BMJ Open

The feasibility of A-E, TEAM and SAGAT for measurement of situational awareness and team performance in critical care simulation training

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029412
Article Type:	Research
Date Submitted by the Author:	25-Feb-2019
Complete List of Authors:	Hultin, Magnus; Umeå Universitet Medicinska fakulteten, Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine Jonsson, Karin; Umeå Universitet Medicinska fakulteten, Departments of Surgical and Perioperative Sciences and Nursing Härgestam, Maria; Umea Universitet Medicinska fakulteten, Department of Nursing Lindkvist, Marie; Umeå Universitet Medicinska fakulteten, Department of Public Health and Clinical Medicine Brulin, Christine; Umeå Universitet Medicinska fakulteten, Department of Nursing
Keywords:	Teamwork, Situation awareness, Taskwork

The feasibility of A-E, TEAM and SAGAT for measurement of situational awareness and team performance in critical care simulation training

Magnus Hultin, MD PhD, Associate Professor, Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden.
https://orcid.org/0000-0003-2935-7161

Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden
https://orcid.org/0000-0001-7974-6777

Maria Härgestam, RNA PhD, Lecturer, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden

Marie Lindkvist, PhD, Associate Professor, Department of Public Health and Clinical Medicine, Umeå University, S-901 87 Umeå, Sweden

Christine Brulin, RNA PhD, Professor, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden

Abstract 300 words

Main article 4364 words

Keywords
SAGAT, TEAM, resuscitation

Corresponding author
Magnus Hultin, Dept of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, Sweden, magnus.hultin@umu.se, phone: +46-90-786 9109

Abstract

Objectives: To test (1) an A–E checklist based on the Trauma Team Evaluation Tool for interrater reliability, (2) the TEAM instrument for interrater reliability, (3) a Swedish version of the SAGAT instrument for feasibility, and (4) internal consistency.

Design: Cross sectional.

Setting: Two full-scale simulation scenarios designed to train medical students in structured initial resuscitation.

Participants: Fifty-five medical students aged 22–40 years during year 4 clerkship in anaesthesiology and critical care medicine formed 23 different teams. All students answered the SAGAT instrument and, of those, 24 students answered the follow up post simulation questionnaire (PSQ). The TEAM instrument and A–E checklist were scored by four professionals.

Primary and secondary outcome measures: A modified A-E checklist adapted to the scenario and the TEAM instrument used in its original form were scored by independent raters using the video recorded scenarios and thereafter tested for interrater reliability using intraclass correlation (ICC). A Swedish version of the SAGAT instrument with 22 questions was translated and adapted in order to correspond to the scenarios. The feasibility of the SAGAT instrument was tested using the PSQ. SAGAT was tested for internal consistency (Cronbach's alpha and normed χ^2) both at individual level (SAGAT) as well as at team level (TSAGAT).

Results: The ICC (single / average measurements) was 0.54 / 0.83 for the TEAM scale and 0.55 / 0.83 for the A-E scale. The questions in the SAGAT instrument were rated as relevant and related to the scenario in the PSQ by 96% of the participants. Cronbach's alpha for SAGAT/TSAGAT for the two scenarios was 0.80/0.83 vs 0.62/0.76, and normed χ^2 was 1.72 vs 1.62.

Conclusions: The modified A-E checklist, the TEAM instruments as well as SAGAT, both at an individual and at a team level, might purposefully be used in a Swedish context.

Article Summary

Strengths and limitations of this study

- Iterative process used for development of SAGAT questions
- Interrater variability based upon video analyses by raters with different backgrounds
- The participants in the scenarios were homogeneous with some experience as all were medical students
- The TEAM instruments subscales were analysed
- Difficult to test content validity on SAGAT as the questions are case-specific

Funding

This work was supported by Västerbotten County Council, grant no VLL-663801 and VLL-836931, by Norrbotten County Council, grant no NLL-765981, by Alice Lindström's Foundation, and by the Medical Faculty of Umeå University.

Competing interests

None of the authors have any competing interest to declare.

Data availability

The data consisting of deidentified questionnaire responses and ratings of video scenarios as SPSS data files and SPSS syntax files are available upon request from the corresponding author.

Funding

This study was supported by grants from the Medical Faculty of Umeå University and the County Council of Västerbotten.

Author Contributions

MHu, KJ, MHä and CB designed the study. MHu and KJ prepared the initial draft of the paper. All authors were actively engaged in data analysis and in finalizing the paper. All authors have read and approved the final version of the paper.

Introduction

Knowledge about the relation between human errors and patient safety has increased in the last two decades.[1–3] Simulated environments make it possible to improve skills by employing training strategies to prevent errors while simultaneously having an arena for reliable assessments of competencies.[4] Therefore, simulation training is often used by organizations to minimize adverse events and to prevent healthcare errors.[5] This can be accomplished by improving *task performance*, *team performance* and/or *situation awareness*.[6–8]

Optimal task as well as team performance depends on the coordinated activities of a team of individuals.[9,10] In research, it is essential to evaluate task and team performance in order to develop strategies for improvements in clinical practice. To assess task performance in acute care scenarios, checklists are often used to score the adherence to resuscitation protocols and the timing of the task.[11,12] The Trauma Team Evaluation Tool (TTET) was developed by Holcomb for trauma scenarios handled according to protocols based on Advanced Trauma Life Support (ATLS) and validated on US military resuscitation teams from community hospitals.[11] No psychometric data as interrater variability was presented in the original study. In order to be a trustworthy tool for evaluation in other settings, a case specific A–E checklist needs to be developed based upon TTET and tested. The items included and criteria used need to be adapted to the proficiency levels of the participants and the standard operating procedures currently in use.

In addition to measurement of task performance, assessment of team performance is also of importance in order to increase patient safety. Therefore, the TEAM instrument was developed measuring three dimensions of team performance: leadership, teamwork, and task management.[13] The instrument was developed and validated by a team of resuscitation experts from the United Kingdom, Australia and New Zealand for evaluation of simulator-based team-trainings.[13] Recently, the TEAM instrument was also validated for collection of observational ratings of non-technical skills during live resuscitations based on data from two Australian metropolitan emergency departments.[14] The TEAM instrument has not yet been validated for use outside the commonwealth nor has the instrument been tested at a subscale level.

Moreover, and in addition to task and team performance, situation awareness (SA) is a prerequisite for patient safety and the prevention of errors, especially during acute care situations. SA include three levels of ability; (1) perception and attention (*What*, what is going on), (2) comprehension (*So what*, the ability to understand what is going on), and (3) projection (*Now what*, to anticipate and plan for future events).[15] Loss of SA increases the risk of failure in perception, comprehension and/or in the ability to make projections about the situation into the future.[16,17] In order to measure SA in a simulation setting the Situation Awareness Global Assessment Tool (SAGAT) has been developed and later adapted for use in healthcare settings.[7,18,19] One feature of SAGAT is that its use requires the simulation to pause, which might influence the clinical understanding, both by impeding the suspension of disbelief in the simulation setting,[18] and by introducing reflection-on-action.[19] Gardner showed that it was feasible to use SAGAT to measure SA in team training of surgical trainees in advanced cardiac life support.[20] Internationally, SAGAT has been used to study, for example, the effect of sleep deprivation on SA in trauma team training, how SA is associated with surgical trainee team performance, and nurses' clinical judgement of patient deterioration.[20–23] SAGAT can be analysed both at an individual level and at team level by taking the average of the SAGAT scores in the team. A specific application of SAGAT is team SAGAT (TSAGAT) where the different positions in a team answer SAGAT-questions specific to their position. TSAGAT was constructed to account for the teams' shared situation awareness and validated in trauma team education in a Canadian

setting.[24] SAGAT has not to our knowledge previously been used in a Swedish context, and therefore it is of importance to evaluate both the feasibility and the internal consistency of a Swedish version of the instrument.

To summarise, A–E checklists (based on the TTET), TEAM, and SAGAT have been developed in order to evaluate different aspects of teamwork. However, these instruments have neither been translated into Swedish nor tested for feasibility or trustworthiness in a Swedish context. Such studies are necessary so as to further evaluate teamwork in acute care settings and in simulation-based trainings. Thus, the aims of this study are (1) to test an A–E checklist developed by us for interrater reliability, (2) to test the TEAM instrument for interrater reliability, (3) to test a Swedish version of the SAGAT instrument for feasibility, and (4) to test SAGAT for internal consistency.

Method

This cross-sectional study is based on quantitative data collected during video-recorded simulation-based team training sessions.

Setting

Data collection in this study was carried out during videotaped simulation team training at the Clinical Training Centre (CTC) of the Medical Faculty, Umeå university. The week before the team training, the students were asked to watch a twelve-minute video available at the learning platform that introduced both the ABCDE-concept when caring for a patient and the simulation setting.[25] At the start of the trainings session, the students participated in a fifteen-minute live introduction to the simulation laboratory by the operator and the instructor.

Each team trained on four scenarios focusing on assessment and treatment of severely ill patients in an emergency room. The first scenario was a warm up and the last was a teaching summing up. The second and third scenarios were included in the study as case 1 and 2. The order of the two cases was randomized at a group level just before the start of scenario two. Both cases were designed to last 10–15 minutes and were pre-programmed into a Laerdal SimMan simulator to support the standardization of the simulation. Different handlers were added for changes in vital signs, e.g. proportionally increased saturation and blood pressure when oxygen and fluid were administered, respectively.

The cases used were in essence slightly modified versions of the scenarios used by Hogan.[27] Case 1 was a 25-year old male with hypovolemic shock after a traffic incident. Case 2 was set as a 35-year male suffering from pneumothorax with affected vital signs after a traffic incident.

Observations were made during the simulation and field notes were taken by one of the authors (KJ). Two cameras mounted at an angle were used to record videos of the room and one of the views included the patient monitor.

Participants

During March to October 2016, all medical students (n=68) in year 4 doing their clerkship in anaesthesiology and critical care medicine were invited to participate in the study while doing compulsory simulator-based team training. Initially, the students were divided into 27 teams. In total, 55 students (28 males and 27 females) in a total of 23 teams participated in the study (Table 1). Thirteen students in four teams were excluded from the study as one or more in the team did not want to participate in this study.

The different steps in the aim with the participants in the different parts of this study:

- First, the analyses of inter-rater reliability for the TEAM instrument was based on ratings from the videotaped records by four professionals (nurse n=2, physician n=1, and paramedic n=1) (Table 2).

- Second, the analyses of inter-rater reliability for the A-E scale was based on ratings from the videotaped records by four professionals (nurse n=2, physician n=1, and paramedic n=1) (Table 2).
- Third, a post-simulation questionnaire (PSQ) with questions on relevance and impact of the pauses on the training sessions was answered by 24 of the 55 students (Table 1) described above.
- Fourth, the analysis of internal consistency of SAGAT was based on the 55 students described above.

Data collection

In the study, we used a questionnaire including background characteristics, the A–E checklist, the TEAM instrument, the SAGAT instrument, a Post-simulation questionnaire (PSQ), and the time taken for measuring SAGAT. The questionnaire included the informed consent and was answered immediately before the team training started. It had questions on the participants' background characteristics (n=55) such as year of birth, male/female, previous medical training, previous experience of team training, previous experience of human patient simulator-based training, previous experience of CRM, and previous experience of live trauma care.

A–E checklist

In order to measure the completeness of critical tasks in the acute care scenarios, an A–E checklist was used. The original TTET as presented by Holcomb et al.[11] consists of 58 items derived from the ATLS protocol. Each item in the TTET was scrutinized by the authors of this study and the number of items reduced to those reflecting the actions expected to be performed by year 4 medical students in the acute care scenarios. The final list, in Swedish, contained 10 items. Compared to the original scoring system, an additional scoring option was added; performed after a reminder from the instructor. Each item was rated on a Likert scale from 0 to 4 (0 = Not initiated, 1= Performed after reminder from the instructor, 2 = Partly performed, 3 = Performed completely before the end of the simulation, and 4 = Performed consistently during the whole simulation, NA = not applicable).

TEAM instrument

In order to measure team leadership, team work and task management, the TEAM instrument was used unmodified, i.e. in English as developed by Cooper et al.[13] The published internal consistency (Cronbach's alpha) was 0.97, the inter-rater reliability 0.55 as measured by Cohen's Kappa (adjusted for chance), and a mean intraclass correlation coefficient for the 11 items of 0.60. The instrument consists of 11 key behaviours rated on a Likert scale from 0 to 4 (0 = Never/Hardly, 1 = Seldom, 2 = About as often as not, 3 = Often, 4 = Always / Nearly always) and finally a Global rating of the team's overall performance on a 1 to 10 scale. The original TEAM instrument lacks anchors for the 1 to 10 scale, in this study we used 1 = Poor and 10 = Excellent.

The team instrument consists of three subscales: leadership (item 1–2), team work (item 3–9) and task management (item 10–11). An index was constructed as a mean score for respective subscale and for all 11 items in the instrument (mean score for the total TEAM). The global rating of the overall performance was analysed separately.

Procedure in rating the TEAM instrument and A-E check list

The raters in this study used video recordings in the rating procedure of the TEAM scale as well as the A–E check list. The raters had two separate two-hour long meeting to discuss the interpretation of the anchors and items on the A–E checklist and the TEAM instrument.[11] During the first meeting, the discussions were facilitated by a different set of videos with similar scenarios, but of other teams. A total of 6 videos were used for this. During the second meeting, after four of the scenarios were rated by each participant, the raters met again to discuss the interpretation of scales

The raters independently watched the videos of the simulation scenarios included in this study and rated the performances using the A–E checklist and the TEAM instrument. Each video was viewed at least twice by each rater. An online form was used to collect the data from the raters so as to minimize the risk of transfer errors when moving data from paper to an Excel file.

SAGAT

When developing questions for the SAGAT-instrument to a specific scenario, goal directed task analysis was used as originally described by Endsley. Briefly, for each profession, the major goals are identified along with the sub-goals. Then key decisions are identified, and SA requirements are defined as the dynamic information needed to achieve the major goals, as opposed to static information such as rules and guidelines.[26] The samplings (questions) are then matched against the SA requirements. The original recommendation by Endsley was 30–60 samplings (questions) for within-subjects studies for each of the three SA levels; (1) perception and attention, (2) comprehension, and (3) projection.[15] When Gardner et al. validated a SAGAT instrument in a study of medical trainees, each questionnaire consisted of three questions for each level of SA at each freeze.[20,26,27] SAGAT with three questions at level 1, one at level 2 and three at level 3.

In the present study, the SAGAT instrument was refined and adapted to the scenario and expected skills level of the students according to the process used by Hogan et al. [27] First, the Hogan SAGAT instrument was translated into Swedish by the authors of this study. In accordance with the original instrument, the targeted learning objectives for the training scenarios were then formulated and next the specific goals for each simulation was set. An iterative process was used to reformulate the SAGAT questions in Swedish, using a separate group of 6 professionals, all registered nurses, working both in the clinical context and in the teaching context. The final sets of SAGAT questions are shown in Tables 4A and 4B. Lawshe advocated the use of professionals to identify whether an item is important in a certain context. [28] Content validity index is the fraction of professionals that rates the item as important. In the present study, the relevance of the SAGAT questions in relation to the scenarios in the team training was reviewed by three professionals (nurse n=1, physician n=2) before being used in the study. All agreed that the questions were relevant, i.e. the content validity index was high.

The answers given by the participants in the SAGAT instrument were classified as incorrect (0) or correct (1) by two of the authors (MHu, KJ). The classifications were discussed and agreed upon by the two authors. For answers on a continuous scale (e.g. systolic blood pressure) a 10% range around the intended correct answer was accepted as correct. One question was removed from the questionnaire, since it became obvious during the classification process that the question had frequently been misinterpreted.

In order to be able to probe SA with the SAGAT instrument, the scenarios were frozen (i.e. paused) twice. The first freeze of the scenario was five minutes into the scenario, unless there was an active task activity or if the team was doing a team re-evaluation.[25] If so, the freeze was postponed for a moment. During the freeze, the patient monitor was switched off and the participants turned away from the patient simulator while individually answering the questions. The second freeze took place according to the same principles after an additional 5 minutes. All participants were allowed to complete the SAGAT questions before the scenario started again.

The length of the freeze was measured using the video recordings. The start of the freeze was defined as the beginning of the sentence “Now we pause the scenario, so that you can answer some questions about the patient case” and the end of the freeze as the end of the sentence “Everyone in place and ready? Now we start again”. The time taken to measure SAGAT was defined as the end-time minus the start time and was measured in seconds. The simulations lasted 14.1 (sd 2.7) minutes and the freeze needed for all team members to answer SAGAT was 2.7 (sd 0.6) minutes (Table 3).

The feasibility of using SAGAT to measures situation awareness was assessed by measuring the length of the freeze (i.e. pause) needed to answer questions and by asking the participants for their perception of the pauses in the scenario and how the pauses affected the training session.

Team SAGAT (TSAGAT) was developed by Crozier et al. based on SAGAT as an assessment tool to evaluate team performance.[24] In Crozier's study, each team consisted of a trauma leader, an airway manager and a nurse and individual SAGAT instruments were developed for the three team positions including both shared knowledge and complimentary knowledge. TSAGAT was calculated as the sum of individual SAGAT scores, and the TSAGAT scores had a high correlation to a traditional checklist (Pearson correlation, r=0.996).[24] However, Salas defined team SA as a dynamic process defined as the team's shared understanding of a situation at a specific point in time [6] and Endsley argued that team SA involves unique activities as information sharing and coordination.[15] Thus, in the present study, in order to measure this efficiently in all team members, all participants in a case received identical SAGAT instruments. TSAGAT was calculated based on the mean scores in the team for each SAGAT item.

Post-simulation questionnaire (PSQ)

In order to measure whether the SAGAT items were considered relevant to the scenarios, and whether pausing the scenarios affected the team training activity, a post-simulation questionnaire (PSQ) was used.[27] The PSQ consists of 13 statements to be rated on a four-point scale ranging from Strongly disagree (1) to Strongly agree (4), where 5 statements concern the SAGAT and the effect of freezing the scenario and 8 statements concern the simulation and the scenario per se. As far as we have found, no data regarding the reliability of PSQ was presented in the original study. In this study, we only use the five questions relating to SAGAT and freezing the scenario. The PSQ was translated into Swedish by a professional translator and the translation was further refined based on iterative discussions within our research group. In the second week after the training session, the PSQ was answered by 24 of the participating students (Table 5). In this study, the four-point scale was dichotomized into disagree (Strongly disagree and Disagree) (1) and agree (Agree and Strongly agree) (2).

Study size

The aim of the study was to evaluate feasibility of A-E, TEAM and SAGAT for use in further studies. For this, the study sample needs to be large enough to allow for calculation of descriptive statistics (means and standard deviations) and for calculation of reliabilities with a fair precision. To be able to assess a medium-large effect size (Cohen's d of 0.6) with t-test with a power of 80% at the 0.05 level would need 45 participants in each group as determined with G*Power.[30]

Statistical analysis

The statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM, Chicago, IL). Interrater reliability for the A–E checklist and the TEAM instrument was determined by intraclass correlation with a two-way random effects model (ICC (2,1) type absolute).[29,31] The results are reported as both single measures and average measures, since the ICC for single measures answers to how equal the individual ratings are, while average measures answers to how reliable the mean values of the individual ratings are. The results of the PSQ are presented by descriptive statistics. SAGAT was calculated both as the mean of each SAGAT question in each team defined as TSAGAT and also at an individual level defined as SAGAT.[24] Internal consistency for the SAGAT instrument and TSAGAT was measured by Cronbach's alpha.[29,31] Internal consistency considered as the extent to which all items measure the same latent variable was investigated by χ^2 and as suggested by Schweizer, a normed χ^2 below 2 was taken as an indication of a good fit.[32]

Ethics

This study was approved by the Ethical Review Board in Northern Sweden (Chairman Anders Iacobaeus, April 7, 2016, decision No. 2016-54-31M). Participation in the study was voluntary. One week before the start of the simulation session, the participants received information both verbally and by mail. Informed consent was signed individually immediately before the start of the training session, and teams were included in the study only if all participants had agreed to participate.

Patient and Public Involvement

Nor patients nor the public was involved in the design or in the data collection for this study.

