

Quality of interdisciplinary postsimulation debriefing: 360° evaluation

Louise Hull,¹ Stephanie Russ,² Maria Ahmed,³ Nick Sevdalis,¹ David J Birnbach⁴

¹Health Service and Population Research Department, Center for Implementation Science, King's College London, London, UK

²Institute for Education in Medical and Dental Sciences, University of Aberdeen, Aberdeen, UK

³Center for Primary Care, NIHR School for Primary Care Research, Manchester Academic Health Science Center, University of Manchester, Manchester, UK

⁴Miller School of Medicine, University of Miami, Miami, Florida, USA

Correspondence to

Dr Louise Hull, Center for Implementation Science, King's College London, London SE5 8AF, UK; louise.hull@kcl.ac.uk

Accepted 5 November 2016

ABSTRACT

Introduction Debriefing is widely perceived to be the most important component of simulation-based training. This study aimed to explore the value of 360° evaluation of debriefing by examining expert debriefing evaluators, debriefers and learners' perceptions of the quality of interdisciplinary debriefings.

Method This was a cross-sectional observational study. 41 teams, consisting of 278 learners, underwent simulation-based team training. Immediately following the postsimulation debriefing session, debriefers and learners rated the quality of debriefing using the validated Objective Structured Assessment of Debriefing (OSAD) framework. All debriefing sessions were video-recorded and subsequently rated by evaluators trained to proficiency in assessing debriefing quality.

Results Expert debriefing evaluators and debriefers' perceptions of debriefing quality differed significantly; debriefers perceived the quality of debriefing they provided more favourably than expert debriefing evaluators (40.98% of OSAD ratings provided by debriefers were $\geq +1$ point greater than expert debriefing evaluators' ratings). Further, learner perceptions of the quality of debriefing differed from both expert evaluators and debriefers' perceptions: weak agreement between learner and expert evaluators' perceptions was found on 2 of 8 OSAD elements (learner engagement and reflection); similarly weak agreement between learner and debriefer perceptions was found on just 1 OSAD element (application).

Conclusions Debriefers and learners' perceptions of debriefing quality differ significantly. Both groups tend to perceive the quality of debriefing far more favourably than external evaluators. An overconfident debriefer may fail to identify elements of debriefing that require improvement. Feedback provided by learners to debriefers may be of limited value in facilitating improvements. We recommend periodic external evaluation of debriefing quality.

INTRODUCTION

Debriefing strives to maximise learning, improve future performance and ultimately improve the safety and quality of patient care.¹ Debriefing, defined as a 'discussion between two or more individuals in which aspects of a performance are explored and analysed with the aim of gaining insight that impacts the quality of future clinical practice',² is widely considered a critical component of the learning process,³ with debriefer-facilitated debriefing considered the 'gold standard'.¹ Despite debriefing being well integrated into simulation-based training (SBT) and being considered as the most beneficial part of SBT,³ there is evidence that debriefing is not frequently provided

in the clinical environment,¹ and even when feedback is given trainees perceive it to be inadequate or ineffective.⁴

A recent systematic review of the effectiveness of simulation debriefing in health professionals' education found that post-training debriefing leads to significant improvements in performance in technical and non-technical skills (ie, teamwork, situational awareness, decision-making).¹ Furthermore, research has demonstrated that the positive effects of debriefing are retained several months following SBT and that SBT without debriefer-led debriefing offers little benefit to learners.⁵ To date, no such reviews (systematic or meta-analyses) have been conducted to explore the effectiveness of debriefing in the clinical environment.

The ability to conduct an effective debriefing is increasingly considered an essential skill for clinical educators and simulation instructors alike.⁶ Emphasis has been placed on training clinical educators and simulation instructors in the skill of debriefing. Considering the complexity and significant resources required to successfully implement SBT programmes (eg, investment in simulation facilities/technology, availability of debriefer to teach), and the challenges of conducting debriefings in the clinical environment,⁴ it appears logical to implement quality assurance measures to ensure that clinical educators and simulation instructors are equipped with the necessary skills to consistently provide learners with structured high-quality feedback, in clinical and simulation-based environments. This is an important development, particularly in light of previous research suggesting that performance debriefing is far from optimal in clinical settings.^{4,7}

Further, attempts have been made to define and clarify what constitutes an 'effective' debriefing.⁸ These have centred on the ability of a debriefer to maximise learning as a result of an educational intervention and to improve future clinical performance—that is, transfer of learning from the educational to the clinical setting.⁸ To this effect, numerous frameworks have been designed, for use in the clinical environment and/or in simulation, to structure and improve the quality of debriefings, including the Debriefing Assessment for Simulation in Healthcare (DASH);⁹ the Objective Structured Assessment of Debriefing (OSAD framework)¹⁰ and the Promoting Excellence and Reflective Learning in Simulation (PEARLS) debriefing script.¹¹

These frameworks have the potential to improve the quality of debriefing and in turn improve clinical performance and patient outcomes. It has been suggested that these frameworks can facilitate improvement in debriefing in a number of ways;



CrossMark

To cite: Hull L, Russ S, Ahmed M, et al. *BMJ Stel* 2017;**3**:9–16.

for example, inexperienced debriefers can use them as a guide to identify best practices to follow, and academics can use these frameworks to evaluate different models of debriefing and compare their relative quality and effectiveness.¹⁰ Moreover, debriefing frameworks have the potential to be used more broadly, that is, not only by debriefer providing debriefing and/or by expert debriefing evaluators for quality assurance but also by those being debriefed, providing a 360° evaluation of debriefing quality.

