https://doi.org/10.35662/unine-thesis-2894

UNIVERSITÉ DE NEUCHÂTEL

Dissecting team reflexivity, a prospective observation-based study: Effects of content and quality of reflexivity on resuscitation performance in ad-hoc medical teams

Camille Morgenthaler

Thèse présentée à la Faculté des Sciences pour l'obtention du grade de Docteur ès Sciences

Jury de thèse

Franziska Tschan

Laurenz Meier

Stephan Marsch

Université de Neuchâtel

Institut de Psychologie du Travail et des Organisations

Rue Emile-Argand 11, 2000 Neuchâtel

Défendue le 24 septembre 2020

UNIVERSITÉ DE NEUCHÂTEL

Faculté des Sciences
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel – Suisse
Tél : + 41 (0)32 718 21 00
E-mail : secretariat.sciences@unine.ch

IMPRIMATUR POUR THESE DE DOCTORAT

La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par

Madame Camille MORGENTHALER

Titre:

“Dissecting team reflexivity, a prospective observation-based study: effects of content and quality of reflexivity resuscitation performance in ad-hoc medical teams”

sur le rapport des membres du jury composé comme suit:

- Prof. émérite Franziska Tschan Semmer, directrice de thèse, Université de Neuchâtel, Suisse
- Prof. Laurenz Meier, Université de Neuchâtel, Suisse
- Prof. Stephan Marsch, Universitätsspital Basel, Suisse

Neuchâtel, le 13 avril 2021

Le Doyen, Prof. A. Bangerter

UNIVERSITÉ DE
NEUCHÂTEL

Rectorat
Fbg de l'Hôpital **41**
2000 Neuchâtel
Tél. +41 32 718 10 20
messagerie.rectorat@unine.ch

Déclaration sur l'honneur*

Par la présente, j'affirme avoir pris connaissance des documents d'information et de prévention du plagiat émis par l'Université de Neuchâtel et m'être renseigné-e correctement sur les techniques de citation.

J'atteste par ailleurs que le travail rendu est le fruit de ma réflexion personnelle et a été rédigé de manière autonome.

Je certifie que toute formulation, idée, recherche, raisonnement, analyse ou autre création empruntée à un tiers est correctement et consciencieusement mentionnée comme telle, de manière claire et transparente, de sorte que la source en soit immédiatement reconnaissable, dans le respect des droits d'auteur et des techniques de citations.

Je suis conscient-e que le fait de ne pas citer une source ou de ne pas la citer clairement, correctement et complètement est constitutif de plagiat.

Je prends note que le plagiat est considéré comme une faute grave au sein de l'Université. J'ai pris connaissance des risques de sanctions administratives et disciplinaires encourues en cas de plagiat (pouvant aller jusqu'au renvoi de l'université).

Je suis informé-e qu'en cas de plagiat, le dossier sera automatiquement transmis au rectorat.

Au vu de ce qui précède, **je déclare sur l'honneur ne pas avoir eu recours au plagiat ou à toute autre forme de fraude.**

Nom : MORGENTHALER Prénom : Camille Sophie

Cursus : Doctorat - psychologie du travail et des organisations Faculté d'inscription : Sciences

Lieu et date : Neuchâtel, le 2 septembre 2020 Signature : 

****

Ce formulaire doit être dûment rempli par tout étudiant ou toute étudiante rédigeant un travail substantiel (notamment un mémoire de bachelor ou de master) ou une thèse de doctorat. Il doit accompagner chaque travail remis au professeur ou à la professeure.

*Formulaire largement inspiré de la Directive de la direction 0.3 bis, intitulée Formulaire Code de déontologie en matière d'emprunts, de citations et d'exploitation de sources diverses, de l'Université de Lausanne, du 23 avril 2007 et adapté aux besoins de l'Université de Neuchâtel.

Abstract

This dissertation focuses on two studies investigating a self-led between task reflexivity intervention between two simulated resuscitation tasks. We will begin with a theoretical introduction about the reflexivity process and its effects. We will also position reflexivity in relation the challenges faced by healthcare teams. We then present the two papers at the core of this dissertation. The two papers focus on cardiopulmonary performance improvement, content of reflexivity session, adherence to resuscitation algorithm in the resuscitation phases and discussed during reflexivity, and finally quality of reflexivity. Results of both papers show that a short self-led reflexivity intervention help teams to improve their resuscitation performance. Analyses of the reflexivity process itself allowed to demonstrate that our participants engaged in the reflexivity process and followed the instructions. Quality (especially scope of reflection) impacts performance improvement. In the discussion, we address perspectives and limitations.

Keywords: Team training, group reflection, group reflexivity, team briefings, medical simulation, cardiopulmonary resuscitation, action teams

Résumé

Cette thèse est composée de deux études portant sur une intervention de réflexivité entre deux tâches de réanimation simulées. Nous commencerons par une introduction théorique sur le processus de réflexivité et ses effets. Nous positionnerons également la réflexivité par rapport aux défis rencontrés par les équipes de soins. Nous présenterons ensuite les deux articles qui sont au cœur de cette these. Ils portent sur l'amélioration des performances de réanimation cardio-pulmonaires, le contenu de sessions de réflexivité, l'adhésion à l'algorithme de réanimation dans les phases de réanimation ainsi que les éléments de cet algorithme discutés pendant la réflexivité, et enfin la qualité de la réflexivité. Les résultats des deux articles montrent qu'une courte intervention de réflexivité auto-guidée aide les équipes à améliorer leur performance en réanimation. Les analyses du processus de réflexivité lui-même ont permis de démontrer que nos participants se sont engagés dans le processus de réflexivité et ont suivi les instructions. La qualité (en particulier l'étendue de la réflexion) a un impact sur l'amélioration des performances. Dans la discussion, nous abordons les perspectives et les limites.

Mots-clés: Formation des équipes, réflexion de groupe, réflexivité de groupe, briefings d'équipe, simulation médicale, réanimation cardiopulmonaire, action teams

Ever tried. Ever failed. No matter. Try Again. Fail again. Fail better.

Samuel Beckett

Acknowledgments

I can never find words strong enough to express all my gratitude, but here is a non-exhaustive list of the people who have contributed to this work and supported me.

Above all, I thank my thesis director, Franziska Tschan, for her patience, support and advices. She has been most generous in helping me complete this PhD. I really enjoyed our work sessions behind her screen: her enthusiasm for the data was always very inspiring.

She could cheer me up when I did not know where to tackle a problem, or when the results did not go in the direction I wanted (showing that reflexivity was a miracle process). She once quoted a famous advertising slogan “Ça ne va pas mieux. Mais plus longtemps”, referring to the mixed results of reflexivity intervention in our project; she thus taught me the sad and difficult, even cruel, reality of research, where not everything is exclusively black or white and conclusions must be nuanced. She once told me that she felt like Dalí's “Soft Watches”, and I believe she was able to put a name to a condition that I frequently experienced during my thesis.

I also thank all the people who collaborated on the project at one time or another and who helped me, in a more or less direct way, to complete this work. In particular, I would like to thank the members of my team who spent hours and hours behind their screens coding our videos.

I would also like to thank all my family and friends:

Pierre, who suffered my moods,

And Linda, who's been through them before him and always has good advices.

My family, especially my sister Sarah, for those horseback ridings that freed my mind,

my friends,

especially Julie to reread my work a few days before the end of her pregnancy,

who have also lived the ups and downs with me and who have encouraged me again ang again.

My team, Sandra, Simon, Eliane and Jasmin, as well as my fellow PhD students, Julie, Daniela, Cyril, Denise and Ottilie, for their help, their encouragement and support in difficulties, all the writing sessions in common, and especially the complaint sessions in common.

Thanks to Garance, the most formidable secretary,

And Françoise, who is also the most formidable secretary,

for their unfailing support and wise assistance through the intricacies of administration.

Thanks also to Robin, who collaborated with me for his Master's thesis. This was not a simple supervision of a Master's thesis, but a real collaboration. His help was invaluable from a methodological, coding and analysis point of view and allowed me to carry out my own work.

Many thanks to Ottilie and Amy for their wise rereading of my awkward English.

Thanks to all the people who shared my working space, especially during the coronavirus crisis; a little change and freshness helped my concentration. Being surrounded was a real comfort in these troubled times.

I would also like to take this opportunity to apologize to all the people who have suffered my -sometimes- terrible mood.

Thanks to Freddie Mercury and Chopin, in particular, but also to hundreds of other artists for their musical talents and whose wonderful musical works have always accompanied my reflection.

Infinite thanks to the tons of chocolate I devoured.

And thank also to the coronavirus, perhaps a providential opportunity or a spiritual message to force me to devote myself to my task.

Foreword

After these years of thesis that have not always been easy, I can finally allow myself a short reflection focused around my work. Indeed, according to the Cambridge dictionary, reflection is commonly understood as a “serious and careful thought”. I have learned a lot, for sure. I am disappointed that I do not have more compelling evidence showing the wonders of reflexivity, but I am convinced that this process is useful and that I could demonstrate part of its potential in this dissertation. I am still not sure if – or how – I can now answer the terrible but classic question “what is your thesis about?”, but at least now I have this work as a response.

Table of contents

Tables and figures

Theoretical background

Paper 1

Paper 2

Dissecting team reflexivity, a prospective observational study:

Effects of content and quality of reflexivity on resuscitation performance in ad-hoc medical teams

This work is a paper-based dissertation including two papers. The first paper was accepted the 21st of December 2018. The second paper is a chapter of the dissertation that will serve as basis for one or two scientific papers to be submitted.

The project was supported by the Swiss National Foundation with the grant #149734.

Others co-authors are members of the research project which involved different teams from Neuchâtel university, Bern university and Basel university hospital (simulation center).

My first aim is to present in more detail the teams that I studied for a little more than these 4 years of thesis. Then, I am going to develop a theoretical background about reflexivity and key concepts related to reflexivity. I will try to systematically link these central themes to the medical world. I also say a few words about the methodology used in our research. Hopefully, this first theoretical part will help to apprehend the two papers in this dissertation.

References to the theoretical introduction and discussion are at the end of the dissertation (p. 195). References for the two papers presented are at the end of each paper. Some papers may therefore be referenced in several reference lists. The **Appendices** containing the coding manual used for both papers are presented after the references (from p. 215). Figures and tables are numbered independently in the dissertation and in the papers.

Through the two papers presented, I had the opportunity to explore a self-led reflexivity intervention in detail and in the specific context of medical simulations with medical students. I focus on content of reflexivity, as well as its detailed impact on team performance in cardiopulmonary resuscitation.

In the general discussion, I will then attempt to draw some general conclusions and recommendations for research as well as practice and highlight salient themes.

1. Why study reflexivity?

Why pay so much attention to a process that showed inconsistent results in past research? Reflexivity is such a short and easy process, that request no specific material and little cost (Kündig et al., 2019). If it has a potential to benefit the teams, then it is worthwhile to investigate a little further. It is indeed necessary to better understand this phenomenon in order to propose it to the teams, more notably medical teams.

Health care teams typically belong to complex decision-making groups (West, 1996), all means and intervention that can improve their work and help them to react more quickly and efficiently are valuable, especially taking into account the sometimes complicated working conditions (cf. section “Teams in healthcare” p. 28 and “Action teams” p. 29). Indeed, West (2000, p. 23) considers that modern teams (perhaps to this day even more so than at the time he wrote this book) “operate in increasingly uncertain and complex situations; where there is a multiplicity of factors to take into account in the decision-making process”. It is therefore essential to think about how best to support them.

Moreover, “complex decision-making groups depend on learning that cannot occur without reflection” (West, 2000, p. 1) is also the case for medical teams. West (2000) also considers that learning in teams is a process that does not happen spontaneously, time must be allowed for this purpose, and reflexivity is a possible mean. Edmondson et al. (2001) explain that learning is crucial for medical teams in particular, but also for many other teams: Adapting to new technologies, or new forms of business processes for example, is very disruptive. Teams have to go through technical but also organizational learning. Learning requires time until the team is perfectly at ease, but this time is sometimes not available.

Finally, medical errors are still a central problem in the health domain resulting sometimes in minor injuries up to cases of severe disability or even the death of thousands of people around the world. In 1999, the report “To err is human” (Kohn et al., 1999) estimated

about 100,000 deaths per year resulting from medical errors in the USA alone, (44000 to 98000 deaths per year according to Weingart et al., 2000), medical errors are thus the 8th cause of death in North America (Grober & Bohnen, 2005). The Federal Office of Public Health in Switzerland extrapolates that for Switzerland, this could represent 2000 to 3000 deaths annually (Vincent & Staines, 2019). However, it is difficult to find the real data for Switzerland. Interestingly, the main cause of error in the health domain is poor care coordination according to Schwappach (2014). Diagnostic errors represent an important error type with possible severe consequences (Zwaan et al., 2010), with information transfer problems as one of the main causes. It seems therefore that humans are responsible for many mistakes in the medical world. Reflexivity may be one influence to alleviate the problem. Moreover, reflexivity as an intervention can be very easily adapted and beneficial in other areas.

There is a potential that a reflexivity intervention allows to improve work and collaboration in medical teams, and fosters team learning. If reflexivity can contribute to patient safety, it is worth continuing to study it. However, there are inconsistencies and gaps regarding reflexivity that must be addressed through research.

2. Theoretical background

In this section, I will first shortly present the type of teams my research applies to. I will then introduce the concept of reflexivity and discuss research related to reflexivity as an intervention with a particular emphasis on effects of a reflexivity intervention on performance, as this was the topic of study 1 (Kündig et al., 2019). The second important focus of my dissertation is to evaluate the content, process and quality of the reflection process itself as this is the topic of the second paper.

Reflexivity is one of the central themes of this dissertation. It is discussed in detail from page 32 (section "The concept of reflexivity"), but here is a definition of reflexivity that can be kept in mind during the first few pages of reading: "…the extent to which group members overtly reflect upon the group's objectives, strategies and processes, and adapt them to current or anticipated endogenous or environmental circumstances" (West, 1996, p. 559). Reflexivity can be used as an intervention, like it was done in the papers included in this dissertation, and it is believed to increase team performance, even if not all research results are consistent.

We now return to the section with a presentation of the teams we worked with to study reflexivity.

Teams

Work teams are defined as "interdependent collection of individuals who share responsibilities for specific outcomes for their organizations" (Sundstrom et al., 1990, p. 120). Reser et al. (2018, p. 55) propose another interesting definition of teams that includes the concept of reflexivity: "In a team, all members have shared goals, there is interdependency and reflexivity". Both definitions are related to teams with temporal and member stability.

Teams in healthcare

My research is about teams in healthcare. At first glance, the definitions presented above apply to those teams as well. However, many healthcare teams work in specific contexts that influence their stability, their membership but also the team processes. Healthcare teams have some typical characteristics that need to be considered when studying them. St.Pierre et al. (2011) present a list of elements grouped in the following three dimensions (a) mission and goals, (b) performance, and (c) membership, that characterize teams in the medical world.

Regarding mission and goals, all members of a team should be oriented toward accomplishing a defined and timed objective. They should also respect a standard of performance. For the dimension of performance, teams should be oriented to their work, which most often has a definite start and end moment defined by tasks and missions. Teams in healthcare most often communicate and interact in real time. Moreover, several teams may work in parallel on the same case and must coordinate their actions. For certain routine tasks, algorithms can script this coordination. But if teams work ad-hoc (e.g., for an emergency), it is often principles and rules that guide their teamwork. For healthcare teams, good decision-making is thus, because often decisions are critical for patient outcomes.

Regarding the third dimension, membership, it is essential that members consider themselves as members of the team. As team composition changes often (for example due to shiftwork), this is an important and nontrivial aspect. The domain of healthcare is highly hierarchical, and team members must understand and respect the hierarchical structures of a team and the roles of leader and follower. During collaboration, a team's mission and objectives are subordinated to individual goals. It seems important to me to understand the issues and the functioning of medical teams, particularly ad-hoc teams, in order to highlight the interest of interventions that can help them work together, like reflexivity. Indeed, these

teams work under extreme and particular conditions with heavy responsibilities, with the primary objective to care and save lives, and any initiative to help them should be encouraged. This introduction can also help to clarify the context where the population (will) work and the perspective that we chose to study these teams for this research.

Action teams

In the healthcare setting, action teams (similar to "extreme action teams" according to Hollenbeck et al., 2012) can frequently be found. Action teams are often ad-hoc teams, i.e. teams that form around a task and work as a team for one or a few task cycles only (Devine et al., 1999; St.Pierre et al., 2011; Tschan et al., 2009), and generally disband after task fulfillment. This is opposite to temporally stable and, with regards to team membership, stable teams that are either working on different tasks in the same group composition or perform the same task in a cyclical way. Action teams are composed of "highly skilled members that cooperate to perform urgent, unpredictable, interdependent, and highly consequential tasks while simultaneously coping with frequent changes in team composition and training their teams' novice members" (Klein et al., 2006, p. 590). Sundstrom et al. (1990) also use the term action teams for teams working in the operating rooms, intensive care, emergency medicine, and trauma or resuscitation teams and emphasize that – in addition to the above mentioned characteristics, such teams often have to integrate different professional cultures while working under stress (Kolbe et al., 2019). Outside of the medical context, classical examples of action teams are cockpit crews, fire fighters, military teams and those working at nuclear power plants and off-shore oil platforms (St.Pierre et al., 2011). As action teams cannot profit from team familiarity and thus developed routines, they so face particular challenges in order to perform well, which I will develop later.

The teams studied in this dissertation are ad-hoc action teams of medical students without prior collaboration experience with each other. They are confronted with

cardiopulmonary resuscitation (CPR) tasks. In the next section, I summarize research related to CPR teams.

Teamwork and taskwork in cardiopulmonary resuscitation

In our research, we focused on a simulated CPR task. CPR is necessary if a patient suffers a cardiac arrest (patient unresponsive, no breathing or agonal breathing, no pulse, Soar et al., 2015). In a nutshell, in case of a cardiac arrest, a team tries to substitute circulation and assure oxygenation of the brain (with chest compressions and ventilation) while at the same time trying to restart the heart of the patient (e.g., with defibrillation and medications).

What exactly is the team challenge of a CPR task? According to Klein et al. (2006, p. 590), the task of an emergency trauma team is "to stabilize, diagnose, and treat the patient as quickly as possible. Errors or delays in this process may result in the death of the patient; quick and appropriate treatment is likely to save the patient's life".

The first necessity for teams treating a patient suffering an in-hospital cardiac arrest is to perform early CPR and defibrillate manually as quickly as possible, ideally within three minutes, according to the 2015 resuscitation guidelines (Soar et al., 2015). Time is critical, as Valenzuela et al. (1997) found that survival rate decreased by 10% per minute that CPR is delayed and 13% per minute that the first defibrillation is delayed.

Besides technical proficiency, there are also collaborative factors at play when it comes to CPR that are of special interest for my dissertation. Several studies have shown a relationship between team collaboration quality and resuscitation (Cooper, 2001; Hunziker et al., 2010; Hunziker et al., 2009; Tschan et al., 2006). Moreover, nursing staff and junior doctors (or medical students like in the population of our studies) often find it difficult to ask for help, which is essential to deal with the various tasks that need to be undertaken quickly (Soar et al., 2015).

Team performance in cardiopulmonary resuscitation

St.Pierre et al. (2011) reviewed different theoretical models of team performance and applied them to healthcare teams. They identify the following shared dimension: "They share an understanding that defines team performance as the result of how (process, throughput) a team utilizes its human and technical resources given a specific situational and task context (input factors). Results of team performance (output) in healthcare are first of all safe patient care, but also to minimize errors, working climate, and team-member satisfaction" (p. 201). This definition is in accordance with other team performance definitions in the literature (Hackman & Katz, 2010; McGrath, 1984; McGrath, 1991) by including aspects of organizational goal fulfillment, but also team member satisfaction.

Given that measuring performance in resuscitation teams has particular challenges, it was necessary to define specific team performance measures for this task. The ultimate goal of resuscitation is the survival of the patient and resuscitation performance could theoretically be measured as percent survival. However, survival after a cardiac arrest certainly depends more on the patient's condition than on the measures taken by a resuscitation team. Outcome performance measures are thus not optimal for measuring team performance for such a task. Previous research has used rather simple indicators of resuscitation performance, for example the calculation of proportion of time where teams are providing active and adequate care for the patient, or the delay until a certain action starts (e.g., defibrillation) (Castelao et al., 2011; Hunziker et al., 2011; Kündig et al., 2019; Marsch et al., 2005; Tschan et al., 2006). Although these performance measures are compatible with the resuscitation algorithm, they fall short of looking at specific actions. We thus suggest, in addition to the well-known measure, to develop a performance measure that refers to adherence to the resuscitation algorithm and measures how well the team performs resuscitation. Note that this too is a measure of team process performance and not team outcome performance.

The Adult Advanced Life support algorithm published by the European resuscitation councils (Nolan et al., 2010; Soar et al., 2015) is a guideline that defines in a very precise manner the acts that must be undertaken to treat a patient with cardiac arrest. For example, the algorithm recommends that one milligram adrenaline should be administered every three to five minutes until spontaneous circulation returns, and the first dose of one milligram should be administered after the third shock (Soar et al., 2015). These very detailed guidelines enable precise and observable process performance measures. The measures developed for certain analyses of my dissertation are based on the 2010 guidelines because they were in effect at the time of data collection (Nolan et al., 2010). Although the resuscitation guidelines are regularly adapted, the changes in 2015 do not affect the parameters we have considered for the performance coding systems. Details on the content of the performance measures used and developed for my studies can be found in the **Appendices** (from p. 215).

In the next sections, I now develop the concept of reflexivity and other similar concepts.

The concept of reflexivity

An important part of the two studies in this dissertation is that we experimentally tested the effects of a reflexivity intervention in resuscitation teams. In this section, I present the concept of reflexivity and similar concepts, and I discuss the effect of reflexivity on performance and on group processes. Note that I used reflexivity as an intervention in an experiment.

Team reflexivity is a concept developed by Michael West and defined as "…the extent to which group members overtly reflect upon the group's objectives, strategies and processes, and adapt them to current or anticipated endogenous or environmental circumstances" (West, 1996, p. 559). Reflexivity is thus an opportunity for the team to collectively review the strategies they used, their goals and their work processes. Discussion should be based on the

past and future performance in order to share the relevant information and allow adaptation (Schmutz & Eppich, 2017).

West (1996) was originally interested in reflexivity because he considered it beneficial to team performance. Initially, West introduced the concept of reflexivity as an adaptation of the reflection-in-action concept (describing that in uncertain situations, people reflect while acting) of Schön (1938) to complex decision making groups (West, 1996). He states that during such a reflection process "intentions are formed . . . courses of action can be contemplated, and decisions may be reached about contemplated actions . . . decisions may be converted into action" (West, 1996, p. 560).

In this conceptualization, reflexivity was seen as a characteristic, a way to act or a routine in a stable team (Vashdi et al., 2007). For Schippers et al. (2003), the opposite of reflexivity are habitual routines. Indicators of a reflective team are the extent it deliberately considers goals, strategies, plans, processes, or the environment, in other words, the extent to be prepared to adapt to change. Reflexivity as a characteristic (Schippers et al., 2014) is not trivial, because very often, teams do not reflect spontaneously (Hackman & Kaplan, 1974; Schippers et al., 2013). According to Gabelica et al. (2014), naturally occurring reflection might arise in a team as a response to an immediate need, for example when the team members are confused, feel lost, observe ineffective behaviors, have misunderstanding or in case of uncertainty. However, Schippers et al. (2018) suggests that such "natural breakpoints" may only exist and offer an opportunity for reflexivity for teams working under time constraints. They thus advise teams to reflect at predetermined points in time in order to support the emergence of a culture of team reflexivity.

Thus, the initial approach to reflexivity related to existing teams with collaboration experience working on complex decision-making tasks and developing a routine of recurring reflection. West developed a questionnaire (filled in by team members) to measure task (and

social, not discussed here) reflexivity (Carter & West, 1998; Facchin et al., 2006). Most research on the effects of team reflexivity have used a questionnaire approach. I discuss the findings of this line of research in the section "Measuring reflexivity" p. 68.

However, apart from reflection-in-action, Schön also defined reflection-on-action, this is reflection after the action is done, in order to learn for future similar tasks. Although not mentioned in the original chapter by West (1996), many scholars started to use the concept of reflexivity in this sense and developed interventions based on this notion. They introduced within or between task reflexivity sessions as an intervention. On the one hand, this allows for the experimental study of reflexivity (for an early overview of this research, see Moreland & McMinn, 2010). On the other hand, reflexivity as an intervention could be especially interesting for ad-hoc teams as discussed above. Such teams do not have the opportunity to build reflexivity routines because the team composition often changes, and ad-hoc teams generally collaborate only for a short period of time. However, it should be mentioned that reflexivity as an intervention requires an investment of time and effort for the team (Schippers et al., 2013).

In our studies, we used reflexivity as an "intervention to reflect" between two similar tasks in newly formed medical emergency teams.

Reflexivity, (de)briefings, and other related concepts

In this section, I distinguish between reflexivity, briefings, debriefings and other similar concepts.

Reflexivity interventions vs (de)briefings. Although reflexivity was used as an intervention in experimental team psychology, the main idea (teams reflecting before, between or after tasks) has been long known and practiced as briefing and debriefing. From a historical point of view (Eppich & St.Pierre, 2011; Tannenbaum & Cerasoli, 2013) briefing and debriefing are considered standard practice in many high-stakes environments (e.g., civil aviation,

military command), and emerged later in the field of medicine, the latter is of particular interest for this dissertation Nowadays, briefings are very often done in many medical fields (Kolbe et al., 2015).

Generally, one can define briefing as the process of spending time before (briefing) or after (debriefing) a specific team task to discuss aspects of the team process (Allen et al., 2018; Bedwell et al., 2012; Eppich & St.Pierre, 2011; Schmutz & Eppich, 2017, 2019; Vashdi et al., 2007), with the idea to facilitate team and individual learning while discussing something that is still fresh in mind or imminent (Leonard et al., 2004). Whereas in debriefings teams or individuals use post-performance communication as a basis for learning and improvement, in briefings, they use pre-performance communication for the same goals.

Briefings have effects on group processes and emergent states (definition in section "Reflexivity, group processes and emergent states" p. 566). Team debriefings help group members to develop a shared mental model of a task and a vision of their work in the future (Smith-Jentsch et al., 2008), similar effects hold for briefings. For example, in one study, teams in the briefing condition showed higher agreement regarding the patient's state (indicating a shared mental model) than teams in the non-briefing condition (Steinemann et al., 2016). A pre-briefing intervention for nursing students prior to a CPR task also enhanced psychological safety and reduced anxiety (Roh et al., 2018). Briefings have also been found to improve patient safety in surgery (Paull et al., 2010), to help teams to identify potential defects in the operating room, (e.g., equipment, instrumentation, communication or safety issues, Bandari et al., 2012), to improve non-technical skills, (like situation awareness, Savoldelli et al., 2006), and to increase information sharing (Goldenhar et al., 2013).

Briefings have an impact on performance as well. One study showed increased CPR performance for teams attending a briefing session (Roh et al., 2018), and operative debriefings reduced surgery delays (Porta et al., 2013). Smith-Jentsch et al. (2008)

demonstrated a performance increase with self-correction briefings and debriefings led by trained facilitators. According to the meta-analysis made by Tannenbaum and Cerasoli (2013), teams who engage in debriefs perform better than control groups without debriefing. The authors estimate that well-conducted debriefs can improve team and individual performance by approximately 20 to 25%. “Well-conducted debriefs” are debriefs with sufficient structure and guidance, for example with the help of multimedia. Salas et al. (2008) listed 12 recommendations to conduct effective briefings, for example “support feedback with objective indicators of performance” (p. 520). The OSAD tool (for Objective Structured Assessment of Debriefing) provides a way to evaluate the quality of briefings with seven dimensions, notably *learning environment*, *learner engagement* or *diagnosis* (Arora et al., 2012). An effective debriefing should maximise learning (Arora et al., 2012; Hull et al., 2016). The question of quality therefore plays a role in the effectiveness of briefings, and this is also the case for reflexivity, I will address this point in section “Quality of reflexivity” (p. 49).

Given the similarity between reflexivity as an intervention and briefings, it is important to look at how the relationship between them has been discussed. Schmutz and Eppich (2019) consider that team reflexivity, when happening after action, represents the classical debriefing. Allen et al. (2018, p. 511) make the link between debriefing and reflexivity by stating that debriefs are intended to induce team reflexivity. However, in contrast to briefings and debriefings that may take many different forms, reflexivity interventions that are based on West’s initial concept (West, 1996) may be more specific in their process than briefings, because the reflexivity concept suggests a specific structure of the reflective discussion. This structure includes looking back at prior performance, evaluating positive and negative aspects of the team process during prior performance and engaging in

specific action planning (Konradt et al., 2016; Schmutz, Kolbe, et al., 2018). We thus consider reflexivity (interventions) as a specific kind of debriefing.

Other concepts similar to reflexivity. Besides the similarities of reflexivity interventions with (de)briefings, there are other concepts and interventions similar to reflexivity. For example, in the team training literature, interventions known as after-action review or after-event review (Villado & Arthur Jr, 2013; West, 2012), self-correction training (Kolbe et al., 2015; Weaver et al., 2014) or guided team self-correction (Smith-Jentsch et al., 2008) have similarities with reflexivity interventions in that they all describe interventions where teams engage in a reflective communication outside of regular task-work. Schippers et al. (2018, p. 6) indeed note many other "constructs that relate to or assume reflexivity take place" and, in addition to the above mentioned concepts also add action learning/reflection, action team learning, team learning, execution monitoring, constructive confrontation norms, critical (thought) norm, task debate, ambidexterity and heedful interrelating as bearing similarities with reflexivity. Allen et al. (2018) also referenced different terms used to designate debriefing activities, like "huddles" or "hot wash".

In summary, all these concepts have basic similarities ("take a time-out to think about your actions"), but the names vary by field. For example, after-action review is mostly used in a military, fire service, and first responder context; after-event-review in fire service and first responders contexts; briefing and debriefing in an aviation and healthcare context (Allen et al., 2018).

I conclude that there are many similar and even equivalent denominations and concepts to reflexivity in the literature. For research, this may not be favorable and it can impede understanding by creating different research streams (Schippers et al., 2018). However, many of the concepts were not developed by researchers, but were developed and used by practitioners in different fields and context. It is an illusion to try to get people in

many fields to abandon their specific terms, team research will have to live with the different denominations. It is, however, important that the aim and the structure of the reflection session are considered when comparing findings from different fields and backgrounds (Moreland & McMinn, 2010). So far, I did not discuss the potential effects of reflexivity on performance outcomes, this is the theme of next section.

Does reflexivity enhance team performance?

Again, we have to distinguish studies that investigated reflexivity as a team process habit from studies using reflexivity as an intervention. Reflective teams are considered as performing better (Cotard & Michinov, 2020; Villado & Arthur Jr, 2013; West & Sacramento, 2011) and work together more effectively (Hedman, 2016). High reflexivity habits were also found to be positively associated with other indicators of team performance including innovation, conflict management, error-reduction and team efficacy (Schippers et al., 2015; Tjosvold et al., 2003; Vashdi et al., 2013; Villado & Arthur Jr, 2013). Reflexivity habits also influence member satisfaction and commitment to one's team (Schippers et al., 2003), increase openness for communication and cohesion (Villado & Arthur Jr, 2013), enhance information sharing and learning (De Dreu, 2007), as well as feedback seeking (West, 2000), and positively influence innovation (Chen, Liu, Zhang, et al., 2019; Marques-Quinteiro et al., 2019; Schippers et al., 2015; Tjosvold et al., 2003). Thus, reflexivity does not only improve outcomes directly related to the task at hand.

Historically, the study of Carter and West (1998) is one of the first studies that shows positive effects of reflexivity used as an intervention on performance for television production teams. Later, Konradt et al. (2015) and Rasker et al. (2000) also showed positive effects of a guided reflexivity intervention on performance. Reflexivity intervention was shown to increase innovation as well (Reiter-Palmon et al., 2018). Chen et al. (2016) demonstrated a beneficial effect of reflexivity intervention on well-being. Compared to a control condition,

teams participating in a reflexivity intervention experienced an increase in team control and team support, which resulted in a significantly greater reduction in the emotional exhaustion, cynicism and reduced personal achievement dimensions, considered as indicators of burnout.

In their research, Schippers et al. (2014) argue that team reflexivity can help counteract team information-processing failures by increasing information exchange. Reflexivity favors the decision-making process in teams working in a demanding, knowledge-intensive environment, notably medical teams. The authors conclude from their review that teams can be taught to improve their effectiveness by reflecting at strategic moments and this process contributes to team learning.

If we focus on the medical domain, reflective teams perform better too when regarding quality of performance (Urbini et al., 2018). Reflexivity interventions have also been shown to improve performance (Kündig et al., 2019) and to provide an opportunity for team learning (Vashdi et al., 2007). They also helped to reduce fall rates[1] in a hospital (Reiter-Palmon et al., 2018). In the same vein, with another medical team example, Vashdi et al. (2013) showed that a reflexivity intervention focused on the task helped to shorten surgery duration and reduced the risk of errors by encouraging mutual assistance behaviors or workload sharing.

However, some mixed results have tempered the enthusiasm of researchers regarding reflexivity. Indeed, some studies found a negative relationship between reflective habits and team effectiveness (De Dreu, 2002, 2007). Other research also suggested that reflexivity might be efficient only under certain circumstances, for example only for low-performing teams (Schippers et al., 2013; Schippers et al., 2015). Previous research also showed no significant effect of a reflexivity intervention (Wiedow & Konradt, 2010), or no effect of

[1] "A fall is an unintended downward displacement of a patient's body to the ground or other object" (Reiter-Palmon et al., 2018, p. 1).

reflexivity as tool to support diagnosis (Lambe et al., 2018). There is therefore an inconsistency of results about the effects of reflexivity in research.

Throughout this theoretical background, I will discuss possible explanations for the inefficiency of reflexivity. But the main arguments in the papers lie in the idea that a necessary level of depth of reflexivity is not achieved or that people do not engage in reflexivity process at all (Moreland & McMinn, 2010). In empirical studies with reflexivity as an intervention, authors also suggest that participants cannot identify their mistakes during reflexivity (Gurtner et al., 2007; Lambe et al., 2018). There is finally the problem of fear of mutual criticism that keeps them from discussing during the intervention (Kolbe et al., 2015).

In our studies, we focused on potential performance improvement as outcomes after reflexivity.

Reflexivity and team learning

We will now turn to another positive effect of team reflexivity. The reflexivity process (used as an intervenion and as a workig habit) facilitates information processing in teams and is seen as strongly related to team learning (Schippers et al., 2014; Schippers et al., 2018; Vashdi et al., 2007; West, 2000). Indeed, some authors characterize team learning as composed of "reflection and action" (Edmondson, 1999, p. 353), thus including reflection as part of their definition of team learning. Other authors like Wilson et al. (2007) also consider reflection as a prerequisite to group learning and state that "group discussion about performance discrepancies that reflect past, present, and future scenarios increases the probability of group learning" (p. 1047).

It is evident that a clear distinction between reflection and learning has not been made in the literature (and may be difficult). Schippers et al. (2008, p. 1) talk about "vital self-regulatory process" to qualify the team reflexivity process and consider that conscious

reflexivity also has self-regulatory functions and thus enables teams to learn more from previous experiences.

In summary, many authors see a close link or even an overlap between reflexivity and team learning. In the next sessions, I will summarize research related to the process of reflexivity.

The reflexivity process: When, how and about what?

In this section, I will discuss the contents of the reflexivity process and its ideal process as well as the moments when it can take place.

The structure and content of a reflexivity session – What should be discussed and how. Carter and West (1998), when referring to reflexivity as style of a team, distinguished between social reflexivity and task reflexivity. As said above, we will not discuss social reflexivity. In their scale to measure task reflexivity, items refer to reviewing objectives, strategies, and team processes, as well as the preparation to adapt to evolving circumstances. Reflexivity thus provides an opportunity to discuss different aspects, such as "what happened, why it happened, how things might have turned out differently, and whether task performance could be improved" (Moreland & McMinn, 2010, p. 64). During reflexivity, team members can identify discrepancies between actual and desired circumstances and identify necessary adaptions in order to reach their goals (Schippers et al., 2018; Schmutz & Eppich, 2019; West, 2012). In a similar vein, Schmutz and Eppich (2019) describe three main principles on which reflexivity is based: "(i) Looking back and seeking information (i.e. Can we summarise what we have done so far?); (ii) evaluating information in order to acquire a deeper understanding of processes, situations or actions (i.e. Why did this treatment work/not work?), and (iii) looking forward by planning the action(s) to be taken based on the evaluation made previously (i.e. What are our next steps? What will we do differently next time?)" (Schmutz & Eppich, 2019, p. 1). Gabelica et al. (2014) consider reflexivity as a proactive feedback

process that should include three components: (a) Evaluating or reviewing performance or strategies, (b) looking for alternatives and (c) making decisions. During feedback, if those three steps are addressed, the reflexivity cycle is completed.

Schippers et al. (2018) suggest that team reflexivity is powerful and helpful for teams to set goals correctly. Reflexivity also helps teams to achieve goals by creating a high awareness of the gap between the current situation and the desired results. Schippers et al. (2014) propose that a reflexivity intervention might be especially useful during transition moments in the team's life, i.e. during milestones, midpoints or naturally occurring changes and interruptions.

Temporal aspects – when reflexivity occurs. It is important to specify when reflexivity can and should occur, as a working style as well as an intervention. From a chronological point of view, Schmutz and Eppich (2017) define three moments when reflexivity can occur: pre-action team reflexivity (briefing *before* patient care), in-action team reflexivity (deliberations *during* active patient care), and post-action team reflexivity (debriefing *after* patient care). However, it should be pointed out that the idea of reflexivity during ongoing action is fairly recent and that reflexivity has long been considered to take place mostly after the action (Schmutz, Lei, et al., 2018).

Pre-action reflexivity represents an anticipation in the team of upcoming activities. When occurring before an action, reflexivity allows to save time and to prepare optimally for the upcoming action. Teams can thus anticipate upcoming events, and increase their situational awareness by predicting possible scenarios in advance. In-action reflexivity happens along with the action (action that is thus frequently not prepared for) and implies immediate consequences. It includes assessing whether the team focuses on the right issues and executes the task correctly; and whether the team therefore concentrates on the action in progress in order to adapt it. A team can confirm if actions are correct or adapt them if

necessary in a deliberate way. In-action team reflexivity helps teams to stand back mentally from the task at hand to handle emerging information (Schmutz & Eppich, 2019). According to Schmutz, Lei, et al. (2018), in-action reflexivity is necessary for teams facing extreme environments that are complex, dynamic, uncertain and changing, to help them dealing with situations that suddenly differ from routine. It is therefore particularly relevant for medical teams. The authors also state that this type of reflexivity often emerges as short episodes (up to a few seconds in some cases) due to lack of time, varying between short time-out and absolute interruptions. It is possible that only some of the team members may be able to take part to the discussion if the others are busy with another more urgent or demanding task. In their study, Schmutz, Lei et al. showed that teams engaged more in in-action reflexivity in non-routine tasks than in routine task, and that in-action reflexivity tends to increase within the course of the task.

In my dissertation, I focus on the third type: post-action reflexivity intervention, happening after action. In post-action reflexivity, teams can explicitly review and evaluate during the intervention what went well or not in their past performance, and they also have the opportunity to plan for future similar performance. Indeed, post-action reflexivity incites to self-analyse of own behaviors (Schmutz & Eppich, 2019).

For the specific example of medical teams, whose work often requires fast immediate action because any delay can harm the patient (Hazinski et al., 2015), post-action reflexivity is indeed indicated during emergencies (Kündig et al., 2019). In particular situations, teams could nevertheless have the opportunity to discuss before action, for example while waiting for a patient (Su et al., 2017), to prepare their handling of the situation.

Reflexivity in relation to transition and action phases. Reflexivity can also be thought in relation to the framework and taxonomy of team processes proposed by Marks et al. (2001). This theory presents a model of team process where action phases and transition phases

succeed one another. Teams engage in taskwork (what they are doing) during the action phases and use the transition phases for teamwork (how they are doing the task). Transition phases are “periods of time when teams focus primarily on evaluation and/or planning activities to guide their accomplishment of a team goal or objective”, and action phases are defined as “periods of time when teams are engaged in acts that contribute directly to goal accomplishment (i.e., taskwork)” in the theory of Marks et al. (2001, p. 360).

