

# BMJ Open

BMJ Open is committed to open peer review. As part of this commitment we make the peer review history of every article we publish publicly available.

When an article is published we post the peer reviewers' comments and the authors' responses online. We also post the versions of the paper that were used during peer review. These are the versions that the peer review comments apply to.

The versions of the paper that follow are the versions that were submitted during the peer review process. They are not the versions of record or the final published versions. They should not be cited or distributed as the published version of this manuscript.

BMJ Open is an open access journal and the full, final, typeset and author-corrected version of record of the manuscript is available on our site with no access controls, subscription charges or pay-per-view fees (<http://bmjopen.bmj.com>).

If you have any questions on BMJ Open's open peer review process please email [info.bmjopen@bmj.com](mailto:info.bmjopen@bmj.com)

# BMJ Open

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

Journal:	<i>BMJ Open</i>
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; Umeå Universitet Medicinska fakulteten, Department of Nursing Lindkvist, Marie; Umeå Universitet Medicinska fakulteten, Department of Public Health and Clinical Medicine Burlin, Christine; Umeå Universitet Medicinska fakulteten, Department of Nursing
Keywords:	Teamwork, Situation awareness, Taskwork

SCHOLARONE™  
Manuscripts

1  
2  
3  
4 The feasibility of A-E, TEAM and SAGAT for measurement of situational  
5 awareness and team performance in critical care simulation training  
6  
7  
8  
9  
10

11 Magnus Hultin, MD PhD, Associate Professor, Department of Surgical and Perioperative Sciences,  
12 Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden.

13 <https://orcid.org/0000-0003-2935-7161>  
14

15  
16 Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative Sciences,  
17 Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå, Sweden

18 <https://orcid.org/0000-0001-7974-6777>  
19

20 Maria Härgestam, RNA PhD, Lecturer, Department of Nursing, Umeå University, S-901 87 Umeå,  
21 Sweden  
22

23 Marie Lindkvist, PhD, Associate Professor, Department of Public Health and Clinical Medicine, Umeå  
24 University, S-901 87 Umeå, Sweden  
25

26 Christine Brulin, RNA PhD, Professor, Department of Nursing, Umeå University, S-901 87 Umeå,  
27 Sweden  
28

29  
30  
31 Abstract 300 words  
32

33  
34 Main article 4364 words  
35  
36  
37  
38  
39

40 **Keywords**

41 SAGAT, TEAM, resuscitation  
42  
43  
44  
45

46 **Corresponding author**

47 Magnus Hultin, Dept of Surgical and Perioperative Sciences, Anesthesiology and Critical Care  
48 Medicine, Umeå University, Sweden, magnus.hultin@umu.se, phone: +46-90-786 9109  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

---

## 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  $\chi^2$ ) 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  $\chi^2$  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

1  
2  
3 setting.[24] SAGAT has not to our knowledge previously been used in a Swedish context, and  
4 therefore it is of importance to evaluate both the feasibility and the internal consistency of a Swedish  
5 version of the instrument.  
6

7 To summarise, A–E checklists (based on the TTET), TEAM, and SAGAT have been developed in  
8 order to evaluate different aspects of teamwork. However, these instruments have neither been  
9 translated into Swedish nor tested for feasibility or trustworthiness in a Swedish context. Such studies  
10 are necessary so as to further evaluate teamwork in acute care settings and in simulation-based  
11 trainings. Thus, the aims of this study are (1) to test an A–E checklist developed by us for interrater  
12 reliability, (2) to test the TEAM instrument for interrater reliability, (3) to test a Swedish version of the  
13 SAGAT instrument for feasibility, and (4) to test SAGAT for internal consistency.  
14  
15

## 16 Method

17 This cross-sectional study is based on quantitative data collected during video-recorded simulation-  
18 based team training sessions.  
19  
20

### 21 Setting

22 Data collection in this study was carried out during videotaped simulation team training at the Clinical  
23 Training Centre (CTC) of the Medical Faculty, Umeå university. The week before the team training,  
24 the students were asked to watch a twelve-minute video available at the learning platform that  
25 introduced both the ABCDE-concept when caring for a patient and the simulation setting.[25] At the  
26 start of the trainings session, the students participated in a fifteen-minute live introduction to the  
27 simulation laboratory by the operator and the instructor.  
28  
29

30 Each team trained on four scenarios focusing on assessment and treatment of severely ill patients in an  
31 emergency room. The first scenario was a warm up and the last was a teaching summing up. The  
32 second and third scenarios were included in the study as case 1 and 2. The order of the two cases was  
33 randomized at a group level just before the start of scenario two. Both cases were designed to last 10–  
34 15 minutes and were pre-programmed into a Laerdal SimMan simulator to support the standardization  
35 of the simulation. Different handlers were added for changes in vital signs, e.g. proportionally  
36 increased saturation and blood pressure when oxygen and fluid were administered, respectively.  
37  
38

39 The cases used were in essence slightly modified versions of the scenarios used by Hogan.[27] Case 1  
40 was a 25-year old male with hypovolemic shock after a traffic incident. Case 2 was set as a 35-year  
41 male suffering from pneumothorax with affected vital signs after a traffic incident.  
42

43 Observations were made during the simulation and field notes were taken by one of the authors (KJ).  
44 Two cameras mounted at an angle were used to record videos of the room and one of the views  
45 included the patient monitor.  
46

### 47 Participants

48 During March to October 2016, all medical students (n=68) in year 4 doing their clerkship in  
49 anaesthesiology and critical care medicine were invited to participate in the study while doing  
50 compulsory simulator-based team training. Initially, the students were divided into 27 teams. In total,  
51 55 students (28 males and 27 females) in a total of 23 teams participated in the study (Table 1).  
52 Thirteen students in four teams were excluded from the study as one or more in the team did not want  
53 to participate in this study.  
54  
55

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

- 57 • First, the analyses of inter-rater reliability for the TEAM instrument was based on ratings from  
58 the videotaped records by four professionals (nurse n=2, physician n=1, and paramedic n=1)  
59 (Table 2).  
60

- 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



1  
2  
3 The raters independently watched the videos of the simulation scenarios included in this study and  
4 rated the performances using the A–E checklist and the TEAM instrument. Each video was viewed at  
5 least twice by each rater. An online form was used to collect the data from the raters so as to minimize  
6 the risk of transfer errors when moving data from paper to an Excel file.  
7

## 8 SAGAT 9

10 When developing questions for the SAGAT-instrument to a specific scenario, goal directed task  
11 analysis was used as originally described by Endsley. Briefly, for each profession, the major goals are  
12 identified along with the sub-goals. Then key decisions are identified, and SA requirements are  
13 defined as the dynamic information needed to achieve the major goals, as opposed to static  
14 information such as rules and guidelines.[26] The samplings (questions) are then matched against the  
15 SA requirements. The original recommendation by Endsley was 30–60 samplings (questions) for  
16 within-subjects studies for each of the three SA levels; (1) perception and attention, (2)  
17 comprehension, and (3) projection.[15] When Gardner et al. validated a SAGAT instrument in a study  
18 of medical trainees, each questionnaire consisted of three questions for each level of SA at each  
19 freeze.[20,26,27] SAGAT with three questions at level 1, one at level 2 and three at level 3.  
21

22 In the present study, the SAGAT instrument was refined and adapted to the scenario and expected  
23 skills level of the students according to the process used by Hogan et al. [27] First, the Hogan SAGAT  
24 instrument was translated into Swedish by the authors of this study. In accordance with the original  
25 instrument, the targeted learning objectives for the training scenarios were then formulated and next  
26 the specific goals for each simulation was set. An iterative process was used to reformulate the  
27 SAGAT questions in Swedish, using a separate group of 6 professionals, all registered nurses, working  
28 both in the clinical context and in the teaching context. The final sets of SAGAT questions are shown  
29 in Tables 4A and 4B. Lawshe advocated the use of professionals to identify whether an item is  
30 important in a certain context. [28] Content validity index is the fraction of professionals that rates the  
31 item as important. In the present study, the relevance of the SAGAT questions in relation to the  
32 scenarios in the team training was reviewed by three professionals (nurse n=1, physician n=2) before  
33 being used in the study. All agreed that the questions were relevant, i.e. the content validity index was  
34 high.  
35