Results

A–E and TEAM

Interrater reliability as measured by intraclass correlation was 0.55 (single measures) / 0.83 (average measures) for the A–E checklist and 0.54 / 0.83 for the TEAM-scale. For the TEAM subscales leadership, team work, task management and global rating, the intraclass correlations were 0.36 / 0.70, 0.45 / 0.77, 0.35 / 0.68, and 0.38 /0.72, respectively. The mean item A–E score was 2.64 (SD 0.24), mean TEAM item 1–11 2.30 (SD 0.49), and mean TEAM item 12 (global rating) 4.83 (SD 1.06).

SAGAT

The PSQ showed that the SAGAT questions (Tables 4A and 4B) were considered relevant to the case by 96% of the participants and that 96% considered the questions easy to understand. About three out of four (72%) participants stated that the freeze had no negative impact on their concentration or performance during the simulation session (Table 5).

The internal consistency of SAGAT measured as Cronbach's alpha was 0.80 for case 1 and 0.62 for case 2 (Table 6), and normed χ^2 was 1.72 vs 1.62. For level 1 (perception) Cronbach's alpha was low, 0.06 for case 1 and 0.25 for case 2, but for level 3 it was fair, 0.89 and 0.66, respectively.

For TSAGAT, the internal consistency was good for the instrument as a total and for level 3 (Table 6).

Discussion

The main findings in this study indicate that that both the A–E-checklist based on TTET[11] and the TEAM instrument[13] could be used in a Swedish setting with acceptable interrater reliability, and that it was feasible to use SAGAT[26] to measure SA. The combination of these three measurements allows for analysis of task performance, team performance, and the relation to situation awareness in team trainings.

Scaling-down the comprehensive TTET[11] to a smaller A–E checklist might introduce unintended errors in measurement. To be a sensitive instrument, the items measured need to be relevant for the task, and the anchors of the scale used need to be calibrated to the setting to allow for significant changes to be reflected as difference between control and intervention.[29] In the present study, the developed A–E checklist was perceived by professionals to be relevant to the case. The interrater variability was low, showing that the checklist could reliably be used for scoring task performance. The means of the scores were in the middle of the scale, which indicates that the checklist might be sensitive for differentiating between low and high performers.

For the TEAM instrument, the interrater reliability was 0.55 for single measurements and 0.83 for average measurements. According to Koo and Li, an ICC between 0.50 and 0.75 indicates moderate reliability, and 0.75 and 0.90 indicates good reliability.[31] The ICC for single measurements should be reported if the plan is to use the measurements from only one rater, and ICC for average

measurements if the plan is to use averaged measurements.[31] In this study, we chose to report both values to show the importance of using multiple raters to improve overall reliability. In the original study by Cooper, an interrater reliability for the TEAM instrument of 0.55 and 0.60 was reported as measured by Cohen's kappa (adjusted for chance) and ICC, respectively.[13] Thus the interrater reliability as measured by ICC (single measurements) in this study was equivalent to the ICC obtained in the original setting.

Both the PSQ and the CVI indicate that the SAGAT instrument could be used to construct questions that were considered relevant to the case. In addition, the internal consistency was fair (0.80 and 0.61). When analysing the subscales, level 1, 2 and 3, the internal consistency was low for level 1 and 2, and higher for level 3. This might indicate that the perception of the situation in the groups was scattered and not related to the total score, while the ability to project where the cases were heading was more homogeneously related to the total score. TSAGAT had a higher overall homogeneity, as might be expected when analysing the means of the group SAGAT for each question instead of the individual SAGAT answers.

Measurement of situation awareness using SAGAT requires that the scenario is paused while the participants answer questions probing the three levels of SA: perception, comprehension, and projection.[26] Each pause in this study were less than 3 minutes. The pauses and SAGAT questions might influence SA both in negative direction and in a positive direction. This might disrupt the flow in the simulation, induce stress and impede the suspension of disbelief. It might also facilitate the resolution of the clinical problem in the case by triggering reflection-on-action and reflection-in-action.[19] In the PSQ, the pauses in this study did not cause any adverse reactions on the part of the participants.

Limitations of this study

To obtain a convenient sample to analyse the feasibility of using the A–E checklist, TEAM, SAGAT, and TSAGAT, trauma team trainings with medical students were used to collect questionnaires and video for further analysis. Using teams at different skills levels, instead of having a rather homogeneous group of medical students, might have created wider range of ratings and thus have improved the psychometrics. However, as the purpose was to test the interrater reliability of A–E and TEAM as well as the feasibility and internal consistency of SAGAT, this study most likely does not overestimate the psychometric properties.

As the SAGAT questions are case/specific, the content validity is rather difficult to assess and generalize. A systematic use of content validity index (CVI) as suggested by Lawshe might be a purposeful method.[28] Using professionals to rate whether the item is essential and then calculating CVI as the fraction that rated the item as essential allows for a systematic screening of a larger set of potential items.

The negative result for TSAGAT homogeneity level 2 in one of the cases indicates that there is a structural problem such as too few questions or measurements in that category. Since level 2 is represented by only a pair of questions, it is not unlikely. As pointed out by Tavakol et al., Cronbach's alpha should be considered as a lower bound estimate of reliability.[31] The number of questions in each level in the SAGAT tool needs to be balanced between time taken to administer the test and optimal sampling of a trait.

Conclusion

It was possible to develop a homogeneous SAGAT instrument aimed at measuring situation awareness during trauma team training with patient simulators also in a Swedish setting. The TEAM and A–E

checklist had a high interrater reliability, indicating that the descriptors in the instruments were similarly interpreted by the raters. SAGAT and TSAGAT had fair internal consistencies. To conclude, it is feasible to use TEAM, A-E and SAGAT/TSAGAT for measurement of team performance and situational awareness in critical care simulation training

Tables

Table 1 Background characteristics for medical students (n=55) participating in the study

	Part 1
Age years, m (sd)	25 (4.5)
Male n (%)	28 (51%)
Female n (%)	27 (49%)
Former healthcare education	
None n (%)	44 (%)
Assistant nurse n (%)	2 (%)
Registered nurse n (%)	3 (%)
Red Cross volunteer n (%)	2 (%)
Other n (%)	4 (%)
Former team training	
Yes n (%)	20 (%)
Former simulation experience	
Yes n (%)	32 (%)
Former experience of trauma patients *	
Yes n (%)	20 (37%)

* one missing value

Table 2 Professionals involved as raters of video material

	Nurse 1	Nurse 2	Physician 1	Paramedic 1
A–E rating	x	x	x	x
TEAM rating	x	x	x	x

Nurse 1 Registered nurse with a Master's degree (one year) in Nursing (Critical Care Medicine), 20 years working experience at an ICU, 9 years' experience of human patient simulator team training and also PhD student (author KJ).

Nurse 2 Registered nurse with a Master's degree (one year) in Nursing (Acute Care Medicine), had worked 6 years with prehospital care and was also a year 1 Medical student.

Physician 1 Associate Professor, consultant in Anesthesiology and Critical Care Medicine, and had 14 years' experience of human patient simulator team training (author MHu).

Paramedic 1 The Paramedic had limited experience outside two years working as a paramedic in the Israeli Army and was also a year 1 Medical student.

Table 3 Time (minutes) for scenario freezes to measure situation awareness with SAGAT

Scenario	No of teams	Scenario	1st freeze	2nd freeze	Total time
	n	m (sd)	m (sd)	m (sd)	m (sd)
Case 1	12	14.1 (2.0)	2.7 (0.5)	2.5 (0.6)	19.3 (2.3)
Case 2	11	14.1 (3.4)	2.7 (0.6)	2.8 (0.6)	19.7 (3.8)

Table 4a

SAGAT questions vs situation awareness level, case 1

Freeze 1	Question	Level
1.	What is the patient's respiratory rate?	1
2.	What are the findings from the assessment of A and B?	1
3.	Is the patient sufficiently oxygenated? *Yes or No*	2
4.	What did the patient state as their year of birth?	1
5.	What do you think will happen to the patient's blood pressure in the coming minutes? *Increase/Decrease/No-change*	3
6.	What is wrong with the patient? *State your preliminary diagnosis*	2
7.	How many peripheral IV-cannulas does the patient have?	1
	Presuming a normal course of development, how will the vital signs change in the upcoming 10 minutes?	
8.	*HR Increase/Decrease*	3
9.	*BP Increase/Decrease*	3
10.	*POx Increase/Decrease*	3
11.	*RR Increase/Decrease*	3

Freeze 2	Question	Level
12.	What is the patient's blood pressure?	1
13.	What are your findings under C (Circulation)?	1
14.	What additional examinations/tests do you think are needed?	3
15.	Which medications/drugs do you think are needed?	3
16.	What is wrong with the patient (preliminary diagnosis)?	2
17.	Is there an algorithm for CPR on the wall?	1

18.	What previous diseases has the patient stated? Presuming a normal course of development, how will the vital signs change in the upcoming 10 minutes?	2
19.	*HR Increase/Decrease*	3
20.	*BP Increase/Decrease*	3
21.	*POx Increase/Decrease*	3
22.	*RR Increase/Decrease*	3

Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

Table 4b

SAGAT questions vs SAGAT level, case 2

Freeze 1	Question	Level
1.	What is the patient's respiratory rate?	1
2.	What are your findings under A (Airway) and B (Breathing)?	1
3.	Is the patient sufficiently oxygenated? *Yes or No*	2
4.	What is the sign on the patient necklace?	1
5.	What do you believe will happen to the patient's venous return in the upcoming minutes? *Increase/Decrease/No change*	3
6.	What is wrong with the patient (preliminary diagnosis)?	2
7.	Is a suction device available?	1
	Presuming a normal course of development, how will the vital signs change in the upcoming 10 minutes?	
8.	*HR Increase/Decrease*	3
9.	*BP Increase/Decrease*	3
10.	*POx Increase/Decrease*	3
11.	*RR Increase/Decrease*	3

Freeze 2	Question	Level
12.	What is the patient's respiratory rate?	1
13.	What are your findings under C (Circulation)?	1
14.	What additional examinatons/tests do you think are needed?	3
15.	Which medications/drugs do you think will be needed?	3
16.	What is wrong with the patient (preliminary diagnosis)?	2
17.	What is the patient's blood pressure?	1
18.	Is there a defibrillator available in the room?	1
	Presuming a normal course of development, how will the vital signs change in the upcoming 10 minutes?	

19.	*HR Increase/Decrease*	
20.	*BP Increase/Decrease*	3
21.	*POx Increase/Decrease*	3
22.	*RR Increase/Decrease*	3
		3

Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

The questions are shown translated into English in the table and the original questions in Swedish are available upon request from the authors of the study.

Table 5 The number and proportion (%) of participants answering the post-simulation questionnaire

Question	Agree	Disagree	Don't know
Concerning the SAGAT instrument. in my view:	**n (%)**	**n (%)**	**n (%)**
The introduction was adequate	24 (100%)	0 (0%)	0 (0%)
The questions were clear	23 (96%)	1 (4%)	0 (0%)
The "freezes" in the scenario had an adverse effect on my concentration and performance	7 (29%)	17 (71%)	0 (0%)
The questions were relevant to the way I perceived the scenario	23 (96%)	1 (4%)	0 (0%)
Overall, I am satisfied with SAGAT as a tool for evaluating hands-on skills during a trauma exercise.	19 (79%)	1 (4%)	4 (17%)

Table 6 Internal consistency (Cronbach's alpha) for SAGAT and TSAGAT

SAGAT / TSAGAT	Case 1	Case 2
n	39 / 12	36 / 11
1	0.060 / 0.372	0.246 / 0.420
2	0.321 / -0.018	0.332 / 0.332
3	0.891 / 0.891	0.659 / 0.786
Total (Level 1-3)	0.800 / 0.827	0.620 / 0.759

References

1 Hofmann DA, Mark B. An investigation of the relationship between safety climate and medication errors as well as other nurse and patient outcomes. *Pers Psychol* 2006;**59**:847–69. doi:10.1111/j.1744–6570.2006.00056.x

2 Leonard M. The human factor: the critical importance of effective teamwork and communication in providing safe care. *Qual Saf Health Care* 2004;**13**:i85–90. doi:10.1136/qshc.2004.010033

3 Odell M. Human factors and patient safety: changing roles in critical care. *Australian critical care : official journal of the Confederation of Australian Critical Care Nurses* 2011;**24**:215–7. doi:10.1016/j.aucc.2011.02.001

4 Scalese RJ, Obeso VT, Issenberg BS. Simulation Technology for Skills Training and Competency Assessment in Medical Education. *J Gen Intern Med* 2007;**23**:46–9. doi:10.1007/s11606-007-0283-4

5 LeBlanc VR, Manser T, Weinger MB, *et al.* The study of factors affecting human and systems performance in healthcare using simulation. *Simul Healthc* 2011;**6 Suppl**:S24–9. doi:10.1097/SIH.0b013e318229f5c8

6 Salas E, Prince C, Baker DP, *et al.* Situation Awareness in Team Performance: Implications for Measurement and Training. *Human Factors* 1995;**37**:123–36. doi:10.1518/001872095779049525

7 Wright MC, Taekman JM, Endsley MR. Objective measures of situation awareness in a simulated medical environment. *Qual Saf Health Care* 2004;**13**:i65–71. doi:10.1136/qshc.2004.009951

8 Auerbach M, Roney L, Aysseh A, *et al.* In situ pediatric trauma simulation: assessing the impact and feasibility of an interdisciplinary pediatric in situ trauma care quality improvement simulation program. *Pediatric emergency care* 2014;**30**:884–91. doi:10.1097/PEC.0000000000000297

9 Converse SA, Cannon-Bowers JA, Salas E. Team Member Shared Mental Models: A Theory and Some Methodological Issues. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1991;**35**:1417–21. doi:10.1177/154193129103501917

10 Mathieu J, Heffner T, Goodwin G, *et al.* The influence of shared mental models on team process and performance. *The Journal of applied psychology* 2000;**85**:273–83.

11 Holcomb JB, Dumire RD, Crommett JW, *et al.* Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. *J Trauma* 2002;**52**:1078–85; discussion 1085-6.http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=12045633&retmode=ref&cmd=prlinks

12 Thim T, Krarup, Grove, *et al.* Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012;:117. doi:10.2147/IJGM.S28478

13 Cooper S, Cant R, Porter J, *et al.* Rating medical emergency teamwork performance: development

of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;**81**:446–52. doi:10.1016/j.resuscitation.2009.11.027

14 Cant RP, Porter JE, Cooper SJ, *et al.* Improving the non technical skills of hospital medical emergency teams: The Team Emergency Assessment Measure (TEAM). *Emerg Med Australas* 2016;**28**:641–6. doi:10.1111/1742-6723.12643

15 Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors* 1995;**37**:32–64. doi:10.1518/001872095779049543

16 Jones DG, Endsley MR. Sources of situation awareness errors in aviation. *Aviation, space, and environmental medicine* 1996.

17 Durso FT, Bleckley M, Dattel AR. Does situation awareness add to the validity of cognitive tests? *Human factors* 2006;**48**:721–33. doi:10.1518/001872006779166316

18 Hamstra SJ, Brydges R, Hatala R, *et al.* Reconsidering fidelity in simulation-based training. *Academic medicine : journal of the Association of American Medical Colleges*;**89**:387–92. doi:10.1097/ACM.0000000000000130

19 Schön DA. *The reflective practitioner: How professionals think in action*. The Perseus Books Group 1984.

20 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc* 2016;**12**:1–5. doi:10.1097/SIH.0000000000000181

21 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident's Situational Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3. doi:10.1097/BOT.0000000000000003

22 Lavoie P, Cossette S, Pepin J. Testing nursing students' clinical judgment in a patient deterioration simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today* 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015

23 McKenna L, Missen K, Cooper S, *et al.* Situation awareness in undergraduate nursing students managing simulated patient deterioration. *Nurse Educ Today* 2014;**34**:e27–31. doi:10.1016/j.nedt.2013.12.013

24 Crozier MS, Ting HY, Boone DC, *et al.* Use of human patient simulation and validation of the Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009

25 American College of Surgeons. *ATLS Student Course Manual: Advanced Trauma Life Support*. 9th ed. American College of Surgeons 2013.

26 Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. In: *Situation awareness analysis and measuement*. Lawrence Erlbaum Associates 2000. 1–21.

27 Hogan MP, Pace DE, Hapgood J, *et al.* Use of human patient simulation and the situation awareness global assessment technique in practical trauma skills assessment. *J Trauma* 2006;**61**:1047–

52. doi:10.1097/01.ta.0000238687.23622.89

28 Lawshe CH. A quantitative approach to content validity. *Personnel Psychology* 1975;**28**:563–75. doi:10.1111/j.1744-6570.1975.tb01393.x

29 Nunnally JC. *Psychometric Theory, 2nd edition*. McGraw-Hill 1978.

30 Faul F, Erdfelder E, Buchner A, *et al.* Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behavior research methods* 2009;**41**:1149–60. doi:10.3758/BRM.41.4.1149

31 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine* 2016;**15**:155–63. doi:10.1016/j.jcm.2016.02.012

32 Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Medical Educ* 2011;**2**:53–5. doi:10.5116/ijme.4dfb.8dfd

STROBE Statement—Checklist of items that should be included in reports of ***cross-sectional studies***

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	2
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	4
Objectives	3	State specific objectives, including any prespecified hypotheses	5
Methods			
Study design	4	Present key elements of study design early in the paper	5
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	5
Participants	6	(*a*) Give the eligibility criteria, and the sources and methods of selection of participants	5
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	5-7
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	5-7
Bias	9	Describe any efforts to address potential sources of bias	5-7
Study size	10	Explain how the study size was arrived at	8
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	5-7
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	8
		(*b*) Describe any methods used to examine subgroups and interactions	8
		(*c*) Explain how missing data were addressed	8
		(*d*) If applicable, describe analytical methods taking account of sampling strategy	N.A.
		(*e*) Describe any sensitivity analyses	N.A.
Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	5
		(b) Give reasons for non-participation at each stage	5
		(c) Consider use of a flow diagram	N.A.
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	Table 1 and 2
		(b) Indicate number of participants with missing data for each variable of interest	N.A.
Outcome data	15*	Report numbers of outcome events or summary measures	8-9
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	8-9

		(*b*) Report category boundaries when continuous variables were categorized	N.A
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	N.A.
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	8-9
Discussion			
Key results	18	Summarise key results with reference to study objectives	9
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	10
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	9-10
Generalisability	21	Discuss the generalisability (external validity) of the study results	10
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	

*Give information separately for exposed and unexposed groups.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

Are instruments that measure situation awareness, team and task performance reliable and feasible for a simulation training setting?

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029412.R1
Article Type:	Research
Date Submitted by the Author:	21-May-2019
Complete List of Authors:	Hultin, Magnus; Umeå Universitet Medicinska fakulteten, Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine Jonsson, Karin; Umeå Universitet Medicinska fakulteten, Departments of Surgical and Perioperative Sciences and Nursing Härgestam, Maria; Umea Universitet Medicinska fakulteten, Department of Nursing Lindkvist, Marie; Umeå Universitet Medicinska fakulteten, Department of Public Health and Clinical Medicine Brulin, Christine; Umeå Universitet Medicinska fakulteten, Department of Nursing
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Communication
Keywords:	Patient care team, Resuscitation, Situation Awareness, Task performance and analysis, Teamwork

Are instruments that measure situation awareness, team and task performance reliable and feasible for a simulation training setting?

Magnus Hultin, MD Ph.D., Associate Professor, Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0003-2935-7161

Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0001-7974-6777

Maria Härgestam, RNA Ph.D., Lecturer, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0003-0680-9962

Marie Lindkvist, Ph.D., Associate Professor, Department of Epidemiology and Global Health, Umeå University, S-901 87 Umeå, Sweden

Christine Brulin, RNA Ph.D., Professor emeritus, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden

Abstract 247 words

Main article 4,828 words

Keywords

- Patient Care Team
- Resuscitation
- Situation Awareness
- Task Performance and Analysis
- Teamwork

Corresponding author

Magnus Hultin, Dept of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, Sweden, magnus.hultin@umu.se, phone: +46-90-786 9109

Abstract

Objectives: The assessment of situation awareness, team and task performance in a simulation training session requires reliable and feasible measurement techniques. The objectives of this study were to test an Airways-Breathing-Circulation-Disability-Exposure (ABCDE) checklist and Team Emergency Assessment Measure (TEAM) for interrater reliability, as well as an application of Situation Awareness Global Assessment Technique (SAGAT) for feasibility and internal consistency.