Tesler et al. (2017) propose that reflexivity begins during the transition phase, allowing the team to consider the performance and outcomes from previous episode, and adapt strategy for next action. Konradt et al. (2016) developed a model where team reflexivity can occur during both transition and action phases. Their model considers transition phases the optimal time for reflexivity and action phases compatible with adaptation. After a performance episode, feedback represents an important source for reflection: it provides information about where improvements and changes are necessary.

Reflexive behaviors can change during a task. In their study, Gabelica et al. (2014) observed team communication and noticed that reflexivity behaviors occurred more frequently during the action phase and increased over time. During transitions, reflexivity behaviors were less frequent and regular. Not only can reflexivity vary during the task execution, but reflexivity habits can also evolve over time in line with the team’s evolution: West (1996) considers that the longer a team is together, the less its members will be aware of environmental changes and the less reflexive they will become. It is therefore important to consider solutions to encourage teams to engage in a process of reflexivity.

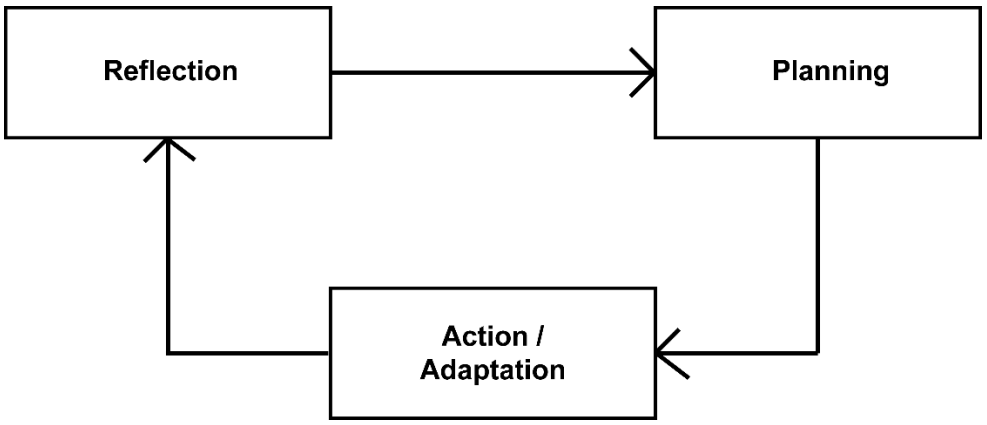
In the next section, I discuss the three steps of the reflexivity process.

The three parts of reflexivity. Reflexivity process is composed of “three stages or components: Reflection, planning, and action or adaptation” (West & Sacramento, 2011, p. 907) that are illustrated in figure 1. Reflection is the identification and evaluation of task-

related issues like objectives, strategies, and processes (Konradt et al., 2015). During the planning phase, action plans are developed in order to achieve the goals identified in the reflection stage (Chen, Liu, Yuan, et al., 2019). Finally, adaptation is composed of the "goal-oriented behaviors being arrange toward to reducing discrepancies between current and desired states" (Chen, Liu, Yuan, et al., 2019, p. 9). We now move on to the dissection of these three individual steps. In practice, it is nevertheless difficult to distinguish between the three components because the groups move from one step to the other constantly (e.g., micro-adjustments); they thus are more interrelated than sequential (Widmer et al., 2009). Similar stages of reflexivity with different denominations can be found, like in the work of Tjosvold (1991) who uses the terms: reflect and learn, prepare and plan, implement and experiment.

Figure 1

The reflective process according to West (2000, p. 4).



Reflection. Reflection refers to "the consideration of work-related issues" (Widmer et al., 2009, p. 3). According to West (2000, p. 4), reflection includes "behaviors such as questioning, planning, exploratory learning, analysis, diversive exploration, making use of knowledge explicitly, planfulness, learning at a meta-level, reviewing past events with self-awareness, and coming to terms over time with a new awareness". Information elaboration is also a basic component in reflection (Schippers et al., 2007), more specifically information

seeking and information evaluating (Konradt et al., 2016). Reflection as the first step not only includes consideration of team goals, strategies, and processes (West, 2000), it also comprises reflection based on the nature of the problem that the team faces (Moreland & Levine, 1992).

Reflection can occur before, during or after task execution (West, 2000). Reflection before task execution is characterized by joint consideration of team goals, strategies, processes, and also the nature of the problem that the team faces (Moreland & Levine, 1992). Reflection during task execution involves reviewing whether the team is still on track, whether the right problem is being addressed, and whether things are done correctly. Reflection after finishing the task is a look back and evaluate how the achievment was executed (Schippers et al., 2007). Reflection thus helps to recognize how current ways of working or making decisions might have become inadequate due to environmental changes (Tjosvold, 1991). West (2000) suggests that reflection phases vary in depth, influencing the quality of the process (I discuss further the concept of quality in section "Quality of reflexivity" p. 49).

Planning. During the reflection phase, previous courses of action are contemplated and intentions are formed; but before resulting in an action, some planning is necessary (West, 2000). During this second step, goals are thus defined as well as the plan and ways to achieve them (Widmer et al., 2009). Planning behaviors can be goal setting, creating an open environment, sharing information related to task requirements, clarifying each team member's roles and responsibilities, discussing relevant environmental characteristics and constraints, prioritize tasks, and determining what types of information the team members possess (Stout et al., 1999, p. 62).

Planning can take place before or during the action phase (Widmer et al., 2009). Stout et al. (1999) report that teams who planned conscientiously prior to performance developed a greater shared mental model of other team members' informational requirements. Team

members were thus able to communicate relevant information to each other before an explicit request during high-workload periods, and they made fewer errors during these stressful periods. The authors conclude that teams that plan better have a higher performance, even with an increased level of workload (e.g., multitasking), and can face several challenges at the same time. This is not negligible, especially for medical action teams (definition section "Action teams" p. 29), because if the team cannot adapt to important task changes, this can reduce their performance. In their study, Weldon et al. (1991) demonstrated that group planning (i.e., "what should be done, who should do it, and how and when it should be done", p. 536) and strategy changes moderated the relationship between goal setting and quality of performance. These results show the importance of this intermediary step for the reflexivity process.

***Action and adaptation*.** Reflection about prior performance and planning alone are not sufficient to lead to changes within a team: Adaptation is necessary as well (West, 2000). It is through reflection that decisions are made, but decisions have to be turned into new actions (West, 1996).

"Actions refers to the goal-directed behaviors relevant to achieving the desired changes" (West, 2000, p. 6). It the phase where teams can implement new objectives, strategies, processes, organizations or environments elaborated in previous phases. This final step allows to assess whether the applied changes are satisfying or whether a new reflexivity cycle should be undertaken ("behavior regulation cycle" Konradt et al., 2016, p. 165). Once actions are implemented, they influence the situation and bring new information to the team, and this information should ideally be integrated in a new reflexivity cycle as an ongoing process until the goals are reached. This spiral of the reflexivity cycle will bring the team to effectiveness (West, 2000).

I detail different moderators of reflexivity in the following sections.

Individual or group reflexivity. Reflexivity exists as an individual or group process (Gurtner et al., 2007). The main difference between individual and group reflexivity is that the second type involves discussion and behaviors that can be observed (Schippers et al., 2018). Our work focused exclusively on the second type (participants in groups of three). The advantage of a group reflexivity session is that each member can contribute to the discussion and share knowledge with the group (Gurtner et al., 2007), which allows the group to discuss new strategies and plan commonly during the reflexivity process.

On the one hand, individual reflexivity might be relevant for group leaders because it gives them the opportunity to plan the task ahead. On the other hand, individual reflexivity does not allow leaders to have an overview of the members' points of view, such as the difficulties previously experienced or their suggestions for improvement. It also reduces the time available to communicate with the team during the reflexivity session if the leader uses part of this session individually (Gurtner et al., 2007).

Surprisingly, Bolinger and Stanton (2014) found that individual reflexivity resulted in a greater improvement in decision-making tasks than group reflexivity. Gurtner et al. (2007) have also demonstrated a beneficial effect of reflexivity greater for the individuals than for the groups in an air-traffic control simulation task. Individuals implemented better strategies in an air-surveillance task. These surprising results might be due to familiarity within the team. Familiarity may indeed influence the quantity and quality of information discussed in the group (Okhuysen, 2001).

For Moreland and McMinn (2010), the “normal” reflection is an individual process, and individual and group reflection might inhibit each other. The authors suggest that individual reflection could be the more important for performance and which cannot happen during group reflection.

Quality of reflexivity. The quality of reflexivity is classically considered to be a question of depth, especially for the first step "reflection" (Schippers et al., 2007). Reflection can vary in three level of depth: shallow, moderate and deep (West, 2000). The first level shallow, consists on reflecting on issues closely linked to the task in hand. There is an example of a question that could be asked in a team that illustrate this level: "Do people think we communicate information about patients well within this team?" (West, 1996, p. 907). The moderate level is a reflection more critical toward the task, for example when providing alternative or improvement solutions. A representative example could be: "Let's think about some alternatives in terms of how we could best improve our product design processes" (West, 1996, p. 907). The last level, deep, happens when there is an evaluation of the norms and values of the organization or the group. Again, West proposes an illustrative example of statement: "So we agree that our communication about patients is hampered by professional divisions within the team" (1996, p. 907).

In 2002, Argyris was using the terms "single-loop-", "double-loop-" and "triple-loop learning" to define three levels of learning that are similar to the concept of three levels of depth (levels of regulation in learning cycles). This pattern defined by Argyris is meant for individuals, especially leaders, to increase their own awareness and competence as well as those of their colleagues in problem detection and correction.

Groups rarely reach a high degree of reflection (Bolinger & Stanton, 2014; Gabelica et al., 2014; Kihlgren et al., 2015; Lauritzen et al., 2016; Moreland & Levine, 1992; West, 2000). Moreover, Moreland and McMinn (2010) invoke the issue of quality to explain why reflexivity is not always effective. They consider that some teams do not reach a sufficiently deep level of reflexivity and as a result cannot benefit from it. Swift and West (1998) actually suggests that at least moderate reflection is needed, and deep reflection is the most useful for team learning.

Distinguishing between quantity and quality of reflexivity. Konradt et al. (2016) assume that the mixed results of reflexivity on performance could be due to the fact that reflexivity effectiveness has always been thought of in terms of quantity and not quality. Indeed, without distinguishing between both aspects, one cannot identify the portions of superficial reflection and those that analyze the situation and its causes in detail (i.e. quality). For example, reflecting for a long duration only on a shallow level represents a risk to miss essential elements that influence task achievement or make erroneous conclusions. Otte et al. (2018) also suggest that a greater reflexivity level can be characteristic in teams who only talk about information already known to each member, which is insufficient for a satisfying reflexivity process. A deeper reflection would help teams to analyse prior activities in a more accurate way and develop a functional plan for future performance. We observe that quality differences might thus influence the learning process as well as the team performance, and this is a more important matter than quantity. In their study notably, Otte et al. (2018) demonstrated that good quality of reflexivity in team predicts performance improvement, whereas a high quantity of reflexivity in team was negatively related to performance improvement.

However, both concepts of quantity and quality are not mutually exclusive: they can complement and replace each other (Otte et al., 2018). A team readdressing a specific topic can for example increase its depth of reflexivity, even if this might need more time (and thus more quantity). The authors suggest that in the long term, reflexivity quantity eventually leads to quality.

The reflexivity process can be positively or negatively influenced by group dynamics. We present a few of those possible influences in the next sections.

Reflexivity and information processing. The reflexivity process involves information processing and decision-making (Konradt et al., 2016; Schippers et al., 2014). This means that

group dynamics can interfere with the reflexivity process, prevent a profitable discussion by limting information processing or inducing information-processing failures, and even lead to errors. Notably, certain heuristics may limit information elaboration, voicing early concerns, as well as the revision of decision making, as it is the case with anchorage bias or confirmation bias. Anchoring bias explains that when something is anchored, it is difficult to see it differently, and confirmation bias is the tendency to find evidence to confirm one's own perspective on a problem (Kolbe et al., 2019). In this section, I mention some other possible sources of interference with information processing and reflexivity, and I then present a theory considering reflexivity as a solution to information processing failures.

The opinion of other group members can influence information processing and decision-making. With four studies, Mojzisch and Schulz-Hardt (2010) demonstrated that knowing the other team members' preferences at the beginning of the discussion impaired the resolution of hidden profiles tasks (decision-making tasks). This was further explained by people paying less attention to the information shared in the discussion when they are already aware of other members' choices.

We can also mention the phenomena of groupthink and the majority rule. The problem of groupthink is that decisions are made with little dicsussion and people dare not speak up about their concerns, because they aim towards reducing conflicts and reaching consensus quickly (Janis, 1972). This happens in groups with high cohesion who want to respond to pressure without tension and make decisions without critical and analytical discussion. Regarding the majority rule, Kolbe et al. (2019) consider that there is a general belief that deciding according to this rule is fair and efficient. Even if this way of deciding needs indeed less time and prevents decision inertia, it still has disadvantages: If a team decides to respect the majority rule, team members tend to focus on compromise and omit to share and discuss other opinions.

During a discussion, group members usually favor discussing information that is already commonly held (Larson Jr & Egan, 2018; Stasser et al., 1995). If participants do not share the individual information they hold, then all relevant elements cannot be addressed in the discussion, and this might impede the functioning of reflexivity and the final decision. Groups also tend to agree on a strategy and make decisions early in their work process, and not question them before the "midpoint transition" of the work (Gersick & Hackman, 1990) or without the occurrence of a major event (Argote et al., 1989). In fact, groups generally do not question their work routines (Schippers et al., 2014). Hackman and Wageman (2005) therefore recommend to incite teams – notalby through management – to engage in a reflexivity process after teammates acquired some experience with the task.

Reflexivity can thus result in more strategy revision and help teams to focus on their experience and benefit from it. However, only encouraging teams to engage in team reflexivity might not be systematically beneficial, because some teams tend to focus on irrelevant topics (Gurtner et al., 2007; Lambe et al., 2018). Teams should be made attentive to the identification of important topics that must be addressed during reflexivity sessions (Otte et al., 2018).

Reflexivity should not be solely depended upon as the the exchange of information. It can also be a tool to prevent information processing failures. Based on a literature review, Schippers et al. (2014) propose a model where team reflexivity has a critical role in the information-processing activity. It can thus help to prevent and reduce some biases in decision-making through conscious reflection, which results in more information procession, in order to reduce errors and enhance adaptation. More precisely, reflexivity could help to avoid: (a) the failure to share and integrate relevant information, (b) failure to elaborate and derive implications from information, and (c) failure to revise and update conclusions (Schippers et al., 2014, p. 731). Moreover, Schippers et al. (2013) show that members of

reflective teams are encouraged to value their own opinions and to practice questioning rather than functioning at a habitual and routine level.

Conditions favoring reflexivity. In order to understand the reflexivity process, it is necessary to define which conditions are favorable or unfavorable to the process, like working climate, task characteristics or leadership. Working context can favor reflexivity and influence the motivation of teams to engage in reflexivity (Schippers et al., 2018). Acutally, as noted by Widmer et al. (2009), psychological safety is an essential condition for members of a group to dare to take risks, which also means discussing sensitive topics during reflexivity. Psychological safety describes "perceptions of the consequences of taking interpersonal risks in a particular context such as a workplace" (Edmondson & Lei, 2014, p. 23). Group members are more likely to engage in reflexivity activities in a cooperative climate if they do not fear peers' reactions (Collins & Smith, 2006). If group members trust each other, they will speak up and express criticisms without worrying that other team members would hold the possible error against them (Edmondson et al., 2007). This represents therefore a particular challenge for new or ad-hoc teams who do not always know the other team members, like it is the case in the medical field.

Moreover, according to Schippers et al. (2018), team members often implicitly assess the importance or the emergency of an issue to decide whether to discuss it or not. A problem perceived as less important or urgent is thus less likely to be addressed. The authors even go so far as to say that the topic of speaking up is similar to the topic of whistleblowing[2], i.e. a calculation between costs and benefits. Pointing a problem might mean "put oneself at risk" (for example being accused of disturbing the work place, of being a spy, of causing a

[2] Whistleblowing is "the disclosure by organization members (former or current) of illegal, immoral, or illegitimate practices under the control of their employers, to persons or organizations that may be able to effect action" (Mesmer-Magnus & Visweraran, 2005, p. 278).

nuisance, or being ignored, according to Mesmer-Magnus & Viswesvaran, 2005; Schippers et al., 2018). It is therefore possible that some topics may be deliberately avoided in the discussion.

Leadership styles can also influence reflexivity in teams. Schippers et al. (2008) were able to demonstrate a model where transformational leadership is related to more reflexivity. Transformational leaders "transform" and motivate their teams, they stimulate them to go beyond self-interest, to achieve results above expectations, sometimes even altering the team members' morals, values or ideals (Bass & Riggio, 2006). To return to the study of Schippers et al. (2008), transformational leaders led their teams to adopt a shared vision, and this shared vision in turn increased reflexivity in teams, which led to higher team performance. There are also interesting results regarding other types of leadership. For example, Hirst et al. (2004) showed that facilitative leaders ("leaders who promote respect and positive relationships between team members, productive conflict resolution, and open expression of ideas and opinions" Hirst et al., 2004, p. 312) simulate reflexivity in their team. In her study alongside healthcare teams, Somech (2006) found that participative leadership (when decisions are made jointly by the leaders and their team, Somech, 2006) was positively associated to reflexivity. She also found the same relationship for directive leadership ("providing the team members with a framework for decision making and action in alignment with the superior's vision" Somech, 2006, p. 35), but only for high functional heterogeneity teams. The behaviors of leaders influence the degree of reflexivity within their teams, but leaders may also deliberately encourage their teams to apply reflexivity. Indeed, West (2012) suggests that leaders can and should promote reflexivity moments in their team, especially during crisis or difficulties, because this might lead to "valuable learning" (p. 260).

Reflexivity might also be beneficial only under certain circumstances related to the task. For example, Schippers et al. (2013) showed that reflexivity was most beneficial for

teams regarding learning and future performance when prior performance was weak. Schippers et al. (2015) also demonstrated more innovation after reflexivity when prior performance was low. West (1996) suggested that reflection may be more useful when the task is more complex or less clearly defined. I have presented conditions that may hinder or encourage reflective behavior. I will now present some explicit interventions that aim to encourage reflexivity.

Interventions to promote reflexivity. Interventions designed to promote and increase reflection behaviors in teams exist. They can be small or simple, but they can be sufficient to make an important difference for teams. For example, interventions can help to establish a collaborative climate that favors reflection of all information available at the moment (Schippers et al., 2018).

Konradt et al. (2015), propose a study with a guided reflexivity intervention that aims to improve team reflective behaviors. They simply use written instructions asking the team to (a) reflect on expert knowledge (b) review based on their first performance (c) reflect on alternative task strategies and use of expert knowledge for future tasks, and (d) plan how to implement these strategies (p. 783). Groups that were in the reflexivity condition showed more reflective behaviors and improved their performance more than groups without. This intervention is actually similar to the one we used in our studies. We also had a self-guided reflexivity session where participants referred to an instruction sheet with questions to discuss without an external person leading the reflexivity (for example “what have you done sufficiently or well in terms of leadership, communication, and coordination?”).

Schippers et al. (2014) also think that creating artificial milestones, taking time-outs, and creating a culture of reflexivity can provide adequate opportunities and enhance team reflexivity. Explicitly including team reflexivity in the team's routine and processes is a way to facilitate reflective behaviors and could in the long term increase reflective habits. The

authors also propose to establish organizational and team practices in order to identify and avoid cognitive shortcomings (called "cognitive repairs") or team information-processing failures. They bring up the idea to implement protocols and cognitive repairs in organizations. For example, they give a concrete and easy example, the "five why". It is simply asking "why" five time in a row in the discussion to go beyond surface reasoning and avoid jumping to conclusions. Cognitive repairs can also take the form of medical protocols in the healthcare world.

These simple interventions are two very simple examples to encourage teams not to stop reflection at a superficial level and to engage in a complete reflexive process. They are very easy and inexpensive to implement, and operate on the principle of making the team explicitly attentive to the opportunity of engaging in reflexivity. The first study of Konradt et al. (2015) used an intervention similar to the reflexivity instructions we used in our studies.

In the next section, I present the relationship between reflexivity and group processes and emergent states.

Reflexivity, group processes and emergent states

Reflexivity as a style and an intervention has effects not only on performance, but also on the group process and on group emergent states[3]. Although in my research, I investigate predominantly the effects of reflexivity on task and process *performance*, I shortly summarize important aspects of the relationship between reflexivity and group process or emergent states that are related to reflexivity – such as shared mental models and transactive memory systems.

[3] "Emergent team phenomena emanate from behaviors of group members, and include emergent states, . . . emergent team phenomena may be (1) *states* such as collective cognition, affect, or other relatively enduring properties, (2) *behavioral patterns* such as conversational routines, conflict resolution, or feedback seeking, or (3) *structures* such as sub-groups or hierarchies" (Waller et al., 2016, p. 565).

This section can help to understand why reflection positively influences team performance and helps to consider its potential effects on other aspects of collaboration.

The typical content of reflexivity (see section “The structure and content of a reflexivity session – What should be discussed and how” p. 41), i.e. that a team reflects on how a task should be done, can be linked to the teamwork – taskwork distinction (Marks et al., 2001). This concept distinguishes taskwork as processes related to tasks itself, whereas teamwork is related to the interaction process and includes communication, coordination, and collaboration. A related concept is “timework”, related to temporal aspects of coordination (Tesler et al., 2017). Reflexivity is part of explicit coordination, it influences emergent states such as shared mental models and transactive memory systems – which then in turn can become a foundation for implicit coordination (Burtscher et al., 2010; Cotard & Michinov, 2020; Marques-Quinteiro et al., 2013; Rico et al., 2008).

In the following sections, I present the concepts of shared mental models and transactive memory systems and relate them to reflexivity.

Shared mental models. A shared mental model (SMM) is a mental model that is – to some degree – shared or common among the team members. A mental model is a structure of organized knowledge. It is often described for specific domains, and serves as background of actions (Rouse & Morris, 1986). Because the content of mental models heavily depends on the specific domain it is applied to, mental models can best be characterized by their functions. Rouse and Morris (1986) describe the purposes of mental models as knowledge that helps *describing*, *explaining and predicting* a system. The term system here is used in a very generic way and may well be related to a task (for example CPR) or to a state (e.g., how does electricity work). Describing, explaining and predicting requires knowledge about the purpose of a system (why does it exist), is function (how it operates), the state (what it is doing) and the form (what it looks like) (Rouse & Morris, 1986).

The notion of mental models has been adapted for teams (Cannon-Bowers & Salas, 2001; Converse et al., 1993), and one of its key points is that the mental model needs to be shared. To ensure good coordination, teams need to be "on the same page", they thus not only profit from a good mental model, but this model needs to be shared. A SMM is described as a mental model that is to a certain degree overlapping among the team members. Teams with SMM have members holding similar conceptions and a common understanding of the taskwork, teamwork and timework aspects mentioned above as well as a shared understanding of "who is responsible for each task and what the different task requirements are" (Stout et al., 1999, p. 61). A kitchen staff in a restaurant knows that they have to serve the customers, they know how they will prepare the different dishes, and they know that they have to serve a whole table at the same time.

There are different SMMs, relating to different aspects. Mathieu et al. (2000) differentiate SMM into shared team mental models and shared task mental models, but shared mental models can also be about other aspects (e.g., environment, cooperation processes).

SMMs have been linked to team effectiveness (DeChurch & Mesmer-Magnus, 2010; Mohammed et al., 2010), because they help team members to anticipate other's needs and thus optimize cooperation (Grossman et al., 2017; Stout et al., 1999).

When engaging in reflection, teams explicitly discuss important aspects of mental models, and this facilitates the development of shared understandings within the team (Gurtner et al., 2007; West, 1996). "Team reflexivity can provide the mechanism by which relevant content can be more fully processed", influencing not only the understanding of a topic, but also the degree of sharedness (Tesler et al., 2017, p. 4). When team members share a mental model, they can anticipate what information is needed during debriefing (Rasker et al., 2000). There is some empirical work showing that reflexivity may enhance SMM. For example, a leader-guided "team self-correction" briefing for Navy officers improved the

accuracy of team mental models compared to a control group with "classical" briefing/debriefing techniques (Smith-Jentsch et al., 2008). The link between reflexivity and performance is believed to depend on the similarity and accuracy of SMMs (Gurtner et al., 2007). Konradt et al. (2015) also showed an effect of reflection on team perforamnce mediated by SMM.

For ad-hoc teams that cannot rely on a common history, building of a SMM might be more difficult. Particularly relevant for this dissertation is that part of a SMM can be present even before an ad-hoc team interacts. Indeed, procedures in medicine are often standardized and algorithms for treatments exist. An example is the algorithms for basic and advanced life support (Nolan et al., 2010; Soar et al., 2015) to be applied when a cardiac arrest occurrs. Such algorithms are "imported" by individuals into the ad-hoc group, and in the best of cases, when group members start collaborating, a SMM already partially exists (Tschan & Semmer, 2001). However, for this kind of group, dedicated time for reflection may be particularly important for the development of an SMM, and thus, performance.

Transactive memory system. Another concept related to shared cognition is the transactive memory system (TMS). In a nutshell, a TMS is the knowledge within the team of "who knows what", or the memory about the structure of specialist knowledge within a team. For example, in a given team, team members know whom to ask regarding how certain decisions have been taken in the past; whose birthday it is, or who can give information where material is kept. TMSs are characterized by two properties: the combination of knowledge held by each member in the group (individual knowledge or specialty domains), and a shared awareness of the skills and knowledge of each member of the team (knowing "who knows what") (Wegner, 1987). TMSs thus combine the memories of individuals into a system (Wegner, 1987). Individuals are specialized in areas according to their relative expertise, skills or previous experiences (Hollingshead, 2001). If other group members know what

others know, they can rely on their colleagues to access specific information. TMSs can only be developed if there is a specialization (individual member's knowlege) within the group. In addition to this, credibility (members' beliefs and trust concerning the reliability of other members' knowledge) is important – if other team members do not trust the person holding a specialized knowledge, they may not use it (Lewis, 2003). The questionnaire of Lewis (2003) also contains the dimension of coordination to describe TMSs. However, in my opinion, this third aspects is more a result of TMSs rather than a definitory aspect.

TMSs are related to performance. Moreland (1999) demonstrated that training people together rather than individually encouraged the formation of a TMS, resulting in a better group performance. Moreland and Myaskovsky (2000) confirm this idea by showing that TMS formation was more important than communication for the group for performance, even when participants were trained individually.

The concept of TMS is also related to reflexivity. Marques-Quinteiro et al. (2019) demonstrated that there is a full mediation effect of reflexivity on the relationship between TMS and team innovation. They conclude that TMS could be a process to push teams to engage in a reflexive action; knowing the expertise field of each members is a support to reflection and helps to benefit from it, and this can in turn help teams to reach high innovation. Cotard and Michinov (2020) conducted a study with police officers. They showed that reflexivity positively influences perceived effectiveness in intervention, and there is a partial mediation of TMS on this relation. They conclude that when teams think collectively, focusing on the task and using an integrated (coordination-related) TMS, they can build a better understanding of the task and its completion.

It seems plausible that by reflecting and collaborating, the members of a group develop their mutual knowledge of each other's areas of expertise and so develop the TMS in

the group. Indeed, group performance tends to improve over time parallel to the development of TMS, which allow the teams to benefit from each member's expertise (Lewis et al., 2007).

The goal of our second study was to identify the content of a reflexivity intervention and to observe its consequences. I therefore present in the next sections the means we used to collect data, namely medical simulations and the development of coding systems.

Methods

In this section, I present two central elements of the methods used in the two papers included in this dissertation, namely medical simulations and the development of coding systems. I also present some of the methods that have been used to measure reflexivity in other research.

Simulations

Medical simulation plays an important role in this work, because it was the main method used to collect the data for the papers presented in this dissertation.

Simulation is used as a tool for training in various fields: aviation, military and medicine, the latter being relevant to this work. In fact, this method makes it possible to explore multiple advanced medical specialties, such as respiratory medicine, emergency medicine, pediatrics, surgery, obstetric, anesthesia, endoscopy and nursing (Moorthy et al., 2005; Riley, 2008). Some skills can be trained individually, and others as a team (Moorthy et al., 2005).

As noted by Schmutz, Kolbe, et al. (2018), environments in the healthcare system can be dynamic, complex, and ambiguous. Teams are often confronted by physical, psychological, and interpersonal demands and challenges. To complete their work quickly and effectively, they must mobilize important resources to face tasks that are often characterized by high uncertainty (incomplete information, high variability among patients). Simulation-based training is a safe way to train teams to handle those complex situations and

environments (Eppich et al., 2011; Moorthy et al., 2005). Schmutz, Kolbe, et al. (2018) insist on the usefulness of simulation training to help teams to think outside the box, but to develop a critical view of ambiguous situations, in order to act accordingly to the conditions and to encourage behaviors such as flexibility.

Riley's Manual of simulation in healthcare (2008) recommend using simulations for physicians' education and training, not only for medical acts but also for non-technical skills, like decision-making or leadership skills training; simulations represent a nice opportunity to learn both skills. The challenge of physicians is indeed working with volatility, uncertainty, complexity, ambiguity, and problems with delayed feedback such as waiting for test results whilst having to make decisions, these situations should also be trained for. Medical simulations are also an opportunity to train situation awareness[4], teamworking, and communication for example (Flin & Maran, 2008). Non-technical skills are key competencies in healthcare, because a number of medical errors are believed to be due to deficiencies in non-technical skills (Bromiley, 2008). These skills have always been considered important for good medical practice, however they are often not explicitly taught in the medical or academic curriculum (Flin & Maran, 2008; St.Pierre et al., 2011). There should nevertheless be a systematic process to identify the skills required for each profession and each team (Bromiley, 2008).

Medical simulation has become increasingly popular to train medical teams in emergency tasks that are difficult to learn in vivo (Marsch, 1998; Salas et al., 2013; Sexton et al., 1998), because "learners retain 90% of skills learnt by doing rather than watching" (Hamdorf & Davies, 2008, p. 79). In order to enable learning, training must be based on

[4] "Perception of the elements of the environment within a volume of time and space, the comprehension of their meaning and the projection of their status in the near future" (Endsley 1988, p. 791).

clearly defined objectives that also allow for the definition of measurable outcomes (Lake, 2008). Moreover, a safe environment promotes learning (Huang & Dongilli, 2008).

Environment in simulation is the key: Because of practical, technical and medical reasons, it is often difficult, if not impossible, to study emergency medical cases “on the field”. This is why working with simulations has several advantages: (a) they allow to investigate emergency cases and medical events that are not planned, (b) it is possible to observe team processes from the very beginning, which is interesting as it is during the early phase of resuscitation that team building occurs (Riley, 2008), (c) one can observe the teams in the simulation without having ethically to intervene if they provide insufficient care (Hunziker et al., 2009), (d) simulations offer the possibility to train professionals without harming patients or learners (Riley, 2008) and trainees can then appreciate the consequences of their mistakes (Moorthy et al., 2005), (e) the self-confidence of learners is not endangered in such cases (Riley, 2008), and finally (f) simulations also offer an opportunity to explore patient variability (Schmutz, Kolbe, et al., 2018).

Tools are available to facilitate learning during simulations, especially technology-facilitator tools. Video recordings or checklists are frequently used in this aim, but they are also useful to provide feedback, assess performance and allow a greater realism (Chin, 2008; Moorthy et al., 2005). Medical equipment is obviously necessary in the case of medical simulations to make the situation realistic and mostly to allow the training of several different medical procedures (Chin, 2008), as well as virtual reality is a new technological tool that offers new and very realistic training possibilities (Moorthy et al., 2005). In our studies, the manikin could blink its eyes, breathe, “speak”, it could be given drugs, and its vital parameters could be monitored.

Figure 2

Manikin and simulation room used in our studies

In summary, simulations were particularly suitable for our research, both relevant to the medical field and to capture non-technical phenomena.

Simulations and reflexivity. Different authors also used simulation for their research in reflexivity as an intervention (e.g., Husebø et al., 2013; Kihlgren et al., 2015; Tesler et al., 2017) or observed as a working style (Gabelica et al., 2014; Kündig et al., 2019; Schmutz, Lei, et al., 2018). Still, reflection and debriefing are also necessary steps after the simulation to enable subsequent learning (Hamdorf & Davies, 2008; Husebø et al., 2013; Tannenbaum & Cerasoli, 2013), because "it helps people reflect on their actions and thinking" (Oriot & Alinier, 2018, p. 3).

In order to be positive, feedback should be formulated in a constructive way. It should include a phase where the trainees give their own evaluation of their performance. Then the trainers give the points that were good and those that could be improved on. They have to express themselves in a sufficiently specific way (Lake, 2008).

In our studies, we have sought to answer some of the questions or gaps raised by the research. Consequently, we had to create our own measures and coding systems since these

were new approaches. In the following section, I present recommendations for the development of coding systems.

Developing coding systems

For our research, we needed different specific coding systems that did not necessarily exist before (used in paper 2). We wanted to measure new aspects that had not been measured in previous research to our knowledge (content discussed in reflexivity, adherence to the resuscitation algorithm in simulations, discussion of aspects related to the algorithm in reflexivity, and quality of reflexivity). We thus developed our own coding systems, based on theoretical concepts while adapting to our data. We relied mainly on the technical recommendations for CPR in place at the moment of the experimentation (Nolan et al., 2010) as well as previous research regarding teamwork (Hunziker et al., 2010; Tschan et al., 2006; Tschan et al., 2011). At the same time, the development of coding systems has been guided by the practical needs of our data. For example, some teams deliberately set the defibrillator to the automatic mode, which is not at all recommended in hospital settings with professionals, so we had to take this fact into account.

We followed the different six steps defined by Tschan et al. (2018) to develop our coding system, beginning with a task analysis (step 1). We used a process similar to Hierarchical Task Analysis (HTA, Gurtner et al., 2007; Tschan et al., 2011), dividing the task into goals and subgoals. We had to define very precisely what we wanted to observe, especially to avoid going back over the coding sequence several times. Consideration should also be given to the possibility of subsequently analyzing the data collected effectively. To take a concrete example of goal analysis, for the coding system "team- and task-related aspect discussed in reflexivity" (paper 2), the main goal was obviously to examine the contents discussed during a post-reanimation reflexivity session (simulation). We defined two main subgoals, team-related communication and task-related communication, then a definition for

each subtask with detailed criteria (e.g. “coordination” for team-related communication, and “cardiac massage” for task-related communication).

Regarding the data source (step 2), we could work with the video recordings of the simulations, from which we wrote complete transcripts.

In step 3, we defined our coding units for the codes. For reflexivity content coding we used “spoken sentences” (“a sequence of a few words conveying a single thought” Weldon et al., 1991, p. 559) as unit. This solution avoids complex grammatical sentence definitions.

Otherwise our coding units were more objectively observable actions (present or absent) defined with precise behaviors, for example for “quality of adherence to algorithm during simulation” (in paper 2, e.g. ventilate with two insufflations).

Step 4 deals with the coding content. We used the framework developed by McGrath and Altermatt (2001). They propose to include the four following aspects: temporal aspect, actors involved, behaviors of interest and eventual qualifiers (quantity or intensity). If we take again the example of team- and task-related aspect discussed in reflexivity coding (paper 2), the time of the video was indicated for each spoken unit. The student saying the phrase was indicated, as well as the categories of content present. Here, the coding was nominal (present – absent), but it was quantified for the reflexivity quality indicators we have developed.

We have a coding manual (step 5, all manuals presented in **Appendices** from p. 215) for each coding system, providing a description of the coding process, details of each individual code with criteria, examples and non-examples. We have carefully updated it to each stage of the development of the coding system.

It took a long time to train the coders (step 6). First it was necessary to become familiar with many aspects of CPR, and therefore technical jargon. Several training sessions were necessary, followed by debriefing and discussion of the differences between coders and

what needed to be clarified, before reaching sufficient inter-rater reliability (<.7, moderate level according to McHugh, 2012).

Tschan et al. (2018) also bring attention to the problem of objectivity; it is sometimes difficult to decide whether an action belongs to one or the other behavior. In such cases, the context must be considered.

Performance measures for cardiopulmonary resuscitationThe papers presented in this dissertation compare team performances in medical simulation for CPR. St.Pierre et al. (2011, p. 201) reviewed different theoretical models of team performance and identify the following common dimension across the different definitions: "they share an understanding that defines team performance as the result of how (process, throughput) a team utilizes its human and technical resources given a specific situational and task context (input factors). Results of team performance (output) in healthcare are first of all safe patient care, but also error incidence, working climate, and team-member satisfaction". This definition focuses on a general and empirical objective towards which medical teams tend to work. But these measures do not give an indication of the work of medical teams: despite ideal management, the patient may not survive due to other factors and vice versa.

In our research, we needed an accurate objective performance measure to compare teams for the technical CPR task. To develop our resuscitation performance measures (paper 1: coordinative performance, time to first defibrillation, paper 1 and 2: hands-on time, paper 2: time to first meaningful measure), we relied on Tschan et al. (2006): We defined performance measures that (a) reflect actual medical performance in a particular situation, (b) can be used in the different phases of the experimentation to allow comparison, (c) are related to medical algorithms and not team collaboration or communication, to allow to test our hypotheses and to distinguish dependent and independent variables (Staw, 1975), and (d) that are not self-reported to avoid biases (Holzman et al., 1995). These measures were coded

through temporal observation (occurrence of an event at a particular time); for example, the occurrence of relevant events whilst the patient had no pulse (hands-on time) or the occurrence of the first defibrillation after the beginning of the simulation (first meaningful measure). The detailed descriptions of these performance measures can be found in the method section of both papers as well as in the **Appendix A** (p. 215).

Research has proposed ways to measure reflexivity, including quality and quantity, by using questionnaires or observation-based methods. Some of them are presented in the next section.

Measuring reflexivity

Several attempts have been made to measure reflexivity, usually as self-reports using questionnaires. Authors have focused on trying to measure the reflexive behaviors in teams, the quality of reflexivity and the quantity of reflexivity. In this section, I give a few examples of the existing instruments, first the questionnaires, then the observation.

Carter and West (1998) elaborated a scale of 16 items, where respondents indicate how their team functioned, in order to evaluate reflective behaviors in the team. The scale examines "the extent to which a team actively reviews its objectives, strategies, and team processes and is prepared to adapt them as necessary to changing circumstances" (p. 588). An example of item is "in this team, we modify our objectives in the light of changing circumstances" (p. 588). This scale also includes the dimension of how the team deals with conflicts ("conflicts are constructively dealt with in this team", p. 588). Finally, it addresses social support among members, well-being within the team and individual development of members.

West (2000, p. 14) proposes a scale to assess reflective behaviors designed for individuals, with statements such as "I rarely change my work strategies" or "I often review the methods I use to get the job done".

Otte et al. (2017) present the REMINT ('Reflection Measure for Individuals and Teams'), a measure of reflection composed of 32 items with statements such as "we often tried to find the cause of problems in our work" or "we made very detailed assessments of how successful our procedures were" (p. 304). The scales focus on quantity and quality of information seeking and information evaluation. West (2000) developed a reflexivity scale with eight questions, as for example "the tam often reviews its objectives" (p. 23). These two tools especially aim towards quantify reflexivity behaviors in a self-reported manner.

In a different approach, Schmutz, Lei, et al. (2018) developed a coding system to measure specific behaviors of in-action reflexivity by behavioral observation (and not with self-report questionnaires) during a simulation study. They were interested in the evolution of those in-action reflective behaviors during the course of actions, focusing on three aspects: "(a) Reviewing and situation assessment (looking back and evaluation of past actions and information concerning patient history), (b) planning ahead (looking forward), and (c) discussing options and inviting input" (p. 756). For each code, they provided definitions and instructions for coders, as well as concrete examples. For the code "reviewing and situation assessment" for example, they give the following definition: "Summarizing or repeating information about patient and work that has been done. Can be seen as an explicit thinking process that is said out loud. This code represents a moment of stepping back and summarizing and evaluating what has been done. The code differentiates itself from single detached expressions which can include similar information but do not have the character of stepping out of the situation (e.g. "pulse is bad", "we already provided adrenalin")" and the following example: "Ok, we have an unresponsive 6-month old boy with fever and increased heart rate. We already administered adrenaline and draw blood for the blood gas analysis and pulse is very high now" (Schmutz, Lei, et al., 2018, p. 756).

