DE-CODE: a coding scheme for assessing debriefing interactions

Julia C Seelandt,1,2 Bastian Grande,2,3 Sarah Kriech,2 Michaela Kolbe2,4

ABSTRACT

Debriefings are crucial for learning during simulation-based training (SBT). Although the quality of debriefings is very important for SBT, few studies have examined actual debriefing conversations. Investigating debriefing conversations is important for identifying typical debriever–learner interaction patterns, obtaining insights into associations between debriefers’ communication and learners’ reflection and comparing different debriefing approaches. We aim at contributing to the science of debriefings by developing DE-CODE, a valid and reliable coding scheme for assessing debriefers’ and learners’ communication in debriefings. It is applicable for both direct, on-site observations and video-based coding.

Methods The coding scheme was developed both deductively and inductively from literature on team learning and debriefing and observing debriefings during SBT, respectively. Inter-rater reliability was calculated using Cohen’s kappa. DE-CODE was tested for both live and video-based coding.

Results DE-CODE consists of 32 codes for debriefers’ communication and 15 codes for learners’ communication. For live coding, coders achieved good inter-rater reliabilities with the exception of four codes for debriefers’ communication and two codes for learners’ communication. For video-based coding, coders achieved substantial inter-rater reliabilities with the exception of five codes for debriefers’ communication and three codes for learners’ communication.

Conclusion DE-CODE is designed as micro-level measurement tool for coding debriefing conversations applicable to any debriefing of SBT in any field (except for the code medical input). It is reliable for direct, on-site observations as well as for video-based coding. DE-CODE is intended for allowing for obtaining insights into what works and what does not work during debriefings and contribute to the science of debriefing.

INTRODUCTION

Debriefing is a core element of team learning and simulation-based training (SBT).1–3 It is an instructor-guided conversation among trainees that aims to understand and understand the relationships among events, actions, thoughts and feeling processes and performance outcomes of the simulation.1,2,4,5 In effective debriefings, learners are encouraged to transfer learning from the simulated setting to the patient care context through reflection.6–7

There are various debriefing approaches available providing advice on how to promote learners’ reflection, for example, the Debriefing with Good Judgment,1 PEARLS,8 The Diamond9 and TeamGAINS.10

In addition, there are techniques available for creating a psychologically safe and engaging setting,11 codebriefing12 and debriefers’ communication such as advocacy inquiry1 and circular questions.12 Though evidence on the effectiveness of debriefings is growing,5,13–15 empirical research evaluating debriefings during SBT is rare, as are studies comparing different debriefing approaches in SBT.16 Even more, a recent meta-analysis on team training in healthcare concluded that training programmes that involved feedback were less effective than programmes without feedback.16 Although debriefing includes much more than giving feedback, this finding is unsettling and calls for further and more detailed research.

Tools have been developed to assess the quality of debriefings, for example, the Debriefing Assessment for Simulation in Healthcare (DASH)17 and the Objective Structured Assessment of Debriefing (OSAD).18,19 These are behavioural marker systems. When using behavioural marker methodology, users rate the overall quality of different behavioural classes (eg, teamwork and communication) rather than single behaviours.20 Both DASH and OSAD have good psychometric qualities17,18 and are extremely useful for developing simulation instructors’ debriefing competencies. However, in a recent study investigating the value of a 360° OSAD-based evaluation of debriefings by examining expert debriefers’ communication and feedback, debriefers and learners, significant differences between these groups were found: debriefers...
perceived the quality of their debriefings more favourably than expert debriefing evaluators.\textsuperscript{21} Also, weak agreement between learner and expert evaluators’ perceptions as well as debriefers’ perceptions were found.\textsuperscript{21} Thus, measuring debriefing quality and processes seems challenging.