Design: Methodological study

Setting: Data collection during team training using full-scale simulation at a university clinical training centre. The video-recorded scenarios were rated independently by four raters.

Participants: 55 medical students aged 22–40 years during their 4th year medical studies, during the clerkship in anaesthesiology and critical care medicine, formed 23 different teams. All students answered the SAGAT questionnaires and, of these students, 24 answered the follow-up post-simulation questionnaire (PSQ). TEAM and ABCDE were scored by four professionals.

Measures: ABCDE and TEAM were tested for interrater reliability. The feasibility of the SAGAT instrument was tested using PSQ. SAGAT was tested for internal consistency both at an individual level (SAGAT) and a team level (TSAGAT).

Results: The intraclass correlation was 0.54/0.83 (single/average measurements) for TEAM and 0.55/0.83 for ABCDE. According to the PSQ, the items in SAGAT were rated as relevant to the scenario by 96% of the participants. Cronbach's alpha for SAGAT/TSAGAT for the two scenarios was 0.80/0.83 vs. 0.62/0.76, and normed χ^2 was 1.72 vs. 1.62.

Conclusions: Task performance, team performance and situation awareness could be purposefully measured, and the reliability of the measurements was good.

Article Summary

Strengths and limitations of this study

- SAGAT could be used to create items in Swedish to probe situation awareness, i.e. in the participants' native language
- TEAM could be used by the raters in its original language (English).
- The developed ABCDE checklist has items that are well-defined concepts. However, the difficulty lies in defining the rubrics for scoring the items.
- It is a weakness that PSQ was only translated into Swedish and not retranslated back into English.
- An interprofessional set of raters with different backgrounds and experiences rated TEAM and ABCDE with similar results.

Funding

This work was supported by Region Västerbotten, grant no. VLL-663801 and VLL-836931, by Region Norrbotten, grant no. NLL-765981, by Alice Lindström's Foundation, and by the Medical Faculty of Umeå University.

Competing interests

None of the authors have any competing interests to declare.

Data availability

The data comprising de-identified questionnaire responses and ratings of video scenarios as SPSS data files and SPSS syntax files are available upon request from the corresponding author.

Author Contributions

MHu, KJ, MHä and CB designed the study. MHu and KJ prepared the initial draft of the paper. MHu, KJ, MHä, ML and CB were all actively engaged in data analysis and in finalising the paper. All authors have read and approved the final version of the paper.

Introduction

Medical errors are the third leading cause of death in the US.[1] Knowledge about the relationship between human errors and patient safety has increased in the last two decades.[2,3] Simulated environments make it possible to improve skills by employing training strategies to prevent errors while simultaneously offering an arena for reliable assessments of skills.[4] Thus, simulation training is often used by organisations to minimise adverse events and prevent healthcare errors.[5,6] This could be accomplished by improving *task performance*, *team performance* or *situation awareness (SA)*.[7–9]

When developing strategies for improving clinical practice, it is essential to evaluate both task and team performance. According to Salas[10] and Kozlowski[11], team performance is a multilevel process that includes the interrelation between individual- and team-level taskwork and teamwork processes. Thus, an optimal task, as well as team performance, depends on the coordinated activities of a team of individuals.[12,13] Checklists are often used to score task performance in acute care scenarios. The lists might include adherence to resuscitation protocols, timing of the task, as well as the time taken to complete the components.[14] The Trauma Team Evaluation Tool (TTET) was developed by Holcomb for trauma scenarios managed according to protocols based on Advanced Trauma Life Support (ATLS) and was tested on US military resuscitation teams from community hospitals.[15] As a pilot study, psychometric data such as reliability and validity were not reported. As TTET was developed for a specific setting, the items and criteria for scoring had to be adapted to the proficiency levels of the participants and the standard operating procedures being used.

Team performance can be measured using the Team Emergency Assessment Measure (TEAM), for example. It measures three dimensions of team performance: leadership,

teamwork and task management.[16,17] The instrument was initially developed and validated for simulator-based team training and was recently validated for the collection of observational ratings of non-technical skills during live resuscitations in emergency departments. [18] Scoring scenarios that use instruments depend on reliable interpretations of the instrument by the raters and it might be even more important to ensure the reliability of such interpretations when using an instrument in a non-native language.

Moreover, and in addition to task and team performance, SA is a prerequisite for patient safety and the prevention of errors, particularly in acute care situations.[9,19] SA includes three levels of ability: (1) perception and attention (*What?*, what's going on), (2) comprehension (*So what?*, the ability to understand what's going on), and (3) projection (*What's next?*) in order to anticipate and plan for future events).[20] In order to measure SA in a simulation setting, the Situation Awareness Global Assessment Tool (SAGAT) has been developed and subsequently adapted for use in healthcare settings.[21,22] One feature of SAGAT is that its use requires the simulation to pause, which might influence clinical understanding, both by impeding the suspension of disbelief in the simulation setting [23] and by facilitating reflection on action.[24] Gardner showed that it was feasible to use SAGAT to measure SA in the team training of surgical trainees in advanced cardiac life support.[25] Globally, SAGAT has been used to study, for example, the effect of sleep deprivation on SA in trauma team training, how SA is associated with surgical trainee team performance, as well as nurses' clinical assessment of patient deterioration.[26–28] SAGAT can be analysed on an individual level and also on a team level. A specific application of SAGAT is team SAGAT (TSAGAT) in which the different members of a team answer SAGAT questions specific to their role. TSAGAT was constructed to account for the teams' shared situation awareness and validated in trauma team training in a Canadian setting.[29] To our knowledge, SAGAT has

not previously been used in a Swedish context and it is therefore important to evaluate both its feasibility and its internal consistency.

To summarise, ABCDE checklists, the TEAM instrument and SAGAT have been developed in order to evaluate different aspects of teamwork. However, these instruments and questionnaires have neither been translated into Swedish nor tested for feasibility or trustworthiness in a Swedish context. Such studies are necessary to enable the evaluation of teamwork in Swedish acute care settings and simulation-based training. Thus, this study aimed to test an ABCDE checklist that we developed for interrater reliability, to test the TEAM instrument for interrater reliability, and to test two Swedish SAGAT questionnaires for feasibility and internal consistency.

Method

This study is based on data collected during simulation-based team training sessions with medical students.

Participants

From March to October 2016, all medical students (n=68) in year 4 undertaking their clerkship in anaesthesiology and critical care medicine were invited to participate in the study while receiving mandatory simulator-based team training. In total, 55 students (81%) participated in the study (Table 1) and 20 of them participated in both scenario A and scenario B.

For interrater reliability of the post-scenario rating of videos from the simulations, four raters from different backgrounds were involved. First, a registered nurse with a Master's Degree (one year) in Nursing (Critical Care Medicine), 20 years' working experiences at an ICU, nine years' experiences of human patient simulator team training and also a PhD student (author KJ). Second, a physician with 20 years' working experience, associate professor and

consultant in Anesthesiology and Critical Care Medicine, with 14 years' experiences of human patient simulator team training (author MHu). Third, a registered nurse with a Master's Degree (one year) in Nursing (Acute Care Medicine), with six years' working experiences of prehospital care and one year as a medical student. Fourth, a paramedic with limited experiences beyond two years working as a paramedic in the Israeli Army and one year as a medical student.

Setting

Data collection in this study was carried out during simulation-based team training at the Clinical Training Centre (CTC) of the Medical Faculty, Umeå University. The briefings, scenarios and debriefings were conducted in Swedish. Two cameras mounted at an angle were used to record videos in the simulation room and one of the views included the patient monitor.

One week before team training, the students were asked to watch a twelve-minute video available at the learning platform that introduced both the ABCDE concept when caring for a patient and the room and equipment to be used during simulation-based training.[30,31] At the start of the training session, the students participated in a 15-minute live introduction to the simulation laboratory presented by the operator and the instructor. Each student group (4–5 students/group) trained on four scenarios that focused on assessment and treatment of severely ill patients in an emergency room. The students all played the role of interns, i.e. physicians who are in training to become licensed, with the attending nurse currently unavailable. The assigned task involved conducting a primary survey and stabilising the patient until more senior staff arrived in 15 minutes. In each scenario, three students were active and 1–2 were observers. After each scenario, a 10–15 minute debriefing session permitted reflection and shared learning. The first scenario was a warm-up, and the last

included a training summary. Case A and B, or in the reverse order, as determined by the toss of a coin immediately before the second scenario, were used as the second and third scenarios. All scenarios were conducted in Swedish, were designed to last 10–15 minutes and were pre-programmed into a Laerdal SimMan simulator in order to support the standardisation of the simulation. In essence, the patient cases used were slightly modified versions of the scenarios used by Hogan.[22] Case A was hypovolemic shock following a traffic incident. Case B was a pneumothorax with affected vital signs following a traffic incident.

Background questionnaire

The background questionnaire included informed consent and was answered immediately before team training started. The questionnaire included questions such as year of birth, male/female, previous medical training, previous experiences of team training, previous experiences of human patient simulator-based training, previous experiences of CRM and previous experiences of live trauma care.

ABCDE checklist

In order to measure the completeness of critical tasks in acute care scenarios, an ABCDE checklist was used. The original TTET comprises 58 items derived from the ATLS protocol.[15] Each item in the TTET was scrutinised by the authors of this study and the number of items reduced to those that reflected the measures expected to be carried out by year four medical students in acute care scenarios. The final list contained ten items. The items are well described in major textbooks and were translated into Swedish in line with the nomenclature being used. Compared to the original scoring system, an additional scoring option was added: performed after a reminder from the instructor. Each item was rated on a Likert scale of 0 to 4 (0 = Not initiated, 1= Performed after a reminder from the instructor, 2 = Partially performed, 3 = Performed completely before the end of the simulation, and 4 = Performed consistently during the whole simulation, NA = not applicable).

TEAM instrument

In order to measure team leadership, teamwork and task management, the TEAM instrument was used unmodified, i.e. in English, as developed by Cooper et al.[16,17] The published internal consistency (Cronbach's alpha) was 0.97 and the interrater reliability was 0.55, as measured by Cohen's Kappa (adjusted for chance) with a mean intraclass correlation coefficient of 0.60 for the 11 items. The instrument comprises 11 critical behaviours rated on a Likert scale of 0 to 4 (0 = Never/Hardly, 1 = Seldom, 2 = About as often as not, 3 = Often, 4 = Always/Nearly always) and, finally, a Global rating of the team's overall performance on a scale of 1 to 10. The original publication with the TEAM instrument had no descriptors for the endpoints. For this study, we used 1 = Poor and 10 = Excellent.

The team instrument comprises three subscales: leadership (items 1–2), teamwork (items 3–9) and task management (items 10–11). An index was constructed as a mean score for the respective subscale and all 11 items in the instrument (mean score for TEAM). The global rating of the overall performance was analysed separately.

Procedure for rating the TEAM instrument and ABCDE checklist

The raters in this study used video recordings in the rating procedure of the TEAM scale and the ABCDE checklist. The raters held two separate two-hour long meetings to discuss the interpretation of the descriptors in the ABCDE checklist and the TEAM instrument. During the first meeting, the discussions were facilitated using a different set of videos with similar scenarios, but from other teams. A total of six videos were used for this. During the second meeting, after four of the scenarios had been rated by each participant, the raters met again to discuss the interpretation of scales.

The raters independently assessed the videos of the simulation scenarios included in this study and rated the performances using the ABCDE checklist and the TEAM instrument. Each video was viewed at least twice by each rater.

SAGAT

When developing SA items for the specific scenarios, goal-directed task analysis was used, as initially described by Endsley.[21] Briefly, for each profession, major goals are identified along with sub-goals. Critical decisions are then identified and SA requirements are defined as the dynamic information needed to achieve the major goals, as opposed to static information such as rules and guidelines. The samplings (items) are then matched against the SA requirements. The original recommendation by Endsley was 30–60 items for within-subject studies for each of the three SA levels: (1) perception and attention, (2) comprehension and (3) projection.[20] When Gardner et al. validated a questionnaire based on SAGAT in a study of medical trainees, each questionnaire comprised three items for each level of SA at each freeze [25], while Hogan et al. developed a questionnaire with three items at level 1, one item at level 2 and three items at level 3.[22]

In the present study, the SA questionnaire was refined and adapted to the scenario and expected skills level of the students according to the process used by Hogan et al. [22] First, the questionnaire was translated into Swedish by the authors of this study as a basis for developing scenario-specific items in Swedish. In accordance with SAGAT, targeted learning objectives for the training scenarios were formulated and then the specific goals for each simulation were set. An iterative process was used to reformulate the items in Swedish, using a separate group of six professionals, all registered nurses, working in both a clinical context and a teaching context. The final sets of SA items are shown in Tables 2A and 2B (author's translation to English). To determine whether an item is essential in a specific context, Lawshe advocated the use of professional assessments.[32]

A content validity index (CVI) was defined as the fraction of professionals who rate the item as important. In the present study, the relevance was reviewed by three professionals (nurse

n=1, physician n=2) before being used in the study. All professionals rated the questions as relevant, i.e. the CVI was 1.0.

The answers given by the participants in the SA questionnaire were classified as incorrect (0) or correct (1). The classifications were discussed and agreed upon by the two authors (KJ and MHu). For answers on a continuous scale (e.g. systolic blood pressure), a 10% range around the intended correct answer was accepted as correct. One question was removed from the questionnaire since it became apparent during the classification process that the question had frequently been misinterpreted.

In order to administer the SA questionnaires, the scenarios were frozen (i.e. paused) twice. The first freeze of the scenario was five minutes into the scenario, unless there was an active task activity or if the team was conducting a team re-evaluation, in which case, the freeze was briefly postponed. During the freeze, the patient monitor was switched off and the participants turned away from the patient simulator while individually answering the questions. The second freeze took place according to the same principles after an additional five minutes. All participants were allowed to complete the questionnaire before the scenario re-started.

The length of the freezes was measured from the video recordings (Table 3). The start of the freeze was defined as the beginning of the sentence "Now we will pause the scenario so that you can answer some questions about the patient case" and the end of the freeze was defined as the end of the sentence "Is everyone in place and ready? Now the scenario will re-start".

Team SAGAT (TSAGAT) was developed by Crozier et al. [29] based on SAGAT as an assessment tool for evaluating team performance. In Crozier's study, each team comprised a trauma leader, airway manager and nurse. Individual SA questionnaires were developed for the three-team roles including both shared knowledge and complementary knowledge. TSAGAT was calculated as the sum of individual SAGAT scores and the TSAGAT scores had a high correlation to a traditional checklist (Pearson correlation, r=0.996). However, Salas

defined team SA as a dynamic process defined as the team's shared understanding of a situation at a specific point in time [33] while Endsley argued that team SA involves unique activities as information sharing and coordination.[20] Thus, in the present study, in order to measure this efficiently in all team members, all participants in a case received identical SA questionnaires. TSAGAT was calculated based on the mean scores of the team for each SA item.

Post-simulation questionnaire

In order to measure whether the items in the SA questionnaire were considered relevant to the scenarios, and whether pausing the scenarios affected the team training activity, a post-simulation questionnaire (PSQ) was used.[22] The PSQ comprises 13 statements to be rated on a four-point scale ranging from Strongly disagree (1) to Strongly agree (4), in which five statements concern the SAGAT and the effect of freezing the scenario and eight statements concern the simulation and the scenario *per se*. To the best of our knowledge, no data regarding the reliability of PSQ were presented in the original study. In this study, we only use the five questions relating to SAGAT and freezing the scenario. The PSQ was translated into Swedish by one of the authors and the translation was further refined based on iterative discussions within our research group. Next, the Swedish PSQ was sent to a professional translation agency together with the original PSQ for verification of the translation. The PSQ was answered in an anonymous web survey during the second week after the simulation training by 24 of the 55 participating students, i.e. the response rate was 44% (Table 4). In this study, the results of the four-point scale were dichotomised into disagree (Strongly disagree and Disagree) and agree (Agree and Strongly agree).

Study size

The aim of the study was to evaluate the feasibility of ABCDE, TEAM and SAGAT for use in further studies. In order to achieve this, the study sample must be large enough to allow for,

with a fair precision, the calculation of descriptive statistics (means and standard deviations) and the calculation of reliabilities. In order to assess a large effect size (Cohen's d of 0.9) with t-test with a power of 80% at the 0.05 level, 16 participants per group would be needed as determined by G*Power.[34] Thus, for the aims of the present study, the inclusion of 50 individuals and 20 teams would suffice.

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM, Chicago, IL). Interrater reliability for the ABCDE checklist and the TEAM instrument was determined by intraclass correlation using a two-way random effects model (ICC (2,1) type absolute).[35,36] The results are reported both as single measures and average measures since the ICC for single measures relates to the reliability of the individual ratings, while average measures relates to the reliability of the mean values. Cronbach's alpha was used to measure the internal consistency of the SAGAT instrument and TSAGAT.[35] Internal consistency considered as the extent to which all items measure the same latent variable was investigated by χ^2 and as suggested by Schweizer, a normed χ^2 below 2 was taken as an indication of a good fit.[37]

Ethics

This study was approved by the Ethical Review Board of Northern Sweden (Chairman: Anders Iacobaeus, 7 April, 2016, decision no. 2016-54-31M).

Patient and Public Involvement

Neither patients nor the public were involved in the design or the data collection for this study.

Results

Scores on ABCDE checklist, TEAM instrument and SA items

The scores on the ABCDE checklist, TEAM instrument and the SA items are shown in Table 5. The mean ABCDE score was 2.6 on a scale of 0–4, the mean TEAM item 1–11 score was

2.3 on a scale of 0–4, the mean TEAM global rating (item 12) was 4.8 on a scale of 0–10 and the mean SA item score was 13 with a maximal score of 22.

Interrater reliability for the ABCDE checklist and TEAM instrument

Interrater reliability as measured by intraclass correlation was 0.55 (single measures)/0.83 (average measures) for the ABCDE checklist and 0.54/0.83 for the TEAM scale. For the TEAM subscales of leadership, team work, task management and global rating, the intraclass correlations were 0.36/0.70, 0.45/0.77, 0.35/0.68 and 0.38 /0.72, respectively.

The feasibility of using SAGAT to measure situation awareness was assessed by measuring the length of the freeze (i.e. pause) needed to answer questions and by asking the participants for their perception of the pauses in the scenario and how the pauses affected the training session.

Feasibility and internal consistency of SAGAT

The PSQ showed that 96% of the participants considered the SA items to be relevant to the case and 96% considered the questions to be easy to understand (Table 4). About three out of four (72%) participants stated that the freeze did not negatively impact their concentration or performance during the simulation session.

The internal consistency of SAGAT measured as Cronbach's alpha was 0.80 for case A and 0.62 for case B (Table 5), and normed χ^2 was 1.72 vs. 1.62. For level 1 (perception), Cronbach's alpha was low, 0.06 for case 1 and 0.25 for case 2, but for level 3 it was fair, 0.89 and 0.66, respectively.

For TSAGAT, the internal consistency was good for the instrument as a total and level 3 (Table 6).

Discussion

Research on how to maximise team performance depends on the availability of sensitive and reliable tools for measuring the impact of an intervention. This study aims to test the usability of an adapted ABCDE checklist, the TEAM instrument, and a questionnaire measuring SA developed according to SAGAT. The original instruments and techniques were in English while the intended use is for rating Swedish acute care teams performing in their native language and cultural context. The main finding of this study was that the developed ABCDE checklist and the TEAM instrument could be used with acceptable interrater reliability and that it was feasible to use SAGAT to measure SA. The scores were in the middle of the scales, indicating that the scales could purposefully be used in a future effect study. The combination of these three measurements could permit analysis of task performance, team performance and the relationship to situation awareness in team training.

Scaling down the comprehensive TTET[15] to a smaller ABCDE checklist could introduce unintended errors in measurements. In order be sensitive to the intended use, the items measured must be relevant to the task and the anchors of the scale used must be calibrated to the setting in order to allow for significant changes to be reflected as the difference between control and intervention.[35] In the present study, the developed ABCDE checklist was perceived by professionals to be relevant to the case. Interrater variability was low, indicating that the checklist could be reliably used for scoring task performance. The means of the scores were in the middle of the scale, which indicates that the checklist might be sensitive to differentiating between low and high performers.

