.

To assess episodic memories, we designed questions relating to our specific scenario. Participants were asked 28 multiple choice questions, some referring to the situation (e.g. 'the patient was lying in a supine position on the floor'); some referring to the actions of the participant (e.g. 'how often did you inject 1 mg of adrenalin?'). To assess the average memory performance, answers were aggregated; the maximum score was 53 points. Item difficulties fell between $p_i = 23$–95; the test was therefore balanced regarding item difficulty (see supplemental material 3 for item analysis).

Statistical analysis

We based the sample size calculation on the analysis of mixed variance with repeated measures (planned effect size of $f = 0.25$, alpha $= 0.05$, power $= 0.95$). The software G*Power calculated a minimum sample size of 18 subjects per group.

For the analysis of between and within group effects of cortisol and sAA levels, mixed design analysis of variance with repeated measures were conducted. Subjective stress levels were compared using independent *t*-test. Declarative knowledge scores were compared prior and after simulation between the groups with mixed analysis of variance to observe between- and within-effects. The analyses concerning performance and memory scores were focussed on posterior between-group effects, thus independent *t*-tests were computed.

As cortisol and sAA data do not follow normal distributions (Rohleder and Nater 2009), we first normalised the data ($\log_e$). We found one outlier concerning cortisol levels; further inspections revealed that cortisol levels lay around 3.5 to 4 SD above the mean. Values $>$ 3 SD above the mean probably reflect sample contamination or an unreported disease; the data for this participant was removed for the cortisol analysis.

Results

Preliminary analysis

We first analysed the groups regarding prior experience with SBT and their prior declarative knowledge to exclude confounded results based on group differences. The groups neither differed regarding experience with SBT, $\chi^2(1) = 2.47$, $p = .12$, nor regarding mean scores on the prior knowledge test $t(39) = -.40$, $p = .69$. We further analysed the effects of group membership on perceived fidelity (presence scale) of the

scenario. There was no significant difference in the ratings between the two groups, $t(39) = -.60$, $p = .55$.

Hypotheses testing

Descriptively, cortisol and sAA levels followed the same pattern (Figure 2 and 3). For the two groups levels increased from the first time of measurement (before SBT (T1)) to the second time of measurement (after SBT (T2)) and went down again towards the end of the testing day (T3).

Hypothesis (1). Inferentially, there was a significant effect of time on cortisol levels, $F(1.45, 54.91) = 34.36$, $p < .001$, $\eta^2 = .48$. Tests of within-subject contrasts revealed a significant difference between before (T1) and after the simulation (T2), $F(1, 38) = 39.05$, $p < .001$, $\eta^2 = .51$. The difference between T2 and T3 was not statistically significant, $F(1, 38) = 0.71$, $p = .41$. Additionally, there was a significant effect of time on sAA levels, $F(2, 76) = 42.74$, $p < .001$, $\eta^2 = .53$. The differences between T1 and T2 ($F(1, 38) = 77.89$, $p = .001$, $\eta^2 = .24$) and between T2 and T3 ($F(1, 38) = 90.24$, $p < .001$, $\eta^2 = .70$) were significant. Thus, the SBT lead to significantly higher stress levels in both groups in spite of the added stimuli.

Hypothesis (2). There was neither a significant interaction between group membership and time on cortisol levels ($F(1.45, 54.91) = .66$, $p = .47$), nor on sAA levels ($F(2, 76) = .86$, $p = .43$). The experimental manipulation did not cause differential responses in physiological stress markers.

Participants within the 'functioning equipment' condition reported higher perceived subjective stress during the simulation ($M = 3.05$, $SD = 0.56$) than participants within the 'equipment failure' condition ($M = 2.68$, $SD = 0.56$). This difference was statistically significant, $t(39) = 2.15$, $p < .05$, $r = .33$.