Typically, debriefings are conversations within teams.\textsuperscript{4, 12, 22} Applying methods of team interaction analysis may thus be a fruitful debriefing measurement avenue. Though behaviourally anchored rating scales are very useful for providing immediate feedback after rating,\textsuperscript{23} they are less suitable for assessing team dynamics or team interactions, mainly because of their static nature that cannot capture dynamic processes.\textsuperscript{24, 25} For studying team processes and interactions, behaviour coding is the method of choice.\textsuperscript{26–29} In contrast to behaviour rating, in which the quality, quantity or degree of a behaviour is assessed, in behaviour coding, the occurrence and timing and, mostly, duration are coded.\textsuperscript{29} Behaviour coding allows for a more descriptive assessment of behaviour as it occurs and for uncovering team patterns and dynamics that become apparent over time. Research following that approach has been able to show via pattern and lag sequential analysis that patterns of behaviours among team members—rather than individual actions of single team members—are what discriminates higher from lower performing teams.\textsuperscript{30, 31} For example, a behavioural observational study on teamwork and communication within surgical teams has shown that more patient-irrelevant and case-irrelevant communication during wound closure is related to worse patient outcomes.\textsuperscript{32} Such findings were of paramount importance for understanding which factors contribute to effective teamwork; similar analysis of debriefing conversations would provide empirical evidence of what works and what does not work during debriefings.

So far, empirical insights into debriefing interaction patterns are scarce. Few studies have examined actual debriefing conversations and how differences in debriefers’ communication influence learners’ outcomes.\textsuperscript{5, 33–36} More knowledge on debriefing interactions is important for addressing research gaps that have been identified in the debriefing literature\textsuperscript{37–40} such as (A) identifying typical debriefer–learner interaction patterns, (B) obtaining insights into associations between debriefer communication and learners’ reflection, (C) targeting faculty development providing feedback based on identified debriefer–learner communication patterns, as well as for (D) comparing different debriefing approaches.

We aim at contributing to debriefing science by developing DE-CODE, a coding scheme for assessing debriefers’ and learners’ communication in debriefings. We will describe the development of DE-CODE and its reliability and content validity for both video-based coding\textsuperscript{41} and direct, on-site observations.\textsuperscript{41} Moreover, DE-CODE is intended to be applicable to any debriefing of SBT in any field.

**METHODS**

**Development of coding scheme**

The coding scheme was developed both deductively and inductively (figure 1). A subteam of the authors (JCS and MK), who have extensive experience in researching team dynamics and behaviour coding,\textsuperscript{12, 15, 30, 32, 40, 62–68} first reviewed the literature on team learning and team debriefings in SBT.\textsuperscript{1–6, 8–10, 17, 18, 33, 34, 49–58} They also watched five videotaped debriefings during SBT and took notes in free-text form about the observed communication. This SBT took place in a large university hospital in Switzerland; debriefers were familiar with the following approaches: Debriefing with Good Judgment,\textsuperscript{59} TeamGAINS,\textsuperscript{44} guided team self-correction\textsuperscript{22} as well as circular questions\textsuperscript{12}; participants were anaesthesia care providers; debriefers and participants knew each other from working together in the clinical setting. Based on the findings from literature review and their notes, JCS and MK extracted possible codes and developed a first version of the coding scheme.

![Figure 1](http://stel.bmj.com/)  
Process for developing the coding scheme.
This version was subsequently discussed with subject matter experts for assuring content validity. These subject matter experts were chosen based on their expertise with SBT, their familiarity with the debriefing approaches mentioned above, their extensive debriefing experience in different medical disciplines and professions as well as different status levels to avoid bias. Three senior consultants (two anaesthetists and one trauma surgeon) and one emergency nurse reviewed the coding scheme and suggested revisions. Based on their feedback, small changes were made (eg, overlapping or inappropriate codes were excluded). We also refined the first version in repeated, iterative cycles by applying it to five videotaped debriefings that took place in the same hospital during another 2-week SBT session. Debriefers were familiar with the approaches mentioned above; participants were anaesthesia care providers. Two psychologists (JCS and MK) discussed remaining difficulties and decided on the final version of the coding scheme. Code definitions and examples are described in online supplementary table 1 (debriefer communication) and online supplementary table 2 (learner communication).