For the TEAM instrument, the interrater reliability was 0.55 for single measurements and 0.83 for average measurements. According to Koo and Li, an ICC between 0.50 and 0.75 indicates moderate reliability, while an ICC between 0.75 and 0.90 indicates good reliability.[36] The ICC for single measurements should be reported if the plan is to use the measurements from

one rater only, and ICC for average measurements if the plan is to use multiple raters. In this study, we chose to report both values in order to show the importance of using multiple raters to improve overall reliability. In the original study by Cooper, an interrater reliability for the TEAM instrument of 0.55 and 0.60 was reported as measured by Cohen's kappa (adjusted for chance) and ICC, respectively.[16] Thus, interrater reliability as measured by ICC (single measurements) in this study was equivalent to the ICC obtained in the original setting.

Both the PSQ and the CVI indicate that the SAGAT instrument could be used to construct questions that were considered relevant to the case. In any case, the internal consistency was fair (0.80 and 0.61). When analysing the subscales, levels 1, 2 and 3, the internal consistency was low for levels 1 and 2, and higher for level 3. This could indicate that the perception of the situation in the groups was diverse and not related to the total score, while the ability to project the direction in which the cases were heading was more homogeneously related to the total score. TSAGAT had a higher overall homogeneity, as might be expected when analysing the means of the group SAGAT for each question, instead of the specific SAGAT answers.

Measuring SA using SAGAT requires that the scenario is paused while the participants answer questions that probe the three levels of SA: perception, comprehension and projection.[21] Each pause in this study lasted less than 3 minutes. The pauses and SAGAT questions could influence SA in both a negative direction and a positive direction. In the negative direction, the flow of the simulation training might be disrupted, stress be induced by being forced to interrupt the case, and the commitment decrease. In the positive direction, the pauses and questions might facilitate the resolution of the clinical problem in the case by triggering and allowing time for reflection. This could be in line with Schön's reflection-on-action and reflection-in-action.[24] According to the PSQ, the pauses in this study did not cause any adverse reactions on the part of the participants.

Limitations of this study

This study concerns the use of instruments and techniques for measuring performance and SA. Team performance depends on the skills of individuals and the ability to work together as a team. To achieve this, both technical and non-technical skills are necessary. The skills are more or less sensitive to context and cultural aspects, for example, leadership is negotiated within the team, and the pattern of closed-loop communication correlates to the leadership strategies being used.[38,39]

The ABCDE checklist was developed based on TTET which, in turn, is based on ATLS.[15] The items in TTET are based on well-known concepts in trauma care, and the ATLS concept has been implemented globally.[31] The items were developed in Swedish. However, as nomenclature in the trauma field is well established because of the ATLS concept, the items and expected skills were strikingly similar to TTET. Difficulties were encountered in the interpretation of what constitutes a completed versus a partially completed item. This might relate to the vast differences in the raters' level of education and previous experience. For this, the raters were training using a separate set of video recordings to improve the accuracy and minimize the variability.

TEAM was not translated into Swedish in order to avoid inducing errors. This is primarily because fluency in English is a requisite for academic studies in Sweden. The ratings based on the original TEAM instrument were consistent between the raters, indicating that this may have been a correct assumption.

The developed set of SAGAT items contained only a few items for measuring each level of SA. The selected number of items were close to what was used by Gardner and Hogan.[22,25] Too few items increase measurement errors, while too many items result in extended freezes. However, in the present study, the probing of level 2 in particular would have benefited by 2–3 more items per measurement.

In the present study, CVI was performed for SAGAT while the development of the ABCDE checklist and the two scenarios relied on the authors' experiences and iterative interactions with clinicians and experts in the field. Both the scenarios, the ABCDE checklists and the developed SAGAT questionnaires, could have benefitted from a full formal CVI by a review panel.

The PSQ was only translated from English and not translated back into Swedish in order to formally check the identity of the items in the final sets. However, the translation was adjusted by a professional translator before being used in the study. Thus, the results of the PSQ can be used for probing the participants in the simulation with regards to their experiences of SAGAT.

The participants in the study comprised year four medical students for the video recordings of simulation-based team-training and the rating of the video material was conducted by four participants representing a wide spectrum of experience and training. It could be argued that it is a limitation that this validation was not performed on a series of critical care teams, for example. However, for testing the reliability as well as the feasibility the techniques and instruments, the simulation-based trainings with medical students was a readily available series of standardized simulations.

The raters in this study included both experienced and inexperienced raters. The study was not designed to systematically test for differences in ratings between raters of different backgrounds. However, the inhomogeneity of the raters is a strength as it indicates that the checklists and instruments can be used by a group of raters with different experiences.

Conclusion

The developed ABCDE checklist and the TEAM instrument had high interrater reliability, the process of using SAGAT during simulation-based training did not affect the participants'

perception of simulation-based training and the developed SAGAT questionnaires had a fair internal consistency. Thus, the measurement of task performance, team performance and situation awareness in future studies in a Swedish simulation-based training setting may be conducted using these techniques.

Table 1 Background characteristics of the medical students (n=55) participating in the study

Age years, m (sd)	25 (4.5)
Male n (%)	28 (51)
Female n (%)	27 (49)
Previous healthcare education	
None n (%)	44 (80)
Assistant nurse (2Y upper secondary school) n (%)	2 (4)
Registered nurse (3Y university) n (%)	3 (5)
Short courses (Red Cross, etc.) n (%)	6 (11)
Previous team training	
Yes n (%)	20 (36)
Previous simulation experience	
Yes n (%)	32 (58)
Previous experience of trauma patients*	
Yes n (%)	20 (37)

* one missing value

Table 2A Situation Awareness items for Case A

Freeze 1	Question	Response scale	Level
1.	What is the patient's respiratory rate?	Free text	1
2.	What are the findings from the assessment of A and B?	Free text	1
3.	Is the patient sufficiently oxygenated?	Yes/No	2
4.	What did the patient state as their year of birth?	Free text	1
5.	What do you think will happen to the patient's blood pressure in the next few minutes?	Incr*/Decr**/No change	3
6.	What is wrong with the patient (preliminary diagnosis)?	Free text	2
7.	How many peripheral IV cannulas does the patient have?	Free text	1
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?		
8.	Heart rate	Increase/Decrease	3
9.	Blood pressure	Increase/Decrease	3
10.	Saturation	Increase/Decrease	3
11.	Respiratory rate	Increase/Decrease	3

Freeze 2	Question	Response scale	Level
12.	What is the patient's blood pressure?	Free text	1
13.	What are your findings under C (Circulation)?	Free text	1
14.	What additional examinations/tests do you think are needed?	Free text	3
15.	Which medications/drugs do you think are needed?	Free text	3
16.	What is wrong with the patient (preliminary diagnosis)?	Free text	2
17.	Is there an algorithm for CPR on the wall?	Yes/No	1
18.	What previous illnesses has the patient described?	Free text	2
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?		
19.	Heart rate	Increase/Decrease	3
20.	Blood pressure	Increase/Decrease	3
21.	Saturation	Increase/Decrease	3
22.	Respiratory rate	Increase/Decrease	3

Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

* Incr = Increase

** Decr = Decrease

Table 2B Situation Awareness items for Case B

Freeze 1	Question	Response scale	Level
1.	What is the patient's respiratory rate?	Free text	1
2.	What are the findings from the assessment of A and B?	Free text	1
3.	Is the patient sufficiently oxygenated?	Yes/No	2
4.	What does the sign on the patient's necklace indicate?	Free text	1
5.	What do you believe will happen to the patient's venous return in the next few minutes?	Incr*/Decr**/No change	3
6.	What is wrong with the patient (preliminary diagnosis)?	Free text	2
7.	Is a suction device available?	Yes/No	1
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?		
8.	Heart rate	Increase/Decrease	3
9.	Blood pressure	Increase/Decrease	3
10.	Saturation	Increase/Decrease	3
11.	Respiratory rate	Increase/Decrease	3

Freeze 2	Question	Response scale	Level
12.	What is the patient's respiratory rate?	Free text	1
13.	What are your findings under C (Circulation)?	Free text	1
14.	What additional examinations/tests do you think are needed?	Free text	3
15.	Which medications/drugs do you think are needed?	Free text	3
16.	What is wrong with the patient (preliminary diagnosis)?	Free text	2
17.	What is the patient's blood pressure?	Free text	1
18.	Is a defibrillator available in the room?	Yes/No	2
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?		
19.	Heart rate	Increase/Decrease	3
20.	Blood pressure	Increase/Decrease	3
21.	Saturation	Increase/Decrease	3
22.	Respiratory rate	Increase/Decrease	3

Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

* Incr = Increase

** Decr = Decrease

The questions are shown translated into English in the table and the original questions in Swedish are available upon request from the authors of the study.

Table 3 Time (minutes) for scenario freezes to measure situation awareness with SAGAT

Scenario	No. of teams n	Scenario m (sd)	1st freeze m (sd)	2nd freeze m (sd)	Total time m (sd)
Case A	12	14.1 (2.0)	2.7 (0.5)	2.5 (0.6)	19.3 (2.3)
Case B	11	14.1 (3.4)	2.7 (0.6)	2.8 (0.6)	19.7 (3.8)

Table 4 Student agreement about usefulness and feasibility of the prospective collection of situation awareness items

Question	Agree	Disagree	Do not know
Concerning the SAGAT questionnaire. In my opinion:	**n (%)**	**n (%)**	**n (%)**
The introduction was adequate	24 (100)	0 (0)	0 (0)
The questions were clear	23 (96)	1 (4)	0 (0)
The "freezes" in the scenario adversely affected my concentration level and my performance	7 (29)	17 (71)	0 (0)
The questions were relevant to the way I perceived the scenario	23 (96)	1 (4)	0 (0)
Overall, I am satisfied with SAGAT as a tool for evaluating hands-on skills during a trauma exercise.	19 (79)	1 (4)	4 (17)

Results from the post-simulation questionnaire

Table 5 Mean scores on ABCDE checklist, TEAM instrument and SAGAT questionnaire

	Case A m (sd)	Case B m (sd)
ABCDE checklist*	2.70 (0.19)	2.58 (0.27)
TEAM instrument*		
Item 1–11	2.32 (0.54)	2.27 (0.46)
Global rating	4.90 (1.13)	4.75 (1.04)
SAGAT questionnaire**	13.95 (4.26)	12.60 (3.35)

*Mean score per item (sd). Max score per item 4.

**Mean score (sd). Max score 22.

Table 6 Internal consistency (Cronbach's alpha) of SAGAT and TSAGAT

SAGAT/TSAGAT	Case A n/n 39/12	Case B n/n 36/11
1	0.060/0.372	0.246/0.420
2	0.321/-0.018	0.332/0.332
3	0.891/0.891	0.659/0.786
Total (Level 1–3)	0.800/0.827	0.620/0.759

References

1 Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ* 2016;**353**:i2139. doi:10.1136/bmj.i2139

2 Kohn LT, Corrigan JM, Donaldson MS. *To Err is Human: Building a Safer Health System*. National Academies Press 2000. doi:10.17226/9728

3 World Health Organization. Multi-professional Patient Safety Curriculum Guide. WHO, Geneva. 2011. https://apps.who.int/iris/bitstream/handle/10665/44641/9789241501958_eng.pdf

4 Scalese RJ, Obeso VT, Issenberg BS. Simulation Technology for Skills Training and Competency Assessment in Medical Education. *J Gen Intern Med* 2007;**23**:46–9. doi:10.1007/s11606-007-0283-4

5 Zendejas B, Brydges R, Wang AT, *et al.* Patient outcomes in simulation-based medical education: a systematic review. *J Gen Intern Med* 2013;**28**:1078–89.

6 Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. *Med Educ* 2012;**46**:1042–54. doi:10.1111/j.1365-2923.2012.04343.x

7 Salas E, Rosen MA, Held JD, *et al.* Performance Measurement in Simulation-Based Training. *Simulat Gaming* 2009;**40**:328–76. doi:10.1177/1046878108326734

8 Rosen MA, Pronovost PJ. Advancing the Use of Checklists for Evaluating Performance in Health Care. *Acad Med* 2014;**89**:963–5. doi:10.1097/acm.0000000000000285

9 Brady PW, Muething S, Kotagal U, *et al.* Improving situation awareness to reduce unrecognized clinical deterioration and serious safety events. *Pediatrics* 2013;**131**:e298-308. doi:10.1542/peds.2012-1364

10 Salas E, Cooke NJ, Rosen MA. On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Hum Factors* 2008;**50**:540–7. doi:10.1518/001872008x288457

11 Kozlowski SW. Enhancing the Effectiveness of Work Groups and Teams: A Reflection. *Perspect*

Psychol Sci 2018;**13**:205–12. doi:10.1177/1745691617697078

12 Converse SA, Cannon-Bowers JA, Salas E. Team Member Shared Mental Models: A Theory and Some Methodological Issues. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1991;**35**:1417–21. doi:10.1177/154193129103501917

13 Mathieu J, Heffner T, Goodwin G, *et al.* The influence of shared mental models on team process and performance. *J Appl Psych* 2000;**85**:273–83.

14 Boyle WA, Murray DJ, Beyatte M, *et al.* Simulation-Based Assessment of Critical Care "Front-Line" Providers*. *Crit Care Med* 2018;**46**:e516–22. doi:10.1097/CCM.0000000000003073

15 Holcomb JB, Dumire RD, Crommett JW, *et al.* Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. *J Trauma* 2002;**52**:1078–85

16 Cooper S, Cant R, Porter J, *et al.* Rating medical emergency teamwork performance: Development of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;**81**:446–52. doi:10.1016/j.resuscitation.2009.11.027

17 Cooper SJ, Cant RP. Measuring non-technical skills of medical emergency teams: An update on the validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2014;**85**:31–3. doi:10.1016/j.resuscitation.2013.08.276

18 Cooper S, Cant R, Connell C, *et al.* Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation* 2016;**101**:97–101. doi:10.1016/j.resuscitation.2016.01.026

19 Schulz CM, Krautheim V, Hackemann A, *et al.* Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. *BMC Anesthesiol* 2016;**16**:4. doi:10.1186/s12871-016-0172-7

20 Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum Factors* 1995;**37**:32–64. doi:10.1518/001872095779049543

21 Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. In: *Situation awareness analysis and measuement*. Lawrence Erlbaum Associates 2000. 1–21.

22 Hogan MP, Pace DE, Hapgood J, *et al.* Use of human patient simulation and the situation awareness global assessment technique in practical trauma skills assessment. *J Trauma* 2006;**61**:1047–52. doi:10.1097/01.ta.0000238687.23622.89

23 Hamstra SJ, Brydges R, Hatala R, *et al.* Reconsidering fidelity in simulation-based training. *Acad Med* 2014;**89**:387–92. doi:10.1097/ACM.0000000000000130

24 Schön DA. *The reflective practitioner: How professionals think in action*. The Perseus Books Group 1984.

25 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc* 2017;**12**:17–21. doi:10.1097/sih.0000000000000181

26 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident's Situational Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3. doi:10.1097/BOT.0000000000000003

27 Lavoie P, Cossette S, Pepin J. Testing nursing students' clinical judgment in a patient deterioration simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today* 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015

28 McKenna L, Missen K, Cooper S, *et al.* Situation awareness in undergraduate nursing students managing simulated patient deterioration. *Nurs Educ Today* 2014;**34**:e27–31. doi:10.1016/j.nedt.2013.12.013

29 Crozier MS, Ting HY, Boone DC, *et al.* Use of human patient simulation and validation of the Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009

30 Thim T, Krarup, Grove, *et al.* Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012;:117. doi:10.2147/IJGM.S28478

31 ATLS Advanced Trauma Life Support Student Course Manual. 2018.

32 Lawshe CH. A quantitative approach to content validity. *Personnel Psychology* 1975;**28**:563–75. doi:10.1111/j.1744-6570.1975.tb01393.x

33 Salas E, Prince C, Baker DP, *et al.* Situation Awareness in Team Performance: Implications for Measurement and Training. *Hum Factors* 1995;**37**:123–36. doi:10.1518/001872095779049525

34 Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers* 1996;**28**:1–11. doi:10.3758/BF03203630

35 Nunnally JC. *Psychometric Theory, 2nd edition*. McGraw-Hill 1978.

36 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *J Chiropr Med* 2016;**15**:155–63. doi:10.1016/j.jcm.2016.02.012

37 Schweizer K. Some guidelines concerning the modeling of traits and abilities in test construction. *Eur J Psychol Assess* 2010;**26**. doi:10.1027/1015-5759/a000001

38 Jacobsson M, Härgestam M, Hultin M, *et al.* Flexible knowledge repertoires: communication by leaders in trauma teams. *Scand J Trauma Resusc Emerg Medicine* 2012;**20**:44. doi:10.1186/1757-7241-20-44

39 Härgestam M, Lindkvist M, Brulin C, *et al.* Communication in interdisciplinary teams: exploring closed-loop communication during in situ trauma team training. *BMJ Open* 2013;**3**:e003525. doi:10.1136/bmjopen-2013-003525

STROBE Statement—checklist of items that should be included in reports of observational studies

	Item No	Recommendation	Page No
Title and abstract	1	(*a*) Indicate the study's design with a commonly used term in the title or the abstract	1
		(*b*) Provide in the abstract an informative and balanced summary of what was done and what was found	3
Introduction			
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported	7-9
Objectives	3	State specific objectives, including any prespecified hypotheses	9
Methods			
Study design	4	Present key elements of study design early in the paper	9-10
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection	9-10
Participants	6	(*a*) *Cohort study*—Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up *Case-control study*—Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls *Cross-sectional study*—Give the eligibility criteria, and the sources and methods of selection of participants	9-10
		(*b*) *Cohort study*—For matched studies, give matching criteria and number of exposed and unexposed *Case-control study*—For matched studies, give matching criteria and the number of controls per case	-
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable	11-15
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group	11-15
Bias	9	Describe any efforts to address potential sources of bias	11-15
Study size	10	Explain how the study size was arrived at	15
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why	16
Statistical methods	12	(*a*) Describe all statistical methods, including those used to control for confounding	16
		(*b*) Describe any methods used to examine subgroups and interactions	16
		(*c*) Explain how missing data were addressed	15
		(*d*) *Cohort study*—If applicable, explain how loss to follow-up was addressed *Case-control study*—If applicable, explain how matching of cases and controls was addressed *Cross-sectional study*—If applicable, describe analytical methods taking account of sampling strategy	-
		(*e*) Describe any sensitivity analyses	

Continued on next page

Results			
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed	9-10,15
		(b) Give reasons for non-participation at each stage	9,15
		(c) Consider use of a flow diagram	-
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders	22
		(b) Indicate number of participants with missing data for each variable of interest	In tables
		(c) *Cohort study*—Summarise follow-up time (eg, average and total amount)	-
Outcome data	15*	*Cohort study*—Report numbers of outcome events or summary measures over time	-
		Case-control study—Report numbers in each exposure category, or summary measures of exposure	-
		Cross-sectional study—Report numbers of outcome events or summary measures	16-17
Main results	16	(*a*) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included	16-17
		(*b*) Report category boundaries when continuous variables were categorized	NA
		(*c*) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period	NA
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses	17
Discussion			
Key results	18	Summarise key results with reference to study objectives	18-19
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias	20-21
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence	21-22
Generalisability	21	Discuss the generalisability (external validity) of the study results	21-22
Other information			
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based	6

*Give information separately for cases and controls in case-control studies and, if applicable, for exposed and unexposed groups in cohort and cross-sectional studies.

Note: An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at http://www.plosmedicine.org/, Annals of Internal Medicine at http://www.annals.org/, and Epidemiology at http://www.epidem.com/). Information on the STROBE Initiative is available at www.strobe-statement.org.