Hypothesis (3). Analysis of variance showed a significant effect of time on knowledge, $F(1, 39) = 101.67$, $p < .001$, $\eta^2 = .72$. The interaction between group and time was not statistically significant, $F(1, 39) = .21$, $p = .65$. Participants of the two groups equally reached higher scores post simulation ('functioning equipment' group: $M_{prä} = 75.5$, $SD_{prä} = 8.61$, $M_{post} = 87.9$, $SD_{post} = 8.00$; 'equipment failure' group: $M_{prä} = 77.33$, $SD_{prä} = 10.84$, $M_{post} = 88.67$, $SD_{post} = 8.22$).

The difference between OSCE scores of the 'functioning equipment' group ($M = 15.58$, $SD = 2.27$) and the 'equipment failure' group ($M = 16.52$, $SD = 2.87$) was not statistically significant, $t(39) = -1.17$, $p = .25$, $r = .18$.

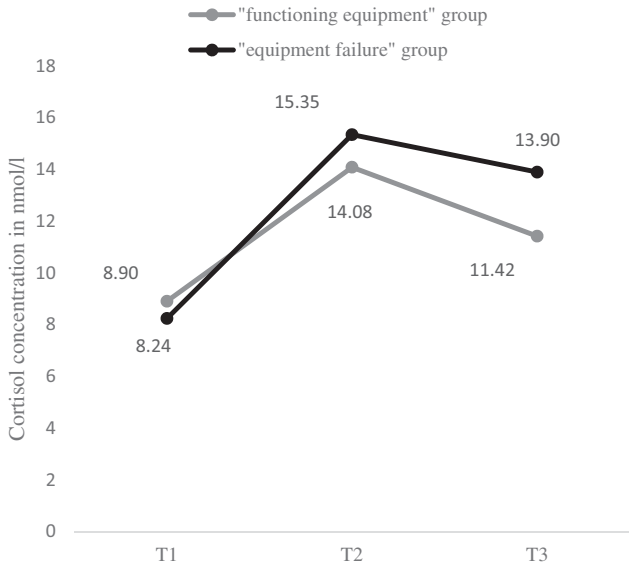

Figure 2. Cortisol level for the two groups before the simulation-based training (T1), right after the training (T2) and 30 min after the training (T3).

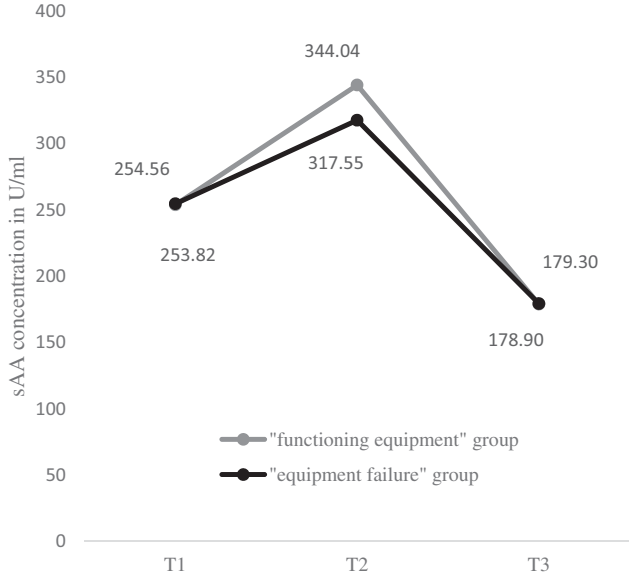

Figure 3. Alpha-amylase level for the two groups before the simulation-based training (T1), right after the training (T2) and 30 min after the training (T3).

Regarding episodic memory, participants within the 'functioning equipment' group reached an average of 30.8 points ($SD = 4.60$), compared to an average of 30.0 points ($SD = 3.08$) for the 'equipment failure' group (the maximum score would have been 53 points). This difference was not statistically significant, $t(39) = .61$, $p = .55$, $r = -.10$.