The 360° evaluation, often referred to as multisource feedback, is an assessment technique that focuses on gathering evaluations on an individual's performance from multiple perspectives, including co-workers, superiors and subordinates.¹² It is thought that negative or discrepant evaluations create awareness and motivate behaviour change.¹² In this sense, 360° evaluation could prove to be a valuable pedagogical strategy for use in clinical medical education as well as continuing professional development.

Exploring the value of 360° debriefing evaluation is important for two reasons: first it is unclear whether different groups of evaluators (learners, debriefers, expert debriefing evaluators) have different perceptions of what constitutes an effective debrief. Second, it has been suggested that widespread implementation of debriefing assessment requires further research to determine how best debriefing frameworks can be used in an efficient and effective manner.¹³ Assessing the quality of debriefing as perceived by learners, debriefers conducting the debriefing and expert (ie, trained) debriefing evaluators will shed light on the need for expert debriefing assessors for optimisation of performance and quality assurance, potentially providing alternative, less resource-intensive options.¹³

The aim of the current study was to address this gap in the evidence base. We explored the value of 360° evaluation of interdisciplinary debriefing by examining learners', debriefers' and expert evaluators' perceptions of the quality of debriefing.

METHODS

Full Institutional Review Board approval and informed consent from participants was obtained before study initiation.

Design and participants

This was an exploratory, cross-sectional observational study (hence without specific a priori hypotheses to be tested), conducted at three simulation centres which are all part of the University of Miami (UM-JMH Center for Patient Safety, Gordon Center for Research in Medical Education, and School of Nursing and Health Studies International Academy for Clinical Simulation and Research). During the study period, 278 learners (150 third year medical students and 128 second year nursing students) participated in 41 interdisciplinary simulation-based team training and debriefing sessions. The number of learners per team debrief ranged from 4 to 8 (2×4 learners, 3×5 learners, 9×6 learners, 15×7 learners, 12×8 learners).

Outcome measures

The quality of debriefing was assessed using the validated OSAD framework. OSAD contains eight core elements of a high-quality debrief. OSAD was selected for a number of reasons. First, OSAD is an evidence-based, end-user informed and psychometrically robust framework: OSAD has been found to be feasible, reliable (inter-rater and test-retest reliability) and valid (face and content).⁹ Second, a number of members of the research team were part of the OSAD development team (MA, LH and NS) and the tool has subsequently been adopted by the research team. Third, two members of the research team (MA

and SR) had been trained to proficiency in using OSAD. OSAD elements and their descriptions are provided in [figure 1](#).

Each debriefing element is rated on a five-point scale ranging from 1 (done very poorly) to 5 (done very well). Descriptive anchors at the lowest point, mid-point and highest point of the scale are provided to guide ratings. The global score for OSAD ranges from a minimum of 8 to a maximum of 40, with higher scores indicating more effective debriefings.

Procedure

All learners participated in a 2-day training programme consisting of lectures, web-based didactic materials and small group activities, followed by two 8 min simulation-based team training scenarios. The first scenario featured a middle-aged hypertensive postsurgical patient with desaturation and chest pain, and the second featured a middle-aged patient with diabetes and chronic obstructive pulmonary disease, new onset desaturation and haemodynamic instability. A 'confederate' nurse was present during all simulations. All scenarios were observed in real time by the debriefers through a one-way mirror. Learners were provided with a very brief (and deliberately inadequate) handover which failed to transfer essential information concerning medication, medical history or the name of the person to call if help was needed. The scenarios were designed to reinforce non-technical skills (communication, situational awareness, teamwork, leadership and calling for help) and clinical skills (development of differential diagnoses, evaluation of patients with worsening conditions). Some simulations involved standardised patients (actors) and some used a mannequin. The key features of this course have been previously reported.^{14 15} Team debriefings occurred within 5 min of the simulation being completed. All debriefing sessions were ~20 min in duration and were video-recorded using Apple iPads, mounted to a wall ensuring a clear view of all learners and the debriefers. Fourteen of the 41 team debriefs were conducted by one debriefer. The remaining debriefing sessions were co-debriefed (conducted by two or more debriefers—both a nursing and physician debriefer). If two debriefers were present, one assumed the role of primary debriefer. In total, debriefing sessions were facilitated by 14 that had all received formal debriefing training and had previously performed at least 50–100 hours of debriefings.

Immediately following the debriefing session, learners and debriefers independently assessed the quality of debriefing using the OSAD tool (the ratings provided were for the quality of the debrief overall, not the individual debriefer). Learners and debriefers were provided with the OSAD booklet, which describes the debriefing framework in detail (available for free download at cpsy.org), at the beginning of the 2-day course and were strongly encouraged to review it prior to the start of the simulation activities. It was also reviewed in depth prior to its use by students during a lecture to the entire group. Learners, debriefers and expert debriefing evaluators used the same version of OSAD to rate the quality of debriefing.