Research has also focused on measuring the quality of reflexivity, that is often viewed as a question of depth (Schippers et al., 2007). Bolinger and Stanton (2014) and Kihlgren et al. (2015) developed coding systems to describe levels of depth of reflexivity by observation. In the study of Bolinger and Stanton (2014), there are three levels described (descriptive reflection, benchmarking reflection and mega-strategic reflection). In the work of Kihlgren et al. (2015), there are five different levels of reflective utterances (from descriptive statements to critical reflection on ethical and political issues). Both studies report that participants rarely reflected in the deepest level; they rather tend to remain on a superficial and descriptive level. Actually, in the results of Kihlgren et al. (2015), participants were novice physicians and they did never reflect in the last two levels of reflexivity.

In reality, few previous studies have assessed the quality of reflexivity, and only few of those studies assessed quality by way of observation (Otte et al., 2017). This is one of the innovations of our study. We developed our own coding system with two different indicators, because we wanted to assess aspects that had not been studied to our knowledge. All the details of the measures we developed are presented in paper 2, and the coding manuals are presented in **Appendices** (from p. 215).

3. Thesis articles

This section presents the two papers included in this dissertation. The first one is a published scientific article, and the second one is a basis for one or two future scientific papers.

Because there are only two papers in this dissertation, I chose two present them together in one chapter, one after the other. For each paper I first briefly introduce the content. The reference information about the paper 1 and date of publication are presented. For paper 2, I explain the context. For each paper comes then the abstract, and finally the article itself. The first paper is presented with a neutral layout different than in the published versions because of copyright reasons. In the table 1 below, I summarize the topics covered in paper 1 and 2.

All the coding systems are presented in the **Appendices** at the end of the dissertation (from p. 215). The literature for each paper is presented after the paper.

Table 1

Summary of themes investigated in paper 1 and paper 2

	Aspects investigated
Paper 1	• Effects of reflexivity on performance – vs control group

Paper 2	• (Effects of reflexivity on performance – vs control group)
	• **Compare performance improvement in different reflexivity conditions**
	• **Engagement in reflexivity, respect of instructions**
	  ◦ Focus of discussion (according to condition)
	  ◦ Time used to reflect
	  ◦ Topics discussed related to the task
	  ◦ Reflexivity temporal structure
	• **Reflexivity process**
	  ◦ Quality of adherence to algorithm
	  ◦ Aspects of adherence to algorithm discussed in reflexivity
	• **Quality of reflexivity**
	  ◦ Scope of reflection
	  ◦ Unresolved disagreements

Paper 1

More than experience: a post-task reflection intervention among team members enhances performance in student teams confronted with a simulated resuscitation task – a prospective randomised trial

The first paper focuses on reflexivity as intervention to help teams increase their performance. Results showed that participants really engaged in short self-led post-action reflexivity, and that these reflexivity sessions enhanced basic resuscitation behaviors (better hands-on score, better coordinative performance) in ad hoc groups compared to control groups. This paper has been accepted for publication in December 2018 and was published in 2019.

Reference

Kündig, P., Tschan, F., Semmer, N. K., Morgenthaler, C., Zimmermann, J., Holzer, E., ... & Marsch, S. (2020). More than experience: a post-task reflection intervention among team members enhances performance in student teams confronted with a simulated resuscitation task—a prospective randomised trial. *BMJ Simulation and Technology Enhanced Learning*, 6(2).

Abstract

Background

Teams that regularly step back from action and deliberately reflect on their performance and strategies show higher performance. Ad hoc emergency teams with changing team composition cannot develop such habits but may engage in short postaction reflection to discuss shortcomings of past performance and potential adaptations of their strategies for future similar tasks. This study aimed to test the effect of a short postaction self-led reflective team briefing on resuscitation performance in a simulator setting in terms of three performance parameters: hands-on time, coordination between chest compression and ventilation, and defibrillation.

Methods

We performed a randomised controlled trial including 56 ad hoc formed teams of three fourth-year medical students each. All groups performed a resuscitation task, followed by a self-guided reflective briefing, based on a general instruction (n=28 teams), or an unrelated discussion session (control condition; n=29), followed by a second resuscitation task in the same team composition.

Results

Adjusted for performance in the first task, teams in the reflection condition showed higher performance gain in the second resuscitation than teams in the control condition (6.21 percentage points (95% CI 1.31 to 11.10, p<0.001)) for basic hands-on performance; 15.0 percentage points (95% CI 2 to 28, p<0.001) for coordinative performance but non-significantly lower performance for defibrillation (−9%, 95% CI −27% to –9%, p=0.312).

Conclusion

Even very short self-led postaction reflective briefings enhance basic resuscitation performance in ad hoc groups but may not influence more complex aspects of the task. We recommend including short self-led team debriefings as part of simulator training.

More than experience: a post-task reflection intervention among team members enhances performance in student teams confronted with a simulated resuscitation task – a prospective randomised trial

Introduction

Many medical emergencies are dealt with by very closely cooperating teams. Performance of such teams depends, among other aspects, on the quality of team collaboration.[1] It is thus imperative to find ways of improving team collaboration in medical emergencies. Team training is an effective means to influence teamwork; it improves outcomes such as medical performance, patient outcomes and patient safety. [2 3]

An important way to support team learning involves team members stepping back from the action, engaging in deliberate reflection about their past performance and their task-related and collaboration-related strategies and discussing strategies for future tasks.[4 5] Team training providing an opportunity to engage in such reflective phases has been found to be particularly effective.[6 7]

However, team learning does not only take place during formal training. Previous studies found that teams that develop habitual routines to regularly engage in reflection show higher performance and more innovation.[5 7 8] As developing routines presupposes membership continuity over time, these findings mostly apply to teams with a history of collaboration[9]; such temporal and compositional team stability is rare in medical settings.[10 11] Particularly in emergency situations, teams often are formed ad hoc and collaborate for only a short period of time and then disband,[12] making it unlikely to develop routines of reflection. Nevertheless, ad hoc teams can engage in short reflection before, during or after collaboration to enhance performance. Studies in other fields,[13] but also in medicine,[14–18] showed mostly positive effects of short reflection phases for ad hoc teams.

In everyday practice, medical teams are encouraged to engage in reflective discussions—often called briefings.[19] A recent overview on reflective briefings in medicine[20] distinguishes briefings by their temporal position with regard to the task. In preaction briefings, teams reflect and discuss before collaboration and reflect about the upcoming tasks and the optimal preparation of the team (eg, team timeouts in surgery[21]; huddles before starting a complex task[22]). In-action reflection takes place during an ongoing task, when the team halts all but life-saving activities and engages in a briefing with the goal of optimising present care or adaptation to changing task conditions. In postaction briefings, teams reflect after the task (eg, mortality and morbidity conferences[23]) and analyse past performance with a focus on future, similar tasks.

For emergency situations, which are the focus of the current study, preaction briefings are possible only under very specific circumstances (e.g., while waiting for a patient to arrive[24]). However, most emergencies require immediate action, and even short delays may harm the patient.[25] In-action briefings may also be very difficult to carry out, particularly if the patient needs constant attention and team members cannot interrupt the action to focus their attention on a briefing. Thus, postaction briefings seem most promising for team learning of ad hoc emergency teams. Previous research shows that reflective postaction briefings have greater impact if they are led by an expert[6 26] or by a team leader,[27] and they yield better results if performance feedback is provided.[17] As objective feedback or an expert adviser are often not available after an emergency, the question arises if postaction briefings relying on team self-guidance can be valuable. There is little empirical evidence that self-led postaction briefings improve performance.

The main goal of the current study is thus to test the effects on performance of a short self-led postaction briefing intervention in ad hoc teams of first responders performing a CPR in a simulator setting.

Primary endpoint is resuscitation performance when treating a patient suffering a cardiac arrest by teams that engaged in a short postaction briefing after a previous, similar task, as compared with teams that did not. As performance is multidimensional,[28–31] it is important to distinguish several components of performance. Based on the resuscitation algorithm,[32] we assessed basic performance (measured as hands-on or flow time), coordinative performance (coordination between chest compression and ventilation) and defibrillation performance.

Methods

Participants and setting

This prospective observational study was conducted in the simulator room of an intensive care unit (ICU) at a midsize European University Hospital. All participants provided signed informed consent. Participants were fourth-year medical students, which are a good representation of medical professionals with a theoretical knowledge base (they all had theoretical knowledge of symptoms of cardiac arrests and cardiopulmonary resuscitation [CPR]) but limited hands-on practice and limited collaboration practice for medical tasks. This work was supported by the Swiss National Science Foundation, grant #149734.

Participants were assigned to teams of three according to their scheduling preferences. Because all medical students were part of the same cohort, members working as a team may have known each other, but they did not have common working experience. The students were blind to the goal of the study.

The patient simulator used consists of a high-fidelity manikin that was controlled remotely by the operator (Meti, Medical Educational Technologies, Inc). It has realistic physiological features, including the ability to talk (through a connection to the control room). The simulator room was equipped as a standard ICU patient room, with one wall as a one-way mirror that allows the observation from the control room.

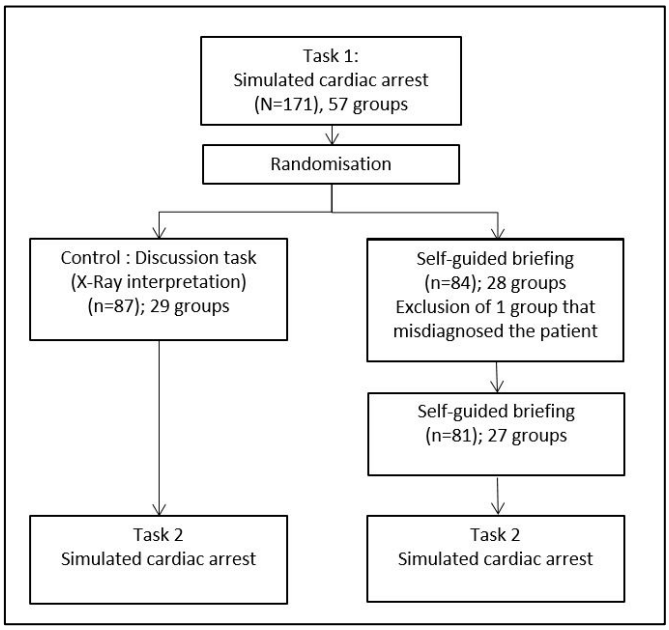
Study design

This is a prospective, randomised controlled trial. The teams were first confronted with a patient suffering a cardiac arrest (task 1; see figure 1). After this, they were randomly assigned to either a discussion task unrelated to resuscitation (control group) or to a self-guided reflective postaction briefing. All teams then performed a second resuscitation task in the same team composition (task 2). We excluded one team from analysis because they misdiagnosed the patient’s condition in the second task and were joined by a confederate to

successfully terminate the task. Of the 56 remaining groups, 29 were in the control condition and 27 in the briefing condition.

Figure1

Study design



Scenario and self-led briefing intervention

On arrival, the students were shown the simulator room and the manikin. They were then asked to wait in the hallway as a team. A confederate nurse stayed in the room throughout both scenarios but intervened only when asked.

Task 1: cardiac arrest

The confederate nurse met the group in the hallway and instructed them to intervene with an emergency in the simulator room. A manual defibrillator was available in the room. The nurse explained that the patient did not feel well after a diagnostic procedure and that he was asked to lie down. On entry of the group, the patient was lying on the bed and was unresponsive, not breathing and with no palpable pulse. A monitor displaying the patient's cardiac rhythm showed ventricular tachycardia. The team had to diagnose the cardiac arrest and initiate resuscitation. The patient did not regain spontaneous circulation regardless of the actions of the team. Three minutes into the resuscitation, the team was interrupted by a confederate.

Briefing intervention

After the interruption, the teams assigned to the self-led briefing condition were handed a sheet with instructions on how to reflect. The instructions were adapted from other studies[16 17 20 33] and contained a suggestion to discuss and look back at the situation just experienced, analysing: (A) what went well, (B) what did not go well and why and (C) discussing strategies on how to proceed, should the team encounter a similar situation again. Teams did not have access to electronic or other sources of information during the briefing. After 3 min of discussion, the team was interrupted by a confederate and asked to proceed to the next patient.

Control group

The teams assigned to the control group where ask to huddle around an X-ray display board displaying a thorax and were given basic patient information about another patient. The team was asked to diagnose the patient's condition (a pneumothorax) and provide treatment recommendations. After 3 min of discussion, the team was interrupted by a confederate and asked to proceed to the next patient. The task of the control group was chosen to be similar in all aspects (talking about a medical problem as a group) with the exception of the content of the discussion. Thus, it can be ruled out that effects of the intervention are actually due to the fact that intervention groups simply spent more time together, independent of the specific content of these discussions.

Task 2: cardiac arrest

After the discussion sessions, the teams were immediately introduced to the next patient. The confederate handed over patient information and told the team that the patient was not feeling well after an ambulatory stress ECG. The team was asked to interview the responsive patient with the aim of providing a recommendation on whether the request of the patient to leave the hospital could be granted. During the interview, the patient complained of pain in the legs (from cycling) and provided standard answers to medical and demographic questions. Two minutes into the interview, the patient said that he felt dizzy, then became unresponsive, closed his eyes, stopped breathing and had no palpable pulse. The monitor showed ventricular tachycardia, and the team had to recognise the situation as a cardiac arrest and start CPR.

Measures

Three performance measures were derived from the in-hospital resuscitation algorithm[32] and adapted to the simulation scenario:

Percent of *hands-on time* (hands-on performance) measures basic CPR efficiency as percentage of time the team provided uninterrupted chest compressions, ventilation or defibrillation on the pulseless manikin. Five seconds before and after defibrillation were counted as hands-on time to account for the shift between chest compressions and defibrillation and to allow for controlling of defibrillation effects.

***Coordination* between chest compressions and ventilation** (coordination performance) was assessed by two independent observers based on video recordings and consisted of two subscores. *Synchronisation* between chest compressions and *ventilation* was rated on a four-point scale (0=chest compressions and ventilation overlapping during the whole procedure; 1=chest compressions and ventilation overlap more often than alternate; 2=mostly alternating, some overlap; 3=chest compression and ventilations never overlap). *Correct application of the 30 (chest compressions) to 2 (ventilations) rule* was coded as 0 (if not observed) and 1 (observed). The two measures were combined with equal weights into one indicator and transformed so that optimal coordination performance yielded a score of 1.

Defibrillation performance was assessed by two independent observers, based on the recordings, and consisted of two subscores: *Compression/ventilation cycles* were coded as incorrect (0) or correct (1) depending on whether they conformed to the number of chest compressions/ventilation cycles between two defibrillations (five according to the resuscitation algorithm in 2010). *Minimising the pause between CPR and defibrillation* was based on the time of interruption between CPR and defibrillation and coded as 0 (if more than 10 s), 1 (between 6 s and 10 s) or 2 (between 0 s and 5 s).[32] The two measures were combined into one indicator with equal weights and adjusted to measure defibrillation performance between 0 (minimal) and 1 (maximal performance). Defibrillation performance was assessed for task 2. For task 1, it was assessed whether (1) or not (0) the team attempted to defibrillate.

Objectives of the study

Primary goal was to assess the effect of the reflective briefing intervention on three aspects of resuscitation performance: hands-on time, coordination between chest compression and ventilation and defibrillation performance.

Statistics

We used analysis of covariance to assess differences in performance between conditions in the second task. To adjust for the effects of having practised this task, we included the respective performance score of task 1 as covariate for hands-on and coordination performance. Because only half of the groups attempted defibrillation in task 1, we included defibrillation attempt, coded as yes (1) or no (0), as covariate when evaluating defibrillation performance in task 2.

Performance measured as percentages (hands-on performance) was arcsine transformed; statistical values are reported based on the transformed variables but descriptive results as percentages and proportions. Analyses were performed with IBM SPSS Statistics for Windows, V.24. $P<0.05$ was considered statistically significant.

Results

Demographic characteristics and gender composition of the teams are displayed in table 1. Six students did not provide data on age. The unbalanced gender team composition reflects the current gender distribution of medical students.

Table 2 summaries the results for all performance measures; figure 2 illustrates the results.

Hands-on performance in the first task was a significant and independent predictor of hands-on performance in the second task. Teams in the briefing condition increased their performance by 16.8 percentage points between task 1 and task 2, while teams in the control condition by 8.6. After adjusting for performance (experience) in the first task, hands-on performance gain was 6.21 percentage points larger for teams in the briefing condition as compared with teams in the control condition (table 2).

Table 1

Participant demographics and team gender composition

Participants (*N* = 168 participants)			
Age	Mean (SD)	24.44	2.36
Females	n (%)	117	69.64
Males	n (%)	21	30.36
Year of study	Mean (SD)	4	0
Team composition (n = 56 groups)			
All female	n (%)	17	30.36
Two females, one male	n (%)	28	50.00
One female, two males	n (%)	10	17.86
All male	n (%)	1	1.79

Table 2

Effect of the briefing intervention on hands-on resuscitation performance, coordinative performance and defibrillator performance

Unadjusted results								Adjusted results							Statistical parameters		
Before intervention				After intervention				After intervention				Difference in performance					
Task 1 control		Task 1 briefing		Task 2 control		Task 2 briefing		Task 2 control		Task 2 briefing		(intervention – control)				df	
M	SD	M	SD	M	SD	M	SD	M	SE	M	SE	M	Lower 95% CI	Upper 95% CI	F		P value
60.04	16.90	56.36	14.82	68.63	13.68	73.16	8.63	67.82	1.69	74.03	1.75	6.21	1.31	11.10			
															32.91†	1	<0.001
															6.09†	1	0.017
0.64	0.35	0.69	0.31	0.65	0.34	0.83	0.26	0.66	0.04	0.81	0.05	0.15	0.02	0.28			
															35.77	1	<0.001
															5.51	1	0.023
65.50%‡		74.10%‡		0.50	0.31	0.47	0.30	0.50	0.06	0.47	0.06	-0.09	-0.27	0.09			
															1.04	1	0.312
															0.19	1	0.663

Note. $N = 56$ teams.

*Adjusted for performance in task 1 (experience).

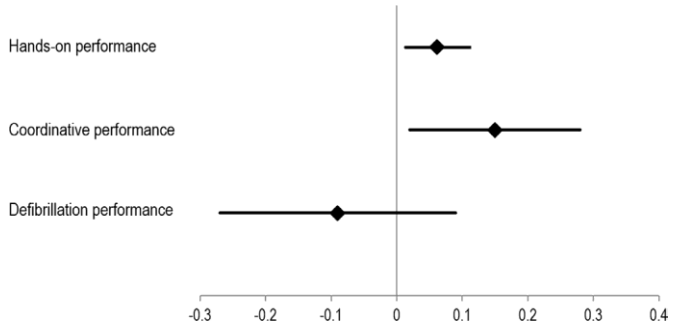
†Statistics are based on arcsine transformation of performance scores.

‡Experience for defibrillation performance represents whether the team defibrillated or not within task 1.

M, Mean.

Figure 2

Illustration of performance gains by briefing for all three performance measures. To allow comparability, hands-on performance is expressed in proportions. Data are displayed as means and 95% CIs of performance gains/losses for teams in the briefing condition as compared with the control condition. Hands-on and coordinative performance gains were significant; defibrillation performance gains were not significant.



Coordination performance in the first task was also a significant and independent predictor of coordination performance in the second task. Teams in the briefing condition increased their coordination performance significantly more than teams in the control condition: after adjusting for coordination performance in the first task, performance gain for teams in the briefing condition was 0.15 points (or 15%) larger than for teams in the control condition (table 2).

Teams in the briefing condition showed a slight, but non-significant, lower defibrillation performance in the second task than teams in the control condition (−0.09 points or −9%). Experience with defibrillation in the first task was not a significant predictor for defibrillation performance in the second task.

Discussion

A 3 min self-led postbriefing after a resuscitation task increased performance in a subsequent, similar task in ad hoc teams of fourth-year medical students as compared with teams in a control group discussing other issues. Results are adjusted for performance in the first task, which therefore cannot explain the results.

The briefing enhanced basic hands-on and coordination performance but not defibrillation performance. Hands-on performance represents basic aspects of resuscitation and is particularly affected if teams have problems initiating the resuscitation or if interruptions occur.[34] Coordination performance also represents basic coordinative aspects of the task, as it measures the correct number of chest compressions and ventilation and their appropriate alternation. We therefore conclude that the briefing positively influenced basic aspects of task performance.

Although experienced task difficulty depends to some degree on the individual, defibrillation performance can be regarded as representing performance of more complex[35] and more advanced aspects of the task; it was not improved by the briefing. Two reasons may explain this result. First, the briefing instructions specifically asked the teams to reflect on their previous experience. As about half of the teams did not attempt defibrillation in task 1, defibrillation may not have been discussed in these teams during the briefing if they followed the instruction to reflect on their past experience. Second, other studies did not find effects of reflective briefings on performance if briefings were not combined with additional information or feedback,[17 36] not led by an expert,[33] or if teams were not trained in reflecting.[19] Other authors also emphasised that the quality of (self-led) debriefings is often low.[37] Such aspects are likely to increase in importance as the complexity of tasks increases. It could thus be that the quality of the briefings was sufficient to increase basic aspects of

performance but not more complex ones. Particularly, the latter aspect can be addressed in the expert-led debriefing after the simulator session.[38]

Self-led briefings without an expert may not only have positive effects, as there is a risk that erroneous information may be discussed and negatively influence subsequent performance. (We thank one of the reviewers for emphasising this aspect.) Unfortunately, we have no information about the content of the reflective briefing and therefore cannot assess the accuracy of the information discussed. As an approximate test to identify potential negative effects of briefings, we compared the proportion of groups with decreasing performance in the second task as compared with the first task across conditions. Hands-on as well as coordinative performance was both worse in 2/27 (7.4%) of teams in the briefing condition but in 7/29 (24.1%; hands-on) and 6/29 (20.7%; coordinative) of teams in the control condition; these differences did not reach conventional statistical significance levels. These results do not rule out negative effects of briefings, but they do indicate that negative effects may be the exception and may be offset by the positive effects of the briefing as compared with no briefing. Note that for defibrillation performance, such an analysis could not be done, as many teams did not defibrillate in the first tasks.

This study has limitations. First, it is an experimental study based on a cohort of fourth-year medical students. Although they represent relatively unexperienced medical professionals quite well, real ad hoc resuscitation teams are likely to be interdisciplinary and composed of professionals with different levels of experience; this limits the generalisability of the results to medical practice. In addition, because participants are from the same cohort, we cannot control for the effects of prior training experiences they may have together. Another limitation is the relatively small sample size. Furthermore, we did not have information on the content of the briefing discussions; therefore, we could not assess the accuracy and level of expertise of the briefing discussion, nor to what degree the teams in the

briefing condition may have addressed specific performance shortcomings in task 1. Further studies should investigate the content of such briefings.

Conclusions

Learning how to reflect on one's own teamwork and task performance is an important skill[39]; in the optimal case, this skill is itself part of a training. This study shows that in simulation training, teams are capable to engage in self-led debriefings and profit from them, although in a limited way, that is, for basic aspects mainly. Evidently, the self-led debriefing is part of the simulation experience and should be discussed in the postsimulation expert-led briefing[33 38] to assure that all relevant aspects were discussed accurately and to discuss the process and potential difficulties of a self-led briefing.

For everyday practice, the results underscore the utility of even very short self-led post-task briefings. The mean length of post-action debriefings in trainings is 17 minutes, according to a meta-analysis.[40] Although they may be more effective than shorter ones, such long briefings constitute major interruptions of ongoing work and therefore are not very likely to be implemented into everyday practice. Our study showed that engaging in a 3 min briefing increased performance on a subsequent similar task by 6 percentage points for basic hands-on performance and by 15 percentage points for coordinative performance. Such a cost-effective and feasible intervention may thus be worthwhile to introduce. The debriefing instructions to look back at the task, analyse what went well and what did not, and to develop plans for a similar task could well be used in everyday practice. To avoid negative effects of inaccurate knowledge, teams should be encouraged to look up information or seek expert advice if they detect knowledge gaps or insecurities during the briefings.

Acknowledgments

We would like to thank Mina Dimitrova, Nadja Jenni and Patrick M Huber for their help in transcribing and coding data; Mark Breuer and Sabine Schweizer for their help in data collection.

Contributors

All authors approved the final version of the manuscript. The first two authors contributed equally to the manuscript. PK: conception, manuscript draft, data collection and coding. FT: conception, manuscript draft, manuscript revision, development of coding and data analysis. NKS: conception and manuscript draft. CM: conception, manuscript draft, coding development, coding and data analysis. JZ and EH: manuscript revision, coding and data analysis. SAH: manuscript revision, coding development, coding and data analysis. SH: manuscript revision and coding development. SUM: conception, manuscript draft, manuscript revision and data collection.

Funding

This work was supported by the Swiss National Science Foundation, grant #149734.

Competing interests

None declared.

Ethical approval

The study was approved by the local ethics committee.

Provenance and peer review

Not commissioned; externally peer reviewed.

References

1 Schmutz J, Manser T. Do team processes really have an effect on clinical performance? A systematic literature review. *Br J Anaesth* 2013;110:529–44.

2 Salas E, DiazGranados D, Weaver SJ, et al. Does team training work? Principles for health care. *Acad Emerg Med* 2008;15:1002–9.

3 Hughes AM, Gregory ME, Joseph DL, et al. Saving lives: A meta-analysis of team training in healthcare. *J Appl Psychol* 2016;101:1266–304.

4 Edmondson AC, Dillon JR, Roloff KS. Three perspectives on team learning: outcome improvement, task mastery, and group process. *The academy of management annals* 2007;1:269–314.

5 Schippers MC, Edmondson AC, West MA. Team reflexivity as an antidote to team information-processing failures. *Small Group Res* 2014;45:731–69.

6 Smith-Jentsch KA, Cannon-Bowers JA, Tannenbaum SI, et al. Guided team self-correction: impacts on team mental models, processes, and effectiveness. *Small Group Res* 2008;39:303–27.

7 Reiter-Palmon R, Kennel V, Allen J, et al. Good catch! using interdisciplinary teams and team reflexivity to improve patient safety. *Group Organ Manag* 2018;43:414–39.

8 Konradt U, Otte KP, Schippers MC, et al. Reflexivity in teams: a review and new perspectives. *J Psychol* 2016;150:153–74.

9 Gersick CJ, Hackman JR. Habitual routines in task-performing groups. *Organ Behav Hum Decis Process* 1990;47:65–97.

10 Arrow H, Mcgrath JE. Membership dynamics in groups at work - a theoretical framework. *Research in Organizational Behavior: An Annual Series of Analytical Essays and Critical Reviews* 1995;17:373–411.

11 Huckman RS, Staats BR. Fluid tasks and fluid teams: the impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management* 2011;13:310–28.

12 Marsch SC, Tschan F, Semmer N, et al. Performance of first responders in simulated cardiac arrests. *Crit Care Med* 2005;33:963–7.

13 Chen J, Bamberger PA, Song Y, et al. The effects of team reflexivity on psychological well-being in manufacturing teams. *J Appl Psychol* 2018;103:443–62.

14 Boet S, Bould MD, Sharma B, et al. Within-team debriefing versus instructor-led debriefing for simulation-based education: a randomized controlled trial. *Ann Surg* 2013;258:53–8.

15 Nederveen Pieterse A, van Knippenberg D, van Ginkel WP. Diversity in goal orientation, team reflexivity, and team performance. *Organ Behav Hum Decis Process* 2011;114:153–64.

16 Gurtner A, Tschan F, Semmer NK, et al. Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organ Behav Hum Decis Process* 2007;102:127–42.

17 Gabelica C, Van den Bossche P, De Maeyer S, et al. The effect of team feedback and guided reflexivity on team performance change. *Learn Instr* 2014;34:86–96.

18 Konradt U, Schippers MC, Garbers Y, et al. Effects of guided reflexivity and team feedback on team performance improvement: The role of team regulatory processes and cognitive emergent states. *Eur J Work Organ Psychol* 2015;24:777–95.

19 Salas E, Klein C, King H, et al. Debriefing medical teams: 12 evidence-based best practices and tips. Jt Comm *J Qual Patient Saf* 2008;34:518–27.

20 Schmutz JB, Eppich WJ. Promoting learning and patient care through shared reflection: a conceptual framework for team reflexivity in health care. *Acad Med* 2017;92:1555–63.

21 Lyons VE, Popejoy LL. Meta-analysis of surgical safety checklist effects on teamwork,vcommunication, morbidity, mortality, and safety. *West J Nurs Res* 2014;36:245–61.

22 Shunk R, Dulay M, Chou CL, et al. Huddle-coaching: a dynamic intervention for trainees and staff to support team-based care. *Acad Med* 2014;89:244–50.

23 de Vos MS, Marang-van de Mheen PJ, Smith AD, et al. Toward best practices for surgical morbidity and mortality conferences: a mixed methods study. *J Surg Educ* 2018;75:33–42.

24 Su L, Kaplan S, Burd R, et al. Trauma resuscitation: can team behaviours in the prearrival period predict resuscitation performance? *BMJ Simul Technol Enhanc Learn* 2017;3:106–10.

25 Hazinski MF, Nolan JP, Aickin R, et al. Part 1: executive summary: 2015 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. *Circulation* 2015;132(16 Suppl 1):S2–39.

26 Hull L, Russ S, Ahmed M, et al. Quality of interdisciplinary postsimulation debriefing: 360° evaluation. *BMJ Simul Technol Enhanc Learn* 2017;3:9–16.

27 Lyubovnikova J, Legood A, Turner N, et al. How authentic leadership influences team performance: the mediating role of team reflexivity. *J Bus Ethics* 2017;141:59–70.

28 Fowlkes J, Dwyer DJ, Oser RL, et al. Event-Based Approach to Training (EBAT). *Int J Aviat Psychol* 1998;8:209–21.

29 Fernandez R, Kozlowski SW, Shapiro MJ, et al. Toward a definition of teamwork in emergency medicine. *Acad Emerg Med* 2008;15:1104–12.

30 Sonnentag S, Frese M. Performance concepts and performance theory. In: Sonnentag S, ed. *Psychological management of individual performance*. Chichester, UK: John Wiley & Sons, 2002:3–25.

31 Shepherd A. *Hierarchial task analysis*. London: Taylor & Francis, 2001.

32 Soar J, Nolan JP, Böttiger BW, et al. European Resuscitation Council Guidelines for Resuscitation 2015: section 3. Adult advanced life support. *Resuscitation* 2015;95:100–47.

33 Schmutz JB, Kolbe M, Eppich WJ. Twelve tips for integrating team reflexivity into your simulation-based team training. *Med Teach* 2018;40:721–7.

34 Tschan F, Vetterli M, Semmer NK, et al. Activities during interruptions in cardiopulmonary resuscitation: a simulator study. *Resuscitation* 2011;82:1419–23.

35 Campbell DJ. Task complexity: a review and analysis. *Acad Manage Rev* 1988;13:40–52.

36 Tesler R, Mohammed S, Hamilton K, et al. Mirror: guided storytelling and team reflexivity's influence on team mental models. *Small Group Res* 2017:104649641772202.

37 Kihlgren P, Spanager L, Dieckmann P. Investigating novice doctors' reflections in debriefings after simulation scenarios. *Med Teach* 2015;37:437–43.

38 Kolbe M, Grande B, Spahn DR. Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting. *Best Pract Res Clin Anaesthesiol* 2015;29:87–96.

39 Kolbe M, Rudolph JW. What's the headline on your mind right now? How reflection guides simulation-based faculty development in a master class. *BMJ Simul Technol Enhanc Learn* 2018;4:126–32.

40 Tannenbaum SI, Cerasoli CP. Do team and individual debriefs enhance performance? A meta-analysis. *Hum Factors* 2013;55:231–45.

Paper 2

Dissecting team reflexivity in ad-hoc medical teams: effects on performance and an analysis of reflexivity content, structure and quality

The second paper combines different themes and aims to fill gaps in the literature. Its genesis is also based on the issues raised by the paper 1. The main goal was to compare the effect of different reflexivity conditions (focused on team-related aspects, task-related aspects, or combination of both team- and task-related aspects). Previous research has not focused on the content discussed during reflexivity sessions. This is why we analyzed the reflexivity session content in order to investigate whether people follow reflexivity instruction and engage in reflexivity. We also aimed to investigate if participants were able to identify their weaknesses and mistakes regarding the resuscitation algorithm, if they addressed them in the reflexivity session, and if they could correct them.

Finally, we were also interested in reflexivity quality. We developed our own quality (unresolved disagreements and scope of reflexivity) indicators in order to know the importance of reflexivity quality on performance increase.

Dissecting team reflexivity in ad-hoc medical teams: effects on performance and an analysis of reflexivity content, structure and quality

Disclaimer

This chapter of my dissertation is the basis for one (maybe two) scientific paper. It is thus structured like a scientific paper, with introduction, method, results and discussion and has its own bibliography. Because this chapter can stand alone, it has redundancies with other chapters of the dissertation.

The author of this dissertation will not be the sole author of the journal publications. As the data for this study was collected at the simulation center of the Basel University Hospital as part of an SNF project, researchers from the Neuchâtel, Bern and Basel research groups have collaborated. The publication agreement states that I will be co-authoring the chapter with Franziska Tschan as equal first authors, other authors include: Simon A. Huber (transcripts, development of coding system, coding), Sabina Hunziker (study conceptualization and data collection), Robin Lüchinger (collaboration development of coding system, coding), Stephan U. Marsch (co-primary investigator, study conceptualization, data collection), Norbert K. Semmer (co-primary investigator SNF project, study conceptualization), and Jasmin Zimmermann (conceptual work and coding). All my co-authors will also work on later versions of the manuscript.

Author Note

This research was supported by a grant from the Swiss National Science Foundation Grant #149734. We thank Patrick M. Huber (transcripts, coding), and Julie Fournier (transcripts, coding), as well as Mark Breuer and Sabine Schweizer for their contributions in conducting the simulations. None of the authors declares conflicts of interest. Data of a control group stem from an earlier project (*citation omitted for peer review*). Camille Morgenthaler is now at the Neuchâtel Hospital Network; Simon Huber is now at Swiss Post

Office; Jasmin Zimmermann is now at the Swiss Council for Accident Prevention; Robin Lüchinger is now at the University Hospital Geneva (HUG). The seminal chapter of Moreland and McMinn (2010) as well as many discussions with Dick Moreland were inspiration for this research, we thus dedicate this paper to the late Richard Moreland.

Correspondence concerning this article should be addressed to Franziska Tschan (franziska.tschan@unine.ch), Université de Neuchâtel, Rue Emile Argand 11; 2000 Neuchâtel, Switzerland.

Abstract

Team reflexivity (the extent to which group members overtly reflect upon the group's objectives, strategies, and processes and adapt them to current or anticipated (…) circumstances) helps teams to improve performance, even if research results are somewhat contradictory, particularly, if reflexivity is used as an intervention in ad-hoc teams.

In this study, we investigated effects of three different reflexivity foci on performance, and analyzed aspects of the reflexivity process in order to investigate how reflexivity can influence team performance.

In an experimental study, we used reflexivity-as-intervention in teams of three medical students between two simulated cardiopulmonary resuscitation tasks. All data are based on observation. Based on task- and collaboration requirement analyses of cardiopulmonary resuscitation, we developed and compared a task-focused, team-focused or combined reflexivity intervention. Answering concerns by Moreland and McMinn (2010) about the competence and willingness of teams to engage in reflexivity, we assessed whether teams followed the reflexivity instructions, whether teams addressed their weaknesses during reflexivity, and whether performance improvement was higher in subtasks with low initial performance. Finally, we investigated to what extent quality of reflection (unresolved disagreements and scope of reflection) influenced performance improvement. Our results show that reflexivity-as-intervention has positive effects on performance compared to a control condition, with teams in the team-focused condition benefitting the most. Teams engaged in reflexivity and followed the instructions with regard to content and temporal sequence of the discussion. However, team only partially their weaknesses in initial performance during reflexivity. Moreover, discussing specific aspects of the task during reflexivity did not improve these aspects subsequently (except for leadership), even when they were low in prior performance. Finally, teams reached rather low quality of reflexivity

even if quality influenced subsequent performance. Teams also experienced few disagreements, but these can also hinder performance. We conclude that reflexivity has an interesting potential for teams, but teams have problems using reflexivity well.

Keywords: Team training, group reflection, group reflexivity, team briefings, medical simulation, cardiopulmonary resuscitation, action teams, cardiopulmonary resuscitation

Introduction

Team reflexivity, defined as "the extent to which group members overtly reflect upon the group's objectives, strategies, and processes and adapt them to current or anticipated (…) circumstances" (West, 1996, p. 599) has been related to higher team performance and/or innovation (Schippers et al., 2015; Widmer et al., 2009). The relationship between team reflexivity and team processes, emergent states and outcomes are studied in field studies that assume that reflexivity is a relatively stable behavior of teams (Carter & West, 1998; Hofhuis et al., 2018; Schippers et al., 2012; Schippers et al., 2007), as well as in experimental or interventional studies (Gurtner et al., 2007; Kneisel, 2020; Konradt et al., 2015). The latter studies use reflexivity-as-intervention: Teams are asked to reflect, and their group process or performance is compared to a control group that does not reflect or reflects differently. In contrast to field studies that conceptualize reflexivity as a behavioral style of real teams that have a history of collaboration along with member and task familiarity, experimental studies are most often conducted in no-history ad-hoc teams.

Studies with a reflexivity-as-intervention paradigm found positive effects of reflexivity on performance (Gurtner et al., 2007; Kneisel, 2020; Konradt et al., 2015; Muller et al., 2008; van Ginkel et al., 2009), but boundary conditions exist (Gabelica, Van den Bossche, De Maeyer, et al., 2014; Tesler et al., 2017); and under certain circumstances, reflexivity even had a negative impact on performance (Nederveen Pieterse et al., 2011). Despite contradicting results on performance, some authors express enthusiasm about reflexivity interventions (Schippers, 2012; Schmutz & Eppich, 2017; Schmutz & Eppich, 2018), whereas others are more critical. The clearest manifestation of skepticism about reflexivity-as-intervention was formulated by Moreland and McMinn (2010). In their own reflexivity-as-intervention study, they did not find any effects on performance, but reported some interesting observations: For example, their teams, asked to reflect between similar

tasks, did not follow the experimenters' instructions during the reflexivity session and did not use the allotted time (a few minutes) to reflect. The authors thus questioned whether teams were willing and able to engage in a meaningful reflection. Other studies questioned the quality of the reflection process: For example, Gabelica and colleagues (2014) found that teams often did not reflect according to the ideal structure that starts with looking back, analyzing past performance (and addressing weaknesses) and then looking forward and plan. Some authors also suggested that the quality of the reflection was not sufficiently high to influence future performance (Moreland & McMinn, 2010; Swift & West, 1998). This critique is not unfounded: Although team members are aware of the necessity to plan and coordinate (Putnam, 1979), groups often fail to switch from the task level to the strategic and coordinative teamwork level (Zellmer-Bruhn et al., 2004). In addition to those difficulties, Moreland and McMinn (2010) suggested that social aspects, such as conflicts and the reluctance to criticize each other, may influence the reflexivity process.

All these criticisms imply that it may be worthwhile to have a closer look at the reflexivity process itself. In the present study, teams of medical students were asked to treat two similar medical emergency tasks and to engage in a reflexivity session between the two tasks. One of the goals of this study was to test effects of three different reflexivity interventions (task-focused, team-focused, combination of both) on performance. Another goal was to analyze the content of the reflexivity session.

Despite the fact that team reflection necessarily involves open discussion (Schippers et al., 2018) and can thus be observed, only few studies have done so and provide detailed analyses of the reflection process itself. We thus "dissected" the reflexivity as well as the team process to investigate whether teams engaged in reflexivity according to the instructions, whether teams addressed their weaknesses during reflexivity (Schippers et al., 2013), and whether performance improvement was particularly high in task domains with

low initial performance. Finally, we investigated whether social aspects (unresolved disagreements) and the quality of reflexivity (scope of information integration) influenced performance gains between task 1 and task 2. To avoid self-report biases, all analyses are based on behavioral observation data.