BMJ Open

The reliability of instruments that measure situation awareness, team and task performance in a simulation setting with medical students

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029412.R2
Article Type:	Research
Date Submitted by the Author:	05-Jul-2019
Complete List of Authors:	Hultin, Magnus; Umeå Universitet Medicinska fakulteten, Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine Jonsson, Karin; Umeå Universitet Medicinska fakulteten, Departments of Surgical and Perioperative Sciences and Nursing Härgestam, Maria; Umea Universitet Medicinska fakulteten, Department of Nursing Lindkvist, Marie; Umeå Universitet Medicinska fakulteten, Department of Public Health and Clinical Medicine Brulin, Christine; Umeå Universitet Medicinska fakulteten, Department of Nursing
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Communication
Keywords:	Patient care team, Resuscitation, Situation Awareness, Task performance and analysis, Teamwork

The reliability of instruments that measure situation awareness, team and task performance in a simulation setting with medical students

Magnus Hultin, MD Ph.D., Associate Professor, Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0003-2935-7161

Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0001-7974-6777

Maria Härgestam, RNA Ph.D., Lecturer, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0003-0680-9962

Marie Lindkvist, Ph.D., Associate Professor, Department of Epidemiology and Global Health, Umeå University, S-901 87 Umeå, Sweden

Christine Brulin, RNA Ph.D., Professor emerita, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden

Abstract 246 words

Main article 4,837 words

Keywords

- Patient Care Team
- Resuscitation

- Situation Awareness
- Task Performance and Analysis
- Teamwork

Corresponding author

Magnus Hultin, Dept of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, Sweden, magnus.hultin@umu.se, phone: +46-90-786 9109

Abstract

Objectives: The assessment of situation awareness, team and task performance in a simulation training session requires reliable and feasible measurement techniques. The objectives of this study were to test an Airways-Breathing-Circulation-Disability-Exposure (ABCDE) checklist and Team Emergency Assessment Measure (TEAM) for interrater reliability, as well as an application of Situation Awareness Global Assessment Technique (SAGAT) for feasibility and internal consistency.

Design: Methodological approach

Setting: Data collection during team training using full-scale simulation at a university clinical training centre. The video-recorded scenarios were rated independently by four raters.

Participants: 55 medical students aged 22–40 years in their 4th year of medical studies, during the clerkship in anaesthesiology and critical care medicine, formed 23 different teams. All students answered the SAGAT questionnaires and, of these students, 24 answered the follow-up post-simulation questionnaire (PSQ). TEAM and ABCDE were scored by four professionals.

Measures: ABCDE and TEAM were tested for interrater reliability. The feasibility of SAGAT was tested using PSQ. SAGAT was tested for internal consistency both at an individual level (SAGAT) and a team level (TSAGAT).

Results: The intraclass correlation was 0.54/0.83 (single/average measurements) for TEAM and 0.55/0.83 for ABCDE. According to the PSQ, the items in SAGAT were rated as relevant to the scenario by 96% of the participants. Cronbach's alpha for SAGAT/TSAGAT for the two scenarios was 0.80/0.83 vs. 0.62/0.76, and normed χ^2 was 1.72 vs. 1.62.

Conclusions: Task performance, team performance and situation awareness could be purposefully measured, and the reliability of the measurements was good.

Article Summary

Strengths and limitations of this study

- SAGAT could be used to create items in Swedish to probe situation awareness, i.e. in the participants' native language
- TEAM could be used by the raters in its original language (English).
- The developed ABCDE checklist has items that are well-defined concepts and the difficulty lies in defining the rubrics for scoring the items.
- It is a weakness that PSQ was only translated into Swedish and not retranslated back into English.
- An interprofessional set of raters with different backgrounds and experiences rated TEAM and ABCDE with similar results.

Funding

This work was supported by Region Västerbotten, grant no. VLL-663801 and VLL-836931, by Region Norrbotten, grant no. NLL-765981, by Alice Lindström's Foundation, and by the Medical Faculty of Umeå University.

Competing interests

None of the authors have any competing interests to declare.

Data availability

The data comprising de-identified questionnaire responses and ratings of video scenarios are available upon request from the corresponding author.

Author Contributions

MHu, KJ, MHä and CB designed the study. MHu and KJ prepared the initial draft of the paper. MHu, KJ, MHä, ML and CB were all actively engaged in data analysis and in finalising the paper. All authors have read and approved the final version of the paper.

Introduction

Medical errors are the third leading cause of death in the US.[1] Knowledge about the relationship between human errors and patient safety has increased in the last two decades.[2,3] Simulated environments make it possible to improve skills by employing training strategies to prevent errors while simultaneously offering an arena for reliable assessments of skills.[4] Thus, simulation training is often used by organisations to minimise adverse events and prevent healthcare errors.[5,6] This could be accomplished by improving *task performance*, *team performance* or *situation awareness (SA)*.[7–9]

When developing strategies for improving clinical practice, it is essential to evaluate both task and team performance. According to Salas[10] and Kozlowski[11], team performance is a multilevel process that includes the interrelation between individual- and team-level taskwork and teamwork processes. Thus, an optimal task, as well as team performance, depends on the coordinated activities of a team of individuals.[12,13] Checklists are often used to score task performance in acute care scenarios. The lists might include adherence to resuscitation protocols, timing of the task, as well as the time taken to complete the components.[14] The Trauma Team Evaluation Tool (TTET) was developed by Holcomb for trauma scenarios managed according to Airways-Breathing-Circulation-Disabilities-Exposure (ABCDE) protocols based on Advanced Trauma Life Support (ATLS) and was tested on US military resuscitation teams from community hospitals.[15] As a pilot study, psychometric data such as reliability and validity were not reported. As TTET was developed for a specific setting, the items and criteria for scoring has to be adapted to the proficiency levels of the participants and the standard operating procedures being used.

Team performance can be measured using the Team Emergency Assessment Measure (TEAM), for example. It measures three dimensions of team performance: leadership, teamwork and task management.[16,17] The instrument was initially developed and validated for simulator-based team training and was recently validated for the collection of observational ratings of non-technical skills during live resuscitations in emergency departments.[18] Using instruments to score scenarios depend on reliable interpretations of the instrument by the raters and it might be even more important to ensure the reliability of such interpretations when using an instrument in a non-native language.

Moreover, and in addition to task and team performance, SA is a prerequisite for patient safety and the prevention of errors, particularly in acute care situations.[9,19] SA includes three levels of ability: (1) perception and attention (*What?*, what's going on), (2) comprehension (*So what?*, the ability to understand what's going on), and (3) projection (*What's next?*) in order to anticipate and plan for future events.[20] In order to measure SA in a simulation setting, the Situation Awareness Global Assessment Tool (SAGAT) has been developed and subsequently adapted for use in healthcare settings.[21,22] One feature of SAGAT is that its use requires the simulation to pause[21], which might influence clinical understanding, both by impeding the suspension of disbelief in the simulation setting [23] and by facilitating reflection on action.[24] Gardner showed that it was feasible to use SAGAT to measure SA in the team training of surgical trainees in advanced cardiac life support.[25] Globally, SAGAT has been used to study, for example, the effect of sleep deprivation on SA in trauma team training, how SA is associated with surgical trainee team performance, as well as nurses' clinical assessment of patient deterioration.[26–28] SAGAT can be analysed on an individual level and also on a team level. A specific application of SAGAT is team SAGAT (TSAGAT) in which the different members of a team answer SAGAT questions specific to their role. TSAGAT was constructed to account for the teams' shared situation awareness and

validated in trauma team training in a Canadian setting.[29] To our knowledge, SAGAT has not previously been used in a Swedish context and it is therefore important to evaluate both its feasibility and its internal consistency.

To summarise, ABCDE checklists, the TEAM instrument and SAGAT have been developed in order to evaluate different aspects of teamwork. However, these instruments and questionnaires have neither been translated into Swedish nor tested for feasibility or trustworthiness in a Swedish context. Such studies are necessary to enable the evaluation of teamwork in Swedish acute care settings and simulation-based training. Thus, this study aimed to test an ABCDE checklist that we developed for interrater reliability, to test the TEAM instrument for interrater reliability, and to test two Swedish SAGAT questionnaires for feasibility and internal consistency.

Method

This study is based on data collected during simulation-based team training sessions with medical students.

Participants and raters

From March to October 2016, all medical students (n=68) in year 4 undertaking their clerkship in anaesthesiology and critical care medicine were invited to participate in the study while receiving mandatory simulator-based team training. In total, 55 students (81%) participated in the study (Table 1) and 20 of them participated in both scenario A and scenario B.

All scenarios were videorecorded and later scored by four raters to allow for calculation of interrater reliability. First, a registered nurse with a Master's Degree (one year) in Nursing (Critical Care Medicine), 20 years' working experiences at an ICU, nine years' experiences of

human patient simulator team training and also a PhD student (author KJ). Second, a physician with 20 years' working experience, associate professor and consultant in Anesthesiology and Critical Care Medicine, with 14 years' experiences of human patient simulator team training (author MHu). Third, a registered nurse with a Master's Degree (one year) in Nursing (Acute Care Medicine), with six years' working experiences of prehospital care and one year as a medical student. Fourth, a paramedic with limited experiences beyond two years working as a paramedic in the Israeli Army and one year as a medical student.

Setting

Data collection in this study was carried out during simulation-based team training at the Clinical Training Centre (CTC) of the Medical Faculty, Umeå University. The briefings, scenarios and debriefings were conducted in Swedish. Two cameras mounted at an angle were used to record videos in the simulation room and one of the views included the patient monitor.

One week before team training, the students were asked to watch a twelve-minute video available at the learning platform that introduced both the ABCDE concept when caring for a patient and the room and equipment to be used during simulation-based training.[30,31] At the start of the training session, the students participated in a 15-minute live introduction to the simulation laboratory presented by the operator and the instructor. Each student group (4–5 students/group) trained on four scenarios that focused on assessment and treatment of severely ill patients in an emergency room. In each scenario, 3-4 students were active and 1–2 were observers. Therefore, only 3-4 participated in the interactive role play in each scenario (the actual simulation). In all, the 55 unique students made up a total of 23 teams with 3-4 participants in each team and 20 students participated in both case A and case B. The students all played the role of interns, i.e. physicians who are in training to become licensed, with the attending nurse currently unavailable. The assigned task involved conducting a primary

survey and stabilising the patient until more senior staff arrived in 15 minutes. After each scenario, a 10–15 minute debriefing session permitted reflection and shared learning. The first scenario was a warm-up, and the last included a training summary. Case A and B, or in the reverse order, as determined by the toss of a coin immediately before the second scenario, were used as the second and third scenarios.

All scenarios were conducted in Swedish, were designed to last 10–15 minutes and were pre-programmed into a Laerdal SimMan simulator in order to support the standardisation of the simulation. In essence, the patient cases used were slightly modified versions of the scenarios used by Hogan.[22] Case A was hypovolemic shock following a traffic incident. Case B was a pneumothorax with affected vital signs following a traffic incident.

Background questionnaire

The background questionnaire included informed consent and was answered immediately before team training started. The questionnaire included questions such as year of birth, male/female, previous medical training, previous experiences of team training, previous experiences of human patient simulator-based training, previous experiences of CRM and previous experiences of live trauma care.

ABCDE checklist

In order to measure the completeness of critical tasks in acute care scenarios, an ABCDE checklist was used. The original TTET comprised 58 items derived from the ATLS protocol.[15] Each item in the TTET was scrutinised by the authors of this study and the number of items reduced to those that reflected the measures expected to be carried out by year four medical students in the acute care scenarios in this study. Some items were also slightly modified to better reflect The final list contained ten items reflecting the key elements of the ATLS primary survey, i.e. the core management of ABCDE: airway assessed,

airway secured, saturation assessed, oxygen applied, ventilation assessed, ventilation optimised, pulses checked (radial – femoral – carotid), venous access and infusions established, neurological disabilities checked (consciousness, pupils), and full/complete exposure. The items are well described in major textbooks and were translated into Swedish in line with the nomenclature being used. Compared to the original scoring system, an additional scoring option was added: performed after a reminder from the instructor. Each item was rated on a rating scale of 0 to 4 (0 = Not initiated, 1= Performed after a reminder from the instructor, 2 = Partially performed, 3 = Performed completely before the end of the simulation, and 4 = Performed consistently during the whole simulation, NA = not applicable).

TEAM instrument

In order to measure team leadership, teamwork and task management, the TEAM instrument was used unmodified, i.e. in English, as developed by Cooper et al.[16,17] The published internal consistency (Cronbach's alpha) was 0.97 and the interrater reliability was 0.55, as measured by Cohen's Kappa (adjusted for chance) with a mean intraclass correlation coefficient of 0.60 for the 11 items. The instrument comprises 11 critical behaviours rated on a Likert scale of 0 to 4 (0 = Never/Hardly, 1 = Seldom, 2 = About as often as not, 3 = Often, 4 = Always/Nearly always) and, finally, a Global rating of the team's overall performance on a scale of 1 to 10. The original publication with the TEAM instrument had no descriptors for the endpoints. For this study, we used 1 = Poor and 10 = Excellent.

The team instrument comprises three subscales: leadership (items 1–2), teamwork (items 3–9) and task management (items 10–11). An index was constructed as a mean score for the respective subscale and all 11 items in the instrument (mean score for TEAM). The global rating of the overall performance was analysed separately.

Procedure for rating the TEAM instrument and ABCDE checklist

The raters (n=4) in this study used video recordings in the rating procedure of the TEAM scale and the ABCDE checklist. The raters held two separate two-hour long meetings to discuss the interpretation of the descriptors in the ABCDE checklist and the TEAM instrument. During the first meeting, the discussions were facilitated using a different set of videos with similar scenarios, but from other teams. A total of six videos were used for this. During the second meeting, after four of the scenarios had been rated by each participant, the raters met again to discuss the interpretation of scales.

The raters independently assessed the videos of the simulation scenarios included in this study and rated the performances using the ABCDE checklist and the TEAM instrument. Each video was viewed at least twice by each rater.

SAGAT

When developing SA items for the specific scenarios, goal-directed task analysis was used, as initially described by Endsley.[21] Briefly, for each profession, major goals are identified along with sub-goals. Critical decisions are then identified and SA requirements are defined as the dynamic information needed to achieve the major goals, as opposed to static information such as rules and guidelines. The samplings (items) are then matched against the SA requirements. The original recommendation by Endsley was 30–60 items for within-subject studies for each of the three SA levels: (1) perception and attention, (2) comprehension and (3) projection.[20] When Gardner et al. validated a questionnaire based on SAGAT in a study of medical trainees, each questionnaire comprised three items for each level of SA at each freeze [25], while Hogan et al. developed a questionnaire with three items at level 1, one item at level 2 and three items at level 3.[22]

In the present study, the SA questionnaire was refined and adapted to the scenario and expected skills level of the students according to the process used by Hogan et al. [22] First,

the questionnaire was translated into Swedish by the authors of this study as a basis for developing scenario-specific items in Swedish. In accordance with SAGAT, targeted learning objectives for the training scenarios were formulated and then the specific goals for each simulation were set. An iterative process was used to reformulate the items in Swedish, using a separate group of six professionals, all registered nurses, working in both a clinical context and a teaching context. The final sets of SA items are shown in Tables 2A and 2B (author's translation to English) with 11 items in each freeze. To determine whether an item is essential in a specific context, Lawshe advocated the use of professional assessments such as a content validity index (CVI) defined as the fraction of professionals who rate the item as important.[32] In the present study, the relevance was reviewed by three professionals (nurse n=1, physician n=2) before being used in the study. All professionals rated the questions as relevant, i.e. the CVI was 1.0.

The answers given by the participants in the SA questionnaire were classified as incorrect (0) or correct (1). The classifications were discussed and agreed upon by the two authors (KJ and MHu). For answers on a continuous scale (e.g. systolic blood pressure), a 10% range around the intended correct answer was accepted as correct. One question was removed from the questionnaire since it became apparent during the classification process that the question had frequently been misinterpreted.

In order to administer the SA questionnaires, the scenarios were frozen (i.e. paused) twice. The first freeze of the scenario was five minutes into the scenario, unless there was an active task activity or if the team was conducting a team re-evaluation, in which case, the freeze was briefly postponed. During the freeze, the patient monitor was switched off and the participants turned away from the patient simulator while individually answering the questions. The second freeze took place according to the same principles after an additional five minutes. All participants were allowed to complete the questionnaire before the scenario re-started.

The length of the freezes was measured from the video recordings (Table 3). The start of the freeze was defined as the beginning of the sentence "Now we will pause the scenario so that you can answer some questions about the patient case" and the end of the freeze was defined as the end of the sentence "Is everyone in place and ready? Now the scenario will re-start". Team SAGAT (TSAGAT) was developed by Crozier et al. [29] based on SAGAT as an assessment tool for evaluating team performance. In Crozier's study, each team comprised a trauma leader, airway manager and nurse. Individual SA questionnaires were developed for the three-team roles including both shared knowledge and complementary knowledge. TSAGAT was calculated as the sum of individual SAGAT scores and the TSAGAT scores had a high correlation to a traditional checklist (Pearson correlation, r=0.996). However, Salas defined team SA as a dynamic process defined as the team's shared understanding of a situation at a specific point in time [33] while Endsley argued that team SA involves unique activities as information sharing and coordination.[20] Thus, in the present study, in order to measure this efficiently in all team members, all participants in a case received identical SA questionnaires. TSAGAT was calculated based on the mean scores of the team for each SA item.

Post-simulation questionnaire

In order to measure whether the items in the SA questionnaire were considered relevant to the scenarios, and whether pausing the scenarios affected the team training activity, a post-simulation questionnaire (PSQ) was used.[22] The PSQ comprises 13 statements to be rated on a four-point scale ranging from Strongly disagree (1) to Strongly agree (4), in which five statements concern the SAGAT and the effect of freezing the scenario and eight statements concern the simulation and the scenario *per se*. To the best of our knowledge, no data regarding the reliability of PSQ were presented in the original study. In this study, we only use the five questions relating to SAGAT and freezing the scenario. The PSQ was translated

into Swedish by one of the authors and the translation was further refined based on iterative discussions within our research group. Next, the Swedish PSQ was sent to a professional translation agency together with the original PSQ for verification of the translation. The PSQ was answered in an anonymous web survey during the second week after the simulation training by 24 of the 55 participating students, i.e. the response rate was 44% (Table 4). In this study, the results of the four-point scale were dichotomised into disagree (Strongly disagree and Disagree) and agree (Agree and Strongly agree).

Study size

The aim of the study was to evaluate the feasibility of ABCDE, TEAM and SAGAT for use in further studies. In order to achieve this, the study sample must be large enough to allow for, with a fair precision, the calculation of descriptive statistics (means and standard deviations) and the calculation of reliabilities. In order to assess a large effect size (Cohen's d of 0.9) with t-test with a power of 80% at the 0.05 level, 16 participants per group would be needed as determined by G*Power.[34] Thus, for the aims of the present study, the inclusion of 50 individuals and 20 teams would suffice.

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM, Chicago, IL). Interrater reliability for the ABCDE checklist and the TEAM instrument was determined by intraclass correlation using a two-way random effects model (ICC (2,1) type absolute).[35,36] ICC is reported both as single measures and average measures since the ICC for single measures relates to the reliability of the individual ratings, while average measures relates to the reliability of the mean values. Cronbach's alpha was used to measure the internal consistency of the SAGAT and TSAGAT.[35] Internal consistency considered as the

extent to which all items measure the same latent variable was investigated by χ^2 and as suggested by Schweizer, a normed χ^2 below 2 was taken as an indication of a good fit.[37]

Ethics

This study was approved by the Ethical Review Board of Northern Sweden (April 7, 2016, decision no. 2016-54-31M).

Patient and Public Involvement

Neither patients nor the public were involved in the design or the data collection for this study.

Results

Scores on ABCDE checklist, TEAM instrument and SA items

Fifty-five participants participated in the study, combined into 23 teams with 3-4 participants in each team, running either case A or B. The scores on the ABCDE checklist, TEAM instrument and the SA items are shown in Table 5. The mean ABCDE score was 2.6 on a scale of 0–4, the mean TEAM item 1–11 score was 2.3 on a scale of 0–4, the mean TEAM global rating (item 12) was 4.8 on a scale of 0–10. The mean correct score per item ranged from 0.25 to 0.95 on a scale of 0 or 1, and the mean total SA item score per participant was 13 with a maximal score of 22.

Interrater reliability for the ABCDE checklist and TEAM instrument

Interrater reliability as measured by intraclass correlation was 0.55 (single measures)/0.83 (average measures) for the ABCDE checklist and 0.54/0.83 for the TEAM scale. For the TEAM subscales of leadership, team work, task management and global rating, the intraclass correlations were 0.36/0.70, 0.45/0.77, 0.35/0.68 and 0.38 /0.72, respectively.