Discussion

The purpose of this study was to investigate effects of added stressors on different learning outcomes, i.e. performance, declarative learning and episodic memory. The methodological design included a randomised experimental setting; one group of medical students performed cardiopulmonary resuscitation on a patient simulator, the second group encountered a defect defibrillator (added stimuli) in the otherwise identical scenario. As expected, SBT lead to significant increases in biological stress markers for all participants, which supports studies that emphasise the stressful nature of SBT in medical settings (Hagemann et al. 2017; Keitel et al. 2011; Müller et al. 2009). Beyond, we did not find the failed equipment to lead to enhanced stress levels: participants of both groups showed increased biological stress-levels independent of group allocation. Paradoxically, participants who encountered the added stimuli (equipment failure) subjectively reported less stress.

We first want to address the result of non-significant differences in physiological stress levels, as it signifies a failed experimental manipulation. This entails and prohibits answering our subsequent hypothesis on the effects of stress on learning and performance, which substantially comprises the interpretation of our study. The result constitutes a major limitation and we discuss it explicitly in the 'limitations' section (see below). Yet, we are convinced that the result of the failed manipulation is in itself conclusive and of merit – especially combined with the paradoxical finding regarding subjective stress levels. That is, although we must refrain from any conclusions regarding the effects of stress on performance and learning, we want to offer a discussion on the results regarding physiological and subjective stress levels, which we believe to be of value for future studies.

Regarding physiological stress levels, the hypothesis was not supported in that there were no group differences. Yet the result did not occur because there was *no* increase in stress, on the contrary, *both* groups showed significant increases in stress levels. We want to suggest the possibility of a 'ceiling effect' for our targeted population. Dickerson and Kemeny (2004) reported in their meta-analysis of acute stressors and cortisol response an overall mean effect size of $d = 0.31$. The highest effect sizes are reported for tasks including public speaking/cognitive task combinations ($d = 0.87$). Effect sizes for other type of tasks fell between $d = 0.20$–0.39. Following the same calculation reported in the meta-analysis, we find an effect size of $d = 0.84$ for our cortisol data. It is conceivable that the added stimuli (equipment failure) did not lead to the desired effect, because stress levels were already high and thus did not allow much room to rise. This would be in line with research hinting at the 'inherently' stressful nature of simulations (Keitel et al. 2011; Müller et al. 2009). It also adds another dimension to the question regarding the failed manipulation for the experimental group: following our explanation of a ceiling effect, the question might not (only) be why the stimuli did not provoke enough stress in the experimental group, it might also be whether low stress conditions for the control group and thus targeted sample exist.

The result of the high physiological stress levels is especially intriguing combined with the result on subjective stress levels: participants of the 'functioning equipment' group reported to have experienced greater stress compared to the 'equipment failure' group. We draw on attribution theories to explain this paradox. Attribution theories differentiate if people trace their failures or successes back to their own personal acts or abilities or to external reasons (Davis and Davis 1972). In the context of our study, it is possible that – especially because we assessed subjective stress retrospectively – participants within the 'equipment failure' group were able to attribute their struggles and stress to the failing device. Participants of the 'functioning equipment' group in contrast were not 'offered' an external cause. They did not have the chance of attributing their stress to anything else but their skills and actions, which in turn might have led to a bigger subjective perception of stress. DeMaria and Levine (2013) discussed that high stress during simulations might motivate students to study harder after SBT. They therefore suggest adding stressors to the simulation to support this 'motivational boost'. Our study adds concern to this argumentation, as it hints at the possibility of 'attributional excuses' for students. If the simulation scenario offers students a possibility to trace back their struggles to anything else rather than their own skills, it might reduce educational value.

Limitations

We want to address limitations of the present study to facilitate adequate integration of our research. First, our experimental manipulation did not elicit different stress levels, making the interpretation of our last hypothesis based on group differences meaningless. As described above, we did not rely on stimuli for which previous studies demonstrated significant effects (e.g.