Two OSAD evaluators (SR and MA), who had previously been trained to proficiency in applying the framework and had established inter-rater reliability with each other in using OSAD to assess clinical team debriefs,¹⁰ also assessed the quality of the debriefing sessions from the video recordings. Unlike debriefers, expert debriefing evaluators had not received formal training in providing feedback within SBT sessions (although they did have experience of providing feedback/debriefing in clinical and medical education scenarios). Expert debriefing evaluators, unlike debriefers, had received extensive training in using OSAD and were trained to proficiency in using the framework.

Element	Element description
Approach	Manner in which the facilitator conducts the debriefing session, their level of enthusiasm and positivity when appropriate, showing interest in the learners by establishing and maintaining rapport and finishing the session on an upbeat note.
Learning environment	Introduction of the simulation/learning session to the learner(s) by clarifying what is expected of them during the debriefing, emphasising ground rules of confidentiality and respect for others, and encouraging the learners to identify their own learning objectives.
Learner engagement	Active involvement of all learners in the debriefing discussions, by asking open questions to explore their thinking and using silence to encourage their input, without the facilitator talking for most of the debriefing, to ensure that deep rather than surface learning occurs.
Reaction	Establishing how the simulation/learning session impacted emotionally on the learners.
Reflection	Self-reflection of events that occurred in the simulation/learning session in a step by step factual manner, clarifying any technical clinical issues at the start, to allow ongoing reflection from all learners throughout the analysis and application phases, linking to previous experiences.
Analysis	Eliciting the thought processes that drove a learner's actions, using specific examples of observable behaviours, to allow the learner to make sense of the simulation/learning session events.
Diagnosis	Enabling the learner to identify their performance gaps and strategies for improvement, targeting only behaviours that can be changed, and thus providing structured and objective feedback on the simulation/learning session.
Application	Summary of the learning points and strategies for improvement that have been identified by the learner(s) during the debrief and how these could be applied to change their future clinical practice.

Figure 1 The Objective Structured Assessment of Debriefing (OSAD) framework.

The sessions were split evenly between the two experts; five of the sessions (12% of total) were assessed by both experts and submitted to further inter-rater reliability analysis to quality assure these data.

Statistical analyses

All analyses were conducted using SPSS V.22.0 (SPSS, Chicago, Illinois, USA). We computed descriptive statistics (means and SDs) for each OSAD element (approach, learning environment, engagement of learners, reaction, reflection, analysis, diagnosis and application) as evaluated by the three groups (ie, expert evaluators, debriefers and learners). As each learner in the debriefing sessions completed OSAD, we calculated the average rating for each element. We calculated Global OSAD scores by summing the eight element ratings. We assessed inter-rater reliability between expert debriefing evaluators using intraclass correlation coefficients (ICC; absolute agreement/single measures type). To assess whether discrepancies exist between expert evaluators, debriefers and learners' perceptions of the quality of debriefing we calculated ICCs between their scores. In addition, we computed the absolute difference between expert debriefing evaluators and debriefers' perceptions of the quality of debriefing, and expressed this using descriptive statistics (frequency counts and percentages).

RESULTS

Expert evaluation of debriefing quality

Inter-rater reliability

Inter-rater reliability between the two expert debriefing evaluators was perfect across all OSAD elements (ICCs 1.00, absolute agreement).

Overall quality of debriefing: expert debriefing evaluators, debriefers and learners

Figure 2 displays the overall quality of debriefing sessions as perceived by expert debriefing evaluators, debriefers and learners. The range of Global OSAD scores provided by expert debriefing evaluators, debriefers and learners ranged from 29 to 38, 25 to 40 and 30 to 40, respectively.

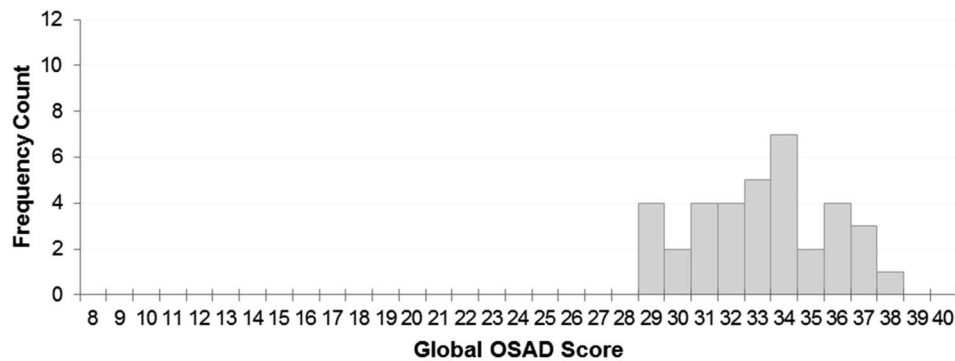
Quality of debriefing elements as perceived by expert debriefing evaluators, debriefers and learners

Figure 3 displays the quality of debriefing across the eight OSAD elements (approach, learning environment, learner engagement, reaction, reflection, analysis, diagnosis and application) as evaluated by experts, debriefers and learners. Six of the eight OSAD elements were rated by expert debriefing evaluators above the OSAD midpoint (3), indicating that the quality of the majority of debriefing elements was typically 'done well'. In comparison, debriefers and learners rated all OSAD elements above the OSAD mid-point.