Feasibility and internal consistency of SAGAT

The PSQ showed that 96% of the participants considered the SA items to be relevant to the case and 96% considered the questions to be easy to understand (Table 4). About three out of four (72%) participants stated that the freeze did not negatively impact their concentration or performance during the simulation session.

The internal consistency of SAGAT measured as Cronbach's alpha was 0.80 for case A and 0.62 for case B (Table 5), and normed χ^2 was 1.72 vs. 1.62. For level 1 (perception), Cronbach's alpha was low, 0.06 for case 1 and 0.25 for case 2, but for level 3 it was fair, 0.89 and 0.66, respectively.

For TSAGAT, the internal consistency as measured by Cronbach's alpha was good for the SAGAT questionnaire (All levels) with 0.83 and 0.76 for case A and B, respectively, and for level 3 it was 0.89 and 0.79, respectively (Table 6).

Discussion

Research on how to maximise team performance depends on the availability of sensitive and reliable tools for measuring the impact of an intervention. This study aimed to test the usability of three instruments and techniques developed in English-speaking contexts for rating in a Swedish setting with teams performing in their native language and cultural context. The main finding of this study was that the adapted ABCDE checklist and the TEAM instrument could be used with acceptable interrater reliability and that it was feasible to use SAGAT to measure SA. The scores were in the middle of the scales indicating that the scales could purposefully be used in a future effect study providing that the test group also score in the sensitive range of these instruments. The combination of these three measurements,

ABCDE, TEAM and SAGAT, could permit analysis of task performance, team performance and the relationship to situation awareness in team training.

Scaling down the comprehensive TTET[15] to a smaller ABCDE checklist could introduce unintended errors in measurements. In order be sensitive to the intended use, the items measured must be relevant to the task and the anchors of the scale used must be calibrated to the setting in order to allow for significant changes to be reflected as the difference between control and intervention.[35] In the present study, the developed ABCDE checklist was perceived by professionals to be relevant to the case. Interrater variability was low, indicating that the checklist could be reliably used for scoring task performance. The means of the scores were in the middle of the scale, which indicates that the checklist might be sensitive to differentiating between low and high performers.

For the TEAM instrument, the interrater reliability was 0.55 for single measurements and 0.83 for average measurements, which are similar to the results reported by Cooper et al where the interrater reliability for the TEAM instrument was 0.55 as measured by Cohen's kappa and 0.60 as measured by ICC [16], and also to the results reported by McKay et al where the ICC were 0.59 – 0.88 for the different items.[38] According to Koo and Li, an ICC between 0.50 and 0.75 indicates moderate reliability, while an ICC between 0.75 and 0.90 indicates good reliability.[36] Thus, the averaged measurements from four raters had a good reliability. In the original TEAM publication by Cooper et al, the performance improved significantly from novice learners to experts, and in this study the participants were a homogenous set of medical students, and as such to be considered as novice learners.[16] The medical students in the present study had an average TEAM item score of 2.3 of a maximum 4, which translates to showing the desired behavior a bit more often than not. The national learning objectives requires a systematic training in leadership and followership which might explain the rather high score. [39]

Both the PSQ and the CVI indicate that SAGAT could be used to construct questions that were considered relevant to the case. In both cases, the internal consistency was fair (0.80 and 0.61). When analysing the subscales, levels 1, 2 and 3, the internal consistency was low for levels 1 and 2, and higher for level 3. This could indicate that the perception of the situation in the groups was diverse and not related to the total score, while the ability to project the direction in which the cases were heading was more homogeneously related to the total score. TSAGAT had a higher overall homogeneity, as might be expected when analysing the means of the group SAGAT for each question, instead of the specific SAGAT answers.

The feasibility of using SAGAT to measure situation awareness was assessed by measuring the length of the freeze (i.e. pause) needed to answer questions and by asking the participants for their perception of the pauses in the scenario and how the pauses affected the training session. According to Edsley, when measuring SA using SAGAT, the scenario is frozen while the participants answer questions that probe the three levels of SA: perception, comprehension and projection.[21] Each pause in this study lasted less than 3 minutes. The pauses and SAGAT questions could influence SA in both a negative direction and a positive direction. In the negative direction, the flow of the simulation training might be disrupted, stress be induced by being forced to interrupt the case, and the commitment decrease. According to the PSQ, the majority of the participants did not perceive that the pauses adversely affected concentration and performance. Contrary, the pauses and questions might facilitate the resolution of the clinical problem in the case by triggering and allowing time for reflection, fully in line with Schön's reflection-on-action.[24]

Limitations of this study

The participants in the study comprised year four medical students for the video recordings of simulation-based team-training and the rating of the video material was conducted by four participants representing a wide spectrum of experience and training. It could be argued that it

is a limitation that this validation was not performed on a series of critical care teams, for example. However, for testing the reliability as well as the feasibility of the checklist, techniques and instruments, the simulation-based trainings with medical students was a readily available series of standardized simulations.

The raters in this study included both experienced and inexperienced raters. The study was not designed to systematically test for differences in ratings between raters of different backgrounds. However, the inhomogeneity of the raters is a strength as it indicates that the checklists and instruments might be used by a group of raters with different experiences. Difficulties were encountered in the interpretation of what constitutes partially performed versus performed completely in the ABCDE checklist. This could relate to the vast differences in the raters' level of education and previous experience. To resolve this, the raters were trained using a separate set of video recordings to improve the accuracy and minimize the variability. TEAM was not translated into Swedish in order to avoid inducing errors. This was possible as fluency in English is a prerequisite for academic studies in Sweden. The ratings based on the original TEAM instrument were consistent between the raters, indicating that this may have been a correct assumption. The developed set of SAGAT items contained only a few items for measuring each level of SA. The selected number of items were close to what was used by Gardner and Hogan.[22,25] Too few items increase measurement errors, while too many items result in extended freezes. However, in the present study, the probing of level 2 in particular would have benefited by 2–3 more items per measurement.

In the present study, CVI for SAGAT was measured according to Lawshe [32] while the development of the ABCDE checklist and the two cases relied on the authors' experiences and iterative interactions with clinicians and experts in the field. Both the cases, the ABCDE checklists and the developed SAGAT questionnaires, could have benefitted from a full formal CVI by a review panel.

The PSQ was only translated from English and not translated back into Swedish in order to formally check the identity of the items in the final sets. However, the translation was adjusted by a professional translator before being used in the study. Thus, the results of the PSQ can be used for probing the participants in the simulation with regards to their experiences of SAGAT.

This study focused on the quantification of performance during a simulation based training. The transferability of the studied behaviours into a real world setting is an intriguing question for further studies.

Conclusion

In this setting with medical students, the developed ABCDE checklist and the TEAM instrument had high interrater reliability, the process of using SAGAT questionnaires during simulation-based training did not negatively affect the participants' evaluation of simulation-based training and the developed SAGAT questionnaires had a fair internal consistency. Thus, the measurement of task performance, team performance and situation awareness in future studies in a Swedish simulation-based training setting may be conducted using these techniques.

Table 1 Background characteristics of the medical students (n=55) participating in the study

Age years, m (sd)	25 (4.5)
Male n (%)	28 (51)
Female n (%)	27 (49)
Previous healthcare education	
None n (%)	44 (80)
Assistant nurse (2Y upper secondary school) n	2 (4)
(%)	3 (5)
Registered nurse (3Y university) n (%)	6 (11)
Short courses (Red Cross, etc.) n (%)	
Previous team training	
Yes n (%)	20 (36)

Previous simulation experience	
Yes n (%)	32 (58)
Previous experience of trauma patients*	
Yes n (%)	20 (37)

* one missing value

Table 2A Situation Awareness items for Case A. Response scale, SA level and fraction correct answers.

Freeze 1	Question	Response scale	Level	Correct
1.	What is the patient's respiratory rate?	Free text	1	0.26
2.	What are the findings from the assessment of A and B?	Free text	1	0.51
3.	Is the patient sufficiently oxygenated?	Yes/No	2	0.49
4.	What did the patient state as their year of birth?	Free text	1	0.36
5.	What do you think will happen to the patient's blood pressure in the next few minutes?	Increase/Decrease/No change	3	0.74
6.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	0.69

7.	How many peripheral IV cannulas does the patient have?	Free text	1	0.95
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?			
8.	Heart rate	Increase/Decrease	3	0.85
9.	Blood pressure	Increase/Decrease	3	0.82
10.	Saturation	Increase/Decrease	3	0.59
11.	Respiratory rate	Increase/Decrease	3	0.74

Freeze 2	Question	Response scale	Level	Correct
12.	What is the patient's blood pressure?	Free text	1	0.77
13.	What are your findings under C (Circulation)?	Free text	1	0.82
14.	What additional examinations/tests do you think are needed?	Free text	3	0.67
15.	Which medications/drugs do you think are needed?	Free text	3	0.49
16.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	0.95
17.	Is there an algorithm for CPR on the wall?	Yes/No	1	0.26
18.	What previous illnesses has the patient described?	Free text	2	0.21

	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?			
19.	Heart rate	Increase/Decrease	3	0.74
20.	Blood pressure	Increase/Decrease	3	0.74
21.	Saturation	Increase/Decrease	3	0.59
22.	Respiratory rate	Increase/Decrease	3	0.69

The questions are shown translated into English in the table and the original questions in Swedish are available upon request from the authors of the study.

Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

Table 2B Situation Awareness items for Case B

Freeze 1	Question	Response scale	Level	Correct
1.	What is the patient's respiratory rate?	Free text	1	0.28
2.	What are the findings from the assessment of A and B?	Free text	1	0.64
3.	Is the patient sufficiently oxygenated?	Yes/No	2	0.78
4.	What does the sign on the patient's necklace indicate?	Free text	1	0.36
5.	What do you believe will happen to the patient's venous return in the next few minutes?	Increase/Decrease/No change	3	0.61
6.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	0.58
7.	Is a suction device available?	Yes/No	1	0.69

	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?			
8.	Heart rate	Increase/Decrease	3	0.78
9.	Blood pressure	Increase/Decrease	3	0.58
10.	Saturation	Increase/Decrease	3	0.78
11.	Respiratory rate	Increase/Decrease	3	0.72

Freeze 2	Question	Response scale	Level	Correct
12.	What is the patient's respiratory rate?	Free text	1	0.33
13.	What are your findings under C (Circulation)?	Free text	1	0.69
14.	What additional examinations/tests do you think are needed?	Free text	3	0.92
15.	Which medications/drugs do you think are needed?	Free text	3	0.22
16.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	0.83
17.	What is the patient's blood pressure?	Free text	1	0.53
18.	Is a defibrillator available in the room?	Yes/No	2	0.50
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?			
19.	Heart rate	Increase/Decrease	3	0.72

20.	Blood pressure	Increase/Decrease	3	0.72
21.	Saturation	Increase/Decrease	3	0.69
22.	Respiratory rate	Increase/Decrease	3	0.56

The questions are shown translated into English in the table and the original questions in Swedish are available upon request from the authors of the study.

Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

Table 3 Time (minutes) for scenario freezes to measure situation awareness with SAGAT

Case	No. of teams n	Scenario m (sd)	1st freeze m (sd)	2nd freeze m (sd)	Total time m (sd)
A	12	14.1 (2.0)	2.7 (0.5)	2.5 (0.6)	19.3 (2.3)
B	11	14.1 (3.4)	2.7 (0.6)	2.8 (0.6)	19.7 (3.8)

Table 4 Student agreement about usefulness and feasibility of the prospective collection of situation awareness items

Question	Agree/ strongly agree	Disagree/ strongly disagree	Do not know
Concerning the SAGAT questionnaire. In my opinion:	n (%)	n (%)	n (%)
The introduction was adequate	24 (100)	0 (0)	0 (0)
The questions were clear	23 (96)	1 (4)	0 (0)
The "freezes" in the scenario adversely affected my concentration level and my performance	7 (29)	17 (71)	0 (0)
The questions were relevant to the way I perceived the scenario	23 (96)	1 (4)	0 (0)

Overall, I am satisfied with SAGAT as a tool for evaluating hands-on skills during a trauma exercise.	19 (79)	1 (4)	4 (17)

Results from the post-simulation questionnaire

Table 5 Mean scores on ABCDE checklist, TEAM instrument and SAGAT questionnaire

	Case A m (sd)	Case B m (sd)
ABCDE checklist*	2.70 (0.19)	2.58 (0.27)
TEAM instrument		
Item 1–11*	2.32 (0.54)	2.27 (0.46)
Global rating**	4.90 (1.13)	4.75 (1.04)
SAGAT	13.95	12.60 (3.35)
questionnaire***	(4.26)	

*Mean score per item (sd). Max score per item 4.

**Mean score (sd). Max score 11.

***Mean score (sd). Max score 22.

Table 6 Internal consistency (Cronbach's alpha) of SAGAT and TSAGAT

SAGAT/TSAGAT	Case A n/n 39/12	Case B n/n 36/11
Level 1	0.060/0.372	0.246/0.420
Level 2	0.321/-0.018	0.332/0.332
Level 3	0.891/0.891	0.659/0.786
Total (Level 1–3)	0.800/0.827	0.620/0.759

Internal consistency calculated for the questions measuring the three levels of SA and for the questionnaire in total.

n/n, number of individuals answering the SAGAT questionnaires / number of teams

References

1 Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ* 2016;**353**:i2139. doi:10.1136/bmj.i2139

2 Kohn LT, Corrigan JM, Donaldson MS. *To Err is Human: Building a Safer Health System*. National Academies Press 2000. doi:10.17226/9728

3 World Health Organization. Multi-professional Patient Safety Curriculum Guide. WHO, Geneva. 2011.https://apps.who.int/iris/bitstream/handle/10665/44641/9789241501958_eng.pdf

4 Scalese RJ, Obeso VT, Issenberg BS. Simulation Technology for Skills Training and Competency Assessment in Medical Education. *J Gen Intern Med* 2007;**23**:46–9. doi:10.1007/s11606-007-0283-4

5 Zendejas B, Brydges R, Wang AT, *et al.* Patient outcomes in simulation-based medical education: a systematic review. *Journal of General Internal Medicine* 2013;**28**:1078–89.

6 Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. *Med Educ* 2012;**46**:1042–54. doi:10.1111/j.1365-2923.2012.04343.x

7 Salas E, Rosen MA, Held JD, *et al.* Performance Measurement in Simulation-Based Training. *Simulat Gaming* 2009;**40**:328–76. doi:10.1177/1046878108326734

8 Rosen MA, Pronovost PJ. Advancing the Use of Checklists for Evaluating Performance in Health Care. *Acad Med* 2014;**89**:963–5. doi:10.1097/acm.0000000000000285

9 Brady PW, Muething S, Kotagal U, *et al.* Improving situation awareness to reduce unrecognized clinical deterioration and serious safety events. *Pediatrics* 2013;**131**:e298-308. doi:10.1542/peds.2012-1364

10 Salas E, Cooke NJ, Rosen MA. On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Hum Factors J Hum Factors Ergonomic Soc* 2008;**50**:540–7. doi:10.1518/001872008x288457

11 Kozlowski SW. Enhancing the Effectiveness of Work Groups and Teams: A Reflection. *Perspect Psychol Sci* 2018;**13**:205–12. doi:10.1177/1745691617697078

12 Converse SA, Cannon-Bowers JA, Salas E. Team Member Shared Mental Models: A Theory and Some Methodological Issues. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1991;**35**:1417–21. doi:10.1177/154193129103501917

13 Mathieu J, Heffner T, Goodwin G, *et al.* The influence of shared mental models on team process and performance. *The Journal of applied psychology* 2000;**85**:273–83.

14 Boyle WA, Murray DJ, Beyatte M, *et al.* Simulation-Based Assessment of Critical Care "Front-Line" Providers*. *Critical Care Medicine* 2018;**46**:e516–22. doi:10.1097/CCM.0000000000003073

15 Holcomb JB, Dumire RD, Crommett JW, *et al.* Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. *J Trauma* 2002;**52**:1078–85.

16 Cooper S, Cant R, Porter J, *et al.* Rating medical emergency teamwork performance: Development of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;**81**:446–52.

doi:10.1016/j.resuscitation.2009.11.027

17 Cooper SJ, Cant RP. Measuring non-technical skills of medical emergency teams: An update on the validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2014;**85**:31–3. doi:10.1016/j.resuscitation.2013.08.276

18 Cooper S, Cant R, Connell C, *et al.* Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation* 2016;**101**:97–101. doi:10.1016/j.resuscitation.2016.01.026

19 Schulz CM, Krautheim V, Hackemann A, *et al.* Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. *BMC Anesthesiol* 2016;**16**:4. doi:10.1186/s12871-016-0172-7

20 Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors* 1995;**37**:32–64. doi:10.1518/001872095779049543

21 Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. In: *Situation awareness analysis and measuement*. Lawrence Erlbaum Associates 2000. 1–21.

22 Hogan MP, Pace DE, Hapgood J, *et al.* Use of human patient simulation and the situation awareness global assessment technique in practical trauma skills assessment. *J Trauma* 2006;**61**:1047–52. doi:10.1097/01.ta.0000238687.23622.89

23 Hamstra SJ, Brydges R, Hatala R, *et al.* Reconsidering fidelity in simulation-based training. *Acad Med* 2014;**89**:387–92. doi:10.1097/ACM.0000000000000130

24 Schön DA. *The reflective practitioner: How professionals think in action*. The Perseus Books Group 1984.

25 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc J Soc Simul Healthc* 2017;**12**:17–21. doi:10.1097/sih.0000000000000181

26 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident's Situational Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3. doi:10.1097/BOT.0000000000000003

27 Lavoie P, Cossette S, Pepin J. Testing nursing students' clinical judgment in a patient deterioration simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today* 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015

28 McKenna L, Missen K, Cooper S, *et al.* Situation awareness in undergraduate nursing students managing simulated patient deterioration. *Nurs Educ Today* 2014;**34**:e27–31. doi:10.1016/j.nedt.2013.12.013

29 Crozier MS, Ting HY, Boone DC, *et al.* Use of human patient simulation and validation of the Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009

30 Thim T, Krarup, Grove, *et al.* Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012;:117. doi:10.2147/IJGM.S28478

31 ATLS Advanced Trauma Life Support Student Course Manual. American College of Surgeons. 2018.

32 Lawshe CH. A quantitative approach to content validity. *Pers Psychol* 1975;**28**:563–75. doi:10.1111/j.1744-6570.1975.tb01393.x

33 Salas E, Prince C, Baker DP, *et al.* Situation Awareness in Team Performance: Implications for Measurement and Training. *Hum Factors J Hum Factors Ergonomics Soc* 1995;**37**:123–36. doi:10.1518/001872095779049525

34 Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers* 1996;**28**:1–11. doi:10.3758/BF03203630

35 Nunnally JC. *Psychometric Theory, 2nd edition*. McGraw-Hill 1978.

36 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine* 2016;**15**:155–63. doi:10.1016/j.jcm.2016.02.012

37 Schweizer K. Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment* 2010;**26**. doi:10.1027/1015-5759/a000001

38 McKay A, Walker ST, Brett SJ, *et al.* Team performance in resuscitation teams: comparison and critique of two recently developed scoring tools. *Resuscitation* 2012;**83**:1478–83. doi:10.1016/j.resuscitation.2012.04.015

39 Lindgren S, Brännström T, Hanse E, *et al.* Medical education in Sweden. *Medical teacher*;**33**:798–803. doi:10.3109/0142159X.2011.570816

BMJ Open

The reliability of instruments that measure situation awareness, team and task performance in a simulation setting with medical students

Journal:	*BMJ Open*
Manuscript ID	bmjopen-2019-029412.R3
Article Type:	Research
Date Submitted by the Author:	06-Aug-2019
Complete List of Authors:	Hultin, Magnus; Umeå Universitet Medicinska fakulteten, Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine Jonsson, Karin; Umeå Universitet Medicinska fakulteten, Departments of Surgical and Perioperative Sciences and Nursing Härgestam, Maria; Umea Universitet Medicinska fakulteten, Department of Nursing Lindkvist, Marie; Umeå Universitet Medicinska fakulteten, Department of Public Health and Clinical Medicine Brulin, Christine; Umeå Universitet Medicinska fakulteten, Department of Nursing
Primary Subject Heading:	Medical education and training
Secondary Subject Heading:	Communication
Keywords:	Patient care team, Resuscitation, Situation Awareness, Task performance and analysis, Teamwork

The reliability of instruments that measure situation awareness, team and task performance in a simulation setting with medical students

Magnus Hultin, MD Ph.D., Associate Professor, Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0003-2935-7161

Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0001-7974-6777

Maria Härgestam, RNA Ph.D., Lecturer, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden. https://orcid.org/0000-0003-0680-9962

Marie Lindkvist, Ph.D., Associate Professor, Department of Epidemiology and Global Health, Umeå University, S-901 87 Umeå, Sweden

Christine Brulin, RNA Ph.D., Professor Emerita, Department of Nursing, Umeå University, S-901 87 Umeå, Sweden

Abstract 246 words

Main article 4736 words

Keywords

- Patient Care Team
- Resuscitation

- Situation Awareness
- Task Performance and Analysis
- Teamwork

Corresponding author

Magnus Hultin, Dept of Surgical and Perioperative Sciences, Anesthesiology and Critical Care Medicine, Umeå University, Sweden, magnus.hultin@umu.se, phone: +46-90-786 9109

Abstract

Objectives: The assessment of situation awareness, team and task performance in a simulation training session require reliable and feasible measurement techniques. The objectives of this study were to test an Airways-Breathing-Circulation-Disability-Exposure (ABCDE) checklist and Team Emergency Assessment Measure (TEAM) for interrater reliability, as well as an application of Situation Awareness Global Assessment Technique (SAGAT) for feasibility and internal consistency.