We developed specific, mutually exclusive codes that are used when communication may be assigned to only one code. This requires less cognitive load because it provides coders with clear rules for code assignment.

Training phase for observers

The five observers underwent an extensive training procedure (between 30 hours and 34 hours for the full coding scheme with all codes) prior to analysing debriefings. All observers held at least a bachelor’s degree in psychology. First, observers were provided with literature on team debriefings, SBT and behaviour coding to familiarise themselves with the context, setting and method. Second, after studying the literature, observers participated in an observing role in SBT, including the debriefings. Third, observers were introduced to DE-CODE and the coding software. Fourth, with a debriefing expert (JCS), they watched and discussed one videotaped debriefing. Fifth, observers independently coded three to five videotaped debriefings. Experts familiar with coding and conducting debriefings (JCS and MK) reviewed these codings and provided extensive, written feedback. Sixth, discrepancies were resolved by discussions and further explanations. Observers performed live coding after having coded at least 30 videotaped debriefings.

Participants

Participants were 168 anaesthesia care providers (82 men and 86 women) from a large teaching hospital in Switzerland including 25 attending anaesthetists, 74 resident anaesthetists, 57 anaesthesia nurses and 12 participants having another educational background. The mean age was 36.40 years (SD = 8.75), and participants mean work experience was 7.4 years (SD = 7.98) ranging between 0 and 35 years.

Procedure

To test whether DE-CODE is applicable and reliable for both video-based and on-site coding, we applied it to 50 videotaped and 12 live debriefings performed during SBT. The videotaped debriefings were collected during two 2-week SBT sessions in a large university hospital in Switzerland.

All simulation scenarios used in this study included critical situations during common patient treatment. They were designed based on critical incident reports and were specially written for SBT. They contained the management of medication (over) doses, anaesthesia inductions in critically ill patients and respiratory problems during anaesthesia induction; leadership and re-evaluation were important components of learning objectives. The participants acted in their usual roles.

Videotaped debriefings were coded 2 months after they had been recorded. During live coding, observers sat in close proximity to the debriefing table. They were able to watch and listen to the debriefing but did not interfere with it at all. Observers were fully blinded to each other’s scoring during live coding.

Observations for both video and live coding started once all learners and debriefers were seated around the table in the debriefing room. They ended with the announcement of a questionnaire for evaluating the debriefing. Participants were informed about coding before SBT and before debriefings started; participation in the SBT was voluntary. Observers did not know debriefers and participants further than from their observation role.

Data coding

To code debriefing data in a way that would allow for further frequency, duration, co-occurrence, sequential and pattern information to be derived, we applied timed event-based coding. Events, also called coding units, were defined as units (ie, mostly sentences) to which a respective code of (online supplementary tables 1 and 2) could be assigned. That is, these predefined communication behaviours were coded each time they occurred involving logging the onset and offset of the event and assigning the respective DE-CODE code. Events were coded based on a similar procedure for video-based and direct, on-site coding. Interact coding software (for video-based coding, figure 2) or the corresponding iOS app (for direct, on-site coding, figure 3) was used for this process. To test for inter-rater reliability, every fourth video-taped debriefing was coded independently by two observers, and all direct, on-site codings were independently performed by two observers.

Statistical analysis

Inter-rater reliability was calculated using Cohen’s kappa (κ). κ is a coefficient of inter-rater agreement for nominal scales and ranges from −1.00 to +1.00 with values lower than 0.41 considered as fair, between 0.60 and 0.80 as substantial and above 0.81 as almost perfect agreement. We calculated Cohen’s κ for the occurrence versus non-occurrence of each code for every 1 min segment of the coded period. Statistical analyses were performed using SPSS V22.0 software.