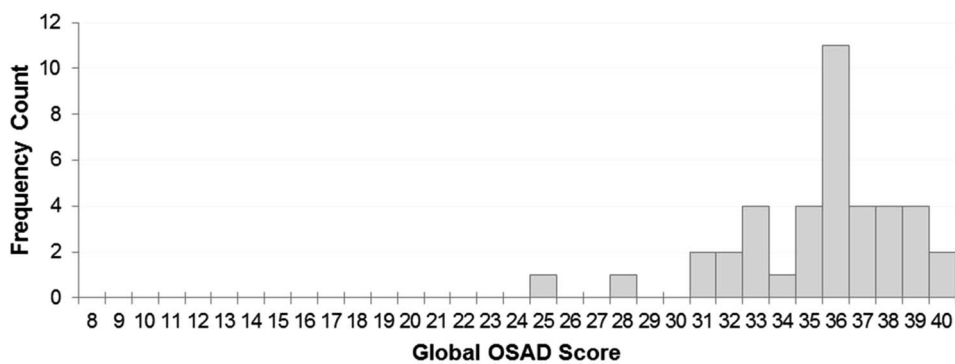
ICCs between expert debriefing evaluators, debriefers and learners' OSAD evaluations across the eight OSAD elements are displayed in table 1. All ICCs between expert and debriefer ratings were low and non-significant indicating an overall lack of agreement. Except for reflection where weak agreement was found. Similarly, all ICCs between expert and learner evaluation indicated a lack of agreement, except for learner engagement and reflection where weak agreement between evaluations was found.

The previous ICC analysis revealed a lack of agreement in OSAD evaluations between the expert debriefing evaluators and

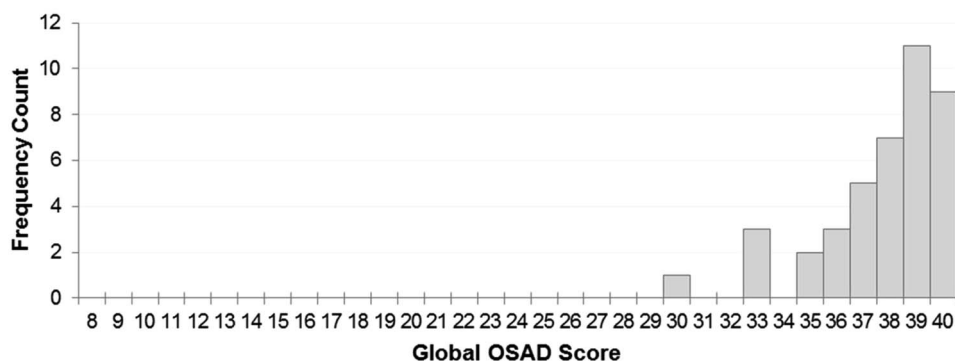
Expert Debriefing Evaluator



Debriefeer



Learner



Note: Learner Global OSAD scores were calculated by summing the 8 mean element ratings provided by the group of learners and rounded up or down to the nearest whole number.

Figure 2 Distribution of the quality of debriefing as perceived by expert debriefing evaluators, debriefers and learners. OSAD, Objective Structured Assessment of Debriefing.

the debriefers. To further examine the lack of agreement, we calculated frequency counts, along with the corresponding percentages, of the absolute difference between expert debriefing evaluators and debriefers evaluations. The absolute difference in debriefers versus expert evaluators' scoring of debriefing quality was grouped into three agreement/accuracy categories:

- ▶ *Matching perceptions*: no difference between debriefers and expert debriefing evaluators' ratings;
- ▶ *Perceptions differ slightly*: debriefers' ratings ± 1 point apart from expert debriefing evaluators' ratings;
- ▶ *Perceptions differ remarkably*: debriefers' ratings ± 2 points or more apart from expert debriefing evaluators' ratings.

For example, if the expert debriefing evaluator rated an element 1 (done very poorly) and the debriefer rated 5 (done very well) this would yield an absolute difference of -4 , and

therefore would be categorised as 'perceptions differ remarkably' as the debriefer's rating is more than 2 points apart from the expert evaluator's rating.

In our analysis, 128/327 (39.14%) of debriefers' and expert debriefing evaluators' perceptions of debriefing quality matched across the eight OSAD elements. In total, 132/327 (40.37%) of debriefers' and expert debriefing evaluators' perceptions of debriefing quality differed slightly, and 67/327 (40.49%) of debriefers' and experts debriefing evaluators' perception of debriefing quality differed remarkably, across the eight OSAD elements. Table 2 displays a tendency for debriefers to perceive the quality of debriefing more favourably; 134 (40.98%) of debriefers' ratings were higher than those provided by the expert debriefing evaluators. In contrast, only 65 of these ratings (19.88%) were lower than those provided by

Figure 3 Quality of debriefing as perceived by expert debriefing evaluators, debriefers and learners across OSAD elements. OSAD, Objective Structured Assessment of Debriefing.