DeMaria et al. 2016; Dias and Scalabrini-Neto 2016, 2017). The selection of the stimuli was initially based on various enumerations of possible stressors proposed by Arora et al. (2010), DeMaria and Levine (2013) and Wetzel et al. (2006). For the piloting we then relied on subjective reporting of assistant doctors. While we are convinced that the field benefits from the empirical investigation of different stimuli serving as stressors, the choice proved not to be helpful with regards to answering our last hypothesis. On the one hand, this underlines the need for accumulative empirical evidence on the interaction of the simulation environment with the characteristics of certain subgroups of learners (Issenberg et al. 2011), as it demonstrates that only because a certain condition is listed as a theoretical stressor, is does not necessarily have to be a stressor in a specific setting for a specific population. Future empirical research on this is needed, especially with regards to advices for practitioners (Cook et al. 2013). On the other hand, the results underline the need to include physiological measures of stress in the piloting of new simulation scenarios. Although we acknowledge the exact replication of a simulation scenario as extremely difficult, we believe the replication of the piloting with the added measure of physiological data to be worthwhile. First, this can help to elaborate on the above discussed idea of whether it was for the stimuli to not elicit *enough stress* or rather for the 'functioning equipment' condition to produce *too much* stress. Second, a replication study can then answer our initial hypothesis regarding effects of stress on learning outcomes and performance.

Third, we want to address the objective of investigating declarative knowledge acquisition. As the participants were familiar with the theories beforehand, we expected the simulation to be useful in order to consolidate and thus enhance their declarative knowledge, as described in previous studies (Anderson and Warren 2011). In accordance with these assumptions, our results demonstrate declarative knowledge acquisition for both groups, independent of stress levels. However, there are numerous studies that emphasise the usefulness of SBT for the acquisition of procedural rather than declarative knowledge (Alinier 2007). Within this context, simulation is aimed at improving implementation of knowledge, not knowledge per se. Against this background, the comparison of procedural and declarative knowledge acquisition would be another interesting research realm for future studies. Fourth, a convenience sample was used; it is perceivable that only highly motivated students took part, leading to self-selection bias. Fifths, Wetzel et al. (2006) point at the importance of experience for performances in stressful environments. Our results are only applicable for medical students and cannot be generalised to medical staff or professionals.

Implications and conclusion

Taken together, the results of our study first and foremost stress the importance of considering specific characteristics of the learners (e.g. experience) when designing (non-) stressful simulation environments. Our results highlight the need for future studies on different proposed stressors for different target populations, to further elaborate on the question what works for whom (Cook et al. 2013). In this context our study underlines the need to include physiological measures of stress in the piloting of different simulation scenarios – or to rely on already validated stressors when addressing a specific hypothesis that necessitates group differences. The result of the high stress levels in both groups combined with the high effect sizes call for research on the question of whether low stress conditions for inexperienced medical students exist. Combined with the paradox regarding subjective stress ratings, it also raises the question, if additional stimuli exhibit an unwanted effect regarding attribution of performance. We propose the exploration of attributional consequences on long-term learning as an interesting research realm.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Ackermann, A. D. 2009. "Investigation of Learning Outcomes for the Acquisition and Retention of CPR Knowledge and Skills Learned with the Use of High-Fidelity Simulation." *Clinical Simulation in Nursing* 5 (6): e213–e222. doi:10.1016/j.ecns.2009.05.002.

Alinier, G. 2007. "A Typology of Educationally Focused Medical Simulation Tools." *Medical Teacher* 29 (8): e243–e250. doi:10.1080/01421590701551185.

Anderson, J. M., and J. B. Warren. 2011. "Using Simulation to Enhance the Acquisition and Retention of Clinical Skills in Neonatology." *Seminars in Perinatology* 35 (2): 59–67. doi:10.1053/j.semperi.2011.01.004.

Andreatta, P. B., M. Hillard, and L. P. Krain. 2010. "The Impact of Stress Factors in Simulation-Based Laparoscopic Training." *Surgery* 147 (5): 631–639. doi:10.1016/j.surg.2009.10.071.