Design: Methodological approach

Setting: Data collection during team training using full-scale simulation at a university clinical training centre. The video-recorded scenarios were rated independently by four raters.

Participants: 55 medical students aged 22–40 years in their 4th year of medical studies, during the clerkship in anaesthesiology and critical care medicine, formed 23 different teams. All students answered the SAGAT questionnaires and, of these students, 24 answered the follow-up post-simulation questionnaire (PSQ). TEAM and ABCDE were scored by four professionals.

Measures: ABCDE and TEAM were tested for interrater reliability. The feasibility of SAGAT was tested using PSQ. SAGAT was tested for internal consistency both at an individual level (SAGAT) and a team level (TSAGAT).

Results: The intraclass correlation was 0.54/0.83 (single/average measurements) for TEAM and 0.55/0.83 for ABCDE. According to the PSQ, the items in SAGAT were rated as relevant to the scenario by 96% of the participants. Cronbach's alpha for SAGAT/TSAGAT for the two scenarios was 0.80/0.83 vs. 0.62/0.76, and normed χ^2 was 1.72 vs. 1.62.

Conclusions: Task performance, team performance and situation awareness could be purposefully measured, and the reliability of the measurements was good.

Article Summary

Strengths and limitations of this study

- SAGAT could be used to create items in Swedish to probe situation awareness, i.e. in the participants' native language
- TEAM could be used by the raters in its original language (English).
- The developed ABCDE checklist has items that are well-defined concepts, and the difficulty lies in defining the rubrics for scoring the items.
- It is a weakness that PSQ was only translated into Swedish and not retranslated back into English.
- An interprofessional set of raters with different backgrounds and experiences rated TEAM and ABCDE with similar results.

Funding

This work was supported by Region Västerbotten, grant no. VLL-663801 and VLL-836931, by Region Norrbotten, grant no. NLL-765981, by Alice Lindström's Foundation, and by the Medical Faculty of Umeå University.

Competing interests

None of the authors has any competing interests to declare.

Data availability

The data comprising de-identified questionnaire responses and ratings of video scenarios are available upon request from the corresponding author.

Author Contributions

MHu, KJ, MHä and CB designed the study. MHu and KJ prepared the initial draft of the paper. MHu, KJ, MHä, ML and CB were all actively engaged in data analysis and in finalising the paper. All authors have read and approved the final version of the paper.

Introduction

Medical errors are the third leading cause of death in the US.[1] Knowledge about the relationship between human errors and patient safety has increased in the last two decades.[2,3] Simulated environments make it possible to improve skills by employing training strategies to prevent errors while simultaneously offering an arena for reliable assessments of skills.[4] Thus, simulation training is often used by organisations to minimise adverse events and prevent healthcare errors.[5,6] This could be accomplished by improving *task performance*, *team performance* or *situation awareness (SA)*.[7–9]

When developing strategies for improving clinical practice, it is essential to evaluate both task and team performance. According to Salas[10] and Kozlowski[11], team performance is a multilevel process that includes the interrelation between individual- and team-level taskwork and teamwork processes. Thus, an optimal task, as well as team performance, depends on the coordinated activities of a team of individuals.[12,13] Checklists are often used to score task performance in acute care scenarios. The lists might include adherence to resuscitation protocols, the timing of the task, as well as the time taken to complete the components.[14] The Trauma Team Evaluation Tool (TTET) was developed by Holcomb for trauma scenarios managed according to Airways-Breathing-Circulation-Disabilities-Exposure (ABCDE) protocols based on Advanced Trauma Life Support (ATLS) and was tested on US military resuscitation teams from community hospitals.[15] As a pilot study, psychometric data such as reliability and validity were not reported. As TTET was developed for a specific setting, the items and criteria for scoring have to be adapted to the proficiency levels of the participants and the standard operating procedures being used.

Team performance can be measured using the Team Emergency Assessment Measure (TEAM), for example. It measures three dimensions of team performance: leadership, teamwork and task management.[16,17] The instrument was initially developed and validated for simulator-based team training and was recently validated for the collection of observational ratings of non-technical skills during live resuscitations in emergency departments.[18] Using instruments to score scenarios depend on reliable interpretations of the instrument by the raters, and it might be even more important to ensure the reliability of such interpretations when using an instrument in a non-native language.

Moreover, and in addition to task and team performance, SA is a prerequisite for patient safety and the prevention of errors, particularly in acute care situations.[9,19] SA includes three levels of ability: (1) perception and attention (*What?*, what's going on), (2) comprehension (*So what?*, the ability to understand what's going on), and (3) projection (*What's next?*) in order to anticipate and plan for future events.[20] In order to measure SA in a simulation setting, the Situation Awareness Global Assessment Tool (SAGAT) has been developed and subsequently adapted for use in healthcare settings.[21,22] One feature of SAGAT is that its use requires the simulation to pause[21], which might influence clinical understanding, both by impeding the suspension of disbelief in the simulation setting [23] and by facilitating reflection on action.[24] Gardner showed that it was feasible to use SAGAT to measure SA in the team training of surgical trainees in advanced cardiac life support.[25] Globally, SAGAT has been used to study, for example, the effect of sleep deprivation on SA in trauma team training, how SA is associated with surgical trainee team performance, as well as nurses' clinical assessment of patient deterioration.[26–28] SAGAT can be analysed on an individual level and also on a team level. A specific application of SAGAT is team SAGAT (TSAGAT) in which the different members of a team answer SAGAT questions specific to their role. TSAGAT was constructed to account for the teams' shared situation awareness and

validated in trauma team training in a Canadian setting.[29] To our knowledge, SAGAT has not previously been used in a Swedish context, and it is therefore important to evaluate both its feasibility and its internal consistency.

To summarise, ABCDE checklists, the TEAM instrument and SAGAT have been developed in order to evaluate different aspects of teamwork. However, these instruments and questionnaires have neither been translated into Swedish nor tested for feasibility or trustworthiness in a Swedish context. Such studies are necessary to enable the evaluation of teamwork in Swedish acute care settings and simulation-based training. Thus, this study aimed to test an ABCDE checklist that we developed for interrater reliability, to test the TEAM instrument for interrater reliability, and to test two Swedish SAGAT questionnaires for feasibility and internal consistency.

Method

This study is based on data collected during simulation-based team training sessions with medical students.

Participants and raters

From March to October 2016, all medical students (n=68) in year 4 undertaking their clerkship in anaesthesiology and critical care medicine were invited to participate in the study while receiving mandatory simulator-based team training. In total, 55 students (81%) participated in the study (Table 1) and 20 of them participated in both scenario A and scenario B.

All scenarios were video recorded and later scored by four raters to allow for calculation of interrater reliability. First, a registered nurse with a Master's Degree (one year) in Nursing (Critical Care Medicine), 20 years' working experiences at an ICU, nine years' experiences of

human patient simulator team training and also a PhD student (author KJ). Second, a physician with 20 years' working experience, associate professor and consultant in Anesthesiology and Critical Care Medicine, with 14 years' experiences of human patient simulator team training (author MHu). Third, a registered nurse with a Master's Degree (one year) in Nursing (Acute Care Medicine), with six years' working experiences of prehospital care and one year as a medical student. Fourth, a paramedic with limited experiences beyond two years working as a paramedic in the Israeli Army and one year as a medical student.

Setting

Data collection in this study was carried out during simulation-based team training at the Clinical Training Centre (CTC) of the Medical Faculty, Umeå University. The briefings, scenarios and debriefings were conducted in Swedish. Two cameras mounted at an angle were used to record videos in the simulation room, and one of the views included the patient monitor.

One week before team training, the students were asked to watch a twelve-minute video available at the learning platform that introduced both the ABCDE concept when caring for a patient and the room and equipment to be used during simulation-based training.[30,31] At the start of the training session, the students participated in a 15-minute live introduction to the simulation laboratory presented by the operator and the instructor. Each student group (4–5 students/group) trained on four scenarios that focused on assessment and treatment of severely ill patients in an emergency room. In each scenario, 3-4 students were active, and 1–2 were observers. Therefore, only 3-4 participated in the interactive role-play in each scenario (the actual simulation). In all, the 55 unique students made up a total of 23 teams with 3-4 participants in each team and 20 students participated in both case A and case B. The students all played the role of interns, i.e. physicians who are in training to become licensed, with the attending nurse currently unavailable. The assigned task involved conducting a primary

survey and stabilising the patient until more senior staff arrived in 15 minutes. After each scenario, a 10–15-minute debriefing session permitted reflection and shared learning. The first scenario was a warm-up, and the last included a training summary. Case A and B, or in the reverse order, as determined by the toss of a coin immediately before the second scenario, were used as the second and third scenarios.

All scenarios were conducted in Swedish, were designed to last 10–15 minutes and were pre-programmed into a Laerdal SimMan simulator in order to support the standardisation of the simulation. In essence, the patient cases used were slightly modified versions of the scenarios used by Hogan.[22] Case A was hypovolemic shock following a traffic incident. Case B was a pneumothorax with affected vital signs following a traffic incident.

Background questionnaire

The background questionnaire included informed consent and was answered immediately before team training started. The questionnaire included questions such as year of birth, male/female, previous medical training, previous experiences of team training, previous experiences of human patient simulator-based training, previous experiences of CRM and previous experiences of live trauma care.

ABCDE checklist

In order to measure the completeness of critical tasks in acute care scenarios, an ABCDE checklist was used. The original TTET comprised 58 items derived from the ATLS protocol.[15] Each item in the TTET was scrutinised by the authors of this study and the number of items reduced to those that reflected the measures expected to be carried out by year four medical students in the acute care scenarios in this study. Some items were also slightly modified to reflect the current ATLS standards. The final list contained ten items reflecting the key elements of the ATLS primary survey, i.e. the core management of

ABCDE: airway assessed, airway secured, saturation assessed, oxygen applied, ventilation assessed, ventilation optimised, pulses checked (radial – femoral – carotid), venous access and infusions established, neurological disabilities checked (consciousness, pupils), and full/complete exposure. The items are well described in major textbooks and were translated into Swedish in line with the nomenclature being used. Compared to the original scoring system, an additional scoring option was added: performed after a reminder from the instructor. Each item was rated on a rating scale of 0 to 4 (0 = Not initiated, 1= Performed after a reminder from the instructor, 2 = Partially performed, 3 = Performed completely before the end of the simulation, and 4 = Performed consistently during the whole simulation, NA = not applicable). Based on all items in the ABCDE checklist an index was constructed as a mean score ranging from 0.0 to 4.0.

TEAM instrument

In order to measure team leadership, teamwork and task management, the TEAM instrument was used unmodified, i.e. in English, as developed by Cooper et al.[16,17] The published internal consistency (Cronbach's alpha) was 0.97, and the interrater reliability was 0.55, as measured by Cohen's Kappa (adjusted for chance) with a mean intraclass correlation coefficient of 0.60 for the 11 items. The instrument comprises 11 critical behaviours rated on a Likert scale of 0 to 4 (0 = Never/Hardly, 1 = Seldom, 2 = About as often as not, 3 = Often, 4 = Always/Nearly always), which are summed into a TEAM score ranging from 0-44, and finally, a global rating of the team's overall performance on a scale of 1 to 10. The original publication with the TEAM instrument had no descriptors for the endpoints. For this study, we used 1 = Poor and 10 = Excellent.

The TEAM instrument comprises three subscales: leadership (items 1–2), teamwork (items 3–9) and task management (items 10–11). Indexes were constructed based on the items in the subscales and on all items in the instrument, ranging from 0.0 to 4.0.

Procedure for rating the TEAM instrument and ABCDE checklist

The raters (n=4) in this study used video recordings in the rating procedure of the TEAM scale and the ABCDE checklist. The raters held two separate two-hour long meetings to discuss the interpretation of the descriptors in the ABCDE checklist and the TEAM instrument. During the first meeting, the discussions were facilitated using a different set of videos with similar scenarios, but from other teams. A total of six videos were used for this. During the second meeting, after four of the scenarios had been rated by each participant, the raters met again to discuss the interpretation of scales.

The raters independently assessed the videos of the simulation scenarios included in this study and rated the performances using the ABCDE checklist and the TEAM instrument. Each video was viewed at least twice by each rater.

SAGAT

When developing SA items for the specific scenarios, goal-directed task analysis was used, as initially described by Endsley.[21] Briefly, for each profession, major goals are identified along with sub-goals. Critical decisions are then identified, and SA requirements are defined as the dynamic information needed to achieve the major goals, as opposed to static information such as rules and guidelines. The samplings (items) are then matched against the SA requirements. The original recommendation by Endsley was 30–60 items for within-subject studies for each of the three SA levels: (1) perception and attention, (2) comprehension and (3) projection.[20] When Gardner et al. validated a questionnaire based on SAGAT in a study of medical trainees, each questionnaire comprised three items for each level of SA at each freeze [25], while Hogan et al. developed a questionnaire with three items at level 1, one item at level 2 and three items at level 3.[22]

In the present study, the SA questionnaire was refined and adapted to the scenario and expected skills level of the students according to the process used by Hogan et al. [22] First,

the questionnaire was translated into Swedish by the authors of this study as a basis for developing scenario-specific items in Swedish. In accordance with SAGAT, targeted learning objectives for the training scenarios were formulated, and then the specific goals for each simulation were set. An iterative process was used to reformulate the items in Swedish, using a separate group of six professionals, all registered nurses, working in both a clinical context and a teaching context. The final sets of SA items are shown in Tables 2A and 2B (author's translation to English) with 11 items in each freeze. To determine whether an item is essential in a specific context, Lawshe advocated the use of professional assessments such as a content validity index (CVI) defined as the fraction of professionals who rate the item as important.[32] In the present study, the relevance was reviewed by three professionals (nurse n=1, physician n=2) before being used in the study. All professionals rated the questions as relevant, i.e. the CVI was 1.0.

The answers given by the participants in the SA questionnaire were classified as incorrect (0) or correct (1). The classifications were discussed and agreed upon by the two authors (KJ and MHu). For answers on a continuous scale (e.g. systolic blood pressure), a 10% range around the intended correct answer was accepted as correct. One question was removed from the questionnaire since it became apparent during the classification process that the question had frequently been misinterpreted.

In order to administer the SA questionnaires, the scenarios were frozen (i.e. paused) twice. The first freeze of the scenario was five minutes into the scenario, unless there was an active task activity or if the team was conducting a team re-evaluation, in which case, the freeze was briefly postponed. During the freeze, the patient monitor was switched off, and the participants turned away from the patient simulator while individually answering the questions. The second freeze took place according to the same principles after an additional

five minutes. All participants were allowed to complete the questionnaire before the scenario re-started.

The length of the freezes was measured from the video recordings (Table 3). The start of the freeze was defined as the beginning of the sentence "Now we will pause the scenario so that you can answer some questions about the patient case" and the end of the freeze was defined as the end of the sentence "Is everyone in place and ready? Now the scenario will re-start".

Team SAGAT (TSAGAT) was developed by Crozier et al. [29] based on SAGAT as an assessment tool for evaluating team performance. In Crozier's study, each team comprised a trauma leader, airway manager and nurse. Individual SA questionnaires were developed for the three-team roles, including both shared knowledge and complementary knowledge. TSAGAT was calculated as the sum of individual SAGAT scores, and the TSAGAT scores had a high correlation to a traditional checklist (Pearson correlation, r=0.996). However, Salas defined team SA as a dynamic process defined as the team's shared understanding of a situation at a specific point in time [33] while Endsley argued that team SA involves unique activities as information sharing and coordination.[20] Thus, in the present study, in order to measure this efficiently in all team members, all participants in a case received identical SA questionnaires. In this study, to account for slight differences in number of team members between the teams, TSAGAT was calculated as the mean SA score in each team.

Post-simulation questionnaire

In order to measure whether the items in the SA questionnaire were considered relevant to the scenarios, and whether pausing the scenarios affected the team training activity, a post-simulation questionnaire (PSQ) was used.[22] The PSQ comprises 13 statements to be rated on a four-point scale ranging from Strongly disagree (1) to Strongly agree (4), in which five statements concern the SAGAT and the effect of freezing the scenario and eight statements

concern the simulation and the scenario *per se*. To the best of our knowledge, no data regarding the reliability of PSQ were presented in the original study. In this study, we only use the five questions relating to SAGAT and freezing the scenario. The PSQ was translated into Swedish by one of the authors, and the translation was further refined based on iterative discussions within our research group. Next, the Swedish PSQ was sent to a professional translation agency together with the original PSQ for verification of the translation. The PSQ was answered in an anonymous web survey during the second week after the simulation training by 24 of the 55 participating students, i.e. the response rate was 44% (Table 4). In this study, the results of the four-point scale were dichotomised into Disagree (Strongly disagree and Disagree) and Agree (Agree and Strongly agree).

Study size

The aim of the study was to evaluate the feasibility of ABCDE, TEAM and SAGAT for use in further studies. In order to achieve this, the study sample must be large enough to allow for, with fair precision, the calculation of descriptive statistics (means and standard deviations) and the calculation of reliabilities. In order to assess a large effect size (Cohen's d of 0.9) with t-test with a power of 80% at the 0.05 level, 16 participants per group would be needed as determined by G*Power.[34] Thus, for the aims of the present study, the inclusion of 50 individuals and 20 teams would suffice.

Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM, Chicago, IL). Interrater reliability for the ABCDE checklist and the TEAM instrument was determined by intraclass correlation using a two-way random-effects model (ICC (2,1) type absolute).[35,36] ICC is reported both as single measures and average measures since the ICC for single measures relates to the reliability of the individual ratings, while average measures

relate to the reliability of the mean values. Cronbach's alpha was used to measure the internal consistency of the SAGAT and TSAGAT.[35] Internal consistency considered as the extent to which all items measure the same latent variable was investigated by χ^2, and as suggested by Schweizer, a normed χ^2 below 2 was taken as an indication of a good fit.[37]

Ethics

This study was approved by the Ethical Review Board of Northern Sweden (April 7, 2016, decision no. 2016-54-31M).

Patient and Public Involvement

Neither patients nor the public was involved in the design or the data collection for this study.

Results

Descriptives of ABCDE checklist, TEAM instrument and SA items

Fifty-five participants participated in the study, combined into 23 teams with 3-4 participants in each team, running either case A or B. The ABCDE mean item score was 2.6. The mean TEAM score was 25.3, and the mean TEAM global rating was 4.8. The mean SA score per item ranged from 0.25 to 0.95, and the mean SA score per participant was 13 (Table 5).

Interrater reliability for the ABCDE checklist and TEAM instrument

Interrater reliability, as measured by intraclass correlation, was 0.55 (single measures)/0.83 (average measures) for the ABCDE checklist and 0.54/0.83 for the TEAM scale. For the TEAM subscales of leadership, team work, task management and global rating, the intraclass correlations were 0.36/0.70, 0.45/0.77, 0.35/0.68 and 0.38 /0.72, respectively.