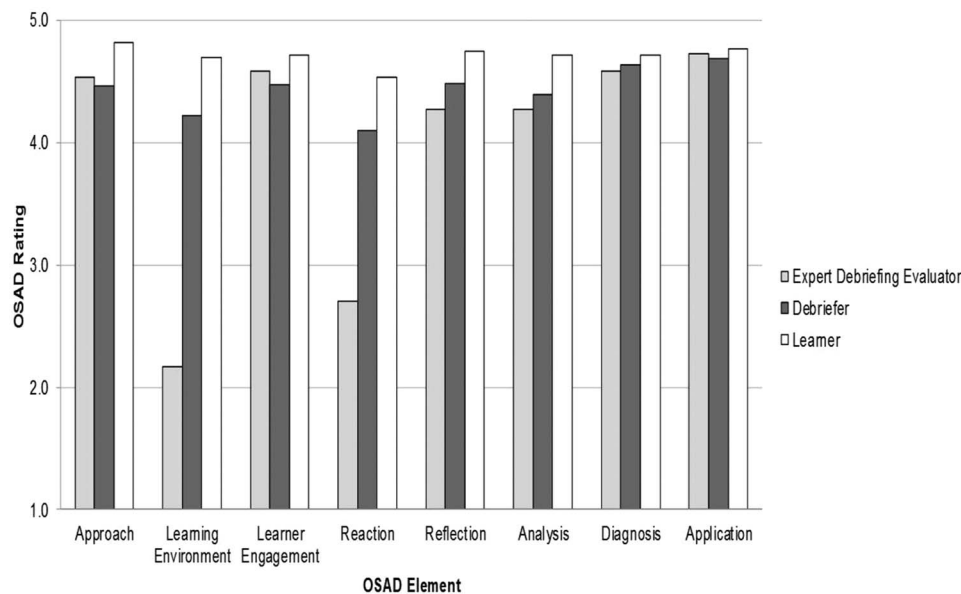


Table 1 ICCs between expert debriefing evaluators, debriefers and learners evaluations across OSAD elements

OSAD element	Learners and debriefers	Learners and expert debriefing evaluators	Debriefers and debriefing expert evaluators
Approach	0.15 ^{NS}	0.28*	0.12 ^{NS}
Learning environment	0.20*	0.01 ^{NS}	0.01 ^{NS}
Learner engagement	0.23 ^{NS}	0.45**	0.22 ^{NS}
Reaction	0.13 ^{NS}	0.03 ^{NS}	-0.10 ^{NS}
Reflection	0.14 ^{NS}	0.43**	0.35**
Analysis	0.14 ^{NS}	0.21*	0.14 ^{NS}
Diagnosis	0.11 ^{NS}	0.26*	-0.02 ^{NS}
Application	0.40**	0.18 ^{NS}	0.16 ^{NS}

*p<0.05; **p<0.01.

ICC interpretation: lack of agreement 0.00–0.30; weak agreement 0.31–0.50; moderate agreement 0.51–0.70; strong agreement 0.71–0.90; very strong agreement 0.91–1.00 (LeBreton and Senter).¹⁶ ICC, intraclass correlation; NS, not statistically significant; OSAD, Objective Structured Assessment of Debriefing.

the expert debriefing evaluators. Looking in more detail, debriefers and expert debriefing evaluators' perceptions of the quality of debriefing differed across the debriefing elements assessed. For example, whereas 'learning environment' tended to be perceived far more favourably by the debriefers compared with the experts debriefing evaluators (97.6% of debriefers ratings ≥ 1 point from expert debriefing evaluators rating), 'application' tended to be similarly viewed (61.0% of ratings no difference between debriefers and expert debriefing evaluators)—if rather underscored by the debriefers compared with how it is viewed by the experts debriefing evaluators (22.0% of expert debriefing evaluators rating ≥ 1 point from debriefers rating).

DISCUSSION

The purpose of this study was to explore the value of 360° evaluation of debriefing by examining expert debriefing evaluators, debriefers and learners' perceptions of the quality of

interdisciplinary debriefings. Our results indicate that while there was some agreement, expert evaluators and faculty debriefers' perceptions of the quality of debriefing differ significantly. Debriefers tended to perceive the quality of their debriefings more favourably (40.98% of evaluations) rather than less favourably (19.88% of evaluations) compared with expert debriefing evaluators. Similarly, learners tended to perceive the quality of debriefing they receive much more favourably than expert debriefing evaluators. These findings raise concerns regarding the value of 360° evaluation of debriefing, in the simulated environment and also in clinical practice, in facilitating improvements in the quality of debriefing and suggest that the level of insight into the debriefing process as expressed by those participating in it and compared with external observers requires further exploration.