36  
37 The answers given by the participants in the SAGAT instrument were classified as incorrect (0) or  
38 correct (1) by two of the authors (MHu, KJ). The classifications were discussed and agreed upon by  
39 the two authors. For answers on a continuous scale (e.g. systolic blood pressure) a 10% range around  
40 the intended correct answer was accepted as correct. One question was removed from the  
41 questionnaire, since it became obvious during the classification process that the question had  
42 frequently been misinterpreted.  
43

44  
45 In order to be able to probe SA with the SAGAT instrument, the scenarios were frozen (i.e. paused)  
46 twice. The first freeze of the scenario was five minutes into the scenario, unless there was an active  
47 task activity or if the team was doing a team re-evaluation.[25] If so, the freeze was postponed for a  
48 moment. During the freeze, the patient monitor was switched off and the participants turned away  
49 from the patient simulator while individually answering the questions. The second freeze took place  
50 according to the same principles after an additional 5 minutes. All participants were allowed to  
51 complete the SAGAT questions before the scenario started again.  
52

53  
54 The length of the freeze was measured using the video recordings. The start of the freeze was defined  
55 as the beginning of the sentence “Now we pause the scenario, so that you can answer some questions  
56 about the patient case” and the end of the freeze as the end of the sentence “Everyone in place and  
57 ready? Now we start again”. The time taken to measure SAGAT was defined as the end-time minus  
58 the start time and was measured in seconds. The simulations lasted 14.1 (sd 2.7) minutes and the  
59 freeze needed for all team members to answer SAGAT was 2.7 (sd 0.6) minutes (Table 3).  
60

1  
2  
3 The feasibility of using SAGAT to measure situation awareness was assessed by measuring the  
4 length of the freeze (i.e. pause) needed to answer questions and by asking the participants for their  
5 perception of the pauses in the scenario and how the pauses affected the training session.  
6

7 Team SAGAT (TSAGAT) was developed by Crozier et al. based on SAGAT as an assessment tool to  
8 evaluate team performance.[24] In Crozier's study, each team consisted of a trauma leader, an airway  
9 manager and a nurse and individual SAGAT instruments were developed for the three team positions  
10 including both shared knowledge and complimentary knowledge. TSAGAT was calculated as the sum  
11 of individual SAGAT scores, and the TSAGAT scores had a high correlation to a traditional checklist  
12 (Pearson correlation,  $r=0.996$ ).[24] However, Salas defined team SA as a dynamic process defined as  
13 the team's shared understanding of a situation at a specific point in time [6] and Endsley argued that  
14 team SA involves unique activities as information sharing and coordination.[15] Thus, in the present  
15 study, in order to measure this efficiently in all team members, all participants in a case received  
16 identical SAGAT instruments. TSAGAT was calculated based on the mean scores in the team for each  
17 SAGAT item.  
18  
19

### 20 Post-simulation questionnaire (PSQ)

21 In order to measure whether the SAGAT items were considered relevant to the scenarios, and whether  
22 pausing the scenarios affected the team training activity, a post-simulation questionnaire (PSQ) was  
23 used.[27] The PSQ consists of 13 statements to be rated on a four-point scale ranging from Strongly  
24 disagree (1) to Strongly agree (4), where 5 statements concern the SAGAT and the effect of freezing  
25 the scenario and 8 statements concern the simulation and the scenario per se. As far as we have found,  
26 no data regarding the reliability of PSQ was presented in the original study. In this study, we only use  
27 the five questions relating to SAGAT and freezing the scenario. The PSQ was translated into Swedish  
28 by a professional translator and the translation was further refined based on iterative discussions  
29 within our research group. In the second week after the training session, the PSQ was answered by 24  
30 of the participating students (Table 5). In this study, the four-point scale was dichotomized into  
31 disagree (Strongly disagree and Disagree) (1) and agree (Agree and Strongly agree) (2).  
32  
33  
34

### 35 Study size

36 The aim of the study was to evaluate feasibility of A-E, TEAM and SAGAT for use in further studies.  
37 For this, the study sample needs to be large enough to allow for calculation of descriptive statistics  
38 (means and standard deviations) and for calculation of reliabilities with a fair precision. To be able to  
39 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  
40 would need 45 participants in each group as determined with G\*Power.[30]  
41  
42

### 43 Statistical analysis

44 The statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM,  
45 Chicago, IL). Interrater reliability for the A-E checklist and the TEAM instrument was determined by  
46 intraclass correlation with a two-way random effects model (ICC (2,1) type absolute).[29,31] The  
47 results are reported as both single measures and average measures, since the ICC for single measures  
48 answers to how equal the individual ratings are, while average measures answers to how reliable the  
49 mean values of the individual ratings are. The results of the PSQ are presented by descriptive statistics.  
50 SAGAT was calculated both as the mean of each SAGAT question in each team defined as TSAGAT  
51 and also at an individual level defined as SAGAT.[24] Internal consistency for the SAGAT instrument  
52 and TSAGAT was measured by Cronbach's alpha.[29,31] Internal consistency considered as the  
53 extent to which all items measure the same latent variable was investigated by  $\chi^2$  and as suggested by  
54 Schweizer, a normed  $\chi^2$  below 2 was taken as an indication of a good fit.[32]  
55  
56  
57  
58  
59  
60

## 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  $\chi^2$  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 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

1  
2  
3 measurements if the plan is to use averaged measurements.[31] In this study, we chose to report both  
4 values to show the importance of using multiple raters to improve overall reliability. In the original  
5 study by Cooper, an interrater reliability for the TEAM instrument of 0.55 and 0.60 was reported as  
6 measured by Cohen's kappa (adjusted for chance) and ICC, respectively.[13] Thus the interrater  
7 reliability as measured by ICC (single measurements) in this study was equivalent to the ICC obtained  
8 in the original setting.  
9

10 Both the PSQ and the CVI indicate that the SAGAT instrument could be used to construct questions  
11 that were considered relevant to the case. In addition, the internal consistency was fair (0.80 and 0.61).  
12 When analysing the subscales, level 1, 2 and 3, the internal consistency was low for level 1 and 2, and  
13 higher for level 3. This might indicate that the perception of the situation in the groups was scattered  
14 and not related to the total score, while the ability to project where the cases were heading was more  
15 homogeneously related to the total score. TSAGAT had a higher overall homogeneity, as might be  
16 expected when analysing the means of the group SAGAT for each question instead of the individual  
17 SAGAT answers.  
18  
19

20 Measurement of situation awareness using SAGAT requires that the scenario is paused while the  
21 participants answer questions probing the three levels of SA: perception, comprehension, and  
22 projection.[26] Each pause in this study were less than 3 minutes. The pauses and SAGAT questions  
23 might influence SA both in negative direction and in a positive direction. This might disrupt the flow  
24 in the simulation, induce stress and impede the suspension of disbelief. It might also facilitate the  
25 resolution of the clinical problem in the case by triggering reflection-on-action and reflection-in-  
26 action.[19] In the PSQ, the pauses in this study did not cause any adverse reactions on the part of the  
27 participants.  
28  
29

### 30 Limitations of this study

31 To obtain a convenient sample to analyse the feasibility of using the A–E checklist, TEAM, SAGAT,  
32 and TSAGAT, trauma team trainings with medical students were used to collect questionnaires and  
33 video for further analysis. Using teams at different skills levels, instead of having a rather  
34 homogeneous group of medical students, might have created wider range of ratings and thus have  
35 improved the psychometrics. However, as the purpose was to test the interrater reliability of A–E and  
36 TEAM as well as the feasibility and internal consistency of SAGAT, this study most likely does not  
37 overestimate the psychometric properties.  
38  
39

40 As the SAGAT questions are case/specific, the content validity is rather difficult to assess and  
41 generalize. A systematic use of content validity index (CVI) as suggested by Lawshe might be a  
42 purposeful method.[28] Using professionals to rate whether the item is essential and then calculating  
43 CVI as the fraction that rated the item as essential allows for a systematic screening of a larger set of  
44 potential items.  
45  
46