RESULTS

DE-CODE coding scheme

The final DE-CODE coding schemes consists of 32 codes for coding debriefers’ communication and 15 codes for coding learners’ communication. Codes are organised in five categories that are based on Tobra and Taylor’s four types of speech (ie, framing, advocating, illustrating and inquiring) and an additional category other. Codes, definitions and examples are provided in online supplementary tables 1 and 2, respectively. The complete Coding Manual is available as online supplementary file 1.

Inter-rater reliability

Kappa values for debriefer codes for both video-based and live coding are shown in online supplementary table 1.

For live coding, all Cohen’s κ values were above 0.66 representing substantial to very good agreement except for irony and
humour ($\kappa=0.30$), laughing ($\kappa=0.33$), normalisation ($\kappa=0.46$) and repeating ($\kappa=0.56$). The codes pseudo-observations ($\kappa=1$), knowledge ($\kappa=1$), circular ($\kappa=1$), guess-what-I-am-thinking ($\kappa=1$) and roleplay ($\kappa=1$) received the highest reliability values for live coding.

For video-based coding, all Cohen’s $\kappa$ values were above 0.60 representing substantial agreement except for opinion ($\kappa=0.59$), input simulation ($\kappa=0.59$), laughing ($\kappa=0.58$), structuring ($\kappa=0.59$) and psychological input ($\kappa=0.55$). The codes guess-what-I-am-thinking ($\kappa=1$) and...
Kappa values for learner codes for both video-based and live coding are shown in online supplementary table 2. For live coding, all Cohen’s $\kappa$ values were above 0.68 representing substantial to very good agreement except for *evaluation of learners’ own actions* ($\kappa=0.49$) and *evaluation of team members’ actions* ($\kappa=0.50$). The codes *feelings* ($\kappa=1$), *positive relevance* ($\kappa=1$), *negative relevance* ($\kappa=1$) and *negative evaluation of simulation* ($\kappa=1$) received the highest reliabilities for live coding.

For video-based coding, all Cohen’s $\kappa$ values are above 0.62 representing substantial agreement except for *evaluation of learners’ own actions* ($\kappa=0.38$), *negative relevance* ($\kappa=0.57$) and *expressions of humour* ($\kappa=0.59$). The code *action plan* ($\kappa=1$) received the highest reliability for video-based coding.

**DISCUSSION**

Debriefings are crucial elements of SBT. They help learners to derive meaning from the simulated learning opportunity. Promoting reflection and learning, and ultimately performance and patient care, are goals of debriefings during SBT. While evidence on the effectiveness of debriefings and the number of debriefing approaches are growing, empirical research evaluating debriefings during SBT is rare, as are studies comparing different debriefing approaches in SBT.

With DE-CODE, we contribute to the science of debriefings by providing a macro-level measurement tool for coding debriefing conversations. DE-CODE provides 32 and 15 codes for debriefer and learner communication, respectively. It is intended to be a comprehensive coding scheme for use in debriefing research in any field. Its codes can be used separately, or combined into larger categories, for selected, direct feedback on debriefings and targeted faculty development. If DE-CODE is used for faculty development, selecting a reduced number of codes (eg, 5 or 6) for the behaviours of interest is recommended and would in turn reduce time for training coders and coding. The use of DE-CODE is not limited to a particular debriefing method, findings from literature on team learning and various team debriefing approaches in SBT were included. DE-CODE complements existing debriefing assessment tools that are based on behaviourally anchored rating scales such as the DASH and the OSAD by allowing for measuring the debriefing communication process more descriptively as it occurs and for empirically identifying debriefing communication dynamics. In analogy with teamwork coding schemes such as the act4teams coding scheme and Co-ACT, using DE-CODE may provide the database for performing statistical analyses that can identify interaction patterns that occur above change and relate them with outcomes such as performance or learning indicators. It also may allow for empirically exploring common debriefing issues, for example, how particular debriefer questions may typically trigger certain learner reactions or what debriefing communications may typically be followed by learners verbalising their mental models. These findings would provide important empirical insights into debriefing effectiveness.