This study contributes to the literature in responding to some of the criticisms of Moreland and McMinn (2010), in investigating aspects of the reflexivity process and its relationship with the team processes before, but also after the reflexivity intervention.

Team reflexivity process

To achieve high performance, individuals, teams and organizations profit from reflecting upon their actions (Schön, 1938). The difference between reflection and action in teams is well represented in the distinction between taskwork (action phases, when teams execute tasks) and teamwork (transition phases, when teams engage in evaluation and planning activities) (Marks et al., 2001, p. 360); reflexivity clearly belongs to transition phases.

In his initial work, West (1996) conceptualized team reflexivity as a measurable characteristic of a stable team (Vashdi et al., 2007). Indicators of high team reflexivity are well represented by the questionnaire items used to measure the concept, such as “we talk about different ways in which we can reach our objectives, (…) we examine the implications that changes in the environment may have for the aims of the team (…) we work out what we can learn from past activities”, etc. (Schippers et al., 2007, p. 210). Typical reflective behaviors correspond to those described for transition phases and may include mission analysis, goal specification, strategy formulation, planning, etc. (LePine et al., 2008, p. 276), but also include a critical look back on past performance and strategies. According to a meta-analysis, transition phase behaviors are clearly related to team performance, but also to

member satisfaction (LePine et al., 2008). Note, however, that not all activities of the transition phase fit the definition of reflexivity.

Outcomes of reflexivity

It is not surprising that links between reflexivity and team performance were found in field studies (Carter & West, 1998; Lyubovnikova et al., 2015; Schippers et al., 2008; Tjosvold et al., 2003; Wu et al., 2017), as well as in interventional studies (Gurtner et al., 2007; Kneisel, 2020; Konradt et al., 2015; Kündig et al., 2019; Schmutz, Lei, et al., 2018); however, note that other studies did not confirm a reflexivity-performance link (Moreland & McMinn, 2010; Wiedow & Konradt, 2010). Interestingly, reflexivity was often found to be linked to team or individual innovation or creative problem solving as outcomes (Carmeli et al., 2014; Schippers et al., 2015; Somech, 2006; Tjosvold et al., 2004; Urbach et al., 2010; Wong et al., 2007), indicating that reflexivity is particularly useful for finding new, innovative solutions. Some authors postulated a link between reflexivity and team member well-being; however, the results of the few studies including measures of team member well-being are contradictory with one study not finding effects (Carter & West, 1998) and a more recent field experiment showing that reflexivity reduced burnout (Chen et al., 2018).

Boundary conditions

Field studies identified several boundary conditions for reflexivity. Some conditions facilitate or hinder the process of reflexivity itself, while others moderate the influence of reflexivity on outcomes (Widmer et al., 2009).

Interpersonal factors

One important influence on the reflexivity process itself are interpersonal factors, for example mutual trust, high psychological safety (De Jong & Elfring, 2010), team cohesion (Lee, 2008), and positive group affective tone (Shin, 2014). Such interpersonal factors allow the team to have a more open and less conflicting discussion (Carmeli et al., 2014; Schippers

et al., 2018). Furthermore, social and interactional skills (Hoegl & Parboteeah, 2006; Wong et al., 2007) facilitate the reflection process; one study found that only teams with cooperative conflict management and high reflexivity reached higher performance (Tjosvold et al., 2003). Studies investigating interpersonal factors all underscore that the reflexivity discussion needs team members to express themselves frankly and openly and that psychological safety needs to be high (Edmondson, 1999); but also that that team members need good interpersonal and conflict management skills to be able to engage in a constructive controversy (Hackman & Wageman, 2005; Johnson et al., 2000).

Work and task conditions

Other boundary conditions refer to specific aspect of work or of the task. Reflexivity is more useful if teams must deal with challenges that can be adequately addressed by reflection. This is the case for teams with high diversity or heterogeneity, for example high minority dissent (De Dreu, 2002), functional heterogeneity (Somech, 2006) or high cognitive diversity (Chen et al., 2019; Yang et al., 2020). During reflection, those teams may be able to identify, address and integrate different approaches or viewpoints. Reflexivity is also more useful for teams that are confronted with adverse conditions, for example high work demands or poor physical environment; such teams have been found to be more innovative than reflexive teams without those adverse conditions (Schippers et al., 2012). Adverse conditions may be a motivator for change; and if a team takes time to reflect, it can find novel solutions that help to address especially pressing problems. Similarly, high reflexivity was particularly useful for teams with low initial performance (Schippers et al., 2013), again indicating that reflexivity is especially useful for teams in clear need of change or improving. However, to implement or plan for implementing solutions, teams need decision latitude – as evidenced by a study showing that reflexivity only led to more innovation if job control was high (Urbach et al., 2010).

Knowledge

Another important boundary condition for effects of reflexivity is team member knowledge. If team members do not have enough knowledge to discuss and solve their problems, no information integration can occur (Lee, 2008); thus, task familiarity facilitates reflexivity (Wu et al., 2017). In fact, the main gain of reflexivity may be that information and knowledge is mobilized and elaborated (van Ginkel & van Knippenberg, 2009), processed and shared (De Dreu, 2007; Schmutz & Eppich, 2017) and integrated (Schippers et al., 2018), and that biases and common decision-making failures are minimized (Schippers et al., 2014). An experimental study found more information elaboration in reflexivity conditions (van Ginkel et al., 2009), another study found more information elaboration particularly for more diverse groups (Nederveen Pieterse et al., 2011).

Quality of the reflection process

If information integration is one of the main processes for reflexivity to be successful, the degree to which this information integration happens and thus the quality of the reflection process is important. Only few studies have directly assessed the reflection process itself and its quality. Swift and West (1998) suggested to distinguish shallow, moderate and deep reflexivity as different reflexivity qualities, suggesting that deeper and more extensive information integration indicates better reflection quality. However, assuring high quality information integration may be difficult for teams. First, teams seem to be generally reluctant to plan and discuss their strategies (Hackman et al., 1976; Hackman & Kaplan, 1974). Second, performance pressure may shorten strategy discussions (Weingart, 1992), because such discussions may be seen as jeopardizing performance. Third, teams may not be able to engage in high quality discussions (Kolbe et al., 2015). For example, one experimental study reported that teams often discussed very general aspects of collaboration not related to the actual task and team during a team-reflexivity session (Gurtner et al., 2007). Further concern

about the quality of discussion comes from analyses of debriefing processes where high-quality debriefings with high information integration were rare, even in the presence of an external expert (Kihlgren et al., 2015).

Reflexivity-as-intervention

Experimental studies that use reflexivity-as-interventions allow to investigate the reflexivity processes more specifically, because boundary conditions can be controlled for. However, results from field studies may not apply to experimental studies and vice versa. Experimental studies are most often done with ad-hoc teams with zero team familiarity, and often also low task familiarity. Such teams could not yet develop a habit to reflect, reflection is imposed to them. Applying results from field study to experimental studies thus are problematic.

Reflexivity intervention and performance

Some studies with reflection-as-intervention showed positive effects on team performance (Gurtner et al., 2007; Kneisel, 2020; Konradt et al., 2015; Kündig et al., 2019; Schmutz, Lei, et al., 2018); however, note that other studies found no or even a negative reflexivity-performance link (Moreland & McMinn, 2010; Nederveen Pieterse et al., 2011; Wiedow & Konradt, 2010). There is, however, evidence for a positive influence of reflecting on team performance from studies that used reflexivity-as-intervention in field settings: A concept very similar to reflexivity interventions is the so-called team briefing. Briefings are regularly and frequently used in military, aviation, safety or medical settings and are often done in teams with frequently changing membership that are typical for those domains (Arrow et al., 2000). Team briefings are short sessions where group members inform and/or reflect about their tasks, either before, during or after their taskwork. Recent comparisons between reflexivity and team briefings (Allen et al., 2018; Reiter-Palmon et al., 2020) found that not all team briefings include each aspect of reflexivity, but confirmed that a reflexivity-

as-intervention session resembles to team briefings, if they involve active self-learning, are aimed at learning and development and not evaluation, and are related to specific tasks and team experiences (Tannenbaum & Cerasoli, 2013). A recent meta-analysis found clear effects of team briefings on team performance: Overall, the percentage increase of performance was 25% with a Cohen's d of 0.67. The study also showed that if the briefing was facilitated by an expert, effect sizes were three times higher (Cohen's d = 0.75) compared to non-facilitated briefings (Cohen's d = .25), and unstructured briefings were clearly less successful (Cohen's d = .32); note, however, that only one study reported completely unstructured debriefings (Tannenbaum & Cerasoli, 2013).

Focus on critical issues during reflexivity

The finding that facilitated and structured debriefings yield better results again underscores the importance of the guidance provided to teams to help them to reflect (Tannenbaum & Cerasoli, 2013). If an expert debriefer is present, her or she can guide the attention of the team, so that teams reflect on the most important aspects and shortcomings. For the current research, we chose not to use an external facilitator for the debriefing, we thus do not discuss this aspect more deeply. However, focusing the team's attention to important aspects can also be done without an external facilitator. Four studies using reflexivity-as-intervention provided some attention-focusing guidance to teams in instructing them to reflect according to the temporal structure proposed by West (1996) – to reflect and analyze previous performance and to plan for future, similar tasks (Eddy et al., 2013; Gurtner et al., 2007; Konradt et al., 2015; Kündig et al., 2019). In addition to the temporal structure, two studies provided further guidance: The study of Konradt et al. (2015) instructed teams about important team concepts (for example shared mental models) and provided feedback on where the team stood with regard to those concepts. In the study of Eddy and colleagues (2013), teams received an instruction that was specifically tailored to the task's cooperation

and action requirements. Such instructions helped the teams to focus their discussion on the most important task and cooperation needs. In both studies, the combination of reflection and specific feedback or specific focus was directly or indirectly related to performance (Eddy et al., 2013; Konradt et al., 2015).

The present study and hypotheses

In the present study, teams of fourth year medical students were confronted with two subsequent medical emergencies (both witnessed cardiac arrests) in a high-fidelity patient simulator setting. Between the two emergencies, teams were asked to engage in a reflexivity session; three different reflexivity instructions were compared. Details are given in the method section.

This experimental situation differs from more traditional experiments in its high external validity: Participants had specialist medical knowledge (four years of medical school), and according to their curriculum, they had received theoretical (but limited practical) instruction on how to diagnose and treat a cardiac arrest; in addition, the simulator setting allowed for a very realistic situation (life-sized manikin with realistic physiological features). Furthermore, fourth year students are likely to be responsible to treat a cardiac arrest emergency very soon in their future professional life. Another realistic feature of the situation studied is that there is no or very little time to plan ahead (Su et al., 2017), because a witnessed cardiac arrest requires immediate medical attention – even minutes in delay of treatment significantly diminishes survival of the patient (Graham et al., 2015; Larsen et al., 1993). However, team composition is rather atypical for a cardiac arrest situation; it is unlikely that this emergency is treated by a group of three medical students.

The treatment of a sudden cardiac arrest follows a specific treatment algorithm (Avramidis, 2014; Soar et al., 2015) that is standardized and known to medical professionals. For basic and advanced life support, this algorithm specifies sub-tasks (e.g. ventilation and

cardiac massage) and their recommended temporal sequence (e.g., two ventilations alternating with 30 compressions); the algorithm also specifies quality criteria for each of the sub-tasks (e.g., compressions about 100 times per minute). Such a specific algorithm allows to perform a task-analysis with regard to task requirements (Henrickson Parker et al., 2018; Tschan, Semmer, Vetterli, et al., 2011). Furthermore, collaboration requirements can be derived from the algorithm: For cardiopulmonary resuscitation leadership, task distribution and explicit communication have been found to be predictors of performance (Cooper & Wakelam, 1999; Edwards & Siassakos, 2012; Fernandez Castelao et al., 2013; Hunziker et al., 2011; Marsch et al., 2004; Roberts et al., 2014).

As outlined above, when designing a reflexivity intervention, it may be advantageous to direct the team's attention to important aspects to discuss. For the cardiac arrest tasks, both the tasks requirements (task focus) as well as the collaboration requirements (team focus) are candidates to be reflected upon. Team training methods that train task and cooperative skills together show more promising results than trainings focusing on teamwork only (Cannon-Bowers & Salas, 1998; Salas et al., 2008; Salas et al., 2007). It seems thus also likely that teams that discuss both task and team aspects during the reflexivity session may profit more from reflexivity than teams focusing on either task or team aspects only. We thus formulate the following hypothesis:

H1: Teams that reflect on both task and team aspects ("combined condition") show higher performance improvements between task 1 and task 2 than teams that are asked to reflect on either task (task-condition) or coordination (team-condition) aspects only.

Moreland and McMinn (2010) observed that teams were reluctant to engage in reflection and did not follow the instructions. Besides of the general reluctance of teams to engage in teamwork activities (Hackman & Kaplan, 1974; Weingart, 1992), Moreland and McMinn (2010) suggested motivational reasons for their findings. In fact, if a problem is not

judged sufficiently relevant, groups may not engage in developing solutions (Moreland & Levine, 1992). Given that the task used in our research has high external validity and is highly relevant for the participant's future professions, we assume low motivational problems. We therefore assume that teams follow the instructions and formulate the following hypothesis:

H2: Teams follow the written instructions regarding the time allocated for reflection (H2a); arbitrarily, we set a limit of more than 20% of unrelated communication as critical. Teams use the reflexivity session to talk about aspects related to cardiopulmonary resuscitation (H2b).

H3: Teams follow the reflection instruction regarding to the thematic focus: Teams in the task-focused reflexivity condition discuss more about task-related aspects (conditions: task = combination > team); teams in the team-focused condition talk more about team-related aspects (conditions: team=combination > task).

H4: Teams follow the reflexivity instruction regarding the suggested temporal sequence of reflection (analyze past – plan for next performance). We (arbitrarily) define that adhering to the temporal structure is given if the observer rating is at least 2.5 (out of 3).

The core idea of reflexivity is to critically discuss past performance and processes and to develop strategies for future, better performance (West, 1996; Widmer et al., 2009). This implies that teams need to be able to correctly identify their shortcomings and problems. However, research found that problem identification in groups is difficult, and misdiagnoses of problems are common (Moreland & Levine, 1992). However, if reflexivity is to be successful, teams need to identify and discuss their shortcomings. This is a prerequisite for the finding that teams particularly improve in areas of low performance (Schippers et al., 2013). We therefore hypothesize:

H5: During the reflexivity session, teams are more likely to discuss sub-tasks of suboptimal initial team performance than sub-tasks of high performance.

H6a Teams that engage in reflexivity regarding specific sub-tasks will show higher performance increase in those sub-tasks. H6b This effect is particularly pronounced if teams performed more poorly in that sub-task in task 1.

As outlined above, interpersonal aspects have been identified as boundary conditions for the success of high reflexivity in stable teams, and interpersonal problems likely impair performance. Even if disagreements are minor and are related to the task, performance can suffer if disagreements are not resolved. For example, if team members do not agree about what tasks are related to the resuscitation algorithm, or how tasks should be done, and do not resolve such task conflicts before the end of the reflexivity period, subsequent performance will suffer.

H7 If during reflexivity, task or social disagreements are not resolved, performance improvement suffers.

In accordance with previous studies, we assume that the quality of information integration during the reflection process is related to higher performance. Previous

experimental research (Gurtner et al., 2007; Kihlgren et al., 2015) found that team reflection is very often superficial and does not attain high quality. Because future tasks may differ with regards to contexts and situations, a wider scope of reflection that includes planning for contingencies may be particularly useful. We therefore hypothesize:

H8 Teams that adopt a wider scope during reflection increase their performance more than teams that do not.

Method

Participants and setting

Participants were 204 fourth-year medical students, 82 males and 122 females; Mean age was 24.58, SD = 2.25. They worked in (68) teams of three, unit of analysis was the team. Participants were recruited from two different medical schools that follow a similar curriculum. The students were assigned to teams according to their scheduling preferences. Whereas it is possible that members within the same team knew each other, none of the members within a team reported having collaborated on a medical task with another member of the same team. All students had received prior instruction about cardiopulmonary resuscitation (CPR) as part of their regular curriculum. Participant characteristics of the 29 teams serving as control condition for one analysis are described in *(reference omitted for blind review).*

Participation in this study was voluntary and not a part of the curriculum; the students did not receive any compensation but could assist in an extended personalized post-task expert led briefing and instruction after the experiment. Participants were blind to study aims and unaware of the different experimental conditions. All participants provided a written informed consent. The study protocol was approved by the local ethics committee (#85/04).

The study took place at the simulator center at the Intensive Care Unit (ICU) of a mid-sized European University Hospital. A patient simulator, consisting of a high-fidelity manikin (Meti, Medical Educational Technologies, Inc.) was used as patient. The manikin is life sized, can talk, displays realistic physiological features such as blinking and breathing, and it can be auscultated, defibrillated (applying an electrical shock in order to reset the electrical state of the heart and to allow the heart to establish a normal rhythm) and intubated, pulses are manually palpable at different physiological locations, a monitor displays vital signs. The room was equipped like a standard ICU patient room, including a patient monitor and a

medication tray. A defibrillator was placed in the room. The manikin's responses were controlled partly by the software and partly by an operator who changed patient reactions according to the protocol for this study and adapted to the team's interventions.

Study design, procedure, and experimental conditions

This is an experimental study comparing three different reflexivity conditions. Data across conditions were pooled for examining Hypothesis 2-8.

With the exception of the control group (N = 29 teams – data taken from an earlier study) (*reference omitted for blind review*), teams were randomly assigned to one of three reflexivity conditions: Task-focused reflexivity (n = 21), team-focused reflexivity (n = 22), or combination of task and team focused reflexivity (n = 25). The unequal samples size across conditions is due to the random assignment procedure. Sample size was a priori determined with a power analysis G*power (Faul et al., 2007), to test differences between experimental groups, an alpha of 0.05, achieved power of 0.80 and a large effect size.

Procedure

Upon arrival at the simulation center, team members were introduced to each other and received a short presentation related to the simulator manikin and its functioning, as well as the medical equipment in the room. The confederate study nurse who stayed in the room at all times was introduced to the team. The students were informed that the study nurse could be asked for information or for help. The nurse was instructed to give standard answers to questions and to help if asked, but not to intervene proactively. The teams were informed that they had to perform several tasks but did not get preliminary information about the type of tasks, that the task was an emergency situation, or that a reflexivity session would occur. Students were blind to the hypothesis and blind to different experimental conditions.

Each team had to perform a resuscitation task (task 1), followed by a reflexivity session, followed by a second resuscitation task (task 2).

Task 1: Cardiac arrest scenario. Task 1 started after the confederate nurse summoned the team from the hallway to the simulator room to help with a male patient lying on the hospital bed who had suddenly stopped talking and become unresponsive; the confederate did not suggest a diagnosis. The team entered the simulation room and had to diagnose the problem (a sudden cardiac arrest), realize the emergency character of the situation and start CPR. Independent of the actions of the team, the patient did not regain spontaneous circulation. After three minutes, a confederate interrupted the team and asked them to step to a desk.

Reflexivity session (task-focused; team-focused; task- and team-focused (combined)). After the interruption, the team was asked to engage in a reflexivity session. Each team member was handed a sheet containing instructions on how to reflect, participants could first individually study the sheet and then start a three-minute group discussion. The discussion was not facilitated by an external person. Reflexivity time started when a team member initiated the discussion and was interrupted after three minutes, however, teams were allowed to finish a topic under discussion. Teams in the control group *(reference omitted for peer review)* were asked to solve a filler task (interpreting an x-ray of another patient). This filler task was chosen to initiate a dialogue within the team, but prevented team members from reflecting about their earlier performance.

The written reflexivity instructions on the sheet were adapted from previous studies (Gurtner et al., 2007; Konradt et al., 2015; Schmutz, Kolbe, et al., 2018) and suggested the following temporal structure (i) to look back on prior performance (what went well, what did not go well) and (ii) to discuss how to proceed, should the same team have to do a similar task. (Kleiner et al., 2014; Widmer et al., 2009) Three different versions of the reflexivity instructions were provided and constituted the three ***reflexivity conditions***: Teams in the *task-focused reflexivity* condition were asked to concentrate the discussion on their adherence to the resuscitation algorithm and on their technical performance. Teams in the *team-focused*

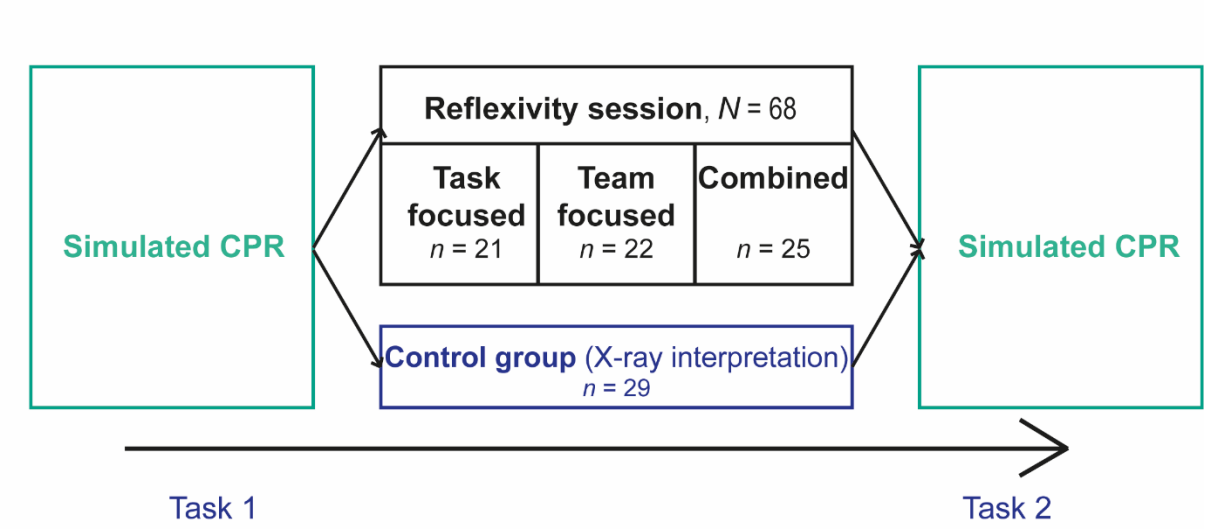
reflexivity condition were asked to specifically discuss aspects of collaboration, such as leadership, communication and team coordination, and teams in the *task and team combined* reflexivity condition were asked to reflect on both the algorithm and on collaborative aspects. For half of the teams in this condition, task aspects were mentioned in the instruction sheet before the team aspects and vice versa.

Task 2: Cardiac arrest scenario. Immediately after the reflexivity session, the team was asked to visit another male patient who, after an ambulatory stress ECG, had signs of a circulatory problem and was now resting on a bed. The team had to interview the patient in order to decide if he could be discharged. When the team arrived, the patient was responsive. During the interview, the patient complained of muscle aches from cycling (which was part of the ECG) and gave scripted answers to medical or demographic questions. Two minutes into the interview, the patient complained that he felt dizzy, and became unresponsive (closing eyes, breathing stopped, no palpable pulses). The attached monitor displaced ventricular tachycardia (an unsynchronized fast heartbeat); the team had to diagnose a cardiac arrest and to perform CPR. The rest of the session unfolded similar to the first task, with the exception that the patient regained spontaneous circulation and consciousness after around three minutes (normally after a defibrillation).

The procedure is illustrated in figure 1 below.

Figure 1

Procedure of the experiment



Note. The control group is not included in all analyses. *N* = 68/97.

Measures

All sessions (task 1, reflexivity session, task 2) were recorded using two cameras from different angles. All data are based on observational coding and transcriptions of the videos.

General resuscitation performance: Hands-on time and time to first meaningful measure

To assess ***resuscitation performance*** of task 1 and task 2, all acts of any team member related to resuscitation (e.g., cardiac massage, ventilation, and defibrillation) were coded second-by-second (cf. Marsch et al., 2005). Two performance measures were derived from this coding: (1) *Hands-on time* or *flow time* measures the percentage of time the team delivered medically sound treatment according to the resuscitation algorithm, if the patient did not have spontaneous circulation. Ventilation was only counted if performed as part of a cycle of ventilation-cardiac massage (ventilation without the "transport" of the oxygenated blood to the brain is ineffective), and a five second gap after each defibrillation was counted as hands-on time to allow the team to assess the effect of defibrillation. Because hands-on performance is a proportional measure, we used the arcsine-transformed score for statistical

analyses, but report percentages for easier understanding. (2) The *time to first meaningful measure* is the time (in seconds) elapsed since the start of the scenario (task 1) or the start of the non-responsiveness of the patient (task 2) and the first medically relevant intervention of the team (either ventilation, cardiac massage, defibrillation, or the intention to defibrillate). Both performance measures follow a fixed coding algorithm; thus, inter-coder reliability is not necessary.

Assessing task- or team-related topics discussed during the reflexivity session

We developed a coding procedure to assess the ***content discussed during the reflexivity sessions***. Team communication was transcribed word by word; the stream of communication was unitized by "spoken sentences" (Futoran et al., 1989; Tschan, Semmer, Vetterli, et al., 2011). Each sentence was then coded according to task-content categories and team-content categories. The coding system allowed for multiple codes for a given sentence. For instance, the statement "it is important that I don't miss the moment to resume ventilation – you should count aloud during cardiac massage" (Group N_17_04) was coded for both task-content (*Ventilation/cardiac massage) and* team-content (*coordination/communication*).

The 16 codes measuring task-related communication were derived from the resuscitation guidelines (Soar et al., 2015). They included communication referring to the (1) *resuscitation algorithm in general*, and (2) to *resuscitation cycles*, to *initial assessment of the patient* such as (3) checking the patient's consciousness, (4) breathing, (5) airways and (6) circulation; to aspects of *basic life support* including (7) cardiac rhythm identification (8) ventilation (9) cardiac massage, (10) synchronization between ventilation and cardiac massage, and (11) patient positioning for CPR; to aspects of *advanced life support* (12) defibrillation, and (13) technical aspects of the defibrillator, (14) medication, and (15) intubation or to (16) *other* task-related aspects. The five codes for team-related content were adapted from previous studies investigating team aspects in resuscitation (Cooper &

Wakelam, 1999; Fernandez Castelao et al., 2011; Hunziker et al., 2010; Tschan et al., 2006) and contained the categories (1) task distribution, (2) leadership, (3) communication, (4) coordination, and (5) (avoiding) interruptions. Communication not related to either task or team was coded as *unrelated communication*. Note that this coding does not include an assessment of the accuracy of the discussion content. Each transcript was independently coded by two coders (CM, SAH, JZ and FTS), differences were resolved by discussion.

Each sentence was then classified as containing task-related discussion (any of the task-related categories coded), team-related discussion (any of the team-related categories coded), both (task and team) or none. Sentences containing simple repetitions or acknowledgements were not counted. We then calculated the sum of task-related and team-related discussion content for each team. Because teams could finish a started topic after the three-minute time-limit for reflection, we adjusted all counts to represent 180 seconds of discussion.

We had to exclude two teams because more than 30% of the communication could not be coded. Reasons for exclusion was that one team talked extremely softly and could not be understood, and equipment failure.

Assessing the temporal sequence of the reflexivity session

Adherence to the temporal structure of the reflexivity session was rated by two independent coders (CM and FTS) who watched the whole reflexivity session and indicated the degree of adherence to the temporal structure on a scale from 0 to 3; 0 was coded if the team discussion did not follow the instruction regarding the temporal structure; 1 was coded if either the analysis of past performance or the planning for future tasks was present, but not both; 2 was coded if analysis of past performance and planning was present, but not in the suggested temporal structure or if teams flipped back and forward between past and future, and 3 was coded if the team followed the instruction.

Coding the quality of adherence to the resuscitation algorithm for task 1 and task 2

In order to relate the content discussed during the reflexivity session the group process in task 1 and task 2, we developed a coding system to measure the ***quality of adherence to the resuscitation algorithm during task 1 and task 2***. Codes were developed based on a task- and coordination requirement analysis of CPR (Shetty et al., 2009; Tschan, Semmer, Vetterli, et al., 2011). The codes related to patient diagnosis, basic and advanced life support based on the 2015 resuscitation algorithm published by the European Resuscitation Council (Soar et al., 2015); the codes for leadership and task distribution are based on publications evaluating nontechnical aspects in resuscitation teams (Cooper & Wakelam, 1999; Fernandez Castelao et al., 2011; Hunziker et al., 2010; Tschan et al., 2006). Twenty observational codes were defined. Unit of coding was the whole resuscitation task; each code was coded once for task 1 and task 2. The codes were then combined into four indices related to the resuscitation algorithm and two indices related to leadership and task distribution (for the extensive coding system, see **Appendix D** p. 231). The following indices were defined (A) The index d*iagnosis* of the patient was composed of five codes: (A.1) check consciousness of the patient, (A.2) check breathing, (A.3) check pulse, (A.4) control if airways are clear[5], and (A.5) ask for help. Each of these aspects were coded as 0 if none of the group members performed this task or if there were important deviations from the resuscitation algorithm and 1 if the task was correctly performed on the patient. The index was calculated as the mean of the five codes, range is thus between 0 and 1. (B) The index *basic life support* was composed of five codes: (B.1.) rhythm of cardiac massage (1 if the team performed cardiac massage with about 100 compressions per minute for more than 50%

[5] For task 2, the index was formed without this code – as the patient was awake and responsive at the beginning of the simulation and how he became unresponsive indicates that airway obstruction was not an issue and teams did not need to check the airway.

of the time they performed cardiac massage; 0 otherwise), (B.2.) counting compressions (1 was coded if at least the three last compressions were counted out loud, 0 otherwise), (B.3) avoiding interruptions (0 was coded if one or more interruptions of more than 10 seconds occurred between cardiac massage/ventilation, 1 if interruptions were less than 10 seconds), (B.4) ventilation (1 was coded if two ventilations were given between cycles of cardiac massage, 0 if less or more); (B.5.) overlap between ventilation and cardiac massage (0 was coded if there was overlap, 1 if not). The index was calculated as the mean of the five codes. (C) The index *defibrillation* was composed of three codes: (C.1) measured whether a warning command before defibrillation was given (1 if yes, otherwise 0), (C.2.) whether the second and following defibrillations occurred at the correct time (0 if too early/late; 1 if after 5 cardiac massage/ventilation cycles or two minutes), (C3) if the interruption to resuming CPR after an unsuccessful defibrillation attempt was less than 10 seconds (1, otherwise 0). The index was calculated as the mean of the three codes. (D) The index *medication* included two codes referring to whether (D.1) epinephrine was administered after the third defibrillation (1, otherwise 0), and (D2) amiodarone was administered (1, otherwise 0). The index was calculated as the mean of the two codes.

Two indices measured adherence to coordination requirements. The index *leadership* (E) contained the codes leadership identification (E.1). If a leader was clearly identified 1 was coded, otherwise 0; and (E.2.) if the leader does not engage in a too involving task (either works hands-off or ventilates) 1 was coded, otherwise 0. The index was calculated as the mean of the two codes. The index *task distribution* (F) assessed if team members took distinguished roles and contained three codes: One was coded if within the team it was clarified who performed ventilation (F.1), cardiac massage (F.2), and defibrillation (F.3), otherwise 0, respectively. The index was calculated as the mean of the tree codes.

Twenty percent of the material was coded independently by two coders (CM and RL), inter- coder reliability was between 0.72 and 1.0 (Cronbach's alpha).

Coding discussion content related to the resuscitation algorithm during the reflexivity session

To assess *which facets of the algorithm and the coordination requirements were discussed during the reflexivity session*, we used the same coding scheme than described above to assess the reflexivity session. For each of the codes, we coded 1 if the team talked about the topic during reflexivity and discussed these topics in accordance with the resuscitation guidelines. We coded 0 if the team did not talk about a coding category or if the discussion of the topic contradicted the resuscitation algorithm and was not corrected by the end of the reflexivity session (e.g., if a team member said during the reflexivity session that a break of 20 second was necessary after a defibrillation and this was not corrected by another team member, 0 was coded for defibrillation C.3; and if a team member said that identifying a leader was not necessary, 0 was coded for leadership E.1; however, if another team member corrected this contention, it was coded as 1). Twenty percent of the material was coded independently by two coders (CM and RL), inter coder reliability was between 0.71 and 1.0.

Coding unresolved disagreements

Unresolved disagreements in the reflexivity session was rated by two independent coders (CM and FTS) who read the transcripts and watched the whole reflexivity session and rated the degree of existing and not resolved disagreements within the team on a scale from 1 to 3; 1 was coded if there was no controversial discussion during the reflexivity session or if disagreements were clearly resolved; 2 was coded if there was one unresolved disagreement and 3 was coded if there were two or more disagreements. Two coders (CM and FTS) coded each of the reflexivity sessions independently and resolved disagreements through discussion.

Coding scope of reflexivity session

The *scope of the reflexivity session* was coded for the whole session on a scale from 1 to 3. The code assesses whether the team members explicitly relate the given task to other experiences or other situations or not, following the suggestion of Swift and West (1998) regarding the quality of reflexivity. Example for wider scope was a team that discussed resuscitation in different contexts (out of hospital cardiac arrest vs. in-hospital cardiac arrest, during the day or night; another team reflected on how a team composed of only two members should do a resuscitation). The code 1 was given if teams never extended the discussion to other situations; 2, if the team discussed one other situation and 3 if the team discussed more than one other situation. Two coders (CM and FTS) coded all reflexivity sessions independently; disagreements were resolved with discussion.

Results

Influence of different reflexivity instructions on resuscitation performance

Hypothesis 1 suggested that teams in the combined task- and team-reflexivity condition showed higher performance improvements than teams in the other two reflexivity conditions.

This was tested using a MANCOVA with task 2 *time to first meaningful measure* and the arcsine-transformed *hands-on performance* as dependent variables; task 1 time to first meaningful measure and the arcsine-transformed hands-on performance as covariates and the three reflexivity-conditions as between subject factor. Table 1 displays descriptive statistics. Multivariate test results indicate no statistically significant difference in performance in task 2 across the different reflexivity conditions when controlling for performances in task 1 $F(4, 124) = 1.694$, $p = .156$, Wilks' $\Lambda = .899$, partial $\eta 2 = .052$. Post-hoc pairwise comparisons based on the covariate-adjusted estimates showed no statistically significant difference between the experimental conditions with regard *to time to first meaningful measure* in task 2. However, for *hands-on performance* in task 2, teams in the team condition performed significantly better than teams in the task condition *($p = .044$)*; there was no difference between the task-condition and the combination-condition ($p = .635$) or the team-condition vs. combination condition ($p = .105$). Hypothesis 1 is thus only very partially supported.

As additional analyses we compared performance of the teams in the three reflexivity conditions with a control condition of 29 teams who performed the same resuscitation tasks but did not engage in a reflexivity session. Data for the control group stem from another study *(citation omitted for blind peer review)*. We estimated a MANCOVA with the performance measures task 2 as dependent variables (time to first meaningful measure; hands-on performance), performance measures at time 1 as covariates and the reflexivity and control conditions as between subject factors. Table 1 displays the descriptive statistics. The

multivariate test indicates statistically significant difference in performance in task 2 after controlling for performance in task 1 ($F(4, 180) = 3.343$, $p = .004$, Wilks' $\Lambda = .810$, partial $\eta 2 = .100$. Post-hoc pairwise comparisons based on covariate adjusted estimates revealed that for time to first meaningful measure, teams in all reflexivity conditions showed a significantly lower time to first meaningful measures (which indicates better performance) ($p_{task\ vs\ control} = .009$; $p_{team\ vs\ control} < .001$; $p_{combi\ vs\ control} = .036$). For hands-on time, teams in the team-condition ($p_{team\ vs\ control} = 0.001$) and teams in the combined condition ($p_{combi\ vs\ control} = .033$) showed significantly higher performance than the control condition; there was no significant difference between teams in the task- and the control condition ($p_{task\ vs\ control} = .279$).

Table 1

Descriptive statistics performance task 1 and task 2

		TMM: Time to first meaningful measure				**Hands-on** performance			
		Task 1		Task 2		Task 1		Task 2	
Condition	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Task condition	21	28.95	8.53	23.81	6.60	69.81	12.86	74.48	10.32
Team condition	22	33.05	16.49	22.37	8.82	66.25	11.00	78.47	9.52
Combined condition	25	37.88	15.66	27.08	9.66	65.50	11.74	75.76	7.15
Control group	29	47.31	17.38	33.83	12.60	60.04	16.90	68.63	13.68

Note: TMM: time to first meaningful measure in seconds (higher scores are lower performance); Descriptive statistics for hands-on performance are in percentages; statistical analyses are based on arcsine transformed variables. Control group data was used from another study (*citation omitted for peer-review*).

Table 2

Correlations among performance variables

Variable	*M*	*SD*	TMM1	TMM2	Hands-on (1)
TMM task 1	33.56	14.44	-		
TMM task 2	24.54	8.66	.149		
Hands-on task 1	67.07	11.84	-.464**	-.203	
Hands on task 2	76.25	9.01	.063	-.374**	.394**

Note: TMM: time to first meaningful measure in seconds (higher scores are lower performance); $N = 68$.

Do teams reflect according to the instruction: Time, topics and temporal sequence

For all analyses related to the reflexivity process itself we pooled the data across the three reflexivity conditions. Given the small performance differences across conditions, this is defendable, nevertheless, for all analyses, the conditions were entered as covariates, contrast condition was the combined condition.

Hypothesis 2a stated that teams would use the three minutes allotted for reflection. This hypothesis is supported by the data (table 3); only three teams reflected less than three minutes (2.3; 2.5, and 2.6 minutes respectively). The mean reflection time was somewhat higher than three minutes because teams were allowed to finish an ongoing discussion related to a specific topic. Reflection duration were slightly different across condition with teams in the combined condition reflecting longest ($F(2, 65) = 4.019$, $p = .021$)). Post-hoc analyses

showed that reflection duration is significantly higher for teams in the combined condition than for teams in the task condition.

Table 3

Duration of reflexivity

Reflexivity condition	*N*	Minutes reflected	SD
Task	22	3.73^{ab}	0.39
Team	21	3.76^{a}	0.95
Combination	25	4.33^{b}	0.95

Note. Time is expressed in minutes and decimals of minutes. Combination = combined task-team condition. Post hoc comparisons between conditions are based on Dunnett T3 correction. Means with the same subscript within a column are not significantly different from each other; $N = 68$ teams.

We expected (H2b) that groups would not digress from discussing CPR related topics, arbitrarily defined by a limit of 20% of communication unrelated to the expected content. This hypothesis was supported by the data. Communication of aspects unrelated to resuscitation was low (1.73 per cent; SD = .37); with 37 teams showing zero unrelated communication; the highest percentage of unrelated communication was 14.5 per cent. There was no significant difference between the experimental conditions ($M_{task\ condition} = 1.41$, SD = 2.66; $M_{team\ condition} = 2.396$, SD = 3.91; $M_{combi\ condition} = 1.700$, SD = 2.29) ($F(63, 2) = 0.867$, F63, 2 = 0.867; $p = .425$ ($N = 66$).

Hypothesis 3 stated that teams reflect according to the focus given during the instruction (discuss task/team/both aspects). Table 4 shows the results that support the hypothesis. A MANOVA was used to compare task-related and team-related contents across conditions. The multivariate result was significant for condition, Pillai's Trace = 0.624, $F(4, 126) = 14.396$, $p < .001$; differences across conditions were significant for task-related topics discussed $F(2) = 16.62$, $p < .001$; partial adjusted $R^2 = 0.33$ and for team-related topics discussed $F(2) = 19.696$, $p < .001$); partial $R^2 = 0.37$. Post-hoc comparisons showed that groups in the team condition discussed significantly more *team-related* contents than groups in the task or the combined conditions, the latter two were not significantly different. For

task-related contents, groups in the task condition discussed significantly more task related aspects than groups in the team or combined condition, the latter two were not significantly different (Table 4, Figure 2).