47 The negative result for TSAGAT homogeneity level 2 in one of the cases indicates that there is a  
48 structural problem such as too few questions or measurements in that category. Since level 2 is  
49 represented by only a pair of questions, it is not unlikely. As pointed out by Tavakol et al., Cronbach's  
50 alpha should be considered as a lower bound estimate of reliability.[31] The number of questions in  
51 each level in the SAGAT tool needs to be balanced between time taken to administer the test and  
52 optimal sampling of a trait.  
53  
54  
55

### 56 Conclusion

57 It was possible to develop a homogeneous SAGAT instrument aimed at measuring situation awareness  
58 during trauma team training with patient simulators also in a Swedish setting. The TEAM and A–E  
59  
60

1  
2  
3 checklist had a high interrater reliability, indicating that the descriptors in the instruments were  
4 similarly interpreted by the raters. SAGAT and TSAGAT had fair internal consistencies. To conclude,  
5 it is feasible to use TEAM, A-E and SAGAT/TSAGAT for measurement of team performance and  
6 situational awareness in critical care simulation training  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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 m (sd)	1 <sup>st</sup> freeze m (sd)	2 <sup>nd</sup> freeze m (sd)	Total time 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)

For peer review only



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? <i>Yes or No</i>	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? <i>Increase/Decrease/No-change</i>	3
6.	What is wrong with the patient? <i>State your preliminary diagnosis</i>	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.	<i>HR Increase/Decrease</i>	3
9.	<i>BP Increase/Decrease</i>	3
10.	<i>POx Increase/Decrease</i>	3
11.	<i>RR Increase/Decrease</i>	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?	2
	Presuming a normal course of development, how will the vital signs change in the upcoming 10 minutes?	
19.	<i>HR Increase/Decrease</i>	3
20.	<i>BP Increase/Decrease</i>	3
21.	<i>POx Increase/Decrease</i>	3
22.	<i>RR Increase/Decrease</i>	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? <i>Yes or No</i>	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? <i>Increase/Decrease/No change</i>	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.	<i>HR Increase/Decrease</i>	3
9.	<i>BP Increase/Decrease</i>	3
10.	<i>POx Increase/Decrease</i>	3
11.	<i>RR Increase/Decrease</i>	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 examinations/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.	<i>HR Increase/Decrease</i>	
20.	<i>BP Increase/Decrease</i>	3
21.	<i>POx Increase/Decrease</i>	3
22.	<i>RR Increase/Decrease</i>	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 n (%)	Disagree n (%)	Don't know n (%)
Concerning the SAGAT instrument, in my view:			
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
<b>Total (Level 1-3)</b>	<b>0.800 / 0.827</b>	<b>0.620 / 0.759</b>

## 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

- 1  
2  
3 of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;**81**:446–52.  
4 doi:10.1016/j.resuscitation.2009.11.027  
5  
6  
7 14 Cant RP, Porter JE, Cooper SJ, *et al.* Improving the non technical skills of hospital medical  
8 emergency teams: The Team Emergency Assessment Measure (TEAM). *Emerg Med Australas*  
9 2016;**28**:641–6. doi:10.1111/1742-6723.12643  
10  
11 15 Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*  
12 1995;**37**:32–64. doi:10.1518/001872095779049543  
13  
14 16 Jones DG, Endsley MR. Sources of situation awareness errors in aviation. *Aviation, space, and*  
15 *environmental medicine* 1996.  
16  
17  
18 17 Durso FT, Bleckley M, Dattel AR. Does situation awareness add to the validity of cognitive tests?  
19 *Human factors* 2006;**48**:721–33. doi:10.1518/001872006779166316  
20  
21 18 Hamstra SJ, Brydges R, Hatala R, *et al.* Reconsidering fidelity in simulation-based training.  
22 *Academic medicine : journal of the Association of American Medical Colleges*;**89**:387–92.  
23 doi:10.1097/ACM.000000000000130  
24  
25  
26 19 Schön DA. *The reflective practitioner: How professionals think in action.* The Perseus Books  
27 Group 1984.  
28  
29 20 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the  
30 SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc* 2016;**12**:1–5.  
31 doi:10.1097/SIH.000000000000181  
32  
33  
34 21 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident's Situational  
35 Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3.  
36 doi:10.1097/BOT.000000000000003  
37  
38  
39 22 Lavoie P, Cossette S, Pepin J. Testing nursing students' clinical judgment in a patient deterioration  
40 simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today*  
41 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015  
42  
43 23 McKenna L, Missen K, Cooper S, *et al.* Situation awareness in undergraduate nursing students  
44 managing simulated patient deterioration. *Nurse Educ Today* 2014;**34**:e27–31.  
45 doi:10.1016/j.nedt.2013.12.013  
46  
47  
48 24 Crozier MS, Ting HY, Boone DC, *et al.* Use of human patient simulation and validation of the  
49 Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team  
50 assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009  
51  
52 25 American College of Surgeons. *ATLS Student Course Manual: Advanced Trauma Life Support.* 9th  
53 ed. American College of Surgeons 2013.  
54  
55 26 Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. In: *Situation*  
56 *awareness analysis and measurement.* Lawrence Erlbaum Associates 2000. 1–21.  
57  
58  
59 27 Hogan MP, Pace DE, Hapgood J, *et al.* Use of human patient simulation and the situation  
60 awareness global assessment technique in practical trauma skills assessment. *J Trauma* 2006;**61**:1047–



1  
2  
3 52. doi:10.1097/01.ta.0000238687.23622.89  
4

5 28 Lawshe CH. A quantitative approach to content validity. *Personnel Psychology* 1975;**28**:563–75.  
6 doi:10.1111/j.1744-6570.1975.tb01393.x  
7

8 29 Nunnally JC. *Psychometric Theory, 2nd edition*. McGraw-Hill 1978.  
9

10 30 Faul F, Erdfelder E, Buchner A, *et al*. Statistical power analyses using G\*Power 3.1: tests for  
11 correlation and regression analyses. *Behavior research methods* 2009;**41**:1149–60.  
12 doi:10.3758/BRM.41.4.1149  
13

14 31 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for  
15 Reliability Research. *Journal of chiropractic medicine* 2016;**15**:155–63.  
16 doi:10.1016/j.jcm.2016.02.012  
17

18 32 Tavakol M, Dennick R. Making sense of Cronbach's alpha. *Int J Medical Educ* 2011;**2**:53–5.  
19 doi:10.5116/ijme.4dfb.8dfd  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

	Item No	Recommendation	Page No
<b>Title and abstract</b>	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
<b>Introduction</b>			
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
<b>Methods</b>			
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.
<b>Results</b>			
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	59
<b>Discussion</b>			
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
<b>Other information</b>			
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](http://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:	<i>BMJ Open</i>
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; Umeå 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
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Communication
Keywords:	Patient care team, Resuscitation, Situation Awareness, Task performance and analysis, Teamwork

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5 Are instruments that measure situation awareness, team and task performance  
6  
7  
8 reliable and feasible for a simulation training setting?  
9  
10  
11  
12  
13

14 Magnus Hultin, MD Ph.D., Associate Professor, Department of Surgical and Perioperative  
15 Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå,  
16  
17 Sweden. <https://orcid.org/0000-0003-2935-7161>  
18  
19

20  
21 Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative  
22 Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå,  
23  
24 Sweden. <https://orcid.org/0000-0001-7974-6777>  
25  
26

27  
28 Maria Härgestam, RNA Ph.D., Lecturer, Department of Nursing, Umeå University, S-901 87  
29 Umeå, Sweden. <https://orcid.org/0000-0003-0680-9962>  
30  
31

32 Marie Lindkvist, Ph.D., Associate Professor, Department of Epidemiology and Global Health,  
33 Umeå University, S-901 87 Umeå, Sweden  
34  
35

36  
37 Christine Brulin, RNA Ph.D., Professor emeritus, Department of Nursing, Umeå University,  
38 S-901 87 Umeå, Sweden  
39  
40