With respect to DE-CODE’s psychometric qualities, we tested its inter-rater reliability for coding both video-taped debriefing as well as on-site, live debriefings. For live coding, coders achieved good inter-rater reliabilities with the exception of four codes for debriefer’s communication (ie, *irony and humour, laughing, normalisation* and *repeating for debriefers’ communication*) as well as for two codes for learners’ communication (ie, *evaluation of learners’ own actions and evaluation of team members’ actions*). We will now discuss the potential challenges of applying these six codes.

First, it seems difficult to code *irony and humour* as well as *laughing* reliably. One explanation could be that irony, humour and laughing are mostly expressed in groups, and several people are involved, which makes it challenging to distinguish and code all group members that are laughing at the same time and to distinguish debriefers from learners. We recommend using these codes for video-based data, which allows coders to watch an interaction multiple times, and when humour in debriefing is of particular interest. Second, the code *normalisation* achieved only fair reliability. Since this code contains the debriefer’s subjective evaluation, one might speculate that coders had difficulty distinguishing it from opinions and suggest emphasising the difference among these codes during observer training. Third, the code *repeating* seemed challenging, too. It could be difficult for coders to distinguish whether the debriefer repeats completely what the learner had articulated from whether he or she repeats it in his or her words, which would require the code *paraphrasing*. Again, emphasising the difference among these codes during observer training seems important. Finally, coding *evaluation of learners’ own actions* and *evaluation of team members’ actions* seems difficult as well. To correctly assign learners’ communication to these two codes, coders must remember who of the participants had been involved in the scenario, resulting in additional cognitive load during coding. Making respective notes prior to the debriefing might provide a remedy.

For video-based coding, coders also achieved good inter-rater reliabilities with the exception of five codes for debriefers’ communication (ie, *opinion, input simulation, laughing, structuring* and *psychological input*) and three codes for learners’ communication (ie, *evaluation of learners’ own actions, negative relevance* and *expressions of humour*). We have discussed the challenges involved in applying these codes for learners’ communication (ie, *evaluation of learners’ own actions and expressions of humour*) above. Regarding input simulation and psychological input, it seems difficult to distinguish these two codes. One explanation could be that all simulation scenarios were based on critical situation containing human factor aspects and psychological phenomena making it difficult for coders to distinguish which communication referred to the scenario design (*input simulation*) and which communication adds information on psychological research and phenomena (*psychological input*). We recommend briefing coders about the scenario design and learning objectives during observer training. In addition, it seemed challenging to reliably code *opinion* and *structuring*. *Structuring* contains debriefers’ communication about what they are now going to talk about, which is part of the debriefers point of view, making it difficult to distinguish *opinion* and *structuring*. Providing more precise coding instructions in the coding manual might provide a remedy.

Overall, the results of reliability testing indicate a promising reliability of DE-CODE, particularly given the coders’ high workload when applying a range of 47 different codes. The results show that live coding leads to similar reliabilities compared with video-based coding. Live coding is less time-consuming because the time needed for coding corresponds to the duration of the debriefing. In contrast, video-based coding is typically more time-consuming; using videos allows for watching sequences several times by going back and forth, which prolongs coding time. As it is reliable, we recommend coding debriefings on-site because it reduces other potential problems of video-taping (eg, costs for cameras in the debriefing room, legal and ethical issues related to filming and data storage).
With respect to validity, support for DE-CODE’s content validity is grounded in its meticulous design process involving extensive literature review, debriefing expert reviews and iterative tests. Of course, more evidence should be sought from its application to a broader range of debriefing settings. Further research is required to test DE-CODE’s predictive validity and determine in how far DE-CODE data will be able to discriminate debriefing styles and outcomes.