Feasibility and internal consistency of SAGAT

The PSQ showed that 96% of the participants considered the SA items to be relevant to the case, and 96% considered the questions to be easy to understand (Table 4). About three out of

four (72%) participants stated that the freeze did not negatively impact their concentration or performance during the simulation session.

The internal consistency of SAGAT measured as Cronbach's alpha was 0.80 for case A and 0.62 for case B (Table 5), and normed χ^2 was 1.72 vs. 1.62. For level 1 (perception), Cronbach's alpha was low, 0.06 for case 1 and 0.25 for case 2, but for level 3, it was fair, 0.89 and 0.66, respectively.

For TSAGAT, the internal consistency as measured by Cronbach's alpha was good for the SAGAT questionnaire (All levels) with 0.83 and 0.76 for case A and B, respectively, and for level 3 it was 0.89 and 0.79, respectively (Table 6).

Discussion

Research on how to maximise team performance depends on the availability of sensitive and reliable tools for measuring the impact of an intervention. This study aimed to test the usability of three instruments and techniques developed in English-speaking contexts for rating in a Swedish setting with teams performing in their native language and cultural context. The main finding of this study was that the adapted ABCDE checklist and the TEAM instrument could be used with acceptable interrater reliability and that it was feasible to use SAGAT to measure SA. The scores were in the middle of the scales indicating that the scales could purposefully be used in a future effect study providing that the test group also score in the sensitive range of these instruments. The combination of these three measurements, ABCDE, TEAM and SAGAT, could permit analysis of task performance, team performance and the relationship to situation awareness in team training.

Scaling down the comprehensive TTET[15] to a smaller ABCDE checklist could introduce unintended errors in measurements. In order be sensitive to the intended use, the items

measured must be relevant to the task and the anchors of the scale used must be calibrated to the setting in order to allow for significant changes to be reflected as the difference between control and intervention.[35] In the present study, the developed ABCDE checklist was perceived by professionals to be relevant to the case. Interrater variability was low, indicating that the checklist could be reliably used for scoring task performance. The means of the scores were in the middle of the scale, which indicates that the checklist might be sensitive for differences between low and high performers.

For the TEAM instrument, the interrater reliability was 0.55 for single measurements and 0.83 for average measurements, which are similar to the results reported by Cooper et al where the interrater reliability for the TEAM instrument was 0.55 as measured by Cohen's kappa and 0.60 as measured by ICC [16], and also to the results reported by McKay et al where the ICC were 0.59 – 0.88 for the different items.[38] According to Koo and Li, an ICC between 0.50 and 0.75 indicates moderate reliability, while an ICC between 0.75 and 0.90 indicates good reliability.[36] Thus, the averaged measurements from four raters had good reliability. In the original TEAM publication by Cooper et al., the performance improved significantly from novice learners to experts, and in this study, the participants were a homogenous set of medical students, and as such to be considered as novice learners.[16] The medical students in the present study had an average TEAM item 1-11 score of 2.3 of a maximum 4, which translates to showing the desired behaviour a bit more often than not. This is in agreement with the 2.49 score reported by Cooper et al. for a group of second-year medical and nursing students rated with TEAM during an interprofessional one-day resuscitation course.[16] The national learning objectives requires systematic training in leadership and followership, which might explain the rather high score. [39]

Both the PSQ and the CVI indicate that SAGAT could be used to construct questions that were considered relevant to the case. In both cases, the internal consistency was fair (0.80 and

0.61). When analysing the subscales, levels 1, 2 and 3, the internal consistency was low for levels 1 and 2, and higher for level 3. This could indicate that the perception of the situation in the groups was diverse and not related to the total score, while the ability to project the direction in which the cases were heading was more homogeneously related to the total score. TSAGAT had a higher overall homogeneity, as might be expected when analysing the means of the group SAGAT for each question, instead of the specific SAGAT answers.

The feasibility of using SAGAT to measure situation awareness was assessed by measuring the length of the freeze (i.e. pause) needed to answer questions and by asking the participants for their perception of the pauses in the scenario and how the pauses affected the training session. According to Endsley, when measuring SA using SAGAT, the scenario is frozen while the participants answer questions that probe the three levels of SA: perception, comprehension and projection.[21] Each pause in this study lasted less than 3 minutes. The pauses and SAGAT questions could influence SA in both a negative direction and a positive direction. In the negative direction, the flow of the simulation training might be disrupted, stress be induced by being forced to interrupt the case, and the commitment decrease. According to the PSQ, the majority of the participants did not perceive that the pauses adversely affected concentration and performance. Contrary, the pauses and questions might facilitate the resolution of the clinical problem in the case by triggering and allowing time for reflection, fully in line with Schön's reflection-on-action.[24]

Limitations of this study

The participants in the study comprised year four medical students for the video recordings of simulation-based team-training and the rating of the video material was conducted by four participants representing a wide spectrum of experience and training. It could be argued that it is a limitation that this validation was not performed on a series of critical care teams, for example. However, for testing the reliability as well as the feasibility of the checklist,

techniques and instruments, the simulation-based training with medical students was a readily available series of standardized simulations.

Difficulties were encountered in the interpretation of what constitutes partially performed versus performed completely in the ABCDE checklist. This could relate to the vast differences in the raters' level of education and previous experience. To improve the accuracy and minimize the variability, the raters were trained using a separate set of video recordings. TEAM was not translated into Swedish in order to avoid inducing errors. This was possible as fluency in English is a prerequisite for academic studies in Sweden. The ratings based on the original TEAM instrument were consistent between the raters, indicating that this may have been a correct assumption.

In the present study, CVI for SAGAT was measured according to Lawshe [32] while the development of the ABCDE checklist and the two cases relied on the authors' experiences and iterative interactions with clinicians and experts in the field. Both the cases, the ABCDE checklists and the developed SAGAT questionnaires, could have benefitted from a full formal CVI by a review panel.

The PSQ was only translated from English and not translated back into Swedish in order to formally check the identity of the items in the final sets. However, the translation was adjusted by a professional translator before being used in the study. Thus, the results of the PSQ can be used for probing the participants in the simulation with regards to their experiences of SAGAT.

This study focused on the quantification of performance during simulation-based training. The transferability of the studied behaviours into a real-world setting is an intriguing question for further studies.

Conclusion

In this setting with medical students, situation awareness, team and task performance could be assessed with techniques that were reliable and feasible. The developed ABCDE checklist and the TEAM instrument had high interrater reliability. The process of using SAGAT questionnaires during simulation-based training did not negatively affect the participants' evaluation of simulation-based training, and the developed SAGAT questionnaires had a fair internal consistency. Thus, the measurement of task performance, team performance and situation awareness may be conducted in future studies in a Swedish simulation-based training setting using these techniques.

Table 1 Background characteristics of the medical students (n=55) participating in the study

Age years, m (sd)	25 (4.5)
Male n (%)	28 (51)
Female n (%)	27 (49)
Previous healthcare education	
None n (%)	44 (80)
Assistant nurse (2Y upper secondary school) n (%)	2 (4)
Registered nurse (3Y university) n (%)	3 (5)
Short courses (Red Cross, etc.) n (%)	6 (11)
Previous team training	
Yes n (%)	20 (36)
Previous simulation experience	
Yes n (%)	32 (58)
Previous experience of trauma patients*	
Yes n (%)	20 (37)

*one missing value

Table 2A Situation awareness items for Case A. Response scale, level and proportion correct (%) (n=39).

Freeze 1	Question	Response scale	Level*	Correct
1.	What is the patient's respiratory rate?	Free text	1	26%
2.	What are the findings from the assessment of A and B?	Free text	1	51%
3.	Is the patient sufficiently oxygenated?	Yes/No	2	49%
4.	What did the patient state as their year of birth?	Free text	1	36%
5.	What do you think will happen to the patient's blood pressure in the next few minutes?	Increase/Decrease/No change	3	74%
6.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	69%
7.	How many peripheral IV cannulas does the patient have?	Free text	1	95%
	Presuming an ordinary course of development, how will vital signs change in			

	the next 10 minutes?			
8.	Heart rate	Increase/Decrease	3	85%
9.	Blood pressure	Increase/Decrease	3	82%
10.	Saturation	Increase/Decrease	3	59%
11.	Respiratory rate	Increase/Decrease	3	74%

Freeze 2	Question	Response scale	Level*	Correct
12.	What is the patient's blood pressure?	Free text	1	77%
13.	What are your findings under C (Circulation)?	Free text	1	82%
14.	What additional examinations/tests do you think are needed?	Free text	3	67%
15.	Which medications/drugs do you think are needed?	Free text	3	49%
16.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	95%
17.	Is there an algorithm for CPR on the wall?	Yes/No	1	26%
18.	What previous illnesses has the patient described?	Free text	2	21%
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?			
19.	Heart rate	Increase/Decrease	3	74%
20.	Blood pressure	Increase/Decrease	3	74%
21.	Saturation	Increase/Decrease	3	59%
22.	Respiratory rate	Increase/Decrease	3	69%

The questions are shown translated into English in the table, and the original questions in Swedish are available upon request from the authors of the study.

*Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

Table 2B Situation awareness items for Case B. Response scale, level and proportion correct (%) (n=36).

Freeze 1	Question	Response scale	Level*	Correct
1.	What is the patient's respiratory rate?	Free text	1	28%
2.	What are the findings from the assessment of A and B?	Free text	1	64%
3.	Is the patient sufficiently oxygenated?	Yes/No	2	78%
4.	What does the sign on the patient's necklace indicate?	Free text	1	36%
5.	What do you believe will happen to the patient's venous return in the next few minutes?	Increase/Decrease/No change	3	61%
6.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	58%
7.	Is a suction device available?	Yes/No	1	69%
	Presuming an ordinary course of development, how will vital signs			

	change in the next 10 minutes?			
8.	Heart rate	Increase/Decrease	3	78%
9.	Blood pressure	Increase/Decrease	3	58%
10.	Saturation	Increase/Decrease	3	78%
11.	Respiratory rate	Increase/Decrease	3	72%

Freeze 2	Question	Response scale	Level*	Correct
12.	What is the patient's respiratory rate?	Free text	1	33%
13.	What are your findings under C (Circulation)?	Free text	1	69%
14.	What additional examinations/tests do you think are needed?	Free text	3	92%
15.	Which medications/drugs do you think are needed?	Free text	3	22%
16.	What is wrong with the patient (preliminary diagnosis)?	Free text	2	83%
17.	What is the patient's blood pressure?	Free text	1	53%
18.	Is a defibrillator available in the room?	Yes/No	2	50%
	Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?			
19.	Heart rate	Increase/Decrease	3	72%
20.	Blood pressure	Increase/Decrease	3	72%
21.	Saturation	Increase/Decrease	3	69%
22.	Respiratory rate	Increase/Decrease	3	56%

The questions are shown translated into English in the table, and the original questions in Swedish are available upon request from the authors of the study.

*Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection

Table 3 Time (minutes) for scenario freezes to measure situation awareness with SAGAT

Case	No. of teams	Scenario	1st freeze	2nd freeze	Total time
	n	m (sd)	m (sd)	m (sd)	m (sd)
A	12	14.1 (2.0)	2.7 (0.5)	2.5 (0.6)	19.3 (2.3)
B	11	14.1 (3.4)	2.7 (0.6)	2.8 (0.6)	19.7 (3.8)

Table 4 Student agreement about usefulness and feasibility of the prospective collection of situation awareness items

Question	Agree/ strongly agree n (%)	Disagree/ strongly disagree n (%)	Do not know n (%)
Concerning the SAGAT questionnaire. In my opinion:			
The introduction was adequate	24 (100)	0 (0)	0 (0)
The questions were clear	23 (96)	1 (4)	0 (0)
The “freezes” in the scenario adversely affected my concentration level and my performance	7 (29)	17 (71)	0 (0)
The questions were relevant to the way I perceived the scenario	23 (96)	1 (4)	0 (0)
Overall, I am satisfied with SAGAT as a tool for evaluating hands-on skills during a trauma exercise.	19 (79)	1 (4)	4 (17)

Results from the post-simulation questionnaire

Table 5 Descriptives of ABCDE checklist, TEAM instrument and SAGAT questionnaire

	Case A m (sd)	Case B m (sd)
ABCDE checklist[1]	2.70 (0.19)	2.58 (0.27)
TEAM instrument		
Sum of items 1-11[2]	25.5 (5.9)	25.0 (5.1)
Mean item score 1-11[1]	2.32 (0.54)	2.27 (0.46)
Mean subscore Leadership (items 1-2)[1]	2.27 (0.77)	2.32 (0.65)
Mean subscore Team Work (items 3-9)[1]	2.28 (0.55)	2.23 (0.51)
Mean subscore Task Management (items 10-11)[1]	2.49 (0.56)	2.38 (0.50)
Global rating[3]	4.90 (1.13)	4.75 (1.04)
SAGAT questionnaire[4]	13.95 (4.26)	12.60 (3.35)

[1]Max score per item 4.

[2]Max score 44.

[3]Max score 11.

[4]Max score 22.

Table 6 Internal consistency (Cronbach's alpha) of SAGAT and TSAGAT

SAGAT/TSAGAT	Case A n/n 39/12	Case B n/n 36/11
Level 1	0.060/0.372	0.246/0.420
Level 2	0.321/-0.018	0.332/0.332
Level 3	0.891/0.891	0.659/0.786
Total (Level 1–3)	0.800/0.827	0.620/0.759

Internal consistency calculated for the questions measuring the three levels of SA and for the questionnaire in total.

n/n, number of individuals answering the SAGAT questionnaires / number of teams

References

1 Makary MA, Daniel M. Medical error—the third leading cause of death in the US. *BMJ* 2016;**353**:i2139. doi:10.1136/bmj.i2139

2 Kohn LT, Corrigan JM, Donaldson MS. *To Err is Human: Building a Safer Health System*. National Academies Press 2000. doi:10.17226/9728

3 World Health Organization. Multi-professional Patient Safety Curriculum Guide. WHO, Geneva. 2011.https://apps.who.int/iris/bitstream/handle/10665/44641/9789241501958_eng.pdf

4 Scalese RJ, Obeso VT, Issenberg BS. Simulation Technology for Skills Training and Competency Assessment in Medical Education. *J Gen Intern Med* 2007;**23**:46–9. doi:10.1007/s11606-007-0283-4

5 Zendejas B, Brydges R, Wang AT, *et al.* Patient outcomes in simulation-based medical education: a systematic review. *Journal of General Internal Medicine* 2013;**28**:1078–89.

6 Gordon M, Darbyshire D, Baker P. Non-technical skills training to enhance patient safety: a systematic review. *Med Educ* 2012;**46**:1042–54. doi:10.1111/j.1365-2923.2012.04343.x

7 Salas E, Rosen MA, Held JD, *et al.* Performance Measurement in Simulation-Based Training. *Simulat Gaming* 2009;**40**:328–76. doi:10.1177/1046878108326734

8 Rosen MA, Pronovost PJ. Advancing the Use of Checklists for Evaluating Performance in Health Care. *Acad Med* 2014;**89**:963–5. doi:10.1097/acm.0000000000000285

9 Brady PW, Muething S, Kotagal U, *et al.* Improving situation awareness to reduce unrecognized clinical deterioration and serious safety events. *Pediatrics* 2013;**131**:e298-308. doi:10.1542/peds.2012-1364

10 Salas E, Cooke NJ, Rosen MA. On Teams, Teamwork, and Team Performance: Discoveries and Developments. *Hum Factors J Hum Factors Ergonomic Soc* 2008;**50**:540–7. doi:10.1518/001872008x288457

11 Kozlowski SW. Enhancing the Effectiveness of Work Groups and Teams: A Reflection. *Perspect Psychol Sci* 2018;**13**:205–12. doi:10.1177/1745691617697078

12 Converse SA, Cannon-Bowers JA, Salas E. Team Member Shared Mental Models: A Theory and Some Methodological Issues. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 1991;**35**:1417–21. doi:10.1177/154193129103501917

13 Mathieu J, Heffner T, Goodwin G, *et al.* The influence of shared mental models on team process and performance. *The Journal of applied psychology* 2000;**85**:273–83.

14 Boyle WA, Murray DJ, Beyatte M, *et al.* Simulation-Based Assessment of Critical Care "Front-Line" Providers*. *Critical Care Medicine* 2018;**46**:e516–22. doi:10.1097/CCM.0000000000003073

15 Holcomb JB, Dumire RD, Crommett JW, *et al.* Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. *J Trauma* 2002;**52**:1078–85.

16 Cooper S, Cant R, Porter J, *et al.* Rating medical emergency teamwork performance: Development of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;**81**:446–52.

doi:10.1016/j.resuscitation.2009.11.027

17 Cooper SJ, Cant RP. Measuring non-technical skills of medical emergency teams: An update on the validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2014;**85**:31–3. doi:10.1016/j.resuscitation.2013.08.276

18 Cooper S, Cant R, Connell C, *et al.* Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation* 2016;**101**:97–101. doi:10.1016/j.resuscitation.2016.01.026

19 Schulz CM, Krautheim V, Hackemann A, *et al.* Situation awareness errors in anesthesia and critical care in 200 cases of a critical incident reporting system. *BMC Anesthesiol* 2016;**16**:4. doi:10.1186/s12871-016-0172-7

20 Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors* 1995;**37**:32–64. doi:10.1518/001872095779049543

21 Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. In: *Situation awareness analysis and measuement*. Lawrence Erlbaum Associates 2000. 1–21.

22 Hogan MP, Pace DE, Hapgood J, *et al.* Use of human patient simulation and the situation awareness global assessment technique in practical trauma skills assessment. *J Trauma* 2006;**61**:1047–52. doi:10.1097/01.ta.0000238687.23622.89

23 Hamstra SJ, Brydges R, Hatala R, *et al.* Reconsidering fidelity in simulation-based training. *Acad Med* 2014;**89**:387–92. doi:10.1097/ACM.0000000000000130

24 Schön DA. *The reflective practitioner: How professionals think in action*. The Perseus Books Group 1984.

25 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc J Soc Simul Healthc* 2017;**12**:17–21. doi:10.1097/sih.0000000000000181

26 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident's Situational Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3. doi:10.1097/BOT.0000000000000003

27 Lavoie P, Cossette S, Pepin J. Testing nursing students' clinical judgment in a patient deterioration simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today* 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015

28 McKenna L, Missen K, Cooper S, *et al.* Situation awareness in undergraduate nursing students managing simulated patient deterioration. *Nurs Educ Today* 2014;**34**:e27–31. doi:10.1016/j.nedt.2013.12.013

29 Crozier MS, Ting HY, Boone DC, *et al.* Use of human patient simulation and validation of the Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009

30 Thim T, Krarup, Grove, *et al.* Initial assessment and treatment with the Airway, Breathing, Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012;:117. doi:10.2147/IJGM.S28478

31 ATLS Advanced Trauma Life Support Student Course Manual. American College of Surgeons. 2018.

32 Lawshe CH. A quantitative approach to content validity. *Pers Psychol* 1975;**28**:563–75. doi:10.1111/j.1744-6570.1975.tb01393.x

33 Salas E, Prince C, Baker DP, *et al.* Situation Awareness in Team Performance: Implications for Measurement and Training. *Hum Factors J Hum Factors Ergonomics Soc* 1995;**37**:123–36. doi:10.1518/001872095779049525

34 Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. *Behavior research methods, instruments, & computers* 1996;**28**:1–11. doi:10.3758/BF03203630

35 Nunnally JC. *Psychometric Theory, 2nd edition*. McGraw-Hill 1978.

36 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine* 2016;**15**:155–63. doi:10.1016/j.jcm.2016.02.012

37 Schweizer K. Some guidelines concerning the modeling of traits and abilities in test construction. *European Journal of Psychological Assessment* 2010;**26**. doi:10.1027/1015-5759/a000001

38 McKay A, Walker ST, Brett SJ, *et al.* Team performance in resuscitation teams: comparison and critique of two recently developed scoring tools. *Resuscitation* 2012;**83**:1478–83. doi:10.1016/j.resuscitation.2012.04.015

39 Lindgren S, Brännström T, Hanse E, *et al.* Medical education in Sweden. *Medical teacher*;**33**:798–803. doi:10.3109/0142159X.2011.570816