41  
42  
43 Abstract 247 words  
44

45 Main article 4,828 words  
46  
47  
48

#### 49 Keywords

- 50 • Patient Care Team
- 51
- 52 • Resuscitation
- 53
- 54 • Situation Awareness
- 55
- 56 • Task Performance and Analysis
- 57
- 58 • Teamwork
- 59
- 60

1  
2  
3  
4  
5 Corresponding author

6 Magnus Hultin, Dept of Surgical and Perioperative Sciences, Anesthesiology and Critical

7  
8 Care Medicine, Umeå University, Sweden, magnus.hultin@umu.se, phone: +46-90-786 9109  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 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 4<sup>th</sup> 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  $\chi^2$  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,

1  
2  
3  
4 teamwork and task management.[16,17] The instrument was initially developed and validated  
5  
6 for simulator-based team training and was recently validated for the collection of  
7  
8 observational ratings of non-technical skills during live resuscitations in emergency  
9  
10 departments. [18] Scoring scenarios that use instruments depend on reliable interpretations of  
11  
12 the instrument by the raters and it might be even more important to ensure the reliability of  
13  
14 such interpretations when using an instrument in a non-native language.  
15  
16 Moreover, and in addition to task and team performance, SA is a prerequisite for patient  
17  
18 safety and the prevention of errors, particularly in acute care situations.[9,19] SA includes  
19  
20 three levels of ability: (1) perception and attention (*What?*, what's going on), (2)  
21  
22 comprehension (*So what?*, the ability to understand what's going on), and (3) projection  
23  
24 (*What's next?*) in order to anticipate and plan for future events).[20] In order to measure SA in  
25  
26 a simulation setting, the Situation Awareness Global Assessment Tool (SAGAT) has been  
27  
28 developed and subsequently adapted for use in healthcare settings.[21,22] One feature of  
29  
30 SAGAT is that its use requires the simulation to pause, which might influence clinical  
31  
32 understanding, both by impeding the suspension of disbelief in the simulation setting [23] and  
33  
34 by facilitating reflection on action.[24] Gardner showed that it was feasible to use SAGAT to  
35  
36 measure SA in the team training of surgical trainees in advanced cardiac life support.[25]  
37  
38 Globally, SAGAT has been used to study, for example, the effect of sleep deprivation on SA  
39  
40 in trauma team training, how SA is associated with surgical trainee team performance, as well  
41  
42 as nurses' clinical assessment of patient deterioration.[26–28] SAGAT can be analysed on an  
43  
44 individual level and also on a team level. A specific application of SAGAT is team SAGAT  
45  
46 (TSAGAT) in which the different members of a team answer SAGAT questions specific to  
47  
48 their role. TSAGAT was constructed to account for the teams' shared situation awareness and  
49  
50 validated in trauma team training in a Canadian setting.[29] To our knowledge, SAGAT has  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 not previously been used in a Swedish context and it is therefore important to evaluate both its  
6  
7 feasibility and its internal consistency.

8  
9 To summarise, ABCDE checklists, the TEAM instrument and SAGAT have been developed  
10  
11 in order to evaluate different aspects of teamwork. However, these instruments and  
12  
13 questionnaires have neither been translated into Swedish nor tested for feasibility or  
14  
15 trustworthiness in a Swedish context. Such studies are necessary to enable the evaluation of  
16  
17 teamwork in Swedish acute care settings and simulation-based training. Thus, this study  
18  
19 aimed to test an ABCDE checklist that we developed for interrater reliability, to test the  
20  
21 TEAM instrument for interrater reliability, and to test two Swedish SAGAT questionnaires  
22  
23 for feasibility and internal consistency.  
24  
25  
26  
27  
28  
29

## 30 Method

31  
32 This study is based on data collected during simulation-based team training sessions with  
33  
34 medical students.

### 35 Participants

36  
37 From March to October 2016, all medical students (n=68) in year 4 undertaking their  
38  
39 clerkship in anaesthesiology and critical care medicine were invited to participate in the study  
40  
41 while receiving mandatory simulator-based team training. In total, 55 students (81%)  
42  
43 participated in the study (Table 1) and 20 of them participated in both scenario A and scenario  
44  
45 B.  
46  
47

48  
49 For interrater reliability of the post-scenario rating of videos from the simulations, four raters  
50  
51 from different backgrounds were involved. First, a registered nurse with a Master's Degree  
52  
53 (one year) in Nursing (Critical Care Medicine), 20 years' working experiences at an ICU, nine  
54  
55 years' experiences of human patient simulator team training and also a PhD student (author  
56  
57 KJ). Second, a physician with 20 years' working experience, associate professor and  
58  
59  
60

1  
2  
3  
4 consultant in Anesthesiology and Critical Care Medicine, with 14 years' experiences of  
5 human patient simulator team training (author MHu). Third, a registered nurse with a  
6 Master's Degree (one year) in Nursing (Acute Care Medicine), with six years' working  
7 experiences of prehospital care and one year as a medical student. Fourth, a paramedic with  
8 limited experiences beyond two years working as a paramedic in the Israeli Army and one  
9 year as a medical student.  
10  
11  
12  
13  
14  
15  
16  
17

### 18 Setting

19 Data collection in this study was carried out during simulation-based team training at the  
20 Clinical Training Centre (CTC) of the Medical Faculty, Umeå University. The briefings,  
21 scenarios and debriefings were conducted in Swedish. Two cameras mounted at an angle were  
22 used to record videos in the simulation room and one of the views included the patient  
23 monitor.  
24  
25  
26  
27  
28  
29  
30

31 One week before team training, the students were asked to watch a twelve-minute video  
32 available at the learning platform that introduced both the ABCDE concept when caring for a  
33 patient and the room and equipment to be used during simulation-based training.[30,31] At the  
34 start of the training session, the students participated in a 15-minute live introduction to the  
35 simulation laboratory presented by the operator and the instructor. Each student group (4–5  
36 students/group) trained on four scenarios that focused on assessment and treatment of  
37 severely ill patients in an emergency room. The students all played the role of interns, i.e.  
38 physicians who are in training to become licensed, with the attending nurse currently  
39 unavailable. The assigned task involved conducting a primary survey and stabilising the  
40 patient until more senior staff arrived in 15 minutes. In each scenario, three students were  
41 active and 1–2 were observers. After each scenario, a 10–15 minute debriefing session  
42 permitted reflection and shared learning. The first scenario was a warm-up, and the last  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 included a training summary. Case A and B, or in the reverse order, as determined by the toss  
5 of a coin immediately before the second scenario, were used as the second and third scenarios.  
6  
7 All scenarios were conducted in Swedish, were designed to last 10–15 minutes and were pre-  
8 programmed into a Laerdal SimMan simulator in order to support the standardisation of the  
9 simulation. In essence, the patient cases used were slightly modified versions of the scenarios  
10 used by Hogan.[22] Case A was hypovolemic shock following a traffic incident. Case B was a  
11 pneumothorax with affected vital signs following a traffic incident.  
12  
13  
14  
15  
16  
17  
18  
19

#### 20 Background questionnaire

21 The background questionnaire included informed consent and was answered immediately  
22 before team training started. The questionnaire included questions such as year of birth,  
23 male/female, previous medical training, previous experiences of team training, previous  
24 experiences of human patient simulator-based training, previous experiences of CRM and  
25 previous experiences of live trauma care.  
26  
27  
28  
29  
30  
31  
32  
33

#### 34 ABCDE checklist

35 In order to measure the completeness of critical tasks in acute care scenarios, an ABCDE  
36 checklist was used. The original TTET comprises 58 items derived from the ATLS  
37 protocol.[15] Each item in the TTET was scrutinised by the authors of this study and the  
38 number of items reduced to those that reflected the measures expected to be carried out by  
39 year four medical students in acute care scenarios. The final list contained ten items. The  
40 items are well described in major textbooks and were translated into Swedish in line with the  
41 nomenclature being used. Compared to the original scoring system, an additional scoring  
42 option was added: performed after a reminder from the instructor. Each item was rated on a  
43 Likert scale of 0 to 4 (0 = Not initiated, 1= Performed after a reminder from the instructor, 2 =  
44 Partially performed, 3 = Performed completely before the end of the simulation, and 4 =  
45 Performed consistently during the whole simulation, NA = not applicable).  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

#### 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



1  
2  
3  
4  
5 n=1, physician n=2) before being used in the study. All professionals rated the questions as  
6  
7 relevant, i.e. the CVI was 1.0.  
8

9 The answers given by the participants in the SA questionnaire were classified as incorrect (0)  
10  
11 or correct (1). The classifications were discussed and agreed upon by the two authors (KJ and  
12  
13 MHu). For answers on a continuous scale (e.g. systolic blood pressure), a 10% range around  
14  
15 the intended correct answer was accepted as correct. One question was removed from the  
16  
17 questionnaire since it became apparent during the classification process that the question had  
18  
19 frequently been misinterpreted.  
20  
21

22  
23 In order to administer the SA questionnaires, the scenarios were frozen (i.e. paused) twice.