Table 4

Amount of task-related and team related content discussed during the reflexivity session across the experimental conditions

Reflexivity condition	*N*	Task-related content discussed (mean) [a]	SD	Team-related content discussed (mean) [a]	SD
Task	21	38.69 $_{b}$	8.35	7.51	5.23
Team	21	23.62	7.19	18.57 $_{b}$	7.10
Combination	24	35.53 $_{b}$	10.72	15.03 $_{b}$	5.11

Note. Combination = combined task-team condition. Post hoc comparisons between conditions are based on Dunnett T3 correction. Means with the same subscript within a column are not significantly different from each other*; N* = 66 teams.

[a] Estimated for a duration of three minutes.

Figure 2

Mean amount of communication related to task (blue bar) and team (red bar) during reflection

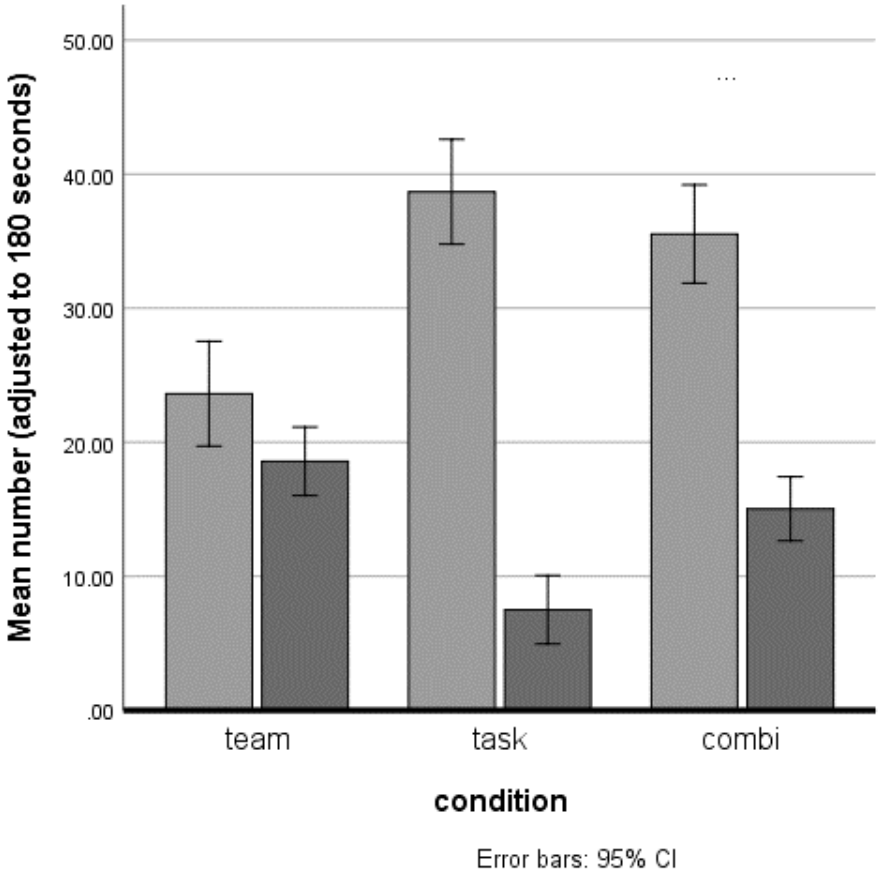


Figure 2 illustrates that although teams followed the instructions, in all conditions, teams focused more on task-related aspects during reflexivity. In an additional analysis, we thus tested this observation with a repeated measurement ANOVA entering team- and task-related discussion as within factor and condition as between factor. Results showed that although the teams overall discussion density was not significantly different (p = .057), teams discussed significantly more task-related than team-related aspects in all conditions (task-condition*:* p < .001; team-condition: p = .031; combined-condition: p < .001). There was also a significant interaction effect (see Figure 2), indicating that the difference between task-related and team-related communication is significantly different between at least two

conditions. A further exploration of the interaction showed that the difference between task- and team-related communication was significantly larger in the task versus the team condition $p < .001$; and the combined condition ($p = .002$).

Hypothesis 4 suggested that the teams in all conditions follow the instruction with regard to the temporal sequencing of the reflexivity discussion (discuss about past – plan for similar performance). The coding ranged from no (0) to optimal (3) temporal sequencing, we arbitrarily defined that following the temporal instructions as 2.5 as threshold. The hypothesis was tested with an ANOVA. Across all conditions, the score evaluating the temporal structure of the reflexivity session was 2.00, SD = 0.86; 95% CI = 1.84-2.61; N = 68 teams. Overall, this is not significantly different from the predefined 2.5 threshold. There were not significant differences between the conditions (task-condition: M=1.90, SD = 0.89, 95% CI = 1.50-2.31; team condition: M=2.23, SD 0.87, 95% CI = 1.84-2.61; combi condition: M=1.88, SD=0.83, 95% CI = 1.54-2.23). Interestingly, however, teams in the task and combi condition were significantly below the defined threshold.

Additional analysis: Relationship of team and task communication and temporal sequence of reflexivity on performance

As additional analysis, we evaluated whether the amount of task-related content and team-related content during reflexivity was related to *hands-on performance* or *time to first meaningful performance measure*. Analyses were hierarchical multiple regressions on hands-on performance task 2 (results see table 5), and on time to first meaningful intervention task 2 (Table 6). In the first model, we entered the covariates (performance task1 and experimental conditions) in the second model we added as predictors the amount of task- and amount of team-related content discussed during reflexivity. We also evaluated whether teams that followed the temporal structure of reflexivity closer showed higher performance (see tables 7 and 8). The analyses show that, controlling for condition and earlier performance, neither the

amount of task related or team-related communication, nor the temporal structure of reflexivity predicted performance.

Table 5

*Regression predicting **hands-on performance** by amount of task- and team-related communication during the reflexivity session*

Hands-on performance task 2	B	95% ci		β	t	P
Model 1: Control variables						
Intercept	0.615					
Hands-on task 1	0.349	0.001	0.159	0.420	3.665	0.001
Condition: team	0.049	0.190	-0.025	0.169	1.325	0.190
Condition: task	-0.037	0.330	-0.111	-0.127	-0.982	0.330
	$F(3, 62) = 5.658$; R2 adj = 0.177, $p = .002$					
Model 2 Predictor variables						
Task-related communication during reflexivity	-0.002	-0.007	0.004	-0.090	-0.577	0.566
Team-related communication during reflexivity	0.001	-0.003	0.004	0.056	0.387	0.700
	$\Delta F(2, 61) = 0.206$; ΔR2 = 0.005, $p = .815$					

Note. $N = 66$; hands-on performance was arcsine transformed, experimental conditions are entered as dummy-variables (contrast condition is combination). Task- and team-related reflexivity are during the reflexivity session and were adjusted to 180 seconds. In model 2, control variables are not separately reported.

Table 6

*Regression predicting **first meaningful measure performance** by amount of task- and team related communication during the reflexivity session*

Time to first meaningful measure task 2	B	95% ci		β	t	P
Model 1: Control variables						
Intercept	25.744					
FMM task 1	0.052	-0.097	0.201	0.089	0.699	0.487
Condition: team	-5.026	-10.150	0.097	-0.274	-1.961	0.054
Condition: task	-3.446	-8.693	1.801	-0.188	-1.313	0.194
	$F(3, 62) = 1.782$; R2 adj = 0.035, $p = .160$					
Model 2 Predictor variables						
Task-related communication during reflexivity	-0.143	-0.519	0.233	-0.122	-0.759	0.451
Team-related communication during reflexivity	-0.202	-0.442	0.038	-0.256	-1.684	0.097
	$\Delta F(2, 61) = 2.026$; ΔR2 = 0.058: $p = 0.141$					

Note. $N = 66$; first meaningful measure: Lower scores indicate higher performance; experimental conditions are entered as dummy-variables (contrast condition is combination). Task- and team-related communication are during the reflexivity session and were adjusted to 180 seconds. In model 2, control variables are not separately reported.

Table 7

*Regression predicting **hands-on performance** by adherence to temporal structure during the reflexivity session*

Hands-on performance task 2	B	95% ci		β	t	P
Model 1: Control variables						
Intercept	0.615					
Hands-on task 1	0.349	0.001	0.159	0.420	3.665	0.001
Condition: team	0.049	0.190	-0.025	0.169	1.325	0.190
Condition: task	-0.037	0.330	-0.111	-0.127	-0.982	0.330
	$F(3, 62) = 5.658$, R2 adj = 0.177, $p = .002$					
Model 2 Predictor variable						
Temporal structure	-0.014	-0.050	0.023	-0.088	-0.760	0.450
	$\Delta F(2, 61) = 0.578$, ΔR2 = 0.007, $p = .450$					

Note. $N = 66$; Hands-on performance was arcsine transformed, experimental conditions are entered as dummy-variables (contrast condition is combi). Temporal structure during reflexivity was coded as 0 (no adherence) to 3 (complete adherence). In model 2, control variables are not separately reported.

Table 8

*Regression predicting **first meaningful measure performance** by adherence to temporal structure during the reflexivity session*

Time to first meaningful measure task 2	B	95% ci		β	t	P
Model 1: Control variables						
Intercept	25.744					
FMM task 1	0.052	-0.097	0.201	0.089	0.699	0.487
Condition: team	-5.026	-10.150	0.097	-0.274	-1.961	0.054
Condition: task	-3.446	-8.693	1.801	-0.188	-1.313	0.194
	$F(3, 62) = 1.782$, R2 adj = 0.035, $p = .160$					
Model 2 Predictor variables						
Temporal structure	-1.228	-3.734	1.279	-0.122	-0.979	0.331
	$\Delta F(2, 61) = 2.026$, ΔR2 = 0.058 $p = .141$					

Note. $N = 66$; first meaningful measure: Lower scores indicate higher performance; experimental conditions are entered as dummy-variables (contrast condition is combi). Temporal structure during reflexivity was coded as 0 (no adherence) to 3 (complete adherence) In model 2, control variables are not separately reported.

Do teams address their weaknesses during the reflexivity session (and does this help?)

Hypothesis 5 suggested that performance in sub-tasks during the first resuscitation influenced the discussion during the reflexivity session in the sense that for sub-tasks teams had weaknesses in task 1 (expressed as low adherence to the resuscitation algorithm), the discussion during the reflexivity session is more complete; in short take the opportunity to discuss their weaknesses.

Table 9 shows means and standard deviations for all indices of *adherence to the resuscitation algorithm* for task 1, the reflexivity session, and task 2, as well as results of a paired t-test comparing adherence to algorithm for task 1 to task 2; figure 3 shows a summary of the results for the different indices. Table 10 shows the results for separate multiple regression analyses on different reflexivity contents. Main predictor was adherence to algorithm in task 1, the team and task condition was entered as a control variable, thus, the combined team-task condition was the contrast condition.

Table 9

Descriptive results for adherence to algorithm indicators, as well as results of task 1 and task 2 comparison

Index	Task 1 *M*	*SD*	Reflexivity *M*	*SD*	Task 2 *M*	*SD*	T (65) t1 vs t2	P
Diagnosis	0.558	0.194	0.270	0.254	0.557	0.266	.022	.982
Basic life Support	0.762	0.229	0.167	0.174	0.855	0.197	-3.446	<.001
Defibrillation	0.646	0.278	0.141	0.203	0.672	0.237	-.712	.479
Medication	0.182	0.242	0.280	0.318	0.280	0.265	-2.610	.011
Leadership	0.220	0.352	0.538	0.423	0.424	0.385	-3.717	<.001
Task distribution	0.424	0.312	0.182	0.305	0.439	0.332	-.280	.170

Note. N = 66; Two teams did not attempt defibrillation at t1; thus, for the index defibrillation N = 64. TMM: time to first meaningful measure in seconds (higher scores are lower performance); Descriptive statistics for hands-on performance are in percentages; F and P values are based on arcsine transformed variables.

Table 10

Separate regression analyses predicting discussion of content during the reflexivity session separately for each index based on adherence to algorithm for the respective index, controlling for experimental conditions

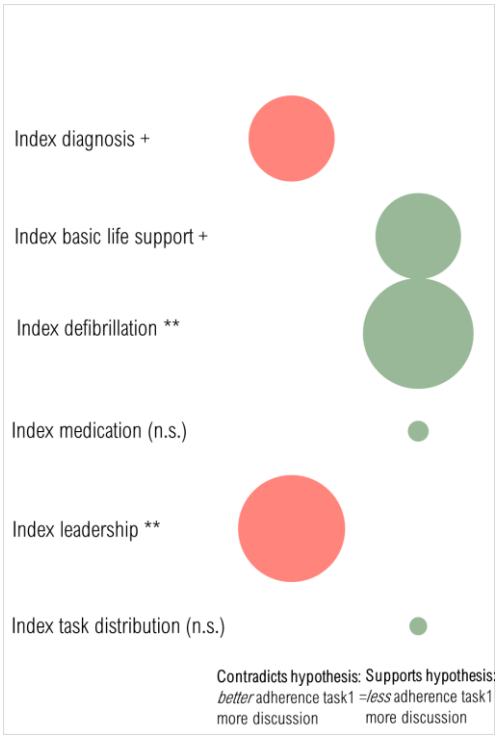
DV Index **Diagnosis** reflexivity	B	95% CI low	95% CI high	t	P	β
Intercept	0.228					
Diagnosis task 1	0.227	-0.19	0.574	1.872	.066	.211
Condition: task	-0.103	-0.241	0.035	-1.489	.142	-.190
Condition: team	-0.254	-0.392	-0.116	-3.672	.001	-.468
	$F(3, 62) = 5.546$R2 adj = 0.173, $p = .002$					
DV Index **Basic life support** reflexivity	B	95% CI low	95% CI high	t	P	β
Intercept	.315					
Basic life support task 1	-.160	-.349	.029	-1.696	.095	-.211
Condition: task	-.020	-.123	.083	-.386	.701	-.054
Condition: team	-.061	-.164	.042	-1.186	.240	-.165
	$F(3, 62) = 1.614$; R2 adj = .0.028, $p = .195$					
DV Index **Defibrillation**[a] reflexivity	B	95% CI low	95% CI high	t	P	β
Intercept	.313					
Defibrillation task 1	-.257	-.435	-.078	-2.881	.005	-.349
Condition: task	-.015	-.135	.106	-.245	.804	-.033
Condition: team	-.007	-.124	.110	-.126	.900	-.017
	$F(3, 60) = 2.851$;R2 adj = .081 $p = .045$					
DV Index **Medication** reflexivity	B	95% CI low	95% CI high	t	P	β
Intercept	.294					
Medication task 1	-.012	-.347	.323	-.072	.943	-.009
Condition: task	.040	-.155	.236	.412	.682	.060
Condition: team	-.078	-.270	.115	-.808	.422	-.115
	$F(3, 62) = .197$; R2 adj = 0, $p = .686$					
DV Index **Leadership** reflexivity	B	95% CI low	95% CI high	t	P	β
Intercept	.557					
Leadership task 1	.326	.076	.576	2.609	.011	.271
Condition: team	-.429	-.637	-.221	-4.121	<.001	-.475
Condition: task	.142	-.070	.353	1.341	.185	.157
	$F(3, 62) = 11.407$; R2 adj = .324, $p < .001$					
DV Index **Task distribution** reflexivity	B	95% CI low	95% CI high	t	P	β
Intercept	.178					
Task distribution task 1	.037	-.208	.283	.304	.762	.038
Condition: task	-.084	-.268	.099	-.919	.362	-.130
Condition: team	.046	-.138	.229	.494	.623	.070
	$F(3, 62) = .650$; R2 adj = -.016, $p = .586$					

$N = 66$; Experimental conditions are entered as dummy-variables (contrast condition is combi).

[a]Two teams did not attempt defibrillation at t1; thus, for the index defibrillation $N = 64$.

Figure 3

Summary of the effects of adherence to algorithm task 1 on subsequent discussion during the reflexivity session



Note. Diameters of circles represent standardized β, n.s. not significant effect; Red circles: contradict the hypothesis, green circles support the hypothesis.

*, ** significant effect; + trend ($p < .10$).

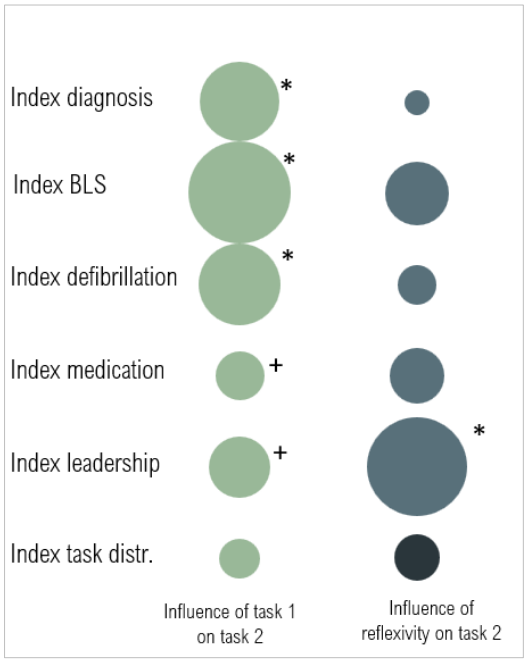
The regression analyses do not support the hypothesis, or at least do not support the hypothesis for all contents (Figure 3). Suboptimal adherence to the algorithm related to basic life support (as a trend) and to defibrillation a is related to a more thorough discussion of these topics whereas better adherence to algorithm in task one for diagnosis (trend) and for leadership leads to more thorough discussion of those topics during the reflexivity sessions. For medication and task distribution, no significant effects were found.

Influences of topics reflected about on subsequent adherence to algorithm

Hypothesis 6a suggested that the discussion of sub-tasks influenced *adherence to the algorithm* for the same sub-tasks in task 2. This hypothesis was tested with separate multiple regression analyses for each indicator and discussion of algorithm during the reflexivity as predictor, adherence of algorithm task 1, and the experimental conditions as control variables. Results are displayed in table 11, and summarized in Figure 4.

Figure 4

Summary and graphical representation of the influence of task 1 and reflexivity on task 2 adherence to algorithm



Note. Diameters of circles represent standardized β of adherence to algorithm task 1 (left) and discussion of algorithm during reflexivity (right) on adherence to algorithm task 2, after controlling for experimental conditions; * significant effect; + trend ($p < .10$). Dark circles represent negative β.

Table 11

Results of the regression analyses predicting adherence to algorithm on task 2 based on discussing the content during reflexivity and adherence to algorithm task 1

DV Index **Diagnosis** task 2	B	95% CI low	95% CI high	t	P	β
Intercept	0.228					
Diagnosis reflexivity	-.022	-.304	.236	-.154	.878	-.021
Diagnosis task 1	.452	.113	.791	2.667	.010	.359
Condition: task	.070	-.086	.227	.896	.374	.124
Condition: team	-.029	-.199	.140	-.345	.732	-.052
	$F(4, 61) = 2.143$; R2 adj = 0.066, $p = .086$					
DV Index **Basic life support** task 2	B	95% CI low	95% CI high	t	P	β
Intercept	.503					
Basic life support reflexivity	.154	-.107	.414	1.180	.243	.136
Basic life support task 1	.436	.238	.634	4.403	<.001	.596
Condition: task	-.021	-.127	.085	-.398	.692	-.050
Condition: team	.001	-.105	.108	.023	.981	.003
	$F(4, 61) = 5.156$; R2 adj = .205, $p = .001$					
DV Index **Defibrillation**[a] task 2	B	95% CI low	95% CI high	t	P	β
Intercept	.499					
Defibrillation reflexivity	.060	-.245	.364	.392	.696	.051
Defibrillation task 1	.333	.109	.558	2.973	.004	.384
Condition: task	.017	-.125	.160	.241	.811	.033
Condition: team	-.019	-.157	.119	-.270	.788	-.036
	$F(4, 59) = 2.455$; R2 adj = .085 $p = .056$					
DV Index **Medication** task 2	B	95% CI low	95% CI high	t	P	β
Intercept	.289					
Medication reflexivity	.016	-.188	.220	.159	.874	.102
Medication task 1	.263	-.006	.533	1.954	.055	.135
Condition: task	-.112	-.269	.046	-1.416	.162	.079
Condition: team	-.081	-.236	.075	-1.038	.303	.078
	$F(4, 61) = 1.818$; R2 adj = 0.048, $p = .137$					
DV Index **Leadership** task 2	B	95% CI low	95% CI high	t	P	β
Intercept	.263					
Leadership reflexivity	.303	.055	.550	2.444	.017	.333
Leadership task 1	.231	-.026	.488	1.799	.077	.211
Condition: team	-.150	-.379	.079	-1.312	.194	-.183
Condition: task	-.014	-.223	.195	-.131	.896	-.017
	$F(4, 61) = 5.653$; R2 adj = .223 $p < .001$					
DV Index **Task distribution** task 2	B	95% CI low	95% CI high	t	P	β
Intercept	.501					
Task distribution reflexivity	-.078	-.347	.190	-.584	.562	-.072
Task distribution task 1	.098	-.161	.358	.758	.451	.093
Condition: task	-.236	-.431	-.040	-2.409	.019	-.334
Condition: team	-.043	-.238	.152	-.444	.659	-.061
	$F(4, 61) = 1.677$; R2 adj = .040, $p = .167$					

$N = 66$; Experimental conditions are entered as dummy-variables (contrast condition is combi).

[a]Two teams did not attempt defibrillation at t1; thus, for the index defibrillation, $N = 64$.

The analyses do not support hypothesis 6a. Discussing specific topics during the reflexivity session does not significantly influence adherence to algorithm in task 2, with the exception of discussion of leadership. The results also evidence that performance in task 1 was a strong predictor of performance in task 2: adherence to algorithm in task 1 significantly predicted task 2 adherence to algorithm for diagnosis, basic life support and defibrillation, as a trend ($p < .10$) for medication and leadership (Figure 4 and Table 11).

Hypothesis 6b suggested that engaging in reflexivity is particularly useful if prior task performance was low. We expected that the influence of reflexivity on task 2 *adherence to algorithm* is particularly strong if adherence to algorithm was low in task 1. This hypothesis was tested running separate moderation analyses, controlling for the experimental condition, using the process macro of Hayes (Hayes, 2017); results are displayed in table 12.

The analyses do not support Hypothesis 6b, none of the interaction effects was significant. Thus, teams that reflected on aspects of the algorithm they were weak during task 1 did not have a higher increase in adherence to algorithm in task 2 than other teams.

Table 12

Results of moderator analysis testing if discussing algorithm-related content during reflexivity enhances adherence to algorithm in task 2 particularly for low adherence to algorithm in task 1

DV Adherence to algorithm **Diagnosis** task 2	B	95% ci Lo	95% ci hi	t	P
Intercept	.333				
Index Diagnosis Reflexivity	-.158	-.892	.577	-.430	.669
Index Diagnosis task 1	.383	-.100	.866	1.586	.118
Diagnosis task1*reflexivity	.237	-.942	1.415	.402	.689
Condition: task	.072	-.086	.230	.916	.363
Condition: team	-.026	-.197	.146	-.298	.767
	$F(5, 60) = 1.723$; R2 = 0.126, $p = .143$				
DV Adherence to algorithm **Basic life support** task 2	B	95% ci Lo	95% ci hi	t	P
Intercept	.498				
Index Basic life support Reflexivity	.177	-.644	.998	.431	.668
Index Basic life support task 1	.443	.147	.739	2.991	.004
Basic life support task1*reflexivity	-.034	-1.159	1.091	-.060	.952
Condition: task	-.021	-.128	.086	-.392	.696
Condition: team	.001	-.106	.109	.020	.984
	$F(5, 60) = 4.082$; R2 =254, $p = .003$				
DV Adherence to algorithm **Defibrillation**[a] task 2	B	95% ci Lo	95% ci hi	t	P
Intercept	.511				
Index Defibrillation Reflexivity	-.203	-.754	.347	-.740	.462
Index Defibrillation task 1	.245	-.027	.517	1.801	.077
Defibrillation task1*reflexivity	.523	-.388	1.434	1.149	.255
Condition: task	.006	-.138	.149	.078	.938
Condition: team	-.022	-.160	.115	-.325	.746
	$F(5, 58) = 2.238$; R2 =.162, $p = .062$				
DV Adherence to algorithm **Medication** task 2	B	95% ci Lo	95% ci hi	t	P
Intercept	.297				
Index Medication Reflexivity	-.029	-.296	.239	-.215	.831
Index Medication task 1	.204	-.149	.557	1.154	.253
Medication task1*reflexivity	.229	-.643	1.102	.526	.601
Condition: task	-.100	-.264	.065	-1.213	.230
Condition: team	-.078	-.235	.079	-.999	.322
	$F(5, 60) = 1.493$; R2 =.111, $p = .206$				
DV Adherence to algorithm **Leadership** task 2	B	95% ci Lo	95% ci hi	t	P
Intercept	.243				
Index Leadership Reflexivity	.350	.060	.640	2.416	.019
Index Leadership task 1	.355	-.112	.823	1.520	.134
Leadership task1*reflexivity	-.204	-.844	.436	-.638	.526
Condition: task	-.149	-.379	.081	-1.295	.200
Condition: team	-.024	-.237	.188	-.299	.819
	$F(5, 60) = 4.559$; R2 =.275, $p < .001$				

(Table 12 continued)

DV Adherence to algorithm **Task distribution**	B	95% ci Lo	95% ci hi	t	P
Intercept	.550				
Index task distribution Reflexivity	-.246	-.697	.205	-1.090	.280
Index task-distribution task 1	.015	-.301	.331	.095	.924
task-distribution task1*reflexivity	.370	-.430	1.170	.925	.359
Condition: task	-.250	-.448	-.052	-2.520	.014
Condition: team	-.070	-.274	.134	-.689	.493
	$F(5, 60) = 1.509$; R2 =.112, $p = .201$				

Note. $N = 66$; Predictors and moderators were not centered, because 0 is an intrinsically meaningful for both the independent variable and the moderator. Experimental conditions are entered as dummy-variables (contrast condition is combi). Moderation analyses are performed using the process macro of Hayes (Hayes, 2017).

[a]Two teams did not attempt defibrillation at t1; thus, for the index defibrillation, $N = 64$.

Interactional aspects and quality of reflection as predictors of performance?

Hypothesis 7 suggested that if content disagreements are not resolved during the reflexivity session, performance could suffer. Hypothesis 8 suggested that if teams discuss a wider scope, they perform better. Table 13 shows the descriptive statistics for *unresolved disagreements* and *scope of discussion* overall and across the experimental conditions, table 14 shows the results of the regression analyses separately for hands-on performance and time to first meaningful measure, controlling for performance task 1 and experimental conditions.

Table 13

Descriptive statistics for disagreements and scope of reflexivity across conditions

Reflexivity condition	*N*	Disagreements (1-3)	SD	Scope of reflexivity (1-3)	SD
Task	21	1.14	.359	1.62	.740
Team	21	1.29	.463	1.43	.598
Combi.	24	1.17	.381	1.46	.658
Overall	66	1.20	.401	1.50	.662

Note. Combi = combined task-team condition. No significant differences between the experimental conditions for disagreements or scope of reflexivity; $N = 66$ teams.

Table 14

Separate regression analyses predicting resuscitation performance (hands-on, time to first meaningful measure) for disagreements during reflection and scope of reflexivity; controlling for performance task 1 and for experimental conditions

DV ***Hands-on performance task 2*** (arcsine-transformed)	B	95% CI low	95% CI high	t	P	β
Intercept	.614					
Disagreements	-.085	-.159	-.012	-2.340	.023	-.253
Scope of reflexivity	.046	.002	.090	2.106	.039	.226
Hands-on performance task 1	.359	.177	.540	3.956	<.001	.431
Condition team	-.026	-.096	.045	-.726	.470	-.089
Condition task	.039	-.031	.109	1.120	.267	.135
	$F(5, 60) = 5.725$; R2 adj =.267, p < .001					
DV ***Time to first meaningful measure task 2***	B	95% CI low	95% CI high	t	P	β
Intercept	29.187					
Disagreements	1.945	-3.156	7.045	.763	.449	.090
Scope of reflexivity	-4.020	-7.090	-.949	-2.619	.011	-.309
Time to first meaningful measure task 1	.056	-.087	.200	.782	.437	.095
Condition team	-3.760	-8.831	1.310	-1.483	.143	-.205
Condition task	-4.312	-9.262	.638	-1.742	.087	-.235
	$F(5, 60) = 2.615$; R2 adj =.111, p = .033					

Note. $N = 66$.

The analyses partially support Hypothesis 7 and fully support Hypothesis 8. More unresolved disagreements during the reflexivity session is related to lower hands-on performance in task 2; and a wider scope of reflexivity is related to higher hands-on performance in task 2 as well as lower time to first meaningful measure.

Discussion

This study had two goals – to compare different reflexivity instructions with each other with regard to performance, and to "dissect" the reflexivity procedure in order to respond to some of Moreland and McMinn's (2010) reserves with regard to reflexivity (for a summary of the results, see table 15).

Table 15

Verbal Summary of all results

Hypothesis	Results
Does reflexivity influence (resuscitation) performance?	
H1 - Teams that are asked to reflect about task aspects as well as team aspects (combination condition) show higher performance improvements than teams that only reflect on one of those aspects	• Time to first meaningful measure: No significant differences among experimental conditions • Hands-on time: team condition < task condition
Additional result Teams that reflect show higher performance than teams in control group	• Time to first meaningful measure: Each reflexivity condition > control condition • Hands-on time: Team and combined condition > control condition
Do teams follow the reflexivity instructions?	
H2a - teams engage in reflexivity	• Yes; teams use allotted time to reflect
H2b - teams discuss topics related to the task	• Yes; very low percentage of unrelated talk
H3 - teams respect the instructed thematic focus	• Talking about task-related aspects: o Task-condition = combined condition > team-condition • Talking about team-related aspects: o Team condition = combined condition > task-condition
Additional result	• In all conditions: more task-related aspects than team-related aspects discussed
H4 - teams follow the temporal structure of reflexivity	• Yes; teams follow the temporal structure • No differences between conditions
Additional analyses – the relationship of content and temporal structure of the reflexivity on team performance	• The amount of task-related or of team-related communication during reflexivity is not related to time of first meaningful measure • The amount of task-related or of team-related communication during reflexivity is not related to hands-on performance • Following the temporal structure of reflexivity is not related to higher performance (first meaningful measure; hands-on)
Process performance (improvements)	
H5 - When reflecting, teams address their weaknesses within the resuscitation algorithm	• Mixed: o Leadership (sig.) and diagnosis (trend) are more discussed if adherence was high in task 1 o Defibrillation (sig.) and basic life support (trend) are more discussed if adherence was low in task 1 o No differences for medication and task distribution
H6a - Adherence to algorithm for topics discussed during reflexivity will be higher in task 2	• No, with the exception of discussion of leadership
Additional result	• Previous process performance influences process performance in task 2. (adherence to algorithm task 1) is a significant predictor for process performance in task 2 for diagnosis, basic life support, defibrillation; is a trend for medication and leadership.
H6b - If aspects of low process performance in task 1 are addressed during reflexivity process, performance for those aspects will be particularly improve in task 2	• No
Quality of the reflexivity process	
H7 - Unresolved disagreements during the reflexivity session hamper resuscitation performance	• Partially: Higher unresolved disagreements lower improvement of hands-on performance in task 2; but not first meaningful measure
H8 - A larger scope during the reflexivity session is related to more resuscitation performance improvement	• Supported: A larger scope during the reflexivity discussion enhances hands-on performance gains, and decrease time to first meaningful measure.

The study results indicate that reflexivity-as-intervention has positive effects on performance, and teams in the team-condition – who were asked to reflect on coordination aspects, showed more pronounced effects on performance. With regard to the reflexivity session itself, contrary to Moreland and McMinn's (2010) study and concerns, teams engaged in reflexivity when asked to, they also discussed according to the focus of the experimental conditions, and they were able to respect the suggested temporal sequence of first looking back and then planning for future performance.

However, three important assumptions of how reflexivity influences performance were at best partially supported: Teams did not consistently focus during reflexivity on their weaknesses in task one. In addition, discussing specific aspects of the task during reflexivity did not improve performance in task 2 regarding those specific aspects – with the exception of leadership. Finally, there were no indications that team that addressed their specific weaknesses improved performance with regard to those specific weaknesses.

In accordance with other studies, we also found that the quality of reflexivity in terms of scope of reflection was rather low, but influenced performance. Similarly, unresolved conflicts were rare, but could hamper performance.

We discuss each of these aspects below.

Effects of different reflexivity interventions on performance were the main goal of the study. Again, note that in all analyses, task 1 performance was controlled for. Contrary to our expectation, teams in the "combination" condition who reflected on task- as well as on collaborative aspects did not outperform teams in the other two conditions, and differences between the conditions were only found for hands-on time performance, but not for time to first meaningful measure.

We first discuss the finding for hands-on time because it constitutes the more general performance measure. We had derived the hypothesis that a combined team and task focus

would be most useful for performance from team training research which found that a combined task-collaboration instruction was more successful than only a task or collaboration instruction (Salas et al., 2008; Salas et al., 2007). That makes sense, as in a given task, collaboration requirements are often different for different sub-tasks: For some sub-tasks, task distribution among team members may be optimal and cooperation will be loose, but for other sub-tasks teams have to engage in a tight cooperation. Undercollaboration – when the task requires tight collaboration and the team does not well collaborate, but also overcollaboration – when task distribution would be more optimal than tight collaboration within the group, can hamper performance. For the resuscitation task studied, tight coordination is particularly important for defibrillation (because all team members have to interrupt working on the patient and stay away for the moment of the shock). Collaboration requirements for the other resuscitation sub-tasks do not necessarily require the whole team to focus on the same task (Tschan, Semmer, Hunziker, et al., 2011). Indeed, a study found that overcollaboration during a simulated resuscitation was observed in many teams (e.g., all team members collaborated to resolve technical issues of the defibrillator and leaving the patient without support during a considerable amount of time) and led to poorer performance (Tschan, Vetterli, et al., 2011). Thus, asking teams to reflect on task and team aspects simultaneously may help them to integrate the two aspects. We have several explanations why we did not find the expected result. First, it could be that the three minutes reflection period was too short to discuss links between task and collaboration aspects thoroughly. An indicator of this is that teams in the combined condition asked more often for extra time to finish a topic than the teams in the other conditions. Second, it is well known that resuscitation knowledge wanes rather quickly, within months (Berden et al., 1993). It could be that teams in the combined condition primarily focused on task-aspects (see below) to first update their knowledge and that team aspects did not get the necessary attention. Third, for

known tasks, discussion of team aspects might be the most useful for enhancing performance, as our results suggest. This argument is corroborated by results from two other studies asking teams to reflect on team aspects and finding positive results on task performance (Eddy et al., 2013; Konradt et al., 2015). Note, however, that none of these studies tested against a task-focused reflexivity condition. Previous studies also found that even generic team-skill training improved task performance on a collaborative technical task, because generic team-skills training positively influenced emergent states such as transactive memory systems which subsequently improved task performance (Prichard & Ashleigh, 2007).

For the performance measure time to first meaningful measure, there was no difference between experimental conditions. For a CPR, rapid intervention is crucial for patient survival. The guidelines allow for an assessment of the patient of no longer than 10 seconds before starting CPR, arguing that immediate intervention is important and that even trained medical "cannot assess the breathing and pulse sufficiently reliably to confirm cardiac arrest" (Soar et al., 2015. p. 105), and that the risk of cardiac massage or ventilation on a patient not in cardiac arrest are not likely to cause harm. A swift start is thus crucial, and resuscitation team performance should be evaluated in this regard. One explanation for the lack of differences between the reflexivity conditions for swift start is that there may have been a ceiling effect; the mean time to first meaningful measure was fast (ca. 25 seconds). Although 25 seconds is more than twice the recommended time, in comparison to other studies, teams in this study started fast: One study reports a mean of 50 seconds (Amacher et al., 2017), another a mean of 77 seconds for medical student samples (Luscher et al., 2010); another study reports 44 seconds for a second resuscitation task after an expert-led debriefing (Hunziker et al., 2010). Twenty-five seconds is a relatively short time to organize a team of three and diagnose a cardiac arrest, it is well possible that this time is close to the attainable best limit, despite the prescription of the algorithm to limit basic diagnostic to 10 seconds.

This could lead to a ceiling effect in performance for all teams, making the detection of differences across experimental conditions difficult.

Reflexivity and team performance improvement

The comparison between the reflexivity conditions and the control group (data from an earlier study) indicated that teams in all reflexivity condition showed higher performance improvements than the not-reflecting control group for performance measure time to first meaningful measure and teams in team-focused and combined condition showed significant better improvement than control group for performance measure hands-on time. Note that performance in task 1 was controlled for, the effect can thus not be due to different performance levels in task 1. Even if it is not central to our hypotheses, this result remains interesting and corroborates earlier research that found effects of reflexivity on performance also in reflection-as-intervention studies (Gurtner et al., 2007; Kneisel, 2020; Konradt et al., 2015; Kündig et al., 2019; Schmutz, Lei, et al., 2018). However, the analyses reported here need to be interpreted with caution, because the data of the control condition was collected earlier, with no random assignment for this study. History effects are possible, even if we are not aware of a change in the medical curriculum during the time of both studies.

Dissecting reflexivity (I): Do teams reflect as expected?

One important concern of Moreland and McMinn (2010) was that teams may not reflect even when asked and given time to do so and may not follow instructions. We analyzed the reflexivity session with regard to several aspects: Whether teams used the time to reflect, to what degree they digressed from the general topic of reflection, whether teams in the different conditions reflected according to the focus of the condition, and whether teams reflected according to the optimal temporal sequence. Our analyses of the reflection process show that the concerns of Moreland and McMinn (2010) were not substantiated in this study: In all conditions, teams used the whole time to reflect, they digressed in less than 3% of the

communication from the general topic; followed the instructions regarding the focus (team, task or combination) of reflections, and overall followed the temporal sequence.

Moreland and McMinn's (2010) concerns were based on two arguments, namely that teams were not motivated or not able to reflect. Although we did not directly measure motivation and ability and can thus not empirically test the hypothesis, our study setting makes motivational and ability problems unlikely: The medical students volunteered for participation in the study, likely because they expected to gain hands-on experience in a domain important for their future profession in a simulator (Escher et al., 2017). The resuscitation task is highly significant for medical doctors, which increases its motivational potential (Hackman & Oldham, 1980). Given that all students had four years of medical training, which includes instructions related to cardiac diseases and resuscitation, we can also rule out that they had enough knowledge to engage in a at least basic discussion.

More discussion of task- than team-related aspects

An interesting result was that although students focused their discussion according to the reflection instruction, in all conditions, participants discussed more task-related than team-related aspects. Groups in the task-focused condition almost exclusively discussed task-related aspects (around five times more than team aspects); but also groups in the team reflexivity condition discussed 1.3 times more task-related contents than team-related content.

There are several possible explanations for the predominance of task-related topics. First, discussing cooperative behavior could be interpersonally sensitive. Cooperative behavior is often seen as influenced by teamwork- or non-technical skills (Andersen et al., 2010; Frankel et al., 2007), and these are often seen as personal characteristics (Hackman, 2002). The reluctance to address cooperative behaviors may thus be out of concerns of mutual criticism (Kolbe et al., 2015). Focusing more on the technical aspects of a

resuscitation may be one way to avoid the sensitive issues related to coordination. Second, problems related to the resuscitation task may be of particularly high significance for the population studied, because the primary goal and instruction in this situation was to treat the patient according to the algorithm, this may be another explanation for the focus on this task for all teams. Third, the participants may have a much clearer mental model of the task aspects than of the team aspects involved. They may simply lack the necessary knowledge needed to identify cues that indicate problems of cooperation (Gurtner et al., 2007). Indeed, as St.Pierre et al. (2011) underlined, healthcare workers are taught highly challenging technical skills and clinical algorithms, but communication and team performance aspects are often neglected in the curriculum.

Dissecting reflexivity (II): Do teams address their weaknesses and does this help?