Arora, S., N. Sevdalis, D. Nestel, M. Woloshynowych, A. Darzi, and R. Kneebone. 2010. "The Impact of Stress on Surgical Performance: A Systematic Review of the Literature." *Surgery* 147 (3): 318–330. doi:10.1016/j.surg.2009.10.007.

Bradley, P. 2006. "The History of Simulation in Medical Education and Possible Future Directions." *Medical Education* 40 (3): 254–262. doi:10.1111/j.1365-2929.2006.02394.x.

Clarke, L. 2018. *Motivation to Learn during Simulation-Based Learning: An Examination of Learner Characteristics in Health Science Students*. Portland: Concordia University Portland.

Cook, D. A., S. J. Hamstra, R. Brydges, B. Zendejas, J. H. Szostek, A. T. Wang, P. J. Erwin, and R. Hatala. 2013. "Comparative Effectiveness of Instructional Design Features in Simulation-Based Education: Systematic Review and Meta-Analysis." *Medical Teacher* 35 (1): e867–e898. doi:10.3109/0142159X.2012.714886.

Cotard, C., and E. Michinov. 2018. "When Team Member Familiarity Affects Transactive Memory and Skills: A Simulation-Based Training among Police Teams." *Ergonomics* 61 (12): 1591–1600. doi:10.1080/00140139.2018.1510547.

Davis, W. L., and D. E. Davis. 1972. "Internal-External Control and Attribution of Responsibility for Success and Failure." *Journal of Personality* 40 (1): 123–136. doi:10.1111/j.1467-6494.1972.tb00653.x.

DeMaria, S., E. R. Silverman, K. A. B. Lapidus, C. H. Williams, J. Spivack, A. Levine, and A. Goldberg. 2016. "The Impact of Simulated Patient Death on Medical Students' Stress Response and Learning of ACLS." *Medical Teacher* 38 (7): 730–737. doi:10.3109/0142159X.2016.1150986.

DeMaria, S., Jr., E. O. Bryson, T. J. Mooney, J. H. Silverstein, D. L. Reich, C. Bodian, and A. I. Levine. 2010. "Adding Emotional Stressors to Training in Simulated Cardiopulmonary Arrest Enhances Participant Performance." *Medical Education* 44 (10): 1006–1015. doi:10.1111/j.1365-2923.2010.03775.x.

DeMaria, S., Jr., and A. I. Levine. 2013. "The Use of Stress to Enrich the Simulated Environment." In *The Comprehensive Textbook of Healthcare Simulation*, edited by A. I. Adam, 65–71. New York: Springer Science + Business Media. doi:10.1007/978-1-4614-5993-4_5.

Dias, R. D., and A. Scalabrini-Neto. 2016. "Stress Levels during Emergency Care: A Comparison between Reality and Simulated Scenarios." *Journal of Critical Care* 33: 8–13. doi:10.1016/j.jcrc.2016.02.010.

Dias, R. D., and A. Scalabrini-Neto. 2017. "Acute Stress in Residents during Emergency Care: A Study of Personal and Situational Factors." *Stress (Stress* 20 (3): 241–248. doi:10.1080/10253890.2017.1325866.

Dickerson, S. S., and M. E. Kemeny. 2004. "Acute Stressors and Cortisol Responses: A Theoretical Integration and Synthesis of Laboratory Research." *Psychological Bulletin* 130 (3): 355–391. doi:10.1037/0033-2909.130.3.355.

Frank, B., and A. Kluge. 2014. "Development and First Validation of the PLBMR for Lab-Based Microworld Research." Proceedings of the International Society for Presence Research, Vienna, Austria. Retrieved from https://ispr.info/presence-conferences/previous-conferences/ispr-2014/.