24  
25 The first freeze of the scenario was five minutes into the scenario, unless there was an active  
26  
27 task activity or if the team was conducting a team re-evaluation, in which case, the freeze was  
28  
29 briefly postponed. During the freeze, the patient monitor was switched off and the participants  
30  
31 turned away from the patient simulator while individually answering the questions. The  
32  
33 second freeze took place according to the same principles after an additional five minutes. All  
34  
35 participants were allowed to complete the questionnaire before the scenario re-started.  
36  
37

38  
39 The length of the freezes was measured from the video recordings (Table 3). The start of the  
40  
41 freeze was defined as the beginning of the sentence “Now we will pause the scenario so that  
42  
43 you can answer some questions about the patient case” and the end of the freeze was defined  
44  
45 as the end of the sentence “Is everyone in place and ready? Now the scenario will re-start”.

46  
47 Team SAGAT (TSAGAT) was developed by Crozier et al. [29] based on SAGAT as an  
48  
49 assessment tool for evaluating team performance. In Crozier’s study, each team comprised a  
50  
51 trauma leader, airway manager and nurse. Individual SA questionnaires were developed for  
52  
53 the three-team roles including both shared knowledge and complementary knowledge.  
54  
55

56  
57 TSAGAT was calculated as the sum of individual SAGAT scores and the TSAGAT scores  
58  
59 had a high correlation to a traditional checklist (Pearson correlation,  $r=0.996$ ). However, Salas  
60

1  
2  
3  
4 defined team SA as a dynamic process defined as the team's shared understanding of a  
5 situation at a specific point in time [33] while Endsley argued that team SA involves unique  
6 activities as information sharing and coordination.[20] Thus, in the present study, in order to  
7 measure this efficiently in all team members, all participants in a case received identical SA  
8 questionnaires. TSAGAT was calculated based on the mean scores of the team for each SA  
9 item.  
10  
11  
12  
13  
14  
15  
16  
17

#### 18 Post-simulation questionnaire

19 In order to measure whether the items in the SA questionnaire were considered relevant to the  
20 scenarios, and whether pausing the scenarios affected the team training activity, a post-  
21 simulation questionnaire (PSQ) was used.[22] The PSQ comprises 13 statements to be rated  
22 on a four-point scale ranging from Strongly disagree (1) to Strongly agree (4), in which five  
23 statements concern the SAGAT and the effect of freezing the scenario and eight statements  
24 concern the simulation and the scenario *per se*. To the best of our knowledge, no data  
25 regarding the reliability of PSQ were presented in the original study. In this study, we only  
26 use the five questions relating to SAGAT and freezing the scenario. The PSQ was translated  
27 into Swedish by one of the authors and the translation was further refined based on iterative  
28 discussions within our research group. Next, the Swedish PSQ was sent to a professional  
29 translation agency together with the original PSQ for verification of the translation. The PSQ  
30 was answered in an anonymous web survey during the second week after the simulation  
31 training by 24 of the 55 participating students, i.e. the response rate was 44% (Table 4). In this  
32 study, the results of the four-point scale were dichotomised into disagree (Strongly disagree  
33 and Disagree) and agree (Agree and Strongly agree).  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53

#### 54 Study size

55 The aim of the study was to evaluate the feasibility of ABCDE, TEAM and SAGAT for use in  
56 further studies. In order to achieve this, the study sample must be large enough to allow for,  
57  
58  
59  
60

1  
2  
3  
4 with a fair precision, the calculation of descriptive statistics (means and standard deviations)  
5  
6 and the calculation of reliabilities. In order to assess a large effect size (Cohen's d of 0.9) with  
7  
8 t-test with a power of 80% at the 0.05 level, 16 participants per group would be needed as  
9  
10 determined by G\*Power.[34] Thus, for the aims of the present study, the inclusion of 50  
11  
12 individuals and 20 teams would suffice.  
13  
14

#### 15 16 Statistical analysis

17  
18 Statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM,  
19  
20 Chicago, IL). Interrater reliability for the ABCDE checklist and the TEAM instrument was  
21  
22 determined by intraclass correlation using a two-way random effects model (ICC (2,1) type  
23  
24 absolute).[35,36] The results are reported both as single measures and average measures since  
25  
26 the ICC for single measures relates to the reliability of the individual ratings, while average  
27  
28 measures relates to the reliability of the mean values. Cronbach's alpha was used to measure  
29  
30 the internal consistency of the SAGAT instrument and TSAGAT.[35] Internal consistency  
31  
32 considered as the extent to which all items measure the same latent variable was investigated  
33  
34 by  $\chi^2$  and as suggested by Schweizer, a normed  $\chi^2$  below 2 was taken as an indication of a  
35  
36 good fit.[37]  
37  
38  
39

#### 40 41 Ethics

42  
43 This study was approved by the Ethical Review Board of Northern Sweden (Chairman:  
44  
45 Anders Iacobaeus, 7 April, 2016, decision no. 2016-54-31M).  
46

#### 47 48 Patient and Public Involvement

49  
50 Neither patients nor the public were involved in the design or the data collection for this  
51  
52 study.  
53

#### 54 55 Results

##### 56 57 Scores on ABCDE checklist, TEAM instrument and SA items

58  
59 The scores on the ABCDE checklist, TEAM instrument and the SA items are shown in Table  
60  
5. The mean ABCDE score was 2.6 on a scale of 0–4, the mean TEAM item 1–11 score was

1  
2  
3  
4  
5 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  
6  
7 the mean SA item score was 13 with a maximal score of 22.

8  
9 Interrater reliability for the ABCDE checklist and TEAM instrument

10 Interrater reliability as measured by intraclass correlation was 0.55 (single measures)/0.83  
11  
12 (average measures) for the ABCDE checklist and 0.54/0.83 for the TEAM scale. For the  
13  
14 TEAM subscales of leadership, team work, task management and global rating, the intraclass  
15  
16 correlations were 0.36/0.70, 0.45/0.77, 0.35/0.68 and 0.38 /0.72, respectively.

17  
18  
19 The feasibility of using SAGAT to measure situation awareness was assessed by measuring  
20  
21 the length of the freeze (i.e. pause) needed to answer questions and by asking the participants  
22  
23 for their perception of the pauses in the scenario and how the pauses affected the training  
24  
25 session.

26  
27  
28 Feasibility and internal consistency of SAGAT

29  
30 The PSQ showed that 96% of the participants considered the SA items to be relevant to the  
31  
32 case and 96% considered the questions to be easy to understand (Table 4). About three out of  
33  
34 four (72%) participants stated that the freeze did not negatively impact their concentration or  
35  
36 performance during the simulation session.

37  
38  
39 The internal consistency of SAGAT measured as Cronbach's alpha was 0.80 for case A and  
40  
41 0.62 for case B (Table 5), and normed  $\chi^2$  was 1.72 vs. 1.62. For level 1 (perception),  
42  
43 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  
44  
45 and 0.66, respectively.