With respect to feasibility, we are aware that DE-CODE has an extensive number of codes. Compared with other debriefing assessment tools such as DASH and OSAD, which require rater training as well, it may seem resource-intense, daunting and effortful. DE-CODE’s intention is neither to substitute these tools nor to be feasible at the expense of its capability to assess debriefing dynamics. In fact, many of the most relevant findings on team communication required the use of extensive behaviour coding schemes, especially in healthcare and also in other high-risk settings. Debriefing is a team process that is by definition dynamic; the lack of research addressing this dynamic has been criticised repeatedly. We think it is important that DE-CODE allows for assessing team debriefings in a way that enables statistically analysing debriefings because recent group research has shown that patterns of behaviours among group members—rather than frequencies of individual group members’ actions—are what discriminate high-performing from low-performing groups. In line with similar approaches, we recommend to use DE-CODE flexibly in accordance with the respective research question. For example, if only selected debriefing behaviours are of interest, only they may be coded.

This study has limitations. DE-CODE’s reliability has so far been exclusively tested during anaesthesia SBT within a single institution that may limit the generalisability of the respective reliability findings. Broader reliability tests in other debriefing settings—even outside of SBT—are necessary to provide more robust reliability results. Selected DE-CODE codes did not yet reach sufficient reliability values (eg, normalisation and repeating). More double observations are required to obtain more data to explore and, ultimately, improve these values. We particularly encourage observations from a variety of observers to reduce observer bias. Similar to other behaviour coding schemes, the application of DE-CODE is limited to trained observers. Observer training (between 30 hours and 34 hours for the complete coding scheme with all codes) will be required prior to its use, especially for live coding. The required observer training will vary according to the desired use of DE-CODE: more time will be required if DE-CODE is used in full version for research purpose and less time will be required if only selected codes of DE-CODE are used for faculty development. It is necessary to consider in advance which device could be applied for coding (eg, tablet apps and software) and how feasible the respective data analysis would be.

DE-CODE has practical and research implications. It may be used to obtain insights into debriefing interaction patterns, for identifying effective debriefing methods and for comparing different debriefing approaches and methods in different work and cultural contexts (eg, interprofessional vs intraprofessional teams, surgical teams vs emergency teams, single debriefer vs code briefing setting and so on). Particularly, knowledge on debriefing interaction patterns is needed for understanding the process of how debriefer communication impacts learning. This feedback can help debriefers to develop and improve their competence in conducting debriefings. In that respect, data obtained with DE-CODE may be valuable for faculty development because it provides insights into what specifically debriefers can do and say during debriefings that helps learners to reflect on their mental models. This may also contribute to defining best practices for debriefings during SBT and further enhance their impact on learning, performance, and patient safety. In addition, the DE-CODE-based empirical evidence about what debriefers communicate are most effective to improve participants’ learning during debriefing conversations could be used to design effective debriefing tools for the clinical setting helping teams to debrief themselves.

Acknowledgements The authors would like to thank Hubert Heckel, Adrian Marty, Valentin Neuhauß, Niels Buse and Michael Hanusch for their help in collecting data, Lynn Häslar and Rebecca Hadley for their help in data coding and Alfonso Scherrer and Andrea Nef for their operational support.

Contributors JCS and MK were involved in the planning, conduct and reporting of this study and developed the coding scheme. JCS and MK prepared the study proposal and created the manuscript. JCS collected data and performed the statistics. BG served as a scientific advisor, helped with data collection and revised the final manuscript. SK collected data and revised the final manuscript.

Funding The research was supported by a grant from the Swiss National Science Foundation (Grant No. 100014_152822).

Competing interests None declared.

Ethics approval The study was approved by the local ethics committee (Kantonale Ethikkommission Zürich (KEK-ZH-No. 2013–0592)).

Provenance and peer review Not commissioned; externally peer reviewed.

© Article author(s) (or their employer(s) unless otherwise stated in the text of the article) 2018. All rights reserved. No commercial use is permitted unless otherwise expressly granted.

REFERENCES