Hagemann, V., F. Herbstreit, C. Kehren, J. Chittamadathil, S. Wolfertz, D. Dirkmann, A. Kluge, and J. Peters. 2017. "Does Teaching Non-Technical Skills to Medical Students Improve Those Skills and Simulated Patient Outcome?" *International Journal of Medical Education* 8: 101–113. doi:10.5116/ijme.58c1.9f0d.

Hamstra, S. J., R. Brydges, R. Hatala, B. Zendejas, and D. A. Cook. 2014. "Reconsidering Fidelity in Simulation-Based Training." *Academic Medicine* 89 (3): 387–392. doi:10.1097/ACM.0000000000000130.

Harden, R. M., and F. A. Gleeson. 1979. "Assessment of Clinical Competence Using an Objective Structured Clinical Examination (OSCE)." *Medical Education* 13 (1): 39–54. doi:10.1111/j.1365-2923.1979.tb00918.x.

Herten, N., T. Otto, D. Adolph, B. M. Pause, R. Kumsta, and O. T. Wolf. 2016. "Enhanced Startle Responsivity 24 Hours after Acute Stress Exposure." *Behavioral Neuroscience* 130 (5): 521–530. doi:10.1037/bne0000156.

Horn, G. P., R. M. Kesler, R. W. Motl, E. T. Hsiao-Wecksler, R. E. Klaren, I. Ensari, M. N. Petrucci, B. Fernhall, and K. S. Rosengren. 2015. "Physiological Responses to Simulated Firefighter Exercise Protocols in Varying Environments." *Ergonomics* 58 (6): 1012–1021. doi:10.1080/00140139.2014.997806.

Hunziker, S., L. Laschinger, S. Portmann-Schwarz, N. K. Semmer, F. Tschan, and S. Marsch. 2011. "Perceived Stress and Team Performance during a Simulated Resuscitation." *Intensive Care Medicine* 37 (9): 1473–1479. doi:10.1007/s00134-011-2277-2.

Issenberg, S. B., C. Ringsted, D. Ostergaard, and P. Dieckmann. 2011. "Setting a Research Agenda for Simulation-Based Healthcare Education: A Synthesis of the Outcome from an Utstein Style Meeting." *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare* 6 (3): 155–167. doi:10.1097/SIH.0b013e3182207c24.

Keitel, A., M. Ringleb, I. Schwartges, U. Weik, O. Picker, U. Stockhorst, and R. Deinzer. 2011. "Endocrine and Psychological Stress Responses in a Simulated Emergency Situation." *Psychoneuroendocrinology* 36 (1): 98–108. doi:10.1016/j.psyneuen.2010.06.011.

Kluge, A., M. Silbert, U. S. Wiemers, B. Frank, and O. T. Wolf. 2019. "Retention of a Standard Operating Procedure under the Influence of Social Stress and Refresher Training in a Simulated Process Control Task." *Ergonomics* 62 (3): 1–15. doi:10.1080/00140139.2018.1542036.

Levine, A. I., S. DeMaria Jr., A. D. Schwartz, and A. J. Sim (Eds.). 2013. *The comprehensive textbook of healthcare simulation*. New York: Springer Science & Business Media.

Merry, A. F. 2007. "Human Factors and the Cardiac Surgical Team: A Role for Simulation." *The Journal of the American Society of Extra-Corporeal Technology* 39: 264–266.

Müller, M. P., M. Hänsel, A. Fichtner, F. Hardt, S. Weber, C. Kirschbaum, S. Rüder, F. Walcher, T. Koch, and C. Eich. 2009. "Excellence in Performance and Stress Reduction during Two Different Full Scale Simulator Training Courses: A Pilot Study." *Resuscitation* 80 (8): 919–924. doi:10.1016/j.resuscitation.2009.04.027.

Nater, U. M., and N. Rohleder. 2009. "Salivary Alpha-Amylase as a Non-Invasive Biomarker for the Sympathetic Nervous System: Current State of Research." *Psychoneuroendocrinology* 34 (4): 486–496. doi:10.1016/j.psyneuen.2009.01.014.