46  
47  
48 For TSAGAT, the internal consistency was good for the instrument as a total and level 3  
49  
50 (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 to 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

1  
2  
3  
4  
5 one rater only, and ICC for average measurements if the plan is to use multiple raters. In this  
6  
7 study, we chose to report both values in order to show the importance of using multiple raters  
8  
9 to improve overall reliability. In the original study by Cooper, an interrater reliability for the  
10  
11 TEAM instrument of 0.55 and 0.60 was reported as measured by Cohen's kappa (adjusted for  
12  
13 chance) and ICC, respectively.[16] Thus, interrater reliability as measured by ICC (single  
14  
15 measurements) in this study was equivalent to the ICC obtained in the original setting.  
16  
17

18 Both the PSQ and the CVI indicate that the SAGAT instrument could be used to construct  
19  
20 questions that were considered relevant to the case. In any case, the internal consistency was  
21  
22 fair (0.80 and 0.61). When analysing the subscales, levels 1, 2 and 3, the internal consistency  
23  
24 was low for levels 1 and 2, and higher for level 3. This could indicate that the perception of  
25  
26 the situation in the groups was diverse and not related to the total score, while the ability to  
27  
28 project the direction in which the cases were heading was more homogeneously related to the  
29  
30 total score. TSAGAT had a higher overall homogeneity, as might be expected when analysing  
31  
32 the means of the group SAGAT for each question, instead of the specific SAGAT answers.  
33  
34

35 Measuring SA using SAGAT requires that the scenario is paused while the participants  
36  
37 answer questions that probe the three levels of SA: perception, comprehension and  
38  
39 projection.[21] Each pause in this study lasted less than 3 minutes. The pauses and SAGAT  
40  
41 questions could influence SA in both a negative direction and a positive direction. In the  
42  
43 negative direction, the flow of the simulation training might be disrupted, stress be induced by  
44  
45 being forced to interrupt the case, and the commitment decrease. In the positive direction, the  
46  
47 pauses and questions might facilitate the resolution of the clinical problem in the case by  
48  
49 triggering and allowing time for reflection. This could be in line with Schön's reflection-on-  
50  
51 action and reflection-in-action.[24] According to the PSQ, the pauses in this study did not  
52  
53 cause any adverse reactions on the part of the participants.  
54  
55  
56  
57  
58  
59  
60

#### 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.

1  
2  
3  
4  
5 In the present study, CVI was performed for SAGAT while the development of the ABCDE  
6  
7 checklist and the two scenarios relied on the authors' experiences and iterative interactions  
8  
9 with clinicians and experts in the field. Both the scenarios, the ABCDE checklists and the  
10  
11 developed SAGAT questionnaires, could have benefitted from a full formal CVI by a review  
12  
13 panel.  
14

15  
16 The PSQ was only translated from English and not translated back into Swedish in order to  
17  
18 formally check the identity of the items in the final sets. However, the translation was  
19  
20 adjusted by a professional translator before being used in the study. Thus, the results of the  
21  
22 PSQ can be used for probing the participants in the simulation with regards to their  
23  
24 experiences of SAGAT.  
25  
26

27  
28 The participants in the study comprised year four medical students for the video recordings of  
29  
30 simulation-based team-training and the rating of the video material was conducted by four  
31  
32 participants representing a wide spectrum of experience and training. It could be argued that it  
33  
34 is a limitation that this validation was not performed on a series of critical care teams, for  
35  
36 example. However, for testing the reliability as well as the feasibility the techniques and  
37  
38 instruments, the simulation-based trainings with medical students was a readily available  
39  
40 series of standardized simulations.  
41  
42

43  
44 The raters in this study included both experienced and inexperienced raters. The study was not  
45  
46 designed to systematically test for differences in ratings between raters of different  
47  
48 backgrounds. However, the inhomogeneity of the raters is a strength as it indicates that the  
49  
50 checklists and instruments can be used by a group of raters with different experiences.  
51  
52

## 53 54 55 Conclusion

56  
57 The developed ABCDE checklist and the TEAM instrument had high interrater reliability, the  
58  
59 process of using SAGAT during simulation-based training did not affect the participants'  
60



1  
2  
3  
4 perception of simulation-based training and the developed SAGAT questionnaires had a fair  
5  
6 internal consistency. Thus, the measurement of task performance, team performance and  
7  
8 situation awareness in future studies in a Swedish simulation-based training setting may be  
9  
10 conducted using these techniques.  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

**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?  Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?	Free text	1
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?		3
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	Scenario	1 <sup>st</sup> freeze	2 <sup>nd</sup> freeze	Total time
<b>o</b>	<b>n</b>	<b>m (sd)</b>	<b>m (sd)</b>	<b>m (sd)</b>	<b>m (sd)</b>
<b>Case A</b>	12	14.1 (2.0)	2.7 (0.5)	2.5 (0.6)	19.3 (2.3)
<b>Case B</b>	11	14.1 (3.4)	2.7 (0.6)	2.8 (0.6)	19.7 (3.8)

For peer review only

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

Question	Agree	Disagree	Do not know
	n (%)	n (%)	n (%)
<b>Concerning the SAGAT questionnaire.</b>			
<b>In my opinion:</b>			
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

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

\*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

<b>SAGAT/TSAGAT</b>	<b>Case A</b>	<b>Case B</b>
	<b>n/n</b>	<b>n/n</b>
	<b>39/12</b>	<b>36/11</b>
<b>1</b>	0.060/0.372	0.246/0.420
<b>2</b>	0.321/-0.018	0.332/0.332
<b>3</b>	0.891/0.891	0.659/0.786
<b>Total (Level 1–3)</b>	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](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*

1  
2  
3  
4  
5 *Psychol Sci* 2018;**13**:205–12. doi:10.1177/1745691617697078  
6  
7

8 12 Converse SA, Cannon-Bowers JA, Salas E. Team Member Shared Mental Models: A Theory and  
9 Some Methodological Issues. *Proceedings of the Human Factors and Ergonomics Society Annual*  
10 *Meeting* 1991;**35**:1417–21. doi:10.1177/154193129103501917  
11  
12

13 13 Mathieu J, Heffner T, Goodwin G, *et al.* The influence of shared mental models on team process  
14 and performance. *J Appl Psych* 2000;**85**:273–83.  
15  
16

17  
18 14 Boyle WA, Murray DJ, Beyatte M, *et al.* Simulation-Based Assessment of Critical Care “Front-  
19 Line” Providers\*. *Crit Care Med* 2018;**46**:e516–22. doi:10.1097/CCM.0000000000003073  
20  
21

22  
23 15 Holcomb JB, Dumire RD, Crommett JW, *et al.* Evaluation of trauma team performance using an  
24 advanced human patient simulator for resuscitation training. *J Trauma* 2002;**52**:1078–85  
25  
26

27  
28 16 Cooper S, Cant R, Porter J, *et al.* Rating medical emergency teamwork performance: Development  
29 of the Team Emergency Assessment Measure (TEAM). *Resuscitation* 2010;**81**:446–52.  
30  
31 doi:10.1016/j.resuscitation.2009.11.027  
32  
33

34  
35 17 Cooper SJ, Cant RP. Measuring non-technical skills of medical emergency teams: An update on the  
36 validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation*  
37 2014;**85**:31–3. doi:10.1016/j.resuscitation.2013.08.276  
38  
39

40  
41 18 Cooper S, Cant R, Connell C, *et al.* Measuring teamwork performance: Validity testing of the  
42 Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation*  
43 2016;**101**:97–101. doi:10.1016/j.resuscitation.2016.01.026  
44  
45

46  
47 19 Schulz CM, Krautheim V, Hackemann A, *et al.* Situation awareness errors in anesthesia and critical  
48 care in 200 cases of a critical incident reporting system. *BMC Anesthesiol* 2016;**16**:4.  
49  
50 doi:10.1186/s12871-016-0172-7  
51  
52

53  
54 20 Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. *Hum Factors*  
55 1995;**37**:32–64. doi:10.1518/001872095779049543  
56  
57

58 21 Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. In: *Situation*  
59 *awareness analysis and measurement*. Lawrence Erlbaum Associates 2000. 1–21.  
60