We tested whether teams identified and discussed their own weaknesses during the reflexivity session (hypothesis 5) and whether the discussion of weaknesses improved performance in task 2. The results are contradictory. For two indices of adherence to the resuscitation algorithm (medication and task-distribution), there was no relationship between low adherence to the algorithm and subsequent discussion. For the indices diagnosis and leadership, better performance in task 1 was related to more, not less, discussion, and only for basic life support and defibrillation, our hypothesis was not rejected. This is not a clear pattern. However, we can speculate that in addition to the mechanism postulated (that teams talk about their weaknesses), another mechanism may play a role: teams, when analyzing past performance, may talk about their successes, too. This corresponds to literature on how feedback should be given – i.e. balancing positive and negative aspects to avoid threatening self-worth with feedback (Krings et al., 2015). The teams may intuitively have tried to balance positive and negative evaluation of their earlier performance.

The hypothesis that teams talk more about their weaknesses was only supported for basic life support as a trend and for defibrillation. Interestingly, those are the indices with the highest performance in task 1 (0.76/1; 0.65/1, respectively). Basic life support and defibrillation are core skills for CPR, this is well known to medical professionals. It may thus be that low performance in a core aspect of the task is particularly salient for a team and may have triggered more discussion (Larson & Egan, 2018). However, other mechanisms such as the common knowledge effect (Gigone & Hastie, 1993) or aspects of psychological safety (Edmondson, 2004) may have played a role. Finally, it is also possible that teams were unable to identify which aspects of the task they performed well or not.

For the reflexivity discussion to be useful, it makes sense to hypothesize that teams improved adherence to the algorithm most for topics that were actually discussed during reflexivity. The analyses revealed two interesting results. First, the overall strongest influence on adherence to algorithm on task 2 was adherence to algorithm on task 1: earlier performance predicted later performance (a significant relationship was found for diagnosis, basic life support and defibrillation, a trend for medication and leadership). In contrast, reflecting about specific sub-task did not influence performance on this sub-task later, with one exception: leadership. Furthermore, we did not find a significant moderation of performance in task 1 on the link between reflexivity and performance in task 2 (aka teams who discuss their weaknesses profit most from reflexivity). In view of what researchers believe to be the mechanisms of reflexivity on performance, the absence of any relationship between reflexivity on specific topics and subsequent performance is a counterintuitive, concerning, but important finding. One explanation is that teams had difficulties to identify aspects of low task performance. The resuscitation guidelines define sub-tasks performance markers in quite some detail. To the extent that the participants notice deviations from the algorithm, it seems obvious that these must then be corrected. However, having to execute

these sub-tasks in a very short time may make it very difficult to keep track of their own actions in detail, to memorize the exact way they were performed, and to relate this to the standards provided by the guideline. Being inexperienced, planning, executing, and remembering their action in a very general sense (e.g., we did perform ventilation) may be the maximum they could process in their (working) memory, whereas remembering details may be beyond their mental capacity in this situation. This interpretation would be in line with the results by Bogenstätter et al. (2009), that participants in a resuscitation had difficulties reporting details of their actions (e.g., the number of defibrillations) correctly. To the extent that such mechanisms are operating, discussing details of their performance might not be helpful, as it may miss the point. By contrast, deficiencies in leadership and coordination may be especially salient, as they tend to induce uncertainty ("should I prepare defibrillation?"), and possibly even anger ("Why doesn't he say what he wants?"). Such incidents might therefore lend themselves better to profiting specifically from discussions, as the implications are not so obvious and require joint deliberation, in contrast to the medical part proper, which is exactly specified.

Dissecting reflexivity (III): overall quality of reflexivity

We tested two aspects of quality of reflexivity with regard to resuscitation performance. *Unresolved disagreements* (to what extent after the reflexivity discussion unresolved conflicts or controversies remained) were negatively related to hands-on performance, but not to time to first meaningful measure. Many of the disagreements during reflection concerned the question of which resuscitation algorithm was the correct for the specific patient. If such strategic disagreements remain before collaborating on the next task, teams may not develop a shared mental model; this can lead to coordination problems and performance deficits. Note, however, that the level of persisting disagreements was very low, with a mean of 1.2 on a 3-point scale.

Interestingly, the most consistent predictor of performance in task 2 was *scope of discussion*: Hands-on performance and time to first meaningful measure performances were significantly higher for teams with a broader scope of reflection. This means the more teams went beyond discussing the first task just experienced and compared to other situations (e.g., an out-of-hospital cardiac arrest) or to other procedures (e.g., a trauma situation), the higher the performance increased. A broader scope of discussion can be a sign of better information integration (Schippers et al., 2018), and corresponds to the level of dialogical reflection described by Kihlgren and colleagues (2015).

Moreland and McMinn (2010) emphasized that a low quality of reflection could be one of the main problems of reflexivity-as-intervention. With an overall mean of 1.5 (on a scale of 1 to 3), the scope of reflection overall was low, as found in other studies (Gurtner et al., 2007; Kihlgren et al., 2015; Moreland & McMinn, 2010).

Conclusions

The results of this study support the general finding that even short reflexivity sessions can have positive effects on performance, and that teams may well be motivated to engage in reflexivity session. However, the results also cast doubt on whether teams are able to identify and discuss their weaknesses. The general assumption that reflecting on own shortcomings and improving particularly in domains of low performance was not supported for specific facets of the task, but it was supported for the important collaborative aspect of leadership. In addition, reflection quality was overall low, but quality was an important influence on performance.

One main conclusion for research is that aspects of quality of reflexivity and the exact mechanisms between earlier performance, reflexivity and later performance are not yet clear. The finding that teams do not more consequently address past performance problems during reflexivity is surprising. Future research thus needs to investigate the link of past performance to reflexivity and to future performance in more detail. Although quality of reflexivity is often discussed in reflexivity research, only very few studies have empirically rated reflection quality, there is need for future research in this field, too. Only a good understanding of the relationship between earlier taskwork, reflexivity and future taskwork can inform about how reflexivity instructions need to be conceived to help teams to reflect.

With regard to applied aspects, briefings in medical teams are common, and they have been found to improve performance. Based on our study, however, we cannot exclude that briefings could be much more effective, if quality could be improved. Kihlgren et al. (2015) showed that even specialist-led briefings do not achieve good quality ratings. Our study clearly showed that overall quality of the reflexivity session was low, but also that quality of reflexivity is important for performance.

Moreland and McMinn (2010) expressed concerns that reflexivity might "feel good" for team members and that they on the one hand overestimate the effects (which may be a reason for the high correlations between reflexivity, team processes and performance for exclusively self-report-based studies). On the other hand, and more important, the good feeling of "we talked about it" could lead to overconfidence with regard to team performance and to an underestimation of aspects to improve.

Limitations & strengths

In this study, we approached reflexivity-as-intervention from different perspectives, but based all data on behavioral observations. We thus do not have data about the subjective experiences of the team members. It would have been particularly interesting to compare observed reflexivity quality with subjective perception of the usefulness of reflexivity by the team members. Another limitation is the sample size of 68 groups, which allows only to detect big effects. Reasons for this are limited resources and the analyses on the group level – for each group, three student participants and a study nurse as well as a technician were necessary. We opted for large effect sizes because of the applied character of the study – in view of an implementation recommendation of reflexivity sessions after emergencies, effects should also be medically meaningful. However, larger sample sizes would allow to detect important smaller effects. The team composition in this study (a homogeneous group of fourth year students) is not realistic for ad-hoc medical teams at hospitals. The study therefore cannot take into consideration the heterogeneity of first-responders at acute emergencies. An important limitation is the generalizability of the results to other tasks than a CPR. It is clear that collaboration requirements need to be adapted to the task (Hackman & Morris, 1975; McGrath, 1984; Tschan & von Cranach, 1996). Kerr (2017) states that tasks – "the most neglected moderator in group research" – are powerful influence of group behavior and group and team research should take this into account.

This study has strengths. With CPR, we studied a realistic and important situation with clear requirements and clear performance standards, and participants were specialists in the sense that they had studied medicine for already four years. Briefings are increasingly common in medicine, and are also recommended for ad-hoc composed teams that form around a medical emergency. Second, behavior observational data is also one of the strengths of this study, because it reduces known biases of self-report (Moreland & McMinn, 2010). The quest to base more studies on behavioral observations in (social) psychological research is particularly important for team research. This study took the recommendation "why don't we study real behaviors" (Agnew et al., 2009, p. 6) very seriously.

References

Agnew, C. R., Carlston, D. E., Graziano, W. G., & Kelly, J. R. (2009). *Then a miracle occurs: Focusing* on *behavior in social psychological theory and research*. Oxford University Press.

Allen, J. A., Reiter-Palmon, R., Crowe, J., & Scott, C. (2018). Debriefs: Teams learning from doing in context. *American Psychologist Journal, 73*(4), 504-516. https://doi.org/10.1037/amp0000246

Amacher, S. A., Schumacher, C., Legeret, C., Tschan, F., Semmer, N. K., Marsch, S., & Hunziker, S. (2017). Influence of gender on the performance of cardiopulmonary rescue teams: A randomized, prospective simulator study. *Critical Care Medicine, 45*(7), 1184-1191. https://doi.org/10.1097/CCM.0000000000002375

Andersen, P. O., Jensen, M. K., Lippert, A., & Østergaard, D. (2010). Identifying non-technical skills and barriers for improvement of teamwork in cardiac arrest teams. *Resuscitation, 81*(6), 695-702. https://doi.org/10.1016/j.resuscitation.2010.01.024

Arrow, H., McGrath, J. E., & Berdahl, J. L. (2000). *Small groups as complex systems: formation, coordination, development and adaptation*. Sage Publications. http://www.loc.gov/catdir/enhancements/fy0656/99050490-d.html

Avramidis, D. (2014). New developments in cardiopulmonary resuscitation. *Hospital Chronicles, 9*(1 Sup), 8-10.

Berden, H., Willems, F. F., Hendrick, J., Pijls, N., & Knape, J. (1993). How frequently should basic cardiopulmonary resuscitation training be repeated to maintain adequate skills? *BMJ: British Medical Journal, 306*(6892), 1576.

Bogenstätter, Y., Tschan, F., Semmer, N. K., Spychiger, M., Breuer, M., & Marsch, S. (2009). How accurate is information transmitted to medical professionals joining a medical emergency? A simulator study. *Human factors, 51*(2), 115-125.

Cannon-Bowers, J. A., & Salas, E. (1998). Team performance and training in complex environments: Recent findings from applied research. *Current Directions in Psychological Science, 7*(3), 83-87.

Carmeli, A., Sheaffer, Z., Binyamin, G., Reiter-Palmon, R., & Shimoni, T. (2014). Transformational leadership and creative problem-solving: The mediating role of psychological safety and reflexivity. *The Journal of Creative Behavior, 48*(2), 115-135. https://doi.org/10.1002/jocb.43

Carter, S. M., & West, M. A. (1998). Reflexivity, effectiveness, and mental health in BBC-TV production Teams. *Small Group Research, 29*, 583-601.

Chen, J., Bamberger, P. A., Song, Y., & Vashdi, D. R. (2018). The effects of team reflexivity on psychological well-being in manufacturing teams. *Journal of Applied Psychology, 103*(4), 443.

Chen, X., Liu, J., Zhang, H., & Kwan, H. K. (2019). Cognitive diversity and innovative work behaviour: The mediating roles of task reflexivity and relationship conflict and the moderating role of perceived support. *Journal of Occupational and Organizational Psychology, 92*(3), 671-694. https://doi.org/10.1111/joop.12259

Cooper, S., & Wakelam, A. (1999, Sep). Leadership of resuscitation teams: 'Lighthouse Leadership'. *Resuscitation, 42*(1), 27-45. https://doi.org/S0300-9572(99)00080-5

De Dreu, C. K. W. (2002). Team innovation and team effectiveness: The importance of minority dissent and reflexivity. *European Journal of Work and Organizational Psychology, 11*(3), 285-298. http://ejournals.ebsco.com/direct.asp?ArticleID=0X9YN33GH0HMYBD8Y7XW

De Dreu, C. K. W. (2007). Cooperative outcome interdependence, task reflexivity, and team effectiveness: A motivated information processing perspective. *Journal of Applied Psychology, 92*(3), 628-638. https://doi.org/Doi 10.1037/0021-9010.92.3.628

De Jong, B. A., & Elfring, T. (2010). How does trust affect the performance of ongoing teams? The mediating role of reflexivity, monitoring, and effort. *Academy of Management Journal, 53*(3), 535-549.

Eddy, E. R., Tannenbaum, S. I., & Mathieu, J. E. (2013). Helping teams to help themselves: Comparing two team-led debriefing methods. *Personnel Psychology, 66*(4), 975-1008. https://doi.org/10.1111/peps.12041

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350-383. http://www.jstor.org/stable/pdfplus/2666999.pdf

Edmondson, A. C. (2004). Learning from failure in health care: frequent opportunities, pervasive barriers. *Quality & Safety in Health Care, 13*, 3-9. https://doi.org/DOI 10.1136/qshc.2003.009597

Edwards, S., & Siassakos, D. (2012). Training teams and leaders to reduce resuscitation errors and improve patient outcome. *Resuscitation, 83*(1), 13-15. https://doi.org/10.1016/j.resuscitation.2011.10.015

Escher, C., Creutzfeldt, J., Meurling, L., Hedman, L., Kjellin, A., & Fellander-Tsai, L. (2017). Medical students' situational motivation to participate in simulation based team training is predicted by attitudes to patient safety. *BMC Medical Education, 17*(1), 37. https://doi.org/10.1186/s12909-017-0876-5

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175-191.

Fernandez Castelao, E., Russo, S. G., Cremer, S., Strack, M., Kaminski, L., Eich, C., Timmermann, A., & Boos, M. (2011). Positive impact of crisis resource management training on no-flow time and team member verbalisations during simulated cardiopulmonary resuscitation: A randomised controlled trial. *Resuscitation, 82*, 1338-1343. https://doi.org/S0300-9572(11)00321-2

Fernandez Castelao, E., Russo, S. G., Riethmüller, M., & Boos, M. (2013). Effects of team coordination during cardiopulmonary resuscitation: A systematic review of the literature. *Journal of Critical Care*.

Frankel, A., Gardner, R., Maynard, L., & Kelly, A. (2007). Using the Communication and Teamwork Skills (CATS) assessment to measure health care team performance. *The Joint Commission Journal on Quality and Patient Safety, 33*(9), 549-558. https://doi.org/10.1016/s1553-7250(07)33059-6

Futoran, G. C., Kelly, J. R., & McGrath, J. E. (1989). TEMPO: A time-based system for analysis of group interaction process. *Basic and Applied Social Psychology, 10*(3), 211-232. https://doi.org/10.1207/s15324834basp1003_2

Gabelica, C., Van den Bossche, P., De Maeyer, S., Segers, M., & Gijselaers, W. (2014). The effect of team feedback and guided reflexivity on team performance change. *Learning and Instruction, 34*, 86-96. https://doi.org/10.1016/j.learninstruc.2014.09.001

Gabelica, C., Van den Bossche, P., Segers, M., & Gijselaers, W. (2014). Dynamics of team reflexivity after feedback. *Frontline Learning Research, 2*(2), 64-91.

Gigone, D., & Hastie, R. (1993). The common knowledge effect: Information sharing and group judgment. *Journal of Personality and Social Psychology, 65*(5), 959.

Graham, R., McCoy, M. A., & Schultz, A. M. (2015). *Strategies to improve cardiac arrest survival: a time to act*. National Academies Press Washington, DC.

Gurtner, A., Tschan, F., Semmer, N. K., & Nagele, C. (2007). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organizational Behavior and Human Decision Processes, 102*(2), 127-142. https://doi.org/DOI10.1016/j.obhdp.2006.05.002

Hackman, J. R. (2002). *Leading teams : setting the stage for great performances*. Harvard Business School Press.

Hackman, J. R., Brousseau, K. R., & Weiss, J. A. (1976). The interaction of task design and group performance strategies in determining group effectiveness. *Organizational Behavior and Human Performance, 16*(2), 350-365.

Hackman, J. R., & Kaplan, R. E. (1974). Interventions into group process: An approach to improving the effectiveness of groups. *Decision Sciences, 5*(3), 459-480.

Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. *Advances in Experimental Social Psychology, 8*, 45-99.

Hackman, J. R., & Oldham, G. R. (1980). *Work redesign*. Addison-Wesley.

Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review, 30*(2), 269-287.

Hayes, A. F. (2017). *Introduction to Mediation, Moderation, and Conditional Process Analysis. A Regression-Based Approach* (2nd ed.). Guilford Press

Henrickson Parker, S., Schmutz, J. B., & Manser, T. (2018). Training needs for adaptive coordination: Utilizing task analysis to identify coordination requirements in three different clinical settings. *Group & Organization Management, 43*(3), 504-527. https://doi.org/10.1177/1059601118768022

Hoegl, M., & Parboteeah, K. P. (2006). Team reflexivity in innovative projects. *R & D Management, 36*(2), 113-125.

Hofhuis, J., Mensen, M., ten Den, L. M., van den Berg, A. M., Koopman-Draijer, M., van Tilburg, M. C., Smits, C. H. M., & de Vries, S. (2018). Does functional diversity increase effectiveness of community care teams? The moderating role of shared vision, interaction frequency, and team reflexivity. *Journal of Applied Social Psychology, 48*(10), 535-548. https://doi.org/10.1111/jasp.12533

Hunziker, S., Buhlmann, C., Tschan, F., Balestra, G., Legeret, C., Schumacher, C., Semmer, N. K., Hunziker, P., & Marsch, S. (2010). Brief leadership instructions improve cardiopulmonary resuscitation in a high-fidelity simulation: a randomized controlled trial. *Critical Care Medicine, 38*(4), 1086-1091. https://doi.org/10.1097/CCM.0b013e3181cf7383

Hunziker, S., Johansson, A. C., Tschan, F., Semmer, N. K., Rock, L., Howell, M. D., & Marsch, S. (2011). Teamwork and leadership in cardiopulmonary resuscitation. *Journal of the American College of Cardiology, 57*(24), 2381-2388. https://doi.org/S0735-1097(11)01072-2

Johnson, D. W., Johnson, R. T., & Tjosvold, D. (2000). Constructive controversy: The value of intellectual opposition. In *The handbook of conflict resolution: Theory and practice.* (pp. 65-85). Jossey-Bass/Wiley.

Kerr, N. L. (2017). The most neglected moderator in group research. *Group Processes & Intergroup Relations, 20*(5), 681-692. https://doi.org/10.1177/1368430217712050

Kihlgren, P., Spanager, L., & Dieckmann, P. (2015). Investigating novice doctors' reflections in debriefings after simulation scenarios. *Medical Teacher, 37*(5), 437-443. https://doi.org/10.3109/0142159X.2014.956054

Kleiner, C., Link, T., Maynard, M. T., & Halverson Carpenter, K. (2014). Coaching to improve the quality of communication during briefings and debriefings. *AORN Journal, 100*(4), 358-368. https://doi.org/10.1016/j.aorn.2014.03.012

Kneisel, E. (2020). Team reflections, team mental models and team performance over time. *Team Performance Management: An International Journal, 26*(1/2), 143-168. https://doi.org/10.1108/tpm-09-2018-0061

Kolbe, M., Grande, B., & Spahn, D. R. (2015). Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting. *Best Practice & Research: Clinical Anaesthesiology, 29*(1), 87-96. https://doi.org/10.1016/j.bpa.2015.01.002

Konradt, U., Schippers, M. C., Garbers, Y., & Steenfatt, C. (2015). Effects of guided reflexivity and team feedback on team performance improvement: The role of team regulatory processes and cognitive emergent states. *European Journal of Work and Organizational Psychology, 24*(5), 777-795. https://doi.org/10.1080/1359432x.2015.1005608

Krings, R., Jacobshagen, N., Elfering, A., & Semmer, N. K. (2015). Subtly offending feedback. *Journal of Applied Social Psychology, 45*(4), 191-202. https://doi.org/10.1111/jasp.12287

Kündig, P., Tschan, F., Semmer, N. K., Morgenthaler, C., Zimmermann, J., Holzer, E., ... & Marsch, S. (2020). More than experience: a post-task reflection intervention among team members enhances performance in student teams confronted with a simulated resuscitation task—a prospective randomised trial. *BMJ Simulation and Technology Enhanced Learning*, 6(2).

Larsen, M. P., Eisenberg, M. S., Cummins, R. O., & Hallstrom, A. P. (1993). Predicting survival from out-of-hospital cardiac arrest: A graphic model. *Annals of Emergency Medicine, 22*(11), 1652-1658. https://doi.org/https://doi.org/10.1016/S0196-0644(05)81302-2

Larson, J. R., & Egan, A. C. (2018). Information Sharing Within Groups in Organizations. In L. Argote & J. M. Levine (Eds.), *The Oxford Handbook of Group and Organizational Learning*. https://doi.org/10.1093/oxfordhb/9780190263362.013.26

Lee, L. T. S. (2008). The effects of team reflexivity and innovativeness on new product development performance. *Industrial Management & Data Systems, 108*(3-4), 548-569. https://doi.org/Doi 10.1108/02635570810868380

LePine, J. A., Piccolo, R. F., Jackson, C. L., Mathieu, J. E., & Saul, J. R. (2008). A meta-analysis of teamwork processes: Tests of a multidimensional model and relationships with team effectiveness criteria. *Personnel Psychology, 61*(2), 273-307.

Luscher, F., Hunziker, S., Gaillard, V., Tschan, F., Semmer, N. K., Hunziker, P. R., & Marsch, S. (2010). Proficiency in cardiopulmonary resuscitation of medical students at graduation: a simulator-based comparison with general practitioners. *Swiss Medical Weekly, 140*(3-4), 57-61. https://doi.org/smw-12735

Lyubovnikova, J., Legood, A., Turner, N., & Mamakouka, A. (2015). How authentic leadership influences team performance: The mediating role of team reflexivity. *Journal of Business Ethics, 141*(1), 59-70. https://doi.org/10.1007/s10551-015-2692-3

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*(3), 356-376.

Marsch, S., Tschan, F., Semmer, N. K., Spychiger, M., Breuer, M., & Hunziker, P. R. (2005). Performance of first responders in simulated cardiac arrests. *Critical Care Medicine, 33*(5), 963-967. https://doi.org/10.1097/01.ccm.0000157750.43459.07

Marsch, S. C., Muller, C., Marquardt, K., Conrad, G., Tschan, F., & Hunziker, P. R. (2004). Human factors affect the quality of cardiopulmonary resuscitation in simulated cardiac arrests. *Resuscitation, 60*(1), 51-56. https://doi.org/10.1016/j.resuscitation.2003.08.004

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Prentice-Hall Englewood Cliffs, NJ.

Moreland, R. L., & Levine, J. M. (1992). Problem identification by groups. In S. Worchel, W. Wood, & J. A. Simpso (Eds.), *Group process and productivity* (pp. 17-47).

Moreland, R. L., & McMinn, J. G. (2010). Group reflexivity and performance. In R. S. Thye & E. Lawler (Eds.), *Advances in group processes* (Vol. 27, pp. 63-95). Emerald Press.

Muller, A., Herbig, B., & Petrovic, K. (2008). The explication of implicit team knowledge and its supporting effect on team processes and technical innovations: An action regulation perspective on team reflexivity. *Small Group Research, 40*(1), 28-51. https://doi.org/10.1177/1046496408326574

Nederveen Pieterse, A., van Knippenberg, D., & van Ginkel, W. P. (2011). Diversity in goal orientation, team reflexivity, and team performance. *Organizational Behavior and Human Decision Processes, 114*(2), 153-164. https://doi.org/http://dx.doi.org/10.1016/j.obhdp.2010.11.003

Prichard, J. S., & Ashleigh, M. J. (2007). The effects of team-skills training on transactive memory and performance. *Small Group Research, 38*(6), 696-726. https://doi.org/10.1177/1046496407304923

Putnam, L. L. (1979). Preference for procedural order in task-oriented small groups. *Communication Monographs, 46*, 193-218.

Reiter-Palmon, R., Leone, S., Murugavel, V., & Allen, J. A. (2020). Fostering effective debriefs: The integral role of team reflexivity. In *Managing Meetings in Organizations* (pp. 93-109). https://doi.org/10.1108/s1534-085620200000020005

Roberts, N. K., Williams, R. G., Schwind, C. J., Sutyak, J. A., McDowell, C., Griffen, D., Wall, J., Sanfey, H., Chestnut, A., Meier, A. H., Wohltmann, C., Clark, T. R., & Wetter, N. (2014). The impact of brief team communication, leadership and team behavior training on ad hoc team performance in trauma care settings. *American Journal Of Surgery, 207*(2), 170-178. https://doi.org/10.1016/j.amjsurg.2013.06.016

Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008, Dec). Does team training improve team performance? A meta-analysis. *Human Factors, 50*(6), 903-933. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19292013

Salas, E., Nichols, D. R., & Driskell, J. E. (2007). Testing three team training strategies in intact teams - A meta-analysis. *Small Group Research, 38*(4), 471-488. https://doi.org/Doi 10.1177/1046496407304332

Schippers, M. (2012). Why team reflexivity works. *RSM insight, 12*(4), 18-19.

Schippers, M., Edmondson, A. C., & West, M. A. (2018). Team Reflexivity. In L. Argote & J. M. Levine (Eds.), *The Oxford Handbook of Group and Organizational Learning*. https://doi.org/10.1093/oxfordhb/9780190263362.013.39

Schippers, M., West, M. A., & Dawson, J. F. (2012). Team reflexivity and innovation: The moderating role of team context. *Journal of Management*.

Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. (2007). Reflexivity in teams: A measure and correlates. *Applied Psychology, 56*(2), 189-211. https://doi.org/10.1111/j.1464-0597.2006.00250.x

Schippers, M. C., Den Hartog, D. N., Koopman, P. L., & Van Knippenberg, D. (2008). The role of transformational leadership in enhancing team reflexivity. *Human Relations, 61*(11), 1593-1616.

Schippers, M. C., Edmondson, A. C., & West, M. A. (2014). Team reflexivity as an antidote to team information-processing failures. *Small Group Research, 45*(6), 731-769. https://doi.org/10.1177/1046496414553473

Schippers, M. C., Homan, A. C., & Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on learning and final team performance. *Journal of Organizational Behavior, 34*, 6-23.

Schippers, M. C., West, M. A., & Dawson, J. F. (2015). Team reflexivity and innovation: The moderating role of team context. *Journal of Management, 41*(3), 769-788.

Schmutz, J., & Eppich, W. J. (2017). Promoting learning and patient care through shared reflection: A conceptual framework for team reflexivity in health care. *Academic Medicine, 92*(11), 1555-1563. https://doi.org/10.1097/ACM.0000000000001688

Schmutz, J. B., & Eppich, W. J. (2018). When I say ... team reflexivity. *Medical Education*. https://doi.org/10.1111/medu.13768

Schmutz, J. B., Kolbe, M., & Eppich, W. J. (2018). Twelve tips for integrating team reflexivity into your simulation-based team training. *Medical Teacher, 40*(7), 721-727. https://doi.org/10.1080/0142159X.2018.1464135

Schmutz, J. B., Lei, Z., Eppich, W. J., & Manser, T. (2018). Reflection in the heat of the moment: The role of in-action team reflexivity in health care emergency teams. *Journal of Organizational Behavior, 39*(6), 749-765. https://doi.org/10.1002/job.2299

Schön, D. (1938). *The reflective practitioner*. Basic Books

Shetty, P., Cohen, T., Patel, B., & Patel, V. L. (2009). The cognitive basis of effective team performance: features of failure and success in simulated cardiac resuscitation. *AMIA ... Annual Symposium proceedings. AMIA Symposium, 2009*, 599-603. https://pubmed.ncbi.nlm.nih.gov/20351925 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2815442/

Shin, Y. (2014). Positive group affect and team creativity: Mediation of team reflexivity and promotion focus. *Small Group Research, 45*(3), 337-364. https://doi.org/10.1177/1046496414533618

Soar, J., Nolan, J. P., Bottiger, B. W., Perkins, G. D., Lott, C., Carli, P., Pellis, T., Sandroni, C., Skrifvars, M. B., Smith, G. B., Sunde, K., Deakin, C. D., & Adult advanced life support section, C. (2015). European resuscitation council guidelines for resuscitation 2015: Section 3. Adult advanced life support. *Resuscitation, 95*, 100-147. https://doi.org/10.1016/j.resuscitation.2015.07.016

Somech, A. (2006). The effects of leadership style and team process on performance and Innovation in Functionally Heterogeneous Teams. *Journal of Management, 32*(1), 132-157. https://doi.org/10.1177/0149206305277799

St.Pierre, M., Hofinger, G., Buerschaper, C., & Simon, R. (2011). The key to success: Teamwork. In *Crisis Management in Acute Care Settings* (pp. 195-220). https://doi.org/10.1007/978-3-642-19700-0_11

Su, L., Kaplan, S., Burd, R., Winslow, C., Hargrove, A., & Waller, M. (2017). Trauma resuscitation: can team behaviours in the prearrival period predict resuscitation performance? *BMJ Simulation and Technology Enhanced Learning, 3*(3), 106-110. https://doi.org/10.1136/bmjstel-2016-000143

Swift, T. A., & West, M. A. (1998). *Reflexivity and group processes: Research and practice.* ESRC Centre for Organization and Innovation

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human Factors, 55*(1), 231-245. https://doi.org/10.1177/0018720812448394

Tesler, R., Mohammed, S., Hamilton, K., Mancuso, V., & McNeese, M. (2017). Mirror, mirror: Guided storytelling and team reflexivity's influence on team mental models. *Small Group Research*, 104649641772202. https://doi.org/10.1177/1046496417722025

Tjosvold, D., Hui, C., & Yu, Z. Y. (2003). Conflict management and task reflexivity for team in-role and extra-role performance in China. *International Journal of Conflict Management, 14*(2), 141-163. http://www.emeraldinsight.com/journals.htm?issn=1044-4068&volume=14&issue=2&articleid=1660048&show=pdf

Tjosvold, D., Tang, M. M. L., & West, M. (2004). Reflexivity for team innovation in China - The contribution of goal interdependence. *Group & Organization Management, 29*(5), 540-559. https://doi.org/Doi 10.1177/105960110325911

Tschan, F., Semmer, N. K., Gautschi, D., Hunziker, P., Spychiger, M., & Marsch, S. U. (2006). Leading to recovery: Group performance and coordinative activities in medical emergency driven groups. *Human Performance, 19*(3), 277-304.

Tschan, F., Semmer, N. K., Hunziker, S., & Marsch, S. C. U. (2011). Decisive action vs joint deliberation: Different medical tasks imply different coordination requirements. In V. G. Duffy (Ed.), *Advances in Human Factors and Ergonomics in healthcare* (pp. 191-200). Taylor & Francis.

Tschan, F., Semmer, N. K., Vetterli, M., Gurtner, A., Hunziker, S., & Marsch, S. (2011a). Developing observational categories for group process research based on task and coordination-requirement analysis: Examples from research on medical emergency-driven teams. In M. Boos, M. Kolbe, P. Kappeler, & T. Ellwart (Eds.), *Coordination in Human and Primate Groups* (pp. 93-). Springer.

Tschan, F., Vetterli, M., Semmer, N. K., Hunziker, S., & Marsch, S. C. (2011). Activities during interruptions in cardiopulmonary resuscitation: A simulator study. *Resuscitation, 11*, 1419-1423. https://doi.org/S0300-9572(11)00399-6

Tschan, F., & von Cranach, M. (1996). Group task structure, processes and outcome. In M. West (Ed.), *Handbook of work group psychology* (pp. 95–121). Wiley

Urbach, T., Fay, D., & Goral, A. (2010). Extending the job design perspective on individual innovation: Exploring the effect of group reflexivity. *Journal of Occupational and Organizational Psychology, 83*(4), 1053-1064. https://doi.org/10.1348/096317909x479394

van Ginkel, W., Tindale, R. S., & van Knippenberg, D. (2009). Team reflexivity, development of shared task representations, and the use of distributed information in group decision making. *Group Dynamics: Theory, Research, and Practice, 13*(4), 265-280. https://doi.org/10.1037/a0016045

van Ginkel, W. P., & van Knippenberg, D. (2009). Knowledge about the distribution of information and group decision making: When and why does it work?. *Organizational Behavior and Human Decision Processes, 108*(2), 218-229. http://www.sciencedirect.com/science/article/pii/S0749597808001064

Vashdi, D. R., Bamberger, P. A., Erez, M., & Weiss-Meilik, A. (2007). Briefing-debriefing: Using a reflexive organizational learning model from the military to enhance the performance of surgical teams. *Human Resource Management, 46*(1), 115-142. https://doi.org/10.1002/hrm.20148

Weingart, L. R. (1992). Impact of group goals, task component complexity, effort, and planning on group performance. *Journal of Applied Psychology, 77*(5), 682.

West, M. A. (1996). *Reflexivity and work group effectiveness: A conceptual integration*. John Wiley & Sons, Ltd.

Widmer, P. S., Schippers, M., & West, M. A. (2009). Recent developments in reflexivity research: A review. *Psychology of Everyday Activity, 2-11*.

Wiedow, A., & Konradt, U. (2010). Two-Dimensional Structure of Team Process Improvement: Team Reflection and Team Adaptation. *Small Group Research, 42*(1), 32-54. https://doi.org/10.1177/1046496410377358

Wong, A., Tjosvold, D., & Su, F. (2007). Social face for innovation in strategic alliances in China: the mediating roles of resource exchange and reflexivity. *Journal of Organizational Behavior, 28*(8), 961-978. https://doi.org/Doi 10.1002/Job.468

Wu, W.-Y., Amaya Rivas, A., & Chen, Y.-C. (2017). The role of team reflexivity as a mediator between project management skills, task familiarity, procedural justice, and product performance. *Journal of Management & Organization, 25*(6), 876-895. https://doi.org/10.1017/jmo.2017.34

Yang, M., Schloemer, H., Zhu, Z., Lin, Y., Chen, W., & Dong, N. (2020). Why and When Team Reflexivity Contributes to Team Performance: A Moderated Mediation Model. *Frontiers in Psychology, 10*, 3044. https://doi.org/10.3389/fpsyg.2019.03044

Zellmer-Bruhn, M. E., Waller, M. J., & Ancona, D. (2004). The effect of temporal entrainment on the ability of teams to change their routines. *Research on Managing Groups and Teams, 6*, 135-158.

4. Overall discussion

In this section, I present a summary of the findings of the two papers included in this dissertation as well as possible reasons why reflexivity effects were not fully positive in our experimentation. I then suggest future research perspectives. Some themes that emerge from the results and that are not directly related to reflexivity are also discussed. Further on, I propose a few implications of our results for practice. Finally, before the conclusion, I discuss the limitations of our projects.

Summary of results

Our studies focus on a reflexivity intervention applied to a medical simulation where medical students engage in a short self-led reflexivity session between two CPR tasks. The first paper demonstrated that while controlling for first performance, teams that engaged in a reflexivity intervention showed higher resuscitation performance gain in the following resuscitation than teams in the control condition (no reflection) for basic hands-on performance and for coordinative performance, but non-significantly lower performance for rapid defibrillation (lower delay means performance increase for this measure). These results indicate that even very short self-led post-action reflective briefings enhance basic resuscitation performance in ad hoc groups. However, they also suggest that the effects of reflexivity might not affect overtly complex aspects of the task.

With our second paper, we wanted to understand in more detail how the reflexivity process was constructed and which aspects of performance are impacted. We therefore delved deeper into the mechanisms of reflexivity here. Firstly, we compared three different reflexivity conditions (task-focused, team-focused and combination of both). For hands-on performance, teams in the team-focused condition improved their performance significantly more. We also found that all reflexivity conditions induced higher performance improvements than the control group for both resuscitation performance measures time to

first meaningful measure and hands-on time. Secondly, we showed that our participants engaged in reflexivity process: they discussed topics related to the task, and that they respected the structure of the reflexivity session. However, teams tend to generally favor task aspects during reflexivity discussion. The amount of task- or team-related communication or structure of reflexivity did not influence subsequent performance.

Thirdly, with a detailed analysis of the resuscitation algorithm, we could show that teams did not only discuss weak aspects of the initial performance; this was the case for defibrillation and basic life support actions, the opposite was found for leadership and diagnosis. Only leadership performance in task 2 was influenced by its discussion during the reflexivity session.

Finally, we tested social and content quality of reflection on performance. *Unresolved disagreements* were significantly related to performance (hands-on) and *scope of discussion* was our most important quality indicator, predicting hands-on time and time to first meaningful measure.

The results are encouraging but not completely conclusive for the effectiveness of reflexivity.

In the next section, we discuss possible explanations for the results we found.

Reflexivity, not yet convincing – reasons why our intervention was not fully successful

We discuss in the following sections possible explanations for our results, i.e. Why reflexivity did not improve all aspects of resuscitation or why it worked only in certain experimental conditions.

Salience of team-related aspects

Alongside our analyses of reflexivity content with different focuses of reflexivity and of CPR performance, we outlined the importance of team-related aspects for CPR, for reflexivity discussion and for medical students. Indeed, the themes related to teamwork stand out as salient in each section of the results. First, teams tend to favor task-related topics in the reflexivity session, even in a reflexivity condition designed to promote discussion of team-related elements. However, we showed that it is possible to increase the amount of team topics discussed during reflexivity with instruction focusing on team aspects (paper 2). Second, team-related topics seem essential in the resuscitation and reflexivity process, because groups in the team-focused reflexivity condition significantly improved more in their second resuscitation performance than groups in the control condition (paper 2). Team-related reflexivity was also more efficient than task-related reflexivity for hands-on time. Third, when looking in detail at the algorithm, we also demonstrated that performance index leadership – a collaborative aspect of performance, is the only index of performance increased by reflexivity (paper 2). The collaborative aspects of the CPR task are essential to the team's performance but they seem difficult to discuss and implement for medical students. Collaboration is thus both a deficiency and a necessity for CPR, reflexivity might improve this specific domain, and this is probably even more important for medicine students who are still learning.

Difficulty to identify mistakes and technical deficienciesThe lack of technical knowledge and experience of our participants (they are still students, even if they already have four years of training) can explain why reflexivity was not fully efficient. I noticed that our participants have particular high difficulties with defibrillation because it is a very complex task. It was detected that they expressed during the reflexivity session that they were not proficient in defibrillation, partly because they were not trained in the use of the manual

defibrillator; we understood from what they said that they had only practiced with the automatic defibrillator (AED). The defibrillator was a major challenge for the participants, even though it is an essential and central tool for CPR ("but the problem was that we have never done this before with an AED", "yeah, but the defi (*defibrillator*) should recommend shock or no shock, you know?"). If students lack such basic knowledge, it could partly explain the lack of effectiveness of reflexivity, because participants did not have the necessary resources to adapt their handling of the defibrillator and acknowledge their mistakes.

Participants were often unsure about the right algorithm to follow and the meaning of the steps symbolized by the acronym: "There were three algorithms that you have to give and there was one, but I don't remember which one, that we did massage, then adrenaline and you know... and then shock, and then…", "so algorithm is what? CABD or…?" or "algorithm is uh... the pulse... yeah, no, no...", "what is "D"?".

If the participants are not able to identify clearly not only the weaknesses of their performance, but also the strengths, reflexivity is useless and performance improvement cannot occur. However, if participants are aware of their weaknesses but do not know how to correct them, they cannot discuss solutions and plans during reflexivity or propose improvements. Teams should be guided to identify important topics such as these that must be addressed (Otte et al., 2018). Moreland and McMinn (2010) think that participants focus on irrelevant topics and thus benefit not from reflexivity. For a population of junior doctors, it may be necessary to allow them to verify their assumptions and decisions to make the most of the reflexivity intervention. A limitation of the studies is thus that we did not take into account the knowledge level of the students.

Confusion about guidelines changesThere is a fairly simple possible explanation for the uncertainty of participants during certain technical tasks. The regular adaptations of the resuscitation guidelines (approximately every five years) follow the actual developments in research. However, it is obvious that these changes can be confusing, for example if a team member is not informed about the changes or has memorized earlier guidelines. European guidelines before 2010 recommended to chain several defibrillations without intermittent cardiac massage and ventilations (Nolan et al., 2010), which is no longer recommended and represents a major change (Soar et al., 2015). If members of the same team work with different guidelines in mind and they do not share a common or accurate mental model of the task (about the way to achieve the task) then this could lead to confusion and errors. This problem of guideline adaptation was particularly visible for the question of the "ABC" sequence: Since 2010 and after years of other instructions, it is now recommended to follow the CAB (Circulation-Airway-Breathing) sequence rather than the ABC sequence (Airway-Breathing-Circulation) (Marsch et al., 2013). Given that our data collection occurred after 2010, we expected participants to systemically use the scheme CAB but, again, we have not observed this CAB sequence unanimously among the participants. We also observed that some participants seem to be trained in other work procedures that are no longer recommended (e.g. MONA, an old algorithm for medicine to manage chest pain that is no longer recommended, Gouda et al., 2016). There might be confusion in teams regarding the guidelines (Marsch et al., 2013), and these technical confusions may also be due to training that is not up to date. It is thus important to ensure that medical staff are regularly informed and retrained for any changes in care recommendations.