Deconstructing the Global OSAD scores to examine specific elements of debriefing, we identified two areas of debriefing in particular where debriefers tended to perceive their performance more positively. These were related to the 'learning environment' within the debriefing session and the analysis of the emotional 'reaction' of the learners to the training event. Establishing a positive learning environment, including clarifying the objectives and learners expectations, is considered critical to creating and structuring an educationally conducive environment.^{17–18} Furthermore, exploring the learners' emotional reaction to the learning experience is considered important to long-lasting learning.¹⁸ The expert debriefing evaluators noted that typically within the debriefing sessions, the debriefer rarely explicitly established the expected learning outcomes for the group as a whole, or for individual learners. Furthermore, emotional reactions to the sessions were not fully acknowledged or explored. This may highlight particular challenges when debriefing groups as opposed to individuals, and may have impacted on the ability of learners to structure their learning and to explore and rationalise negative experiences in the simulation setting. This justified the lower scores allocated by the expert debriefing evaluators for these elements. The debriefers, however, consistently perceived themselves as performing well in these areas. There are a number of potential reasons for these discrepancies. Debriefers may not have sufficiently familiarised themselves with the OSAD user guide, which stipulates key

Table 2 Comparison of expert debriefing evaluators and debriefers' perceptions of debriefing quality

OSAD element	Perceptions differ remarkably Expert debriefing evaluators rating ≥ -2 points from debriefers rating Frequency count (%)	Perceptions differ slightly Expert debriefing evaluators rating -1 point from debriefers rating Frequency count (%)	Matching perceptions No difference between expert debriefing evaluators and debriefers rating Frequency count (%)	Perceptions differ slightly Expert debriefing evaluators rating $+1$ point from debriefers rating Frequency count (%)	Perceptions differ remarkably Expert debriefing evaluators rating $\geq +2$ points from debriefers rating Frequency count (%)
Approach ⁽ⁿ⁼⁴¹⁾	0 (0)	10 (24.4)	20 (48.8)	9 (22.0)	2 (4.9)
Learning environment ⁽ⁿ⁼⁴¹⁾	31 (75.6)	9 (22.0)	1 (2.4)	0 (0)	0 (0)
Engagement of learners ⁽ⁿ⁼⁴⁰⁾	2 (5.0)	5 (12.5)	22 (55.0)	9 (22.5)	2 (5.0)
Reaction ⁽ⁿ⁼⁴¹⁾	24 (58.5)	10 (24.4)	5 (12.2)	2 (4.9)	0 (0)
Reflection ⁽ⁿ⁼⁴¹⁾	0 (0)	7 (17.1)	23 (56.1)	8 (19.5)	3 (7.3)
Analysis ⁽ⁿ⁼⁴¹⁾	1 (2.4)	16 (39.0)	12 (29.3)	11 (26.8)	1 (2.4)
Diagnosis ⁽ⁿ⁼⁴¹⁾	0 (0)	12 (29.3)	20 (48.8)	8 (19.5)	1 (2.4)
Application ⁽ⁿ⁼⁴¹⁾	0 (0)	7 (17.1)	25 (61.0)	9 (22.0)	0 (0)
Total 327	58 (17.74)	76 (23.24)	128 (39.14)	56 (17.13)	9 (2.75)
	Debriefers perceive the quality of debriefing more favourably than expert debriefing evaluators 134 (40.98)		Debriefers and expert debriefing evaluators perceptions of the quality of debriefing match 128 (39.14)	Debriefers perceive the quality of debriefing less favourably than expert debriefing evaluators 65 (19.88)	

OSAD, Objective Structured Assessment of Debriefing.

features (together with exemplar statements) for each domain to produce low, medium or high scores. Furthermore, it may be that debriefers' perceptions were influenced by the lecture which occurred prior or subsequent to the debrief itself (which was not recorded and therefore not accessible to the expert assessors). This, however, is unlikely as all debriefs took place immediately (within 5 min) following the simulation. Finally, more favourable evaluations may simply reflect that debriefers perceived these elements to have been covered well as part of the debriefing, whereas they may in fact have been performed implicitly, in-part, or not at all. Identification of such areas of discrepancy in perceptions can help to identify elements that should be a focus for training in debriefing.

Of interest, we found that the learners in the present study perceived the quality of the debriefing sessions more positively than both debriefers and expert debriefing evaluators. The precise reason(s) for such findings are unclear, but do support previous research findings that debriefing is considered highly desirable¹⁹ and in addition to providing learners with the opportunity to critically reflect on their performance/simulation experience, debriefing can provide emotional and social support.²⁰ The overall lack of variation in learners' scores further replicates previous findings we have obtained in previous studies within a perioperative setting when learners are asked to evaluate learning.⁷ This could potentially be attributed to learners' lack of a multidimensional view on their learning experience (ie, they often do not have adequate range of more and less effective experiences to judge against); this remains a question for further study (see also below).

Our findings also suggest that debriefing elements perceived not to have been completed very well by expert debriefing evaluators (ie, learning environment and reaction) were not deemed to be problematic from the perspective of the learner. The fact that debriefing is so highly regarded by trainees but occurs infrequently in clinical practice¹⁹ may explain why learners viewed their detailed debriefings so positively: to them this was a rather rare occurrence, which was very well received and personally valuable. This is one explanation, however the fact

that learners perceived the quality of all OSAD elements highly may indicate a 'halo' effect in which good (or poor, although not in our case) performance in one area (ie, one elements of debriefing) affects the assessor's judgement in other areas (ie, other debriefing elements) and/or 'leniency' effect, where assessors (ie, learners), for a range of reasons, rate performance far too positively. Another possible explanation is that learners were not able to distinguish between the OSAD elements—that is, OSAD might not be a sensitive measure in assessing learners' evaluations of the quality of debriefing elements. We suggest that future research explores the cognitive biases in ratings (eg, halo and leniency effects, as well as the sensitivity of OSAD when used by learners).