- 1  
2  
3  
4  
5  
6 22 Hogan MP, Pace DE, Hapgood J, *et al.* Use of human patient simulation and the situation  
7 awareness global assessment technique in practical trauma skills assessment. *J Trauma* 2006;**61**:1047–  
8 52. doi:10.1097/01.ta.0000238687.23622.89  
9  
10  
11  
12 23 Hamstra SJ, Brydges R, Hatala R, *et al.* Reconsidering fidelity in simulation-based training. *Acad*  
13 *Med* 2014;**89**:387–92. doi:10.1097/ACM.0000000000000130  
14  
15  
16  
17 24 Schön DA. *The reflective practitioner: How professionals think in action.* The Perseus Books  
18 Group 1984.  
19  
20  
21  
22 25 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the  
23 SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc* 2017;**12**:17–21.  
24 doi:10.1097/sih.0000000000000181  
25  
26  
27  
28 26 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident’s Situational  
29 Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3.  
30 doi:10.1097/BOT.0000000000000003  
31  
32  
33  
34 27 Lavoie P, Cossette S, Pepin J. Testing nursing students’ clinical judgment in a patient deterioration  
35 simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today*  
36 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015  
37  
38  
39  
40  
41 28 McKenna L, Missen K, Cooper S, *et al.* Situation awareness in undergraduate nursing students  
42 managing simulated patient deterioration. *Nurs Educ Today* 2014;**34**:e27–31.  
43 doi:10.1016/j.nedt.2013.12.013  
44  
45  
46  
47 29 Crozier MS, Ting HY, Boone DC, *et al.* Use of human patient simulation and validation of the  
48 Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team  
49 assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009  
50  
51  
52  
53 30 Thim T, Krarup, Grove, *et al.* Initial assessment and treatment with the Airway, Breathing,  
54 Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012;:117.  
55 doi:10.2147/IJGM.S28478  
56  
57  
58  
59 31 ATLS Advanced Trauma Life Support Student Course Manual. 2018.  
60

- 1  
2  
3  
4  
5  
6 32 Lawshe CH. A quantitative approach to content validity. *Personnel Psychology* 1975;**28**:563–75.  
7 doi:10.1111/j.1744-6570.1975.tb01393.x  
8  
9  
10  
11 33 Salas E, Prince C, Baker DP, *et al.* Situation Awareness in Team Performance: Implications for  
12 Measurement and Training. *Hum Factors* 1995;**37**:123–36. doi:10.1518/001872095779049525  
13  
14  
15 34 Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. *Behavior research*  
16 *methods, instruments, & computers* 1996;**28**:1–11. doi:10.3758/BF03203630  
17  
18  
19  
20 35 Nunnally JC. *Psychometric Theory, 2nd edition.* McGraw-Hill 1978.  
21  
22  
23 36 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for  
24 Reliability Research. *J Chiropr Med* 2016;**15**:155–63. doi:10.1016/j.jcm.2016.02.012  
25  
26  
27  
28 37 Schweizer K. Some guidelines concerning the modeling of traits and abilities in test construction.  
29 *Eur J Psychol Assess* 2010;**26**. doi:10.1027/1015-5759/a000001  
30  
31  
32  
33 38 Jacobsson M, Härgestam M, Hultin M, *et al.* Flexible knowledge repertoires: communication by  
34 leaders in trauma teams. *Scand J Trauma Resusc Emerg Medicine* 2012;**20**:44. doi:10.1186/1757-  
35 7241-20-44  
36  
37  
38  
39 39 Härgestam M, Lindkvist M, Brulin C, *et al.* Communication in interdisciplinary teams: exploring  
40 closed-loop communication during in situ trauma team training. *BMJ Open* 2013;**3**:e003525.  
41 doi:10.1136/bmjopen-2013-003525  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

	Item No	Recommendation	Page No
<b>Title and abstract</b>	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
<b>Introduction</b>			
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
<b>Methods</b>			
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) <i>Cohort study</i> —Give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up <i>Case-control study</i> —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 <i>Cross-sectional study</i> —Give the eligibility criteria, and the sources and methods of selection of participants	9-10
		(b) <i>Cohort study</i> —For matched studies, give matching criteria and number of exposed and unexposed <i>Case-control study</i> —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) <i>Cohort study</i> —If applicable, explain how loss to follow-up was addressed <i>Case-control study</i> —If applicable, explain how matching of cases and controls was addressed <i>Cross-sectional study</i> —If applicable, describe analytical methods taking account of sampling strategy	-
		(e) Describe any sensitivity analyses	

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

Continued on next page

For peer review only



1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

<b>Results</b>			
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) <i>Cohort study</i> —Summarise follow-up time (eg, average and total amount)	-
Outcome data	15*	<i>Cohort study</i> —Report numbers of outcome events or summary measures over time	-
		<i>Case-control study</i> —Report numbers in each exposure category, or summary measures of exposure	-
		<i>Cross-sectional study</i> —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
<b>Discussion</b>			
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
<b>Other information</b>			
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](http://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:	<i>BMJ Open</i>
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; Umeå Universitet Medicinska fakulteten, Department of Nursing Lindkvist, Marie; Umeå Universitet Medicinska fakulteten, Department of Public Health and Clinical Medicine Burlin, Christine; Umeå Universitet Medicinska fakulteten, Department of Nursing
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Communication
Keywords:	Patient care team, Resuscitation, Situation Awareness, Task performance and analysis, Teamwork

SCHOLARONE™  
Manuscripts

1  
2  
3  
4 The reliability of instruments that measure situation awareness, team and task performance in  
5 a simulation setting with medical students  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17

18 Magnus Hultin, MD Ph.D., Associate Professor, Department of Surgical and Perioperative  
19 Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå,  
20 Sweden. <https://orcid.org/0000-0003-2935-7161>  
21  
22

23 Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative  
24 Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå,  
25 Sweden. <https://orcid.org/0000-0001-7974-6777>  
26  
27

28 Maria Härgestam, RNA Ph.D., Lecturer, Department of Nursing, Umeå University, S-901 87  
29 Umeå, Sweden. <https://orcid.org/0000-0003-0680-9962>  
30  
31

32 Marie Lindkvist, Ph.D., Associate Professor, Department of Epidemiology and Global Health,  
33 Umeå University, S-901 87 Umeå, Sweden  
34  
35

36 Christine Brulin, RNA Ph.D., Professor emerita, Department of Nursing, Umeå University, S-  
37 901 87 Umeå, Sweden  
38  
39

40  
41  
42  
43  
44  
45  
46  
47  
48 **Abstract** 246 words  
49

50  
51 **Main article** 4,837 words  
52  
53

54  
55 **Keywords**  
56

- 57 • Patient Care Team
- 58 • Resuscitation
- 59
- 60

- 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

For peer review only

## 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 4<sup>th</sup> 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  $\chi^2$  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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 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.

1  
2  
3  
4  
5 Team performance can be measured using the Team Emergency Assessment Measure  
6  
7 (TEAM), for example. It measures three dimensions of team performance: leadership,  
8  
9 teamwork and task management.[16,17] The instrument was initially developed and validated  
10  
11 for simulator-based team training and was recently validated for the collection of  
12  
13 observational ratings of non-technical skills during live resuscitations in emergency  
14  
15 departments.[18] Using instruments to score scenarios depend on reliable interpretations of  
16  
17 the instrument by the raters and it might be even more important to ensure the reliability of  
18  
19 such interpretations when using an instrument in a non-native language.  
20  
21

22  
23 Moreover, and in addition to task and team performance, SA is a prerequisite for patient  
24  
25 safety and the prevention of errors, particularly in acute care situations.[9,19] SA includes  
26  
27 three levels of ability: (1) perception and attention (*What?*, what's going on), (2)  
28  
29 comprehension (*So what?*, the ability to understand what's going on), and (3) projection  
30  
31 (*What's next?*) in order to anticipate and plan for future events.[20] In order to measure SA in  
32  
33 a simulation setting, the Situation Awareness Global Assessment Tool (SAGAT) has been  
34  
35 developed and subsequently adapted for use in healthcare settings.[21,22] One feature of  
36  
37 SAGAT is that its use requires the simulation to pause[21], which might influence clinical  
38  
39 understanding, both by impeding the suspension of disbelief in the simulation setting [23] and  
40  
41 by facilitating reflection on action.[24] Gardner showed that it was feasible to use SAGAT to  
42  
43 measure SA in the team training of surgical trainees in advanced cardiac life support.[25]  
44  
45 Globally, SAGAT has been used to study, for example, the effect of sleep deprivation on SA  
46  
47 in trauma team training, how SA is associated with surgical trainee team performance, as well  
48  
49 as nurses' clinical assessment of patient deterioration.[26–28] SAGAT can be analysed on an  
50  
51 individual level and also on a team level. A specific application of SAGAT is team SAGAT  
52  
53 (TSAGAT) in which the different members of a team answer SAGAT questions specific to  
54  
55 their role. TSAGAT was constructed to account for the teams' shared situation awareness and  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 validated in trauma team training in a Canadian setting.[29] To our knowledge, SAGAT has  
6  
7 not previously been used in a Swedish context and it is therefore important to evaluate both its  
8  
9 feasibility and its internal consistency.