Whilst we are talking about guidelines, they could also serve as a facilitator for the collaborative aspects of medical tasks. We identified that team-related aspects were a major difficulty for medical students who took part in our study. Resuscitation guidelines and

algorithms could be adapted to include collaborative aspects, and be taught with an emphasis on teamwork. For example, one of the first steps in the CPR algorithm could be to "designate a leader in the team" and then "assign responsibility for CPR and ventilation". It could also specify that the leader must remain hands-off if possible, one of the recurring errors in our experimentation.

The question of education in medicine for collaboration

In addition to technical deficiencies, as St.Pierre et al. (2011) and Flin and Maran (2008) underlined, there is a problem with collaboration competencies in medical teams. Healthcare workers are taught highly challenging technical skills and clinical algorithms but communication and team performance aspects are generally neglected in the curriculum (St.Pierre et al., 2011). It is an error to consider that "a team of experts is immediately an expert team" (Kolbe et al., 2019, p. 4); meaning that technical skills alone are not sufficient for healthcare workers, they need to be able to work together efficiently as well. Moreover, a major reason why healthcare workers may not be able to prevent human errors – according to the report "Unmet needs: teaching physicians to provide safe patient care" (2010) – is that professionals are not trained to work in teams and are therefore not prepared to face this critical aspect. There seems to be a general lack in training for the collaboration of healthcare workers, our results also raise the question of the education offered to medical students in Switzerland. Indeed, we observed that participants made errors and are uncertain regarding leadership and collaboration during a resuscitation. There was often no task distribution, communication was unclear and limited, and leadership was often absent.

There is therefore an undeniable need to adapt and structure the training of medical professionals in order to provide them with competencies that will enable them to address all aspects of their work effectively. This effort for collaboration is important regarding our results: we showed that a reflexivity intervention which focused on team-related aspects

helped teams to improve their performance (paper 2). Collaborative aspects are important for the smooth running of medical tasks but seem to represent for the moment a difficulty for medical students.

Quality of reflexivity

Our results (see paper 2) outlined show that having a vast scope of reflection (comparing the situation with other experiences) and fewer disagreements, seem to be essential. However, our participants did not always achieve a high quality level of reflexivity. The fact that participants remain on a superficial level can prevent reflexivity from being useful, and this is one of the major critics against reflexivity (Moreland & McMinn, 2010). A sufficiently detailed apprehension of key aspects is required to reflect, plan and adapt, i.e. engage in a successful reflexivity process. In summary, without discussing relevant topics sufficiently, reflexivity may be useless.

We saw possible reasons that may have affected the effects of reflexivity in our experiment. I would also like to take this opportunity to highlight some of the next perspectives that research could explore in the field of reflexivity in the next section.

Future research on reflexivity

Observing these teams on many occasions drew my attention to many processes that were not at the centre of my research questions. If I had had an opportunity to continue this research, I would have liked to analyze them, and they could be further investigated in future research lined to reflexivity and resuscitation.

The composition of the teams is of particular interest: Simple demographic data (such as gender for example) or team size could have an impact on team performance and reflexivity discussion (as shown in Amacher et al., 2017; Hunziker et al., 2018; Tramèr et al., 2020). The proportion of speech contributed by each participant related to their respective knowledge or experience is also interesting data in itself, and could explain the performance

of the groups; if the students with the best knowledge did not communicate, then the group does not reach its potential in the resuscitation as well as in the reflexivity session.

The theme of leadership, which is in itself a vast field, also brings questions: who are the leaders in our research? Are they the best choice to lead the team? Who dominates in the team (decisions in resuscitation, discussion in reflexivity), according to gender, personality, competencies, etc.? Who interrupts others during the reflexivity session? If the leader relies on erroneous judgment during the resuscitation and reflexivity session, then the team cannot progress even though another member may have the necessary knowledge.

Studying reflexivity with other reflexivity conditions might also be interesting. For example, providing the algorithm as reference to the participants during reflexivity can be a really helpful support for students who are still uncertain with resuscitation. By doing so, the tools or means of support needed to improve and promote the reflexivity process could be identified. Perhaps a longer reflexivity session would also help "junior" participants to check their knowledge and fulfil the complete reflexivity cycle more thoroughly.

It would also be interesting to conduct the same deep investigation of reflexivity with professional groups instead of students to allow generalization of results troughout the whole domain of healthcare. Our participants were clearly young with few practical experiences; professional physicians or nurses might have presented different difficulties or benefited more from reflexivity with their advanced experiences and knowledge. It is maybe also necessary to investigate and understand reflexivity in other tasks and amongst other teams, where the underlying mechanisms could be different. Firefighters, police officers, intervention groups, crisis management, first aiders, or other medical specialties (surgeries, childbirth, psychiatric cases, etc.) could also benefit from reflexivity. This would allow an understanding of reflexivity variations across other fields (Moreland & McMinn, 2010).

One aspect that is missing from our approach and that could help explain our results and deepen our understanding of reflexivity is the experience of the participants. Indeed, with simple questionnaires, participants could indicate how comfortable and safe they feel during each part of the experience. This would help to define whether or not the psychological safety prevents them from acting during simulation and discussing freely.

We came across some possible future research linked to reflexivity. In the next section, I wanted to underline some themes that are a little more distant from reflexivity but which caught my attention.

Emergent themes

In this section, I discuss the themes that emerge from our data and that are not directly related to reflexivity. I find them very interesting in that they could be at the center of other research however, still give another perspective on reflexivity.

Leadership as salient aspect

When the participants focused on the collaborative aspects of the task, they tended to dive into the theme of leadership first. It is indeed a particularly salient aspect of performance if it is insufficient, because it plays an important role in resuscitation (Hunziker et al., 2010). There was often no clear leadership in the first resuscitation performance. What could be simpler than finding out that there was no leader and deciding to designate someone ("we did not really have a leader", or "maybe we should have picked a leader right at the beginning")? In addition, it allows to avoid attacking the other members with criticism, by thinking together about what the leader should say, what tasks to delegate, and who will be the leader in the next simulation ("the leader says, "you do this, and you do that"" or "somebody should have taken charge and said, "do this, do that""). These are, of course, correct remarks, but "safe" to express. The leadership theme alone probably plays an important role in

resuscitation and reflexivity and deserves further investigations and was an aspect of performance that benefited from reflexivity process.

Psychological safety

We consider that psychological safety (definition of psychological safety in section "Conditions favoring reflexivity." p. 53) might have impaired the resuscitation performance. Indeed, we observed repeatedly that participants tried to find a simple task that they could not give up on (such as ventilating the patient) especially during first simulation. Once they had their hands on the ventilation bags, they seemed to be aware that the defibrillator needed to be attended to, but no one was willing to do it because it was the more risky and exposed task. The last person who had their hands free was by default responsible for the defibrillator. Similarly, it seemed difficult for the participants to get the simulation "started". Participants wasted precious minutes watching the suffering patient from a distance, feeling uncomfortable, torn between a desire to avoid the exposure and their duty of care towards the task. Their anxiety probably affected the smooth running of the resuscitations, but above all it undoubtedly had an impact on the themes discussed or not during the reflexivity sessions. The participants certainly did not dare to address all the issues because they were being observed and judged not only by the group of three students but also by the experimental team. We were clearly able to observe insecurities amongst the participants and a form of mistrust against the experimental team during the reflexivity session. They often mentioned the camera lights, they whispered in order to not be heard, they pointed out that what they say is heard and judged, etc. Here a few examples: "So the red light is on", "above all, we are being recorded, the microphone is lit up", "ohhh we are still being recorded by the microphone", or "it's kind of scary that we can be observed like this".

Saving the face

There may have been insecurity between the participants and the experimental team, but there was also a phenomena of insecurity within the group of participants itself. This can be explained with the question of face-saving regarding the language used by participants. “Face” is the public self-image that everyone claims for oneself. It is constantly subject to external influence during the course of social interaction (Brown & Levinson, 1987). People very frequently mean more, when talking, than just the words they say; this can be observed with the use of indirect language (Holtgraves, 2008). In our study, participants most likely used the indirectness of language to express criticism in a polite and non-threatening manner to the others’ face (Goffman, 1955; Holtgraves, 2002), which is for every participants’ best interest. Indeed, we observed off-record politeness (when the face-threatening act that is performed must be inferred as well as the different meanings, Holtgraves, 2008) in our data. Notably, participants frequently expressed criticism by using questions, adding “I don't know” in their sentences or referring as “we”. For example: “Did we give adrenaline at the right time?”, “well, I do not know, if one should... we never looked in his mouth”, or “we did not... we did not even count at the beginning, that was also a mistake”. In this way, they avoid putting the blame on someone, although the “perpetrator” is obvious. Using the “we-form” also shows a great sense of solidarity in the team because everyone tried to keep the team together and save each other’s face.

Women depreciating themselves

I wanted to report here some remarkable sentences voiced by female participants during the reflexivity session. They made critical remarks about themselves, not about the team (no use of “we”, e.g. “yes… I did not know what I had to do… as always”, “I just did not do anything, I just stood there”), questioning their own skills and even their gender (“you are the man, the man just has the lead”). These remarks often come up unexpectedly, as the

team discussed a technical theme – individual responsibility or competency is not the subject at all, and are not constructive for the team ("but I thought it was good, so you [*number 6, another participant*] reacted fast, much faster than me for example"). Usually, these remarks are ignored, no teammate validate these comments. I did not observe any remarks of this nature from the men in the group.

It is interesting to compare these remarks with the findings of Tramèr et al. (2020). They showed that women expressed less leadership statements than men during resuscitation, and they initiated the first resuscitation action less often than men. Women also had lower self-esteem scores, and low self-esteem in turn negatively impacted resuscitation performance. Amacher et al. (2017) also showed a gender difference in resuscitation performance: Teams composed of women only had lower scores in hands-on time and in delay before starting CPR than teams of men. There is therefore a tendency for women to behave "backward" in resuscitation, and it seems that it is also the case during reflexivity discussion. Again, it may be interesting to include team composition or gender as a control variable in future research on reflexivity to better understand this phenomenon. A better understanding of what hinders women is crucial to help them reach their potential and empower them.

After reviewing various theoretical concepts related to this research, we now move on to more concrete aspects by discussing practical recommendations.

Implication for practice

Some simple implementations could facilitate reflexivity and help teams to get maximum benefit from it. Reflexivity might be more helpful if teams are first informed and trained about the reflexivity process (Schmutz & Eppich, 2019), as well as its goals and the topics to be addressed before using it in their daily work. Reflexivity training could help beneficiaries to understand its mechanism and its usefulness and thus reveal the full potential

of this intervention. As we have seen, quality plays its role in terms of scope of discussion and solving disagreements; these principles must be followed during reflexivity. Comparing the situation with other experiences and reaching agreements seem to be essential in the reflexivity process to improve subsequent performance. If people are aware of the importance of all these aspects, they will be able to reach better levels of quality and thus benefit more from the process for their performance. A brief introduction to the concept of reflexivity may be sufficient to help participants make better use of this tool.

Not only should reflexivity be encouraged, but we can also wonder how to encourage teams to talk about collaborative aspects during reflexivity, because they are so central to CPR and influenced by reflexivity intervention. We saw that our instructions increased the total number of team-related aspects discussed, but this may not be enough. It is imperative to insist on the aspects that need to be discussed, through instructions, checklists or leader reminders. An expert or a leader could guide the discussion to help the team to keep the structure and the relevant theme (e.g., collaboration) at the forefront of the discussion (Otte et al., 2018). An expert can also provide feedback after the performance based on collaboration or other difficult and important aspects as a basis for reflexivity. Indeed, Konradt et al. (2016) proposed that adding feedback about a performance episode before a reflexivity intervention is an effective way to induce teams to engage in reflexivity and to improve the future performance by helping them to identify potential improvements. The presence of an expert would also allow people with less experience to verify their decisions and plans, which is especially relevant for inexperienced practitioners, as it was the case in our studies. The different biases (see section “Reflexivity and information processing.” p. 50) that can influence and impair the discussion can be presented to the teams to make them aware of their possible influences and avoid them. It is necessary that teammates share all the important information for successful reflexivity.

Before concluding this work, here are the limitations we encountered in our research.

Limitations

Our studies have several limitations. First, the participants were medical students; they are not representative of typical first responders to a cardiac arrest, even if they have some experience in CPR (Marsch et al., 2004; Marsch et al., 2005); one can indeed imagine that the technical difficulties they experienced are generally not to be found in experienced teams (e.g., decreased attention span on routine tasks). Difficulties might appear in other aspect of the performance and generalizations around other teams are thus limited. Second, the results cannot be generalized to other tasks because we worked with a CPR situation in both papers. Here, also, other aspects of the task may be important for the success of the team and therefore need to be considered for a reflexive approach (choosing the right equipment for a fire team). Indeed, the structure of reflexivity must be considered according to the different fields and backgrounds (Moreland & McMinn, 2010). Third, the restricted sample size also represents another limitation of these studies (56 teams of three in paper 1, 68 and 97 in paper 2). Since they worked in groups of three, this results in a small sample, even if power analysis were satisfactory. Fourth, even if simulation studies allow a high controllability that is necessary to study medical tasks and a pretty realistic context, this experimentation did not take place in a real hospital environment, which leads to a cautious generalization of the results because of the impact on external validity. Fifth, we compared cohorts of different academic years, they may have different educational background and experience that could affect their behavior and performance, even if we are not aware of an official change in the medical curriculum. However, we believe that the potential of reflexivity is there and that it can be adapted to a number of areas.

5. Conclusion

This dissertation helped to answer some of the questions that had been raised in previous research regarding reflexivity, especially the criticism of Moreland and McMinn (2010). Short self-led reflexivity interventions in the studies presented in this work seem promising. Applied in the medical world, it has a potential to improve the quality of physicians' work and as a result patients' health. It is an easy procedure to set up, short and inexpensive, which could be easily adapted for other domains, even for teams that are not working under extreme conditions.

To our knowledge, this is the first time that the content of reflexivity sessions has been examined in such detail with an observation method. It is also the first time that the quality of reflexivity was approached with this perspective of two indicators (unresolved disagreements and scope of reflection). We have developed New coding systems to analyze the content of reflexivity sessions as well as their effects on subsequent simulations. They allow measuring reflexivity by observation and not in a self-reported manner like it was usually the case (Edmondson, 1999; Konradt et al., 2015; Schippers et al., 2007). These measures could be used for research in other domains.

The results of the papers presented in this dissertation are not completely in favor of reflexivity but all hope is not lost: There seems to be a potential for reflexivity to have a beneficial effect on performance, especially if collaborative aspects are discussed. Even if there are still questions remaining about reflexivity, we were able to demonstrate that teams of non-professionals engaged in a reflexivity process and who respected the instruction even without an experimenter to guide them. We also demonstrated the importance of quality of reflexivity. New research perspectives were also proposed under the theme of reflexivity because some inconsistencies were raised from our results.

I began this work with Samuel Beckett's quote "Ever tried. Ever failed. No matter. Try Again. Fail again. Fail better". I hope that this work will highlight a new way to fail/try, making a few less mistakes thank to reflexivity.

6. References

Allen, J. A., Reiter-Palmon, R., Crowe, J., & Scott, C. (2018). Debriefs: Teams learning from doing in context. *American Psychologist, 73*(4), 504.

Amacher, S. A., Schumacher, C., Legeret, C., Tschan, F., Semmer, N. K., Marsch, S., & Hunziker, S. (2017). Influence of gender on the performance of cardiopulmonary rescue teams: a randomized, prospective simulator study. *Critical Care Medicine, 45*(7), 1184-1191.

Argote, L., Turner, M. E., & Fichman, M. (1989). To centralize or not to centralize: The effects of uncertainty and threat on group structure and performance. *Organizational Behavior and Human Decision Processes, 43*(1), 58-74.

Argyris, C. Double-loop learning, teaching, and research. *Academy of Management Learning & Education, 1*(2), 206-218.

Arora, S., Ahmed, M., Paige, J., Nestel, D., Runnacles, J., Hull, L., Darzi, A., & Sevdalis, N. (2012). Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. *Annals of Surgery, 256*(6), 982-988.

Bandari, J., Schumacher, K., Simon, M., Cameron, D., Goeschel, C. A., Holzmueller, C. G., Makary, M. A., Welsh, R. J., & Berenholtz, S. M. (2012). Surfacing safety hazards using standardized operating room briefings and debriefings at a large regional medical center. *The Joint Commission Journal on Quality and Patient Safety, 38*(4).

Bass, B. M., & Riggio, R. E. (2006). In *Transformational leadership*. Psychology press.

Bedwell, W. L., Ramsay, P. S., & Salas, E. (2012). Helping fluid teams work: A research agenda for effective team adaptation in healthcare. *Translational Behavioal Medicine, 2*(4), 504-509. https://doi.org/10.1007/s13142-012-0177-9

Bolinger, A. R., & Stanton, J. V. (2014). The gap between perceived and actual learning from group reflection. *Small Group Research, 45*(5), 539-567. https://doi.org/10.1177/1046496414538322

Bromiley, M. (2008). Have you ever made a mistake. *Royal College of Anaesthetists Bulletin, 48*, 2442-2445.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage* (Vol. 4). Cambridge university press.

Burtscher, M. J., Wacker, J., Grote, G., & Manser, T. (2010). Managing nonroutine events in anesthesia: the role of adaptive coordination. *Human Factors, 52*(2), 282-294. https://doi.org/10.1177/0018720809359178

Cannon-Bowers, J. A., & Salas, E. (2001). Reflections on shared cognition. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 22*(2), 195-202.

Carter, S. M., & West, M. A. (1998). Reflexivity, effectiveness, and mental health in BBC-TV production Teams. *Small Group Research, 29*, 583-601.

Castelao, E. F., Russo, S. G., Cremer, S., Strack, M., Kaminski, L., Eich, C., Timmermann, A., & Boos, M. (2011). Positive impact of crisis resource management training on no-flow time and team member verbalisations during simulated cardiopulmonary resuscitation: a randomised controlled trial. *Resuscitation, 82*(10), 1338-1343.

Chen, J., Bamberger, P. A., Vashdi, D. R., & Song, Y. (2016). The effects of team reflexivity on emotional well-being in manufacturing teams. In *Academy of Management Proceedings* (Vol. 2016, No. 1, p. 16523). Briarcliff Manor, NY 10510: Academy of Management.

Chen, X., Liu, J., Yuan, Y., & Cui, X. (2019). The curvilinear effect of task conflict on idea generation. *International Journal of Conflict Management, 30*(2), 158-179. https://doi.org/10.1108/ijcma-02-2018-0029

Chen, X., Liu, J., Zhang, H., & Kwan, H. K. (2019). Cognitive diversity and innovative work behaviour: The mediating roles of task reflexivity and relationship conflict and the moderating role of perceived support. *Journal of Occupational and Organizational Psychology, 92*(3), 671-694. https://doi.org/10.1111/joop.12259

Chin, C. (2008). Simulators, equipment, and props. In *Manual of Simulation in Healthcare* (pp. 51-62).

Collins, C. J., & Smith, K. G. (2006). Knowledge exchange and combination: The role of human resource practices in the performance of high-technology firms. *Academy of Management Journal, 49*(3), 544-560.

Converse, S., Cannon-Bowers, J. A., & Salas, E. (1993). Shared mental models in expert team decision making. *Individual and group decision making: Current issues, 221*.

Cooper, S. (2001). Developing leaders for advanced life support: evaluation of a training programme. *Resuscitation, 49*(1), 33-38. https://doi.org/S0300-9572(00)00345-2

Cotard, C., & Michinov, E. (2020). Relations entre réflexivité d'équipe, mémoire transactive et efficacité interventionnelle perçue chez les policiers-opérateurs du RAID. *Psychologie du Travail et des Organisations*. https://doi.org/10.1016/j.pto.2020.01.007

De Dreu, C. K. (2002). Team innovation and team effectiveness: The importance of minority dissent and reflexivity. *European Journal of Work and Organizational Psychology, 11*(3), 285-298. https://doi.org/10.1080/13594320244000175

De Dreu, C. K. (2007). Cooperative outcome interdependence, task reflexivity, and team effectiveness: a motivated information processing perspective. *Journal of Applied Psychology, 92*(3), 628-638. https://doi.org/10.1037/0021-9010.92.3.628

DeChurch, L. A., & Mesmer-Magnus, J. R. (2010). Measuring shared team mental models: A meta-analysis. *Group Dynamics: Theory, Research, and Practice*, *14*(1), 1-14. https://doi.org/10.1037/a0017455

Devine, D. J., Clayton, L. D., Philips, J. L., Dunford, B. B., & Melner, S. B. (1999). Teams in organizations: Prevalence, characteristics, and effectiveness. *Small group research, 30*(6), 678-711.

Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. *Administrative Science Quarterly, 44*(2), 350-383.

Edmondson, A. C., Bohmer, R., & Pisano, G. (2001). Speeding up team learning. *Harvard Business Review, 79*, 125-134.

Edmondson, A. C., Dillon, J. R., & Roloff, K. S. (2007). Three perspectives on team learning: Outcome improvement, task mastery, and group process. *The Academy of Management Annals, 1*(1), 269-314.

Edmondson, A. C., & Lei, Z. (2014). Psychological safety: The history, renaissance, and future of an interpersonal construct. *Annual Review of Organizational Psychology and Organizational Behavior, 1*(1), 23-43. https://doi.org/10.1146/annurev-orgpsych-031413-091305

Endsley, M. R. (1988). Situation awareness global assessment technique (SAGAT). In *Proceedings of the IEEE 1988 national aerospace and electronics conference* (pp. 789-795). IEEE.

Eppich, W., Howard, V., Vozenilek, J., & Curran, I. (2011). Simulation-based team training in healthcare. *Simulation in Healthcare, 6 Suppl*, S14-19. https://doi.org/10.1097/SIH.0b013e318229f550

Eppich, W. J., & St.Pierre, M. (2011). Speech is golden: Communication. In *Crisis Management in Acute Care Settings* (pp. 221-251). https://doi.org/10.1007/978-3-642-19700-0_12

Facchin, S., Tschan, F., Gurtner, A., Cohen, D., & Dupuis, A. (2006). Validation of the French version of the team reflexivity scale of Carter and West, 1998. *Psychologie du Travail et des Organisations 12*(4), 291-306.

Flin, R., & Maran, N. (2008). Non-technical skills: identifying, training, and assessing safe behaviours. In *Manual of simulation in healthcare* (pp. 249-264). Oxford University Press.

Gabelica, C., Van den Bossche, P., Segers, M., & Gijselaers, W. (2014). Dynamics of team reflexivity after feedback. *Frontline Learning Research, 4*, 64-91. https://doi.org/10.14786/flr.v2i2.79

Gersick, C. J. G., & Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes, 47*(1), 65-97.

Goffman, E. (1955). On face-work: An analysis of ritual elements in social interaction. *Psychiatry, 18*(3), 213-231.

Goldenhar, L. M., Brady, P. W., Sutcliffe, K. M., & Muething, S. E. (2013). Huddling for high reliability and situation awareness. *BMJ Quality & Safety, 22*(11), 899-906. https://doi.org/10.1136/bmjqs-2012-001467

Gouda, P., Bainey, K., & Welsh, R. (2016). The demise of morphine oxygen nitroglycerin aspirin (MONA). *Canadian Journal of Cardiology*, *32*(12), 1578. e1577.

Grober, E. D., & Bohnen, J. M. A. (2005). Defining medical error. *Canadian Journal of Surgery*, *48*(1), 39.

Grossman, R., Friedman, S. B., & Kalra, S. (2017). Teamwork processes and emergent states. In *The Wiley Blackwell Handbook of the Psychology of Team Working and Collaborative Processes* (pp. 243-269).

Gurtner, A., Tschan, F., Sernmer, N. K., & Nagele, C. (2007). Getting groups to develop good strategies: Effects of reflexivity interventions on team process, team performance, and shared mental models. *Organizational Behavior and Human Decision Processes*, *102*(2), 127-142. https://doi.org/DOI10.1016/j.obhdp.2006.05.002

Hackman, J. R., & Kaplan, R. E. (1974). Interventions into group process: An approach to improving the effectiveness of groups. *Decision Sciences*, *5*(3), 459-480.

Hackman, J. R., & Katz, N. (2010). Group Behavior and Performance. In *Handbook of social psychology*. John Wiley & Sons, Inc. https://doi.org/10.1002/9780470561119.socpsy002032

Hackman, J. R., & Wageman, R. (2005). A theory of team coaching. *Academy of Management Review*, *30*(2), 269-287.

Hamdorf, J. M., & Davies, R. (2008). Teaching a clinical skill. In *Manual of Simulation in Healthcare* (pp. 78-88).

Hazinski, M. F., Nolan, J. P., Aickin, R., Bhanji, F., Billi, J. E., Callaway, C. W., Castren, M., de Caen, A. R., Ferrer, J. M., Finn, J. C., Gent, L. M., Griffin, R. E., Iverson, S., Lang, E., Lim, S. H., Maconochie, I. K., Montgomery, W. H., Morley, P. T., Nadkarni, V. M., Neumar, R. W., Nikolaou, N. I., Perkins, G. D., Perlman, J. M., Singletary, E. M., Soar, J., Travers, A. H., Welsford, M., Wyllie, J., & Zideman, D. A. (2015). Part 1: Executive summary: 2015 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care science with treatment recommendations. *Circulation*, *132*(16 Suppl 1), S2-39. https://doi.org/10.1161/CIR.0000000000000270

Hedman, E. (2016). Leadership team tool for better meaning making. *Journal of Management Development, 35*(5), 592-605. https://doi.org/10.1108/jmd-09-2015-0132

Hirst, G., Mann, L., Bain, P., Pirola-Merlo, A., & Richver, A. (2004). Learning to lead: the development and testing of a model of leadership learning. *The Leadership Quarterly, 15*(3), 311-327. https://doi.org/10.1016/j.leaqua.2004.02.011

Hollenbeck, J.-R., Beersma, B., & Schouten, M. E. (2012). Beyond team types and taxonomies: A dimensional scaling conceptualization for team description. *Academy of Management Review, 37*(1), 82-106. https://doi.org/10.5465/amr.2010.0181

Hollingshead, A. B. (2001). Cognitive interdependence and convergent expectations in transactive memory. *Journal of Personality Social Psychology, 81*(6), 1080.

Holtgraves, T. (2002). Face management and politeness. In *Language as Social Action: Social Psychology* (pp. 37-63).

Holtgraves, T. M. (2008). Face and facework in interpersonal communication. *Face, communication and social interaction*, 192-207.

Holzman, R. S., Cooper, J. B., Gaba, D. M., Philip, J. H., Small, S. D., & Feinstem, D. (1995). Anesthesia crisis resource management: real-life simulation training in operating room crises. *Journal of Clinical Anesthesia, 7*(8), 675-687.

Huang, Y., & Dongilli, T. (2008). Simulation center operations and administration. In *Manual of Simulation in Healthcare* (pp. 29-50).

Hull, L., Russ, S., Ahmed, M., Sevdalis, N., & Birnbach, D. J. (2016). Quality of interdisciplinary postsimulation debriefing: 360° evaluation. *BMJ Simulation and Technology Enhanced Learning, 3*(1), 9-16. https://doi.org/10.1136/bmjstel-2016-000125

Hunziker, S., Buhlmann, C., Tschan, F., Balestra, G., Legeret, C., Schumacher, C., Semmer, N. K., Hunziker, P., & Marsch, S. (2010). Brief leadership instructions improve cardiopulmonary resuscitation in a high-fidelity simulation: a randomized controlled trial. *Critical Care Medicine, 38*(4), 1086-1091. https://doi.org/10.1097/CCM.0b013e3181cf7383

Hunziker, S., Laschinger, L., Portmann-Schwarz, S., Semmer, N. K., Tschan, F., & Marsch, S. (2011). Perceived stress and team performance during a simulated resuscitation. *Intensive Care Medicine, 37*(9), 1473-1479. https://doi.org/10.1007/s00134-011-2277-2

Hunziker, S., O'Connell, K. J., Ranniger, C., Su, L., Hochstrasser, S., Becker, C., Naef, D., Carter, E., Stockwell, D., & Burd, R. S. (2018). Effects of designated leadership and team-size on cardiopulmonary resuscitation: The Basel-Washington SIMulation (BaWaSim) trial. *Journal of Critical Care, 48*, 72-77.

Hunziker, S., Tschan, F., Semmer, N. K., Zobrist, R., Spychiger, M., Breuer, M., Hunziker, P. R., & Marsch, S. (2009). Hands-on time during cardiopulmonary resuscitation is affected by the process of teambuilding: a prospective randomised simulator-based trial. *BMC Emergency Medicine, 9*, 3. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=19216796

Husebø, S. E., Dieckmann, P., Søreide, E., & Friberg, F. (2013). The relationship between facilitators' questions and the level of reflection in postsimulation debriefing. *Empirical Investigations, 8*(3), 135-142.

Janis, I. L. (1972). Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes.

Kihlgren, P., Spanager, L., & Dieckmann, P. (2015). Investigating novice doctors' reflections in debriefings after simulation scenarios. *Medical Teacher, 37*(5), 437-443. https://doi.org/10.3109/0142159X.2014.956054

Klein, K. J., Ziegert, J. C., Knight, A. P., & Xiao, Y. (2006). Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative Science Quarterly, 51*(4), 590-621.

Committee on Quality of Health Care in America. Institute of Medicine. (1999*). To err is human: Building a safer health system*.

Kolbe, M., Grande, B., Marty, A., Manka, R., Taramasso, M., Nietlispach, F., Pomar, J. L., Maisano, F., & Reser, D. (2019). Making Heart Team Discussions Work. *Structural Heart, 3*(2), 100-103. https://doi.org/10.1080/24748706.2019.1572254

Kolbe, M., Grande, B., & Spahn, D. R. (2015). Briefing and debriefing during simulation-based training and beyond: Content, structure, attitude and setting. *Best Practice & Research: Clinical Anaesthesiology, 29*(1), 87-96. https://doi.org/10.1016/j.bpa.2015.01.002

Konradt, U., Otte, K. P., Schippers, M. C., & Steenfatt, C. (2016). Reflexivity in teams: A review and new perspectives. *The Journal of Psychology, 150*(2), 153-174. https://doi.org/10.1080/00223980.2015.1050977

Konradt, U., Schippers, M. C., Garbers, Y., & Steenfatt, C. (2015). Effects of guided reflexivity and team feedback on team performance improvement: The role of team regulatory processes and cognitive emergent states. *European Journal of Work and Organizational Psychology, 24*(5), 777-795. https://doi.org/10.1080/1359432x.2015.1005608

Kündig, P., Tschan, F., Semmer, N. K., Morgenthaler, C., Zimmermann, J., Holzer, E., Huber, S. A., Hunziker, S., & Marsch, S. (2019). More than experience: a post-task reflection intervention among team members enhances performance in student teams confronted with a simulated resuscitation task—a prospective randomised trial. *BMJ Simulation and Technology Enhanced Learning, 6*(2).

Lake, F. (2008). Teaching in clinical settings. In *Manual of simulation in healthcare* (pp. 125-137). Oxford University Press.

Lambe, K. A., Hevey, D., & Kelly, B. D. (2018). Guided reflection interventions show no effect on diagnostic accuracy in medical students. *Frontiers in Psychology, 9:2297*. https://doi.org/10.3389/fpsyg.2018.02297

Larson Jr, J. R., & Egan, A. C. (2018). Information sharing within groups in organizations. In *The Oxford Handbook of Group and Organizational Learning*.

Lauritzen, J., Tschan, F., Zeltner, L., Ostergaard, D., Kruse, N., Rognas, L., & Dieckmann, P. (2016). *In search for reflection in simlation debriefings* SESAM, Lissaabon

Leonard, M., Graham, S., & Bonacum, D. (2004). The human factor: the critical importance of effective teamwork and communication in providing safe care. *Quality and Safety in Health Care, 13*(suppl_1), i85-i90. https://doi.org/10.1136/qshc.2004.010033

Lewis, K. (2003). Measuring transactive memory systems in the field: scale development and validation. *Journal of Applied Psychology, 88*(4), 587.

Lewis, K., Belliveau, M., Herndon, B., & Keller, J. (2007). Group cognition, membership change, and performance: Investigating the benefits and detriments of collective knowledge. *Organizational Behavior and Human Decision Processes, 103*(2), 159-178. https://doi.org/10.1016/j.obhdp.2007.01.005

Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*(3), 356-376.

Marques-Quinteiro, P., Curral, L., Passos, A., Lewis, K., & Gomes, C. (2019). How transactive memory systems and reflexivity relate with innovation in healthcare teams. *Análise Psicológica, 37*(1), 41-51. https://doi.org/10.14417/ap.1519

Marques-Quinteiro, P., Curral, L., Passos, A. M., & Lewis, K. (2013). And now what do we do? The role of transactive memory systems and task coordination in action teams. *Group Dynamics: Theory, Research, Practice, 17*(3), 194.

Marsch, S., Müller, C., Marquardt, K., Conrad, G., Tschan, F., & Hunziker, P. R. (2004). Human factors affect the quality of cardiopulmonary resuscitation in simulated cardiac arrests. *Resuscitation, 60*(1), 51-56.

Marsch, S., Tschan, F., Semmer, N. K., Spychiger, M., Breuer, M., & Hunziker, P. R. (2005). Performance of first responders in simulated cardiac arrests. *Critical Care Medicine, 33*(5), 963-967. https://doi.org/10.1097/01.ccm.0000157750.43459.07

Marsch, S., Tschan, F., Semmer, N. K., Zobrist, R., Hunziker, P. R., & Hunziker, S. (2013). ABC versus CAB for cardiopulmonary resuscitation: a prospective, randomized simulator-based trial. *Swiss Medical Weekly, 143*, w13856. https://doi.org/10.4414/smw.2013.13856

Marsch, S. U. (1998). Team oriented medical simulations In L. Henson & A. Lee (Eds.), *Simulators in Anesthesiology Education* (pp. 51-55). Plenum Press

Mathieu, J. E., Heffner, T. S., Goodwin, G. F., Salas, E., & Cannon-Bowers, J. A. (2000). The influence of shared mental models on team process and performance. *Journal of Applied Psychology, 85*(2), 273-283. https://doi.org/10.1037//0021-9010.85.2.273

McGrath, J. E. (1984). *Groups: Interaction and performance* (Vol. 14). Prentice-Hall Englewood Cliffs, NJ.

McGrath, J. E. (1991). Time, interaction, and performance (TIP) A Theory of Groups. *Small Group Research, 22*(2), 147-174.

McGrath, J. E., & Altermatt, T. W. (2001). Observation and analysis of group interaction over time: Some methodological and strategic choicess. In *Blackwell handbook of social psychology: Group processes* (pp. 525-556).

McHugh, M. (2012). Interrater reliability: the kappa statistic. *Biochemia Medica, 22*(3), 276-282.

Mesmer-Magnus, J. R., & Viswesvaran, C. (2005). Whistleblowing in organizations: An examination of correlates of whistleblowing intentions, actions, and retaliation. *Journal of Business Ethics, 62*(3), 277-297.

Mohammed, S., Ferzandi, L., & Hamilton, K. (2010). Metaphor no more: A 15-year review of the team mental model construct. *Journal of Management, 36*(4), 876-910. https://doi.org/10.1177/0149206309356804

Mojzisch, A., & Schulz-Hardt, S. (2010). Knowing others' preferences degrades the quality of group decisions. *Journal of Personality and Social Psychology, 98*(5), 794-808. https://doi.org/10.1037/a0017627

Moorthy, K., Vincent, C., & Darzi, A. (2005). *Simulation based training* (0959-8138).

Moreland, R. L. (1999). Transactive memory: Learning who knows what in work groups and organizations. *Shared Cognition in Organizations: The Management of Knowledge*, 3.

Moreland, R. L., & Levine, J. M. (1992). In G. p. a. productivity (Ed.).

Moreland, R. L., & McMinn, J. G. (2010). Group reflexivity and performance. In R. S. Thye & E. Lawler (Eds.), *Advances in Group Processes (Vol. 27)* (Vol. 27, pp. 63-95). Emerald Press.

Moreland, R. L., & Myaskovsky, L. (2000). Exploring the performance benefits of group training: Transactive memory or improved communication? *Organizational Behavior and Human Decision Processes, 82*(1), 117-133.

Nolan, J. P., Soar, J., Zideman, D. A., Biarent, D., Bossaert, L. L., Deakin, C., Koster, R. W., Wyllie, J., Bottiger, B., & Group, E. R. C. G. W. (2010). European resuscitation council guidelines for resuscitation 2010 section 1. Executive summary. *Resuscitation, 81*(10), 1219-1276. https://doi.org/10.1016/j.resuscitation.2010.08.021

Okhuysen, G. A. (2001). Structuring change: Familiarity and formal interventions in problem-solving groups. *Academy of Management Journal, 44*(4), 794-808.

Oriot, D., & Alinier, G. (2018). *Pocket book for simulation debriefing in healthcare*. Springer.

Otte, K. P., Konradt, U., Garbers, Y., & Schippers, M. C. (2017). Development and validation of the REMINT: a reflection measure for individuals and teams. *European Journal of Work and Organizational Psychology, 26*(2), 299-313. https://doi.org/10.1080/1359432x.2016.1261826

Otte, K. P., Konradt, U., & Oldeweme, M. (2018). Effective team reflection: The role of quality and quantity. *Small Group Research, 49*(6), 739-766. https://doi.org/10.1177/1046496418804898

Paull, D. E., Mazzia, L. M., Wood, S. D., Theis, M. S., Robinson, L. D., Carney, B., Neily, J., Mills, P. D., & Bagian, J. P. (2010). Briefing guide study: preoperative briefing and postoperative debriefing checklists in the Veterans Health Administration medical team training program. *The American Journal of Surgery, 200*(5), 620-623.

Porta, C. R., Foster, A., Causey, M. W., Cordier, P., Ozbirn, R., Bolt, S., Allison, D., & Rush, R. (2013). Operating room efficiency improvement after implementation of a postoperative team assessment. *Journal of Surgical Research, 180*(1), 15-20.

Rasker, P. C., Post, W. M., & Schraagen, J. M. (2000). Effects of two types of intra-team feedback on developing a shared mental model in Command & Control teams. *Ergonomics, 43*(8), 1167-1189. https://doi.org/10.1080/00140130050084932

Reiter-Palmon, R., Kennel, V., Allen, J. A., & Jones, K. J. (2018). Good catch! Using interdisciplinary teams and team reflexivity to improve patient safety. *Group & Organization Management, 43*(3), 414-439. https://doi.org/10.1177/1059601118768163

Reser, D., Maisano, F., Pomar, J. L., Taramasso, M., & Kolbe, M. (2018). How to make a "Heart Team" a "real" Team to ensure optimal patient care. *EJCM, 6*(2), 32-35.

Rico, R., Sánchez-Manzanares, M., Gil, F., & Gibson, C. (2008). Team implicit coordination processes: A team knowledge–based approach. *Academy of Management Review, 33*(1), 163-184.

Riley, R. H. (2008). *Manual of simulation in healthcare*. OUP Oxford.

Roh, Y. S., Ahn, J.-W., Kim, E., & Kim, J. (2018). Effects of Prebriefing on Psychological Safety and Learning Outcomes. *Clinical Simulation in Nursing, 25*, 12-19. https://doi.org/10.1016/j.ecns.2018.10.001

Rouse, W. B., & Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin, 100*(3), 349.

Salas, E., Klein, C., King, H., Salisbury, M., Augenstein, J. S., Birnbach, D. J., Robinson, D. W., & Upshaw, C. (2008). Debriefing medical teams: 12 evidence-based best practices and tips. *The Joint Commission Journal on Quality and Patient Safety, 34*(9), 518-527.