Regardless of the possible explanations, the fact that learners perceive the quality of debriefing so positively has important implications; feedback provided by learners to debriefers regarding the quality of their debriefing may not be enough to facilitate debriefing improvements—as long as the learners' benchmark within clinical training/practice remains low. We thus recommend periodic external evaluation of debriefing quality, driven by expert-trained faculty, to drive educational excellence.

Limitations

Our study has several limitations. First, our findings that debriefers tend to view their debriefing more positively than expert debriefing evaluators and learners represents a general tendency (ie, the results are aggregated over all debriefers). Owing to a lack of power we were unable to examine whether there were individual differences in the discrepancies between debriefers' perceptions of the quality of their debriefing and those of expert debriefing evaluators or learners. This may have been the case. Second, we were unable to control for the potential effect of team size of learners being debriefed on the quality of briefing. Scientifically, this comparison was outside the scope of the study (hence we did not have adequate power to perform it) and practically any attempt to control learners' team size would have rendered the study unfeasible within our clinical setting. These issues should be explored in future research. Third, it is

possible that learners in our study felt obliged to score the debriefing sessions positively—as a result of a ‘social desirability bias’. This may have influenced the debriefing scores that learners assigned to debriefers and may explain the discrepancies between expert evaluators and learners’ perceptions. Fourth, debriefers and learners may have familiarised themselves to varying degrees with the OSAD framework and this may have affected their assessments of debriefing quality. Fifth, due to the exploratory nature of this study, we did not collect demographic data of the learners (eg, age, gender, ethnicity), thus extrapolation of findings beyond the learners in this study (ie, medical and nursing students) should be treated with caution. Finally, statistically the study may have underestimated the size of the analysed relationships due to low variability in learners’ scores.

Taking the expert debriefing evaluators’ ratings of the quality of debriefing as an external, unbiased quality of debriefing, the findings of our study have several implications. First, our findings emphasise the need for expert debriefing assessors in the quality assurance process; the finding that learners evaluated the quality of debriefing more positively than the expert debriefing evaluators suggest that learners may not offer a suitable benchmark for optimal debriefing (for reasons we have discussed earlier). Second, our findings have implications regarding the identification of debriefing training needs. Debriefers that overestimate their ability to provide a high-quality debrief might miss opportunities to identify their limitations and may, in such circumstances, restrict the learning opportunities for learners. Although there is increasing evidence that postsimulation debriefing leads to significant improvements in technical and non-technical skills, it seems logical that a poor quality debrief is unlikely to yield the same degree of performance improvement. Third, an inability to identify deficiencies in debriefing is likely to hinder mastering the skills of debriefing; debriefers that are overconfident in their debriefing skills are unlikely to seek debriefing training to enhance the quality of their debriefing. This is concerning as previous research suggests that not only do physicians have a limited ability to accurately self-assess, the worst accuracy in self-assessment appears to be among physicians who are the least skilled and those who are the most confident.²¹

Based on our findings, we suggest a number of recommendations for faculty development. Practically, the large degree of variability in Global OSAD scores across the debriefing sessions highlights the potential for debriefers to learn from one another through coupling ‘low-performing’ and ‘high-performing’ debriefers so that low-performing debriefers have the opportunity to observe a high-quality debriefing and receive feedback on their debriefing skills. In addition, we suggest periodic external evaluation of debriefing quality, as discrepant or negative evaluations are likely to create awareness of debriefing elements that require improvement and motivate behaviour change¹² thus driving debriefing excellence. Trained faculty rotating between geographically neighbouring institutions may be a practical way to achieve this in practice; alternatively, remote video-based analysis may offer a less costly option, with the potential added benefit of effectively blinding evaluators to debriefers. Furthermore, all of the debriefers in our study had received formal training in debriefing and were experienced debriefers, having previously performed at least 50–100 hours of debriefings. This suggests that ‘one-off’ training in debriefing skills might not be adequate for detailed assessment and that refresher training might be useful to improve the quality of debriefing.

Finally, although our study explored the value of 360° evaluation in the context of simulation-based debriefing, this study

has implications for the potential value of 360° evaluation as a pedagogical strategy for use in clinical medical education as well as continuing professional development. For example, 360° evaluation could be used as a strategy to trigger, stimulate and encourage reflective practice; this might be useful considering that previous research has found that the tendency and ability to reflect appears to vary across individuals.²²

CONCLUSIONS

Debriefers tend to perceive the quality of debriefing they provide more favourably in comparison to expert debriefing evaluators. Overconfidence in debriefing skills may prevent debriefers from actively seeking debriefing training opportunities. Learners also tend to perceive the quality of the debriefing they receive more positively than expert debriefing evaluators, we thus recommend periodic external evaluation of debriefing quality (through in situ or video-based modalities) to drive educational excellence.