Piquette, D., J. Tarshis, T. Sinuff, R. A. Fowler, R. Pinto, and V. R. Leblanc. 2014. "Impact of Acute Stress on Resident Performance during Simulated Resuscitation Episodes: A

Prospective Randomized Cross-over Study." *Teaching and Learning in Medicine* 26 (1): 9–16. doi:10.1080/10401334.2014.859932.

Prabhu, A., W. Smith, Y. Yurko, C. Acker, and D. Stefanidis. 2010. "Increased Stress Levels May Explain the Incomplete Transfer of Simulator-Acquired Skill to the Operating Room." *Surgery* 147 (5): 640–645. doi:10.1016/j.surg.2010.01.007.

Prieto Saborit, J. A., M. Del Valle Soto, V. González Díez, M. A. Montoliu Sanclement, P. Nistal Hernández, J. Egocheaga Rodríguez, and L. Santos Rodríguez. "Physiological Response of Beach Lifeguards in a Rescue Simulation with Surf." *Ergonomics* 53(9): 1140–1150. doi:10.1080/00140139.2010.502255.

Rauen, C. A. 2004. "Simulation as a Teaching Strategy for Nursing Education and Orientation in Cardiac Surgery." *Critical Care Nurse* 24 (3): 46–51.

Rezmer, J., T. Begaz, R. Treat, and M. Tews. 2011. "Impact of Group Size on the Effectiveness of a Resuscitation Simulation Curriculum for Medical Students." *Teaching and Learning in Medicine* 23 (3): 251–255. doi:10.1080/10401334.2011.586920.

Rohleder, N., and U. M. Nater. 2009. "Determinants of Salivary Alpha-Amylase in Humans and Methodological Considerations." *Psychoneuroendocrinology* 34 (4): 469–485. doi:10.1016/j.psyneuen.2008.12.004.

Scalese, R. J., and R. Hatala. 2013. "Competency Assessment." In *The Comprehensive Textbook of Healthcare Simulation*, edited by A. I. Adam, 135–160. New York: Springer Science + Business Media.

Schwabe, L., and O. T. Wolf. 2013. "Stress and Multiple Memory Systems: From 'Thinking' to 'Doing'." *Trends in Cognitive Sciences* 17 (2): 60–68. doi:10.1016/j.tics.2012.12.001.

Soar, J., J. P. Nolan, B. W. Böttiger, G. D. Perkins, C. Lott, P. Carli, T. Pellis, C. Sandroni, M. B. Skrifvars, G. B. Smith., et al. 2015. "European Resuscitation Council Guidelines for Resuscitation 2015: Section 3. Adult Advanced Life Support." *Resuscitation* 95: 100–147. doi:10.1016/j.resuscitation.2015.07.016.

Valentin, B., O. Grottke, M. Skorning, S. Bergrath, H. Fischermann, D. Rörtgen, M. Mennig, C. Fitzner, M. P. Müller, C. Kirschbaum, et al. 2015. "Cortisol and Alpha-Amylase as Stress Response Indicators during Pre-Hospital Emergency Medicine Training with Repetitive High-Fidelity Simulation and Scenarios with Standardized Patients." *Scandinavian Journal of Trauma, Resuscitation and Emergency Medicine* 23 (1): 1–8. doi:10.1186/s13049-015-0110-6.

Wetzel, C. M., R. L. Kneebone, M. Woloshynowych, D. Nestel, K. Moorthy, J. Kidd, and A. Darzi. 2006. "The Effects of Stress on Surgical Performance." The American Journal of Surgery 191 (1): 5–10. doi:10.1016/j.amjsurg.2005.08.034.

Wolf, O. T. 2017. "Stress and Memory Retrieval: Mechanisms and Consequences." *Current Opinion in Behavioral Sciences* 14: 40–46. doi:10.1016/j.cobeha.2016.12.001.