Salas, E., Paige, J. T., & Rosen, M. A. (2013). Creating new realities in healthcare: the status of simulation-based training as a patient safety improvement strategy. *BMJ Quality & Safety, 22*(6), 449-452. https://doi.org/10.1136/bmjqs-2013-002112

Savoldelli, G. L., Naik, V. N., Park, J., Joo, H. S., Chow, R., & Hamstra, S. J. (2006). Value of debriefing during simulated crisis management: Oral versus video-assisted oral feedback. *Anesthesiology: The Journal of the American Society of Anesthesiologists, 105*(2), 279-285.

Schippers, M. C., Den Hartog, D. N., & Koopman, P. L. (2007). Reflexivity in teams: A measure and correlates. *Applied Psychology, 56*(2), 189-211. https://doi.org/10.1111/j.1464-0597.2006.00250.x

Schippers, M. C., Den Hartog, D. N., Koopman, P. L., & Van Knippenberg, D. (2008). The role of transformational leadership in enhancing team reflexivity. *Human Relations, 61*(11), 1593-1616.

Schippers, M. C., Den Hartog, D. N., Koopman, P. L., & Wienk, J. A. (2003). Diversity and team outcomes: The moderating effects of outcome interdependence and group longevity and the mediating effect of reflexivity. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 24*(6), 779-802.

Schippers, M. C., Edmondson, A. C., & West, M. A. (2014). Team reflexivity as an antidote to team information-processing failures. *Small Group Research, 45*(6), 731-769. https://doi.org/10.1177/1046496414553473

Schippers, M. C., Edmondson, A. C., & West, M. A. (2018). Team reflexivity. In *The Oxford handbook of group and organizational learning* (pp. 174-194). https://doi.org/10.1093/oxfordhb/9780190263362.013.39

Schippers, M. C., Homan, A. C., & van Knippenberg, D. (2013). To reflect or not to reflect: Prior team performance as a boundary condition of the effects of reflexivity on learning and final team performance. *Journal of Organizational Behavior, 34*(1), 6-23. https://doi.org/10.1002/job.1784

Schippers, M. C., West, M. A., & Dawson, J. F. (2015). Team reflexivity and innovation: The moderating role of team context. *Journal of Management, 41*(3), 769-788.

Schmutz, J. B., & Eppich, W. J. (2017). Promoting learning and patient care through shared reflection: A conceptual framework for team reflexivity in health care. *Academic Medicine, 92*(11), 1555-1563. https://doi.org/10.1097/ACM.0000000000001688

Schmutz, J. B., & Eppich, W. J. (2019). When I say… Team reflexivity. *Medical Education, 53*(6), 545-546.

Schmutz, J. B., Kolbe, M., & Eppich, W. J. (2018). Twelve tips for integrating team reflexivity into your simulation-based team training. *Medical Teaching*, 1-7. https://doi.org/10.1080/0142159X.2018.1464135

Schmutz, J. B., Lei, Z., Eppich, W. J., & Manser, T. (2018). Reflection in the heat of the moment: The role of in-action team reflexivity in health care emergency teams. *Journal of Organizational Behavior, 39*(6), 749-765.

Schön, D. (1938). *The reflective practitioner*. Basic Books

Schwappach, D. L. (2014). Risk factors for patient-reported medical errors in eleven countries. *Health Expectations, 17*(3), 321-331.

Sexton, B., Marsch, S., Helmreich, R., Betzendoerfer, D., Kocher, T., & Scheidegger, D. (1998). Participant evaluation of team oriented medical simulation In L. Henson & A. Lee (Eds.), *Simulators in Anesthesiology Education* (pp. 107-108). Plenum Press

Smith-Jentsch, K. A., Cannon-Bowers, J. A., Tannenbaum, S. I., & Salas, E. (2008). Guided team self-correction impacts on team mental models, processes, and effectiveness. *Small Group Research, 39*(3), 303-327.

Soar, J., Nolan, J. P., Bottiger, B. W., Perkins, G. D., Lott, C., Carli, P., Pellis, T., Sandroni, C., Skrifvars, M. B., Smith, G. B., Sunde, K., & Deakin, C. D. (2015). European resuscitation council guidelines for resuscitation 2015: Section 3. Adult advanced life support. *Resuscitation, 95*, 100-147. https://doi.org/10.1016/j.resuscitation.2015.07.016

Somech, A. (2006). The effects of leadership style and team process on performance and innovation in functionally heterogeneous teams. *Journal of Management, 32*(1), 132-157. https://doi.org/10.1177/0149206305277799

St.Pierre, M., Hofinger, G., Buerschaper, C., & Simon, R. (2011). The key to success: Teamwork. In *Crisis Management in Acute Care Settings* (pp. 195-220). https://doi.org/10.1007/978-3-642-19700-0_11

Stasser, G., Stewart, D. D., & Wittenbaum, G. M. (1995). Expert roler and information exchange during discussion: The importance of knowing who knows what. *Journal of Experimental Social Psychology, 31*, 244-265.

Staw, B. M. (1975). Attribution of the "causes" of performance: A general alternative interpretation of cross-sectional research on organizations. *Organizational Behavior Human Performance, 13*(3), 414-432.

Steinemann, S., Bhatt, A., Suares, G., Wei, A., Ho, N., Kurosawa, G., Lim, E., & Berg, B. (2016). Trauma team discord and the role of briefing. *The Journal of Trauma and Acute Care Surgery, 81*(1), 184.

Stout, R. J., Cannon-Bowers, J. A., Salas, E., & M., M. D. (1999). Planning, shared mental models, and coordinated performance: An empirical link is established. *Human Factors, 41*(1), 61–71.

Su, L., Kaplan, S., Burd, R., Winslow, C., Hargrove, A., & Waller, M. J. (2017). Trauma resuscitation: Can team behaviours in the prearrival period predict resuscitation performance? *BMJ Simulation and Technology Enhanced Learning, 3*(3), 106-110. https://doi.org/10.1136/bmjstel-2016-000143

Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist, 45*(2), 120-133. https://doi.org/10.1037/0003-066x.45.2.120

Swift, T. A., & West, M. A. (1998). *Reflexivity and group processes: Research and practice.* ESRC Centre for Organization and Innovation

Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. *Human factors, 55*(1), 231-245. https://doi.org/10.1177/0018720812448394

Tesler, R., Mohammed, S., Hamilton, K., Mancuso, V., & McNeese, M. (2017). Mirror, mirror: Guided storytelling and team reflexivity's influence on team mental models. *Small Group Research, 49*(3), 267-305. https://doi.org/10.1177/1046496417722025

Tjosvold, D. (1991). *Team organization: An enduring competitive advantage*. John Wiley & Sons Inc.

Tjosvold, D., Hui, C., & Yu, Z. Y. (2003). Conflict management and task reflexivity for team in-role and extra-role performance in China. *International Journal of Conflict Management, 14*(2), 141-163. http://www.emeraldinsight.com/journals.htm?issn=10444068&volume=14&issue=2&articleid=1660048&show=pdf

Tramèr, L., Becker, C., Schumacher, C., Beck, K., Tschan, F., Semmer, N. K., Hochstrasser, S., Marsch, S., & Hunziker, S. (2020). Association of self-esteem, personality, stress and gender with performance of a resuscitation team: A simulation-based study. *Plos one, 15*(5), e0233155.

Tschan, F., McGrath, J. E., Semmer, N. K., Arametti, M., Bogenstatter, Y., & Marsch, S. U. (2009). Temporal aspects of processes in ad-hoc groups: A conceptual shema and some research examples. In R. Roe, M. J. Waller, & S. Clegg (Eds.), *Doing time. Advancing temporal research in organizations.* Routledge.

Tschan, F., & Semmer, N. K. (2001). Wenn alle dasselbe denken: Geteilte mentale Modelle und Leistung in der Teamarbeit. *Projektgruppen in Organisationen*, 217-235.

Tschan, F., Semmer, N. K., Gautschi, D., Hunziker, P., Spychiger, M., & Marsch, S. U. (2006). Leading to recovery: Group performance and coordinative activities in medical emergency driven groups. *Human Performance, 19*(3), 277-304.

Tschan, F., Semmer, N. K., Vetterli, M., Gurtner, A., Hunziker, S., & Marsch, S. (2011). Developing observational categories for group process research based on task and coordination-requirement analysis: Examples from research on medical emergency-driven teams. In M. Boos, M. Kolbe, P. Kappeler, & T. Ellwart (Eds.), *Coordination in Human and Primate Groups* (pp. 93-). Springer.

Tschan, F., Zimmermann, J., & Semmer, N. K. (2018). Rules for coding scheme development. In *The Cambridge Handbook of Group Interaction Analysis* (pp. 191-207).

Institute at the National Patient Safety Foundation. (2010). *Unmet needs: Teaching physicians to provide safe patient care*.

Urbini, F., Callea, A., Chirumbolo, A., Talamo, A., Ingusci, E., & Ciavolino, E. (2018). Team performance in the Italian NHS: The role of reflexivity. *Journal of Health Organization and Management*. https://doi.org/10.1108/jhom-07-2017-0180

Valenzuela, T. D., Roe, D. J., Cretin, S., Spaite, D. W., & Larsen, M. P. (1997). Estimating effectiveness of cardiac arrest interventions: a logistic regression survival model. *Circulation, 96*(10), 3308-3313.

Vashdi, D. R., Bamberger, P. A., & Erez, M. (2013). Can surgical teams ever learn? The role of coordination, complexity, and transitivity in action team learning. *Academy of Management Journal, 56*(4), 945-971. https://doi.org/10.5465/amj.2010.0501

Vashdi, D. R., Bamberger, P. A., Erez, M., & Weiss-Meilik, A. (2007). Briefing-debriefing: Using a reflexive organizational learning model from the military to enhance the performance of surgical teams. *Human Resource Management, 46*(1), 115-142. https://doi.org/10.1002/hrm.20148

Villado, A. J., & Arthur Jr, W. (2013). The comparative effect of subjective and objective after-action reviews on team performance on a complex task. *Journal of Applied Psychology, 98*(3), 514.

Office fédéral de la santé publique. (2019). *Améliorer la qualité et la sécurité des soins en Suisse.*

Waller, M. J., Okhuysen, G. A., & Saghafian, M. (2016). Conceptualizing emergent states: A strategy to advance the study of group dynamics. *The Academy of Management Annals, 10*(1), 561-598.

Weaver, S. J., Dy, S. M., & Rosen, M. A. (2014). Team-training in healthcare: a narrative synthesis of the literature. *BMJ Quality & Safety, 23*(5), 359-372.

Wegner, D. M. (1987). Transactive memory: A contemporary analysis of the group mind. In *Theories of group behavior* (pp. 185-208). Springer.

Weingart, N. S., Wilson, R. M., Gibberd, R. W., & Harrison, B. (2000). Epidemiology of medical error. *BMJ, 320*(7237), 774-777.

Weldon, E., Jehn, K. A., & Pradhan, P. (1991). Processes that mediate the relationship between a group goal and improved group performance. *Journal of Personality and Social Psychology, 61*(4), 555-569.

West, M. A. (1996). *Reflexivity and work group effectiveness: A conceptual integration*. John Wiley & Sons, Ltd.

West, M. A. (2000). Reflexivity, revolution and innovation in work teams. teams (Ed.), *Product Development Teams* (Vol. 5).

West, M. A. (2012). *Effective teamwork: Practical lessons from organizational research*. John Wiley & Sons.

West, M. A., & Sacramento, C. A. (2011). Team reflexivity. In J. M. Levine & M. A. Hogg (Eds.), *Encyclopedia of Group Processes and Intergroup Relations* (pp. 907-909). Sage.

Widmer, P. S., Schippers, M. C., & West, M. A. (2009). Recent developments in reflexivity research: A review. *Psychology of Everyday Activity, 2*(2).

Wiedow, A., & Konradt, U. (2010). Two-dimensional structure of team process improvement: Team reflection and team adaptation. *Small Group Research, 42*(1), 32-54. https://doi.org/10.1177/1046496410377358

Wilson, J. M., Goodman, P. S., & Cronin, M. A. (2007). Group Learning *The Academy of Management Review, 32,* 1041-1059.

Zwaan, L., de Bruijne, M., Wagner, C., Thijs, A., Smits, M., van der Wal, G., & Timmermans, D. (2010). Patient record review of the incidence, consequences, and causes of diagnostic adverse events. *Archives of Internal Medicine, 170*(12), 1015-1021.

7. Appendices

Appendix A

Performance in cardiopulmonary resuscitation: hands-on time and time to first meaningful measure

This appendix presents the coding manual “performance in cardiopulmonary resuscitation” used in paper 1 and 2. Coding is for both simulations pre- and post-reflexivity.

Plan of the present action coding guide

I Action coding: Basics

I Action coding: Variables (gross coding)

III Action coding: Global performance measurement (CPR Performance)

IV Action coding: Specific performance measurements

V Action coding: Measures calculation

I Action coding: Basics

1. Basis = video transcription (Excel file) -> Copy transcription in a new sheet, but rename (action coding, then also performance measure/performance, etc.)

Please always start new codes in a new Excel sheet so that intermediate stages are not lost and you can quickly go back to an earlier version (name each sheet so that you know on which you did what).

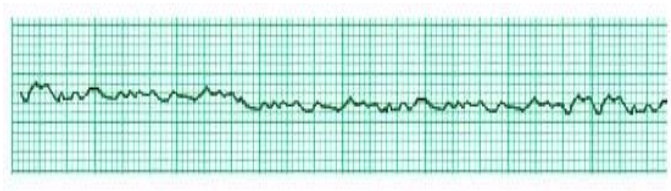
2. Identify Time beginning KT/KF: (whole row) RED highlight this second

Figure 1

KT (Ventricular tachycardia)

Figure 2

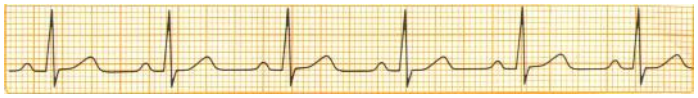
KF (Ventricular fibrillation)



3. Identify time end KT/KF, respectively return of sinus rhythm: (whole row) GREEN highlight this second (follows a few seconds after the last defibrillation).

Figure 3

Sinus rhythm = NORMAL



4. The action coding is done for every second (each row = 1 second) of cardiac arrest, thus code who does what when:

   a. WHO (number of the person see description of the participants in the transcript or number on the T-shirt for students; confederate is marked with 99)

   b. WHAT (each variable is coded in one column, see below)

c. WHEN (each row represents one second. Blank rows mean that none of the following variables were executed).

II Action coding: Variables (gross coding)

Table 1

Resuscitation performance coding

Variable	Description	Coding instructions
PT	**Precordial Thump**: Punch with fist on the breast of patient	If this happens, code with number of the person the second where the patient is "beaten".
V	**Ventilation**: Ventilation of the patient (with mask / ev. mouth to mouth)	**V only (not coordinated with CC):** Is coded to 2 seconds for each time the pump is pressed (bag-valve mask) ->2 sec = tighten + loosen the pump Simply hold the mask on the face is not coded.
		V coordinated with CC: If the group massagse and ventilates in a coordinated manner, V is through-coded (for statistical reasons). This means that if one person holds the mask (more or less) over the face of the patient while someone massages, then both are coded by.
CC	**Chest Compressions**: cardiac massage	It is coded for every second when someone presses on the chest of the patient.
DI	**Defibrillation Intention**: It is about capturing the moment when they say/show that defibrillation is now the next action and that defibrillation will become a major task.	Coding the second someone says "You have to defibrillate". Attention, this statement must result in: • someone takes charge of the defibrillator • or the confederate explains the defibrillator • or the confederate performs the defibrillator Coding the second when someone (not the confederate) goes to get the defibrillator (or brings the defibrillator closer if it is already next to the patient). The group member shows with his behavior that he wants to defibrillate. If it is explicitly communicated, then encode communication. *Background:* *This variable is especially important for students (and family doctors) who cannot operate the defibrillator.* *The confederate then sometimes performs the defibrillation themselves and sometimes not (if not, he explains the defibrillation and this can take a long time). This means that the groups in which the confederate does not perform the defibrillation but only explains the defibrillation are disadvantaged.* *It is therefore a matter of coding the moment when – if they had "mastered" the defibrillation – the defibrillation intention would have quickly led to defibrillation.*

(Table 1 continued)

DM	**Defibrillation moment**: Coding the second where the shock goes off	You actually hear a "clic" or you see them pressing the two buttons of the paddles. Under comments: Enter **Joules** (e.g. 100J, 200J, 360J) (on the same row).
DrA	**Defibrillation related activity**: coding when defibrillation becomes the main task	Preparing the defibrillator - as soon as someone turns their hands to the defibrillator.
I	**Intubation**: Coding when intubation becomes the main task (CC must be interrupted).	Intubation involves the introduction of the tube into the throat and later on also listening (with a stethoscope). Preparation of intubation material is not coded. If several attempts are necessary, code several times. *Some teams – erroneously – do not interrupt the heart massage, then code I and CC simultaneously.* Under comments: code **Intubation and then check intubation** (on the first row for each different subtask).
M	**Medication**: Code the second when a drug is injected (e.g.: lidocaine, amiodarone =cordarone, atropine, adrenalin=epinefrin)	You cannot always tell exactly when they are injecting medication. Help with the coding: • **Communication**: "The medication/adrenalin is in" • **Observation**: If the moment is not clearly visible, code the second before rinsing with NaCl (turn the perfusion) Under comments: **Enter drug name and dose** (on the same row).
Ratio CC:V	**Ratio Chest Compressions: Ventilation** Coding ratio CC to V (the aim is to record a regularity)	The first time CC is executed, code the ratio CC to V. If there are rhythm changes, re-code. Also re-code if another person takes over the CC or V. **Known ratios**: 30:2; 15:2, 5:1 **Other ratios**: • uncoordinated (never the same ratio or CC and V together) • 15: uncoordinated • uncoordinated:2 • or unknown but regular ratio e.g. 60:5 • etc.
Comments		• **For DM:** Code joules (e.g. 100J, 200J, 360J) • **For M:** Drugs: register name and dose • **For I: Intubation and then check intubation** register (on the first row for each different subtask) • **OTHERS in case of special actions, ambiguities, etc.**

III Action coding: Global performance measurement (CPR Performance)

CPR Performance = % of cardiac arrest time in which meaningful measures are performed

(Variable in column= SUPPORT)

First step

Copy the gross action coding into 1 new sheet, rename it (performance, ...)

Delete the following variables (columns) from the gross coding system: DI, M, ratio, comments

That only leaves: PT, V, CC, DM, DrA, I

Create a new column with "Support" *(should be in column "L" on Excel)* and insert the following formula structure into each row of the cardiac arrest (from the red row included to the green row exclusive).

=SI((F140+G140+H140+I140+J140+K140=0);0;1)

- ➔ Meaning of the formula: it calculates whether a number has been coded in any row in the different columns (F for PT, G for V, etc.) (measure coded).
  - If no, then Excel encodes the value "0" in the "support" column
  - If yes, then Excel encodes the value "1" in the "support" column

You can copy the formula from the sample transcript (54-02.xls) (and adjust it to columns, rows). ATTENTION: Formula depends of course on the columns (here F to K), and also on the row (here 140)

Second step

Global performance = % of cardiac arrest time (from the red row inclusive. to the green row exclusive) where the patient is actively supported.

In the column "L" (= support), green row, insert the following formula structure (see example):

=(SOMME(L140:L710))/(NB(L140:L710))

➔ Meaning of the formula: it calculates the sum of everything 1 and 0 in the support column; this sum is divided by the number of rows. It gives the % of time the teams have given active support.

- Here are L140:L710 = Cells between red row inclusive and green row exclusive.

Third step

Please enter the "cardiac arrest time" separately (below the green row). Need to describe data qualitatively (e.g.: mean experimental time for all teams). Time = in seconds

=NB(L140:L710)

IV Action coding: Specific measurements

To be entered in a separate Excel file ***(performance*** *see example "Results_REA2_Regula_MV.xls", PLEASE DO NOT UPDATE THIS FILE).*

Table 2

Resuscitation performance measures

Variable	**Description**	**Calculation (*instructions in italics*)**
Group	Group name	
CPR performance	Link to the Global Performance Measurement	see page 216, step 2 e.g. = (SOMME(L140:L710))/(NB(L140:L710))
CPR performance (in % of time)	CPR performance * 100	Multiply cell "CPR performance" by 100
Duration CPR (sec)	Link to the Global Performance Measurement	see page 217, step 3 e.g. = NB(L140:L710)
Time of CA (cardiac arrest)	Tim in the red row	e.g. 14:58:00
ATTENTION: THE FOLLOWING INFORMATION MUST BE READ FROM THE GROSS CODE (not from the specification sheet)		
Time of 1st treatment (PT, D, CC, M, I)	Time where you start with the first measure: either PT, D, CC, M or I	*Caution: not V if performed alone (without CC)* *If CC (or CPR, resp. CC and V together), die code the **first** second*
Time form CA until 1st treatment (PT, D, CC, M, I) in sec	Time between cardiac arrest and first measure	Cell "Time of 1st treatment" minus cell "Time of CA"
1st treatment	Manual coding which is the first measure	*Either PT, D, CC, M, I oder CPR* *- CPR: if you run CC and V simultaneously*
Time of 1st defibrillation	Time of first defibrillation (DM)	*DM (not DrA)*
Time form CA until 1st defibrillation in sec	Time between cardiac arrest and first defibrillation	Row „Time of 1st defibrillation" minus row "Time of CA"

(Table 2 continued)

Time the 1st cycle (D, M, at least one CPR - or CC - cycle) is completed	Time where they are through with the last measure for the first time. You must have done at least 1 D, M and 1 CC (or CPR) cycle	*Either* D, M, at least one CPR – or CC – cycle *In the case CC (or CPR, resp. CC and V together), code the **last** second oft he cycle* *(If a group should never defibrillate or give medication then code "no value")*
Time form CA until 1st cycle is completed in sec	Time between cardiac arrest and execution of all measures	Cell "Time the 1st cycle is completed" minus cell "Time of CA" *(If a group should never defibrillate or give medication then code "no value")*
End of 1st cycle	Manual coding which is the last measure	*Either D, M, CPR or CC* *(If "no value" here explains what they never did)*

V Action coding: Measures calculation

(Example is in Stud-BS-REA-B1-20.05.2010)

1) create two new columns to the right of comments:

   a. ***no pulse1*** - Simulation 1: code1 for every second the patient has no pulse and until the confederate interrupts

   b. ***no pulse 2*** - Simulation 2: code 1 for every second the patient has no pulse (from the moment the patient says he gets dizzy) until return of cardiac rhythm. With this we measure the time of resuscitation.

2) Insert a column to the left of PT called ***hands-on 1***

   a. Enter a 1 in this column for the first simulation

      i. it has something in CC

      ii. it has something in V, but only if ventilation is connected with cardiac massage (i.e. immediately before CC (max 5 seconds) or between two CCs (there max 5 seconds), if there is no gap between two CCs greater than 5 seconds and ventilation takes place in it, code hands-on time

iii. if there is a DM, 5 seconds after DM, even if there is nothing there, also enter hands-on (we give them the time as hands-on time because there they can check if the defibrillation has worked)

3) Rename the PT column to ***hands-on 2***

a. Enter a 1 in this column for the second simulation

i. it has something in CC

ii. it has something in V, but only if ventilation is connected with cardiac massage (i.e. immediately before CC (max 5 seconds) or between two CCs (there max 5 seconds), if there is no gap between two CCs greater than 5 seconds and ventilation takes place in it, code hands-on time

iii. if there is a DM, 5 seconds after DM, even if there is nothing there, also enter hands-on (we give them the time as hands-on time because there they can check if the defibrillation has worked)

4) Copy the formula row from an old file (you must have both files open to copy the formulas) and paste it after observation: Be careful to make sure that the titles match the above, that is very important

5) First meaningful measure: what the team did first, it can only be V or CC or DM (either V or CC or DM), for simulation 1 and simulation 2

6) Time to first meaningful measure (enter time column FMM minus time column start in excel, see formula row, make it for simulation 1 and simulation 2)

7) Sum of hands-on time (sum of all 1 at hands-on time, see formal row, this should do it automatically)

8) Duration of hands-on time (it should also do this automatically)

9) % hands-on time (sum hands-on time / sum seconds rea), see formula row, it should also do this automatically, otherwise adjust the formula

Appendix B

Adherence to the temporal structure of the reflexivity session

This appendix presents the coding manual "adherence to the temporal structure of the reflexivity session" used in paper 2. Coding is for reflexivity session.

Table 3

Temporal structure codes

	Level	Definition	Comment
Adherence to the structure of the reflexivity process	3	Structure of the reflexivity process mostly respected (retrospect first, followed by outlook)	The structure is coded as fully respected if the group first discussed the retrospective (what went well and what did not), and then the outlook (plan for a similar situation) explicitly. This code indicates the degree to which the group followed the reflexivity instruction sheet.
	2	Structure of the reflexivity process partially respected (retrospect and outlook present)	
	1	Structure of the reflexivity process very partially respected (retrospect or outlook present)	
	0	Structure of the reflexivity process not respected (no retrospect, no outlook), chaotic	

Appendix C

Content of reflexivity session: Task- or team-related topics discussed during reflexivity session

This appendix presents the coding manual “contents discussed in reflexivity session” used in paper 2. Coding is for reflexivity session.

I Coding process

The codes are entered in the file in the same row as the transcribed sentence with corresponding time.

Always enter a 1 when the code is selected. Each row represents one second.

Enter the code and the speaker next to the transcription box

(Define the speakers in the file

Each sentence is coded according to all criteria; multiple coding is possible, for example task-related-content and team-related content in one sentence, or multiple team-related topics or task-related topics occur in one sentence. A 1 is entered in the corresponding column.

- Make three coding runs with the video and the transcription:
    - Run 1: Independent codes
    - Run 2: Task-related coding
    - Run 3: Team-related coding

Ground rule: If several statements obviously belong to the same communication / topic, they are coded / double coded. This also includes e.g. “acknowledgements”.

II Independent codes

Table 3

Content of reflexivity: Independent codes

Code	**Description**	**Example**	**Non-example**
Not-codable	Incomprehensible communication. Demands of the participants based on acoustic incomprehension	*"This is […]", "Then we...", "Oh yes […] out"* *"... what was he [the patient]", "What have you said?"*	
Acknowledgements / filler	Confirmations and/or filler words without further content. Usually these are single words, not whole sentences. However, there can be more than one word.	*"Mhm, yes exactly", "Yes", "Mhm"; "Yes, that is true" „Hmmmmm…", "No"*	*"Yes exactly, you did this"*
Highlight in yellow	Everything related to the instruction sheet, e.g. reading aloud or repeating the task	*"What would we do better?", "What aspects have you adhered to?", "Do we need to discuss this together now?"*	
Do not code	Sentence before the beginning of reflexivity session	*"Have you read the first question?"*	

III Content coding according to the following criteria (task-related coding, codes and descriptions)

Table 4

Content of reflexivity: Task-related communication codes

Code	Description	Example	Non-example
Algorithm	Talk about ABCDE on a meta-level, talk about algorithm in general. This does not include individual contents of A, B, C, D, or E. Algorithm is not double-coded with the individual contents of A, B, C, D, or E. When you talk about REA calling or getting help.	*"Well, does anyone know the algorithm?", "It is more like CABD", "And E is environment", "I do not know exactly what the algorithm is...", "144 we do not have to call because we're in a hospital"*	*A: "We could have maybe looked in the mouth and so..." B: "No, so you checked to see if he was breathing at first?" C: "And then I checked his pulse" All MONA discussions*
Control consciousness	Everything related to the control of consciousness (e.g. talk to the patient, pain stimulus).	*"Painful stimulus on the sternum", "He is unresponsive"*	
Control airways (A)	Checking whether the airways are free or whether one should have checked.	*"We never looked in his mouth"*	
Control breathing (B)	Checking whether the patient is breathing.	*"No, so you checked to see if he was breathing at first?"*	
Circulation (C)	Control circulation (how do you do it, have they done it, concretely), control saturation. This includes control by means of ECG / monitor / screen. Before defibrillation they speak more often of the "rhythm", after defibrillation more often of the "pulse" (circulation).	*"He really did not have any circulation anymore", "First you have to see if there is a rhythm", "We have attached things", "Yes, the thing was connected up wrong, that is why we never had a saturation", "Is he coming back?"*	
Ventilation	How to ventilate, how to use the mask, how to hold the head (e.g. that you have to overstretch), give oxygen, how much, etc. Sometimes they say "CPR" but mean "heart massage" and/or ventilation. Quantity of oxygen is also part of it. Taking the pillow away (so that you can overstretch) is also part of it Attention: "on the head" does not always mean "respiration".	*"The mask is totally difficult to use", "The question is whether you have overstretched it", "And you start with the ventilation"*	*"The leader stays on the head"*
Cycles	Talking about cycles - when during defibrillation are drugs administered - when / how often one has to / can defibrillate or in which intervals. Is often double coded with "defibrillation" or "drugs".	*"That is after about five cycles", "I'm not quite sure if they shock again as quickly as we did", "Yeah, I do not know when you give the medications and what you give exactly"*	
Rhythm	Which heart rhythm is involved, how to recognize it, what differences there are. Before defibrillation they speak more often of "rhythm", after defibrillation more often of "pulse" (circulation). Also question: Is the rhythm from the patient or from the heart massage.	*"I was not sure, it was a ventricular tachycardia?", "Did he have a shockable rhythm?", "It was a ventricular fibrillation"*	
Intubation	Anything to do with intubation. Even if they speak of "tubus".	*"In between you might have had time to put a tubus in it."*	

(Table 4 continued)

Cardiac massage / CPR	When to do cardiac massage, how to do it, how to count compressions (double code with communication), whether to do it (but not synchronization, i.e. 30:2), release during massage, etc. "CPR" and "resuscitation / reanimation" always means cardiac massage.	*"We did CPR well", "So you say when it charges, you can still pump", "And that just says out loud: "You do the CPR, you get the oxygen, I let go, I coordinate the medications"", "CPR we had to do", "And then we start", "We started then"*	*Not synchronization, i.e. 30:2*
Synchronization respiration / cardiac massage	Synchronization between compressions and ventilation (30:2). Often double coded with heart massage and / or ventilation.	*"So sure 30-to-2"*	
Defibrillation	Anything related to defibrillation that is not technical in nature. Defibrillation procedure (charging - say "away" - release) When (at what rhythm, at what time) it is appropriate to defibrillate.	*"We should have defibrillated much earlier", "Should we have stayed off the table?", "You have to say "All gone!"", "A shockable rhythm", "Ventricular fibrillation, is it shockable?"*	*Time: Sentences like "We have to look in the mouth first, then ventilate, then do CPR and only then defibrillate" → is on a higher level → algorithm*
Defibrillation technical	All technical aspects of the defibrillator: setting the joules, phase. When they talk about not knowing the defibrillator (as well as the AED, with which they have more practice).	*"I didn't know if it was mono or biphasic", "You should have hired more Joules", "The defibrillator can be recharged directly", "Too bad we didn't know about the paddles"*	
Drugs	Which and if and when to give medication. This also includes any MONA discussions. What volume to give belongs to the topic of medication.	*"When is this thing coming in (injection)?", "You give a bolus"*	
Bed	Bed height, inclination of the bed frame, hardness of the underlay / mattress, etc.	*"But was the bed at the bottom?"*	
Others	Other content that is part of the task, but for which there is no code. This includes, for example, stories from experiences, diagnostic / prognostic questions or capnography. Is only coded if no content category is coded, i.e. is not double coded (e.g. pupil control, disability, diagnostics). Content that has been coded as "other" is only further coded for collaboration and self-evaluation if it concerns (other) content in the scenario / situation (reflection on the scenario itself is not part of this).	*"We're in a hospital", "I wanted to call the REA team first, but we are the REA team", "If we were on a ward now, we would have called them first, if we were not on the IPS (intensive care unit), I mean", "What is the patient's temperature?"*	*"If he was a pulmonary patient, you'd... still have a pulse", "We never had an automatic defibrillator like this", "144 we don't have to call because we're in a hospital"*
Communication not-related to the task	Anything that is not part of the task, if they deviate from the theme (but no fillers or "I don't know"). Content in this category is no longer coded when collaborating and self-evaluating.	*"I got up very early today", "Simulation is no fun", "You have nice hair"*	

IV Content coding according to the following criteria (team-related coding, codes and descriptions)

Table 5

Content of reflexivity: Team-related communication codes

Code	Description	Example	Non-example
Leadership	General or specific discussion of leadershif Leadership is not double coded with coordination, communication and division of labor, unless one of these three categories is explicitly discussed.	*"Somebody should have taken over the command", "Yes, then I would really say that somebody takes over the command and says... and that also says out loud: "You do the heart massage, you get the oxygen, I let go, I coordinate the medications""*	
Coordination	How to coordinate. Often has a spatiotemporal component. Also code when synchronizing cardiac massage - ventilation - defibrillator (also includes: making sure that everyone is away from the table). When changing roles or tasks (ventilation / cardiac massage / defibrillator). Positioning in the room.	*"One should always be on top", "We should then have walked away from the table", "If we would have had to do it in pairs", "The communication should have been even more", "We do this, then we do that. Now we do this, next we do that"*	
Communication	When it comes to the type of communication (how did you communicate; good, bad, too quiet, too unclear, etc.). Metacommunication about communication. Way of counting during heart massage.	*"We should talk loud and clear", "I think communication was a little bad", "Everyone just said something at one point"*	*"I said, "go away"", "Then in the beginning we also verbalized that it was REA"*
Task distribution	When it comes to who did or is doing what. The focus is on the division of labor. This does not include statements by individuals about what they did exactly (recall – without reference to other tasks).	*"Maybe you could have done that quickly", "You do that, you there", "That would have just been so, really saying out loud who does what", "It might have been cooler if you'd divided it up more clearly", "It just happened we have someone breathing and someone doing CPR"*	*"You gave orders afterwards", "And then I checked the pulse"; "I tried to overstretch", "I looked inside (the mouth)", "You took the defibrillator"*

(Table 5 continued)

Interruptions; speed	Avoid interruptions, work quickly. There are often signal words like “fast”, “slow”, “direct”, “immediate”, etc. Can also be meant positively.	*“No, you have to go on immediately”, “Then we were too slow”, “And give oxygen directly, yes”, “Taking the pulse was too long, too”, “You started right away, that was good”*	*“I was too quick to massage the heart”*

Appendix D

Quality of adherence to the resuscitation algorithm in simulations and aspects of adherence to the resuscitation algorithm discussed in reflexivity.

This appendix presents the coding manual “quality of adherence to algorithm in resuscitation” used in paper 2 for both simulations (T1 and T2) and the manual “aspects of adherence to the resuscitation algorithm discussed in reflexivity” for reflexivity session. Both coding manuals are presented in the same appendix for a better understanding.

I Coding process

Assign a score for each variable in the two simulations (T1 and T2) and for the reflexivity session.

II Coding system quality of adherence to algorithm for simulations T1 and T2

Table 6

Quality of adherence to algorithm – codes related to resuscitation algorithm

Category	Variable	Level	Codes description
Diagnosis	Check consciousness	1	At least one participant asks the patient questions, shakes him, and pinches him
		0	No check of the state of consciousness is carried out
	Check breathing	1	At least one participant checks if the patient is breathing (ear to mouth, observing the chest, or feeling for expelled air)
		0	No check of the breathing is carried out
	Check airways	1	At least one participant checks whether patient's airway is clear
		0	No check of the airways is carried out
	Check pulse	1	At least one participant checks if the patient has a pulse (feel the pulse)
		0	No check of the pulse is carried out
	Ask for help	1	Team requests assistance outside the simulation room (telephone, resuscitation team)
		0	No information is transmitted outside the room

(Table 6 continued)

Basic life support 8BLS)	Cardiac massage rhythm	1	Cardiac massage rhythm is right (100 compressions/minute) for more than half of the simulation
		0	Cardiac massage rhythm is wrong for more than half of the simulation
	Count compressions in cardiac massage	1	The masseur counts the compressions of cardiac massage (including counting at least the last three compressions)
		0	The masseur does not count the compressions of cardiac massage
	Avoid interruptions during BLS-related actions	1	There is no interruption of 10 seconds or more between cardiac massage and ventilation and vice versa or before defibrillation
		0	There are interruptions of 10 seconds or more between cardiac massage and ventilation and vice versa or before defibrillation
	Ventilation	1	Two ventilations are performed in more than half of the compression - ventilation cycles
		0	A different number of two ventilations are performed in more than half of the compression - ventilation cycles
	Overlap between cardiac massage - ventilation	1	There is no overlap between cardiac massage and ventilation
		0	There is overlap between cardiac massage and ventilation
Defibrillation	Warning for shock	1	Responsible for defibrillator warns the team before shock and waits that the team steps back from the table, there is no risk for the team
		0	Responsible for defibrillator does not warn the team before shock or a participant gets the shock
	Premature defibrillation	1	Participants give the following shocks after two minutes / five massage - ventilation cycles
		0	Participants give the following shocks before two minutes / five massage - ventilation cycles
	Avoid interruptions during defibrillation-related actions	1	There is no interruption of 10 seconds or more after defibrillation to resume CPR more than half the time
		0	There are interruptions of 10 seconds or more after defibrillation to resume CPR
Medication	Adrenalin administration	1	1 mg Adrenaline / adenosine is given and the 3 to 5 minutes interval between administrations is respected
		0	No adrenaline / adenosine is given, or adrenaline/adenosine3 is not correctly administered
	Amiodarone administration	1	300 mg amiodarone is given after the third defibrillation
		0	No amiodarone is given, or amiodarone is not correctly administered

Table 7

Quality of adherence to algorithm – codes related to leadership and task distribution

Category	Variable	Level	Codes description
Leadership	Designate leader	1	The leadership role is clearly identified
		0	No leader identified
	Leader task	1	The leader is mainly in charge of the ventilation or does nothing and has a good overview of the team and the monitor
		0	The leader takes care of cardiac massage or defibrillator
Task distribution	Designate responsible ventilation	1	Role of ventilation is clearly taken
		0	No participant is responsible for ventilation
	Designate responsible cardiac massage	1	Role of cardiac massage is clearly taken or ordered by a participant
		0	No participant is responsible for cardiac massage
	Designate responsible defibrillation	1	Role of defibrillation is clearly taken or ordered by a participant
		0	No participant is responsible for defibrillation

III Coding system adherence to algorithm for reflexivity session (T2)

Table 8

Aspects of adherence to algorithm discussed codes

Category	Variable	Level	Description
Same categories and variables		1	The group discusses the variable correctly
		0	The group talks about the variable incorrectly or plans something wrong or the group does not talk about the variable

Appendix E

Unresolved disagreements in reflexivity

This appendix presents the coding manual "Unresolved disagreements" used in paper 2 to code quality of reflexivity. Coding is for reflexivity session.

Process: Assign a score for the indicator by viewing the video and using the transcript. One run of the video.

Table 9

Unresolved disagreements codes

Quality indicator	Level	Definition	Comment
Unresolved disagreements	3	Disagreements were present and remained controversial	Note that disagreements were coded only if they referred to the correct execution of the task (i.e. to elements of the algorithm) or to team coordination and cooperation (e.g. division of labor; leadership).
	2	Disagreements were present but resolved during the reflection session	
	1	No disagreement	

Appendix F

Scope of the reflexivity

This appendix presents the coding manual "Scope of the reflexivity" used in paper 2 to code quality of reflexivity. Coding is for reflexivity session.

Process: Assign a score for the indicator by viewing the video and using the transcript. One run of the video.

Table 10

Scope of reflexivity codes

Quality indicator	Level	Definition	Comment
Scope of the discussion	3	They compared more than once the actual situation (collaboration and algorithm aspects) to other situations / experience / algorithm / patients (high generalization / reflection scope)	*Example of a comparison with potentially similar other situations into the discussion:* *A group member discussing task allocation in emergency groups and telling the others how this task allocation was done in a group she had participated in, but for a different type of emergency.*
	2	They compared at least once the actual situation (collaboration and algorithm aspects) to other situations / experience / algorithm / patients (medium generalization / reflection scope)	
	1	Participants never explicitly compared the actual situation (collaboration and algorithm aspects) to other situations / experience / algorithm / patients (narrow generalization / reflection scope)	