Contributors LH made substantial contributions to the conception and design of the work, the analysis and interpretation of data, drafting the work, final approval of the version published and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. SR and MA made substantial contributions to the analysis and interpretation of data, revising the draft work critically for important intellectual content, final approval of the version published, and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. NS and DJB made substantial contributions to the conception and design of the work, the analysis and interpretation of data, revising the draft work critically for important intellectual content, final approval of the version published and agrees to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Funding LH and NS’s research was supported by the National Institute for Health Research (NIHR) Collaboration for Leadership in Applied Health Research and Care South London at King’s College Hospital NHS Foundation Trust. LH and NS are members of King’s Improvement Science, which is part of the NIHR CLAHRC South London and comprises a specialist team of improvement scientists and senior researchers based at King’s College London. Its work is funded by King’s Health Partners (Guy’s and St Thomas’ NHS Foundation Trust, King’s College Hospital NHS Foundation Trust, King’s College London and South London and Maudsley NHS Foundation Trust), Guy’s and St Thomas’ Charity, the Maudsley Charity and the Health Foundation. MA is a NIHR Academic Clinical Fellow in Primary Care, Education Associate for the General Medical Council and a Trustee of the Clinical Human Factors Group.

Disclaimer The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Competing interests NS is the Director of London Safety & Training Solutions, which provides team skills training and advice on a consultancy basis in hospitals and training programmes in the UK and internationally.

Ethics approval Full Institutional Review Board approval obtained from University of Miami.

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

- 1 Levett-Jones T, Lapkin S. A systematic review of the effectiveness of simulation debriefing in health professional education. *Nurse Educ Today* 2014;34:e58–63.
- 2 Cheng A, Eppich W, Grant V, et al. Debriefing for technology-enhanced simulation: a systematic review and meta-analysis. *Med Educ* 2014;48:657–66.
- 3 Rall M, Manser T, Howard S. Key elements of debriefing for simulator training. *Eur J Anaesthesiol* 2000;17:516–17.
- 4 Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Med Teach* 2012;34:787–91.
- 5 Savoldelli GL, Naik VN, Park J, et al. Value of debriefing during simulated crisis management: oral versus video-assisted oral feedback. *Anesthesiology* 2006;105:279–85.

- 6 Paige JT, Arora S, Fernandez G, *et al.* Debriefing 101: training faculty to promote learning in simulation-based training. *Am J Surg* 2015;209:126–31.
- 7 Ahmed M, Sevdalis N, Vincent C, *et al.* Actual vs perceived performance debriefing in surgery: practice far from perfect. *Am J Surg* 2013;205:434–40.
- 8 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.
- 9 Brett-Fleegler M, Rudolph J, Eppich W, *et al.* Debriefing assessment for simulation in healthcare: development and psychometric properties. *Simul Healthc* 2012;7:288–94.
- 10 Arora S, Ahmed M, Paige J, *et al.* Objective structured assessment of debriefing: bringing science to the art of debriefing in surgery. *Ann Surg* 2012;256:982–8.
- 11 Eppich W, Cheng A. Promoting Excellence and Reflective Learning in Simulation (PEARLS): development and rationale for a blended approach to health care simulation debriefing. *Simul Healthc* 2015;10:106–15.
- 12 Brett JF, Atwater LE. 360 degree feedback: accuracy, reactions, and perceptions of usefulness. *J Appl Psychol* 2001;86:930–42.
- 13 Cheng A, Grant V, Dieckmann P, *et al.* Faculty development for simulation programs: five issues for the future of debriefing training. *Simul Healthc* 2015;10:217–22.
- 14 Arora S, Hull L, Fitzpatrick M, *et al.* Crisis management on surgical wards: a simulation-based approach to enhancing technical, teamwork, and patient interaction skills. *Ann Surg* 2015;261:888–93.
- 15 Shekhter I, Rosen L, Sanko J, *et al.* A patient safety course for preclinical medical students. *Clin Teach* 2012;9:376–81.
- 16 LeBreton JM, Senter JL. Answers to 20 questions about interrater reliability and interrater agreement. *Organ Res Methods* 2008;11:815–52.
- 17 Rudolph JW, Simon R, Rivard P, *et al.* Debriefing with good judgment: combining rigorous feedback with genuine inquiry. *Anesthesiol Clin* 2007;25:361–76.
- 18 Fanning RM, Gaba DM. The role of debriefing in simulation-based learning. *Simul Healthc* 2007;2:115–25.
- 19 Mullan PC, Kessler DO, Cheng A. Educational opportunities with postevent debriefing. *JAMA* 2014;312:2333–4.
- 20 Gunasingam N, Burns K, Edwards J, *et al.* Reducing stress and burnout in junior doctors: the impact of debriefing sessions. *Postgrad Med J* 2015;91:182–7.
- 21 Davis DA, Mazmanian PE, Fordis M, *et al.* Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;296:1094–102.
- 22 Mann K, Gordon J, MacLeod A. Reflection and reflective practice in health professions education: a systematic review. *Adv Health Sci Educ Theory Pract* 2009;14:595–621.