10  
11 To summarise, ABCDE checklists, the TEAM instrument and SAGAT have been developed  
12  
13 in order to evaluate different aspects of teamwork. However, these instruments and  
14  
15 questionnaires have neither been translated into Swedish nor tested for feasibility or  
16  
17 trustworthiness in a Swedish context. Such studies are necessary to enable the evaluation of  
18  
19 teamwork in Swedish acute care settings and simulation-based training. Thus, this study  
20  
21 aimed to test an ABCDE checklist that we developed for interrater reliability, to test the  
22  
23 TEAM instrument for interrater reliability, and to test two Swedish SAGAT questionnaires  
24  
25 for feasibility and internal consistency.  
26  
27  
28  
29  
30  
31  
32

## 33 Method

34  
35 This study is based on data collected during simulation-based team training sessions with  
36  
37 medical students.  
38  
39

### 40 Participants and raters

41  
42 From March to October 2016, all medical students (n=68) in year 4 undertaking their  
43  
44 clerkship in anaesthesiology and critical care medicine were invited to participate in the study  
45  
46 while receiving mandatory simulator-based team training. In total, 55 students (81%)  
47  
48 participated in the study (Table 1) and 20 of them participated in both scenario A and scenario  
49  
50  
51 B.

52  
53 All scenarios were videorecorded and later scored by four raters to allow for calculation of  
54  
55 interrater reliability. First, a registered nurse with a Master's Degree (one year) in Nursing  
56  
57 (Critical Care Medicine), 20 years' working experiences at an ICU, nine years' experiences of  
58  
59  
60

1  
2  
3  
4 human patient simulator team training and also a PhD student (author KJ). Second, a  
5  
6 physician with 20 years' working experience, associate professor and consultant in  
7  
8 Anesthesiology and Critical Care Medicine, with 14 years' experiences of human patient  
9  
10 simulator team training (author MHu). Third, a registered nurse with a Master's Degree (one  
11  
12 year) in Nursing (Acute Care Medicine), with six years' working experiences of prehospital  
13  
14 care and one year as a medical student. Fourth, a paramedic with limited experiences beyond  
15  
16 two years working as a paramedic in the Israeli Army and one year as a medical student.  
17  
18  
19

## 20 21 Setting

22  
23 Data collection in this study was carried out during simulation-based team training at the  
24  
25 Clinical Training Centre (CTC) of the Medical Faculty, Umeå University. The briefings,  
26  
27 scenarios and debriefings were conducted in Swedish. Two cameras mounted at an angle were  
28  
29 used to record videos in the simulation room and one of the views included the patient  
30  
31 monitor.  
32  
33

34  
35 One week before team training, the students were asked to watch a twelve-minute video  
36  
37 available at the learning platform that introduced both the ABCDE concept when caring for a  
38  
39 patient and the room and equipment to be used during simulation-based training.[30,31] At  
40  
41 the start of the training session, the students participated in a 15-minute live introduction to  
42  
43 the simulation laboratory presented by the operator and the instructor. Each student group (4–  
44  
45 5 students/group) trained on four scenarios that focused on assessment and treatment of  
46  
47 severely ill patients in an emergency room. In each scenario, 3-4 students were active and 1–2  
48  
49 were observers. Therefore, only 3-4 participated in the interactive role play in each scenario  
50  
51 (the actual simulation). In all, the 55 unique students made up a total of 23 teams with 3-4  
52  
53 participants in each team and 20 students participated in both case A and case B. The students  
54  
55 all played the role of interns, i.e. physicians who are in training to become licensed, with the  
56  
57 attending nurse currently unavailable. The assigned task involved conducting a primary  
58  
59  
60

1  
2  
3  
4  
5 survey and stabilising the patient until more senior staff arrived in 15 minutes. After each  
6  
7 scenario, a 10–15 minute debriefing session permitted reflection and shared learning. The first  
8  
9 scenario was a warm-up, and the last included a training summary. Case A and B, or in the  
10  
11 reverse order, as determined by the toss of a coin immediately before the second scenario,  
12  
13 were used as the second and third scenarios.  
14  
15

16 All scenarios were conducted in Swedish, were designed to last 10–15 minutes and were pre-  
17  
18 programmed into a Laerdal SimMan simulator in order to support the standardisation of the  
19  
20 simulation. In essence, the patient cases used were slightly modified versions of the scenarios  
21  
22 used by Hogan.[22] Case A was hypovolemic shock following a traffic incident. Case B was a  
23  
24 pneumothorax with affected vital signs following a traffic incident.  
25  
26  
27

#### 28 Background questionnaire

29  
30 The background questionnaire included informed consent and was answered immediately  
31  
32 before team training started. The questionnaire included questions such as year of birth,  
33  
34 male/female, previous medical training, previous experiences of team training, previous  
35  
36 experiences of human patient simulator-based training, previous experiences of CRM and  
37  
38 previous experiences of live trauma care.  
39  
40  
41

#### 42 ABCDE checklist

43  
44 In order to measure the completeness of critical tasks in acute care scenarios, an ABCDE  
45  
46 checklist was used. The original TTET comprised 58 items derived from the ATLS  
47  
48 protocol.[15] Each item in the TTET was scrutinised by the authors of this study and the  
49  
50 number of items reduced to those that reflected the measures expected to be carried out by  
51  
52 year four medical students in the acute care scenarios in this study. Some items were also  
53  
54 slightly modified to better reflect The final list contained ten items reflecting the key  
55  
56 elements of the ATLS primary survey, i.e. the core management of ABCDE: airway assessed,  
57  
58  
59  
60

1  
2  
3  
4  
5 airway secured, saturation assessed, oxygen applied, ventilation assessed, ventilation  
6  
7 optimised, pulses checked (radial – femoral – carotid), venous access and infusions  
8  
9 established, neurological disabilities checked (consciousness, pupils), and full/complete  
10  
11 exposure. The items are well described in major textbooks and were translated into Swedish  
12  
13 in line with the nomenclature being used. Compared to the original scoring system, an  
14  
15 additional scoring option was added: performed after a reminder from the instructor. Each  
16  
17 item was rated on a rating scale of 0 to 4 (0 = Not initiated, 1= Performed after a reminder  
18  
19 from the instructor, 2 = Partially performed, 3 = Performed completely before the end of the  
20  
21 simulation, and 4 = Performed consistently during the whole simulation, NA = not  
22  
23 applicable).

#### 24 25 26 27 28 TEAM instrument

29  
30 In order to measure team leadership, teamwork and task management, the TEAM instrument  
31  
32 was used unmodified, i.e. in English, as developed by Cooper et al.[16,17] The published  
33  
34 internal consistency (Cronbach's alpha) was 0.97 and the interrater reliability was 0.55, as  
35  
36 measured by Cohen's Kappa (adjusted for chance) with a mean intraclass correlation  
37  
38 coefficient of 0.60 for the 11 items. The instrument comprises 11 critical behaviours rated on  
39  
40 a Likert scale of 0 to 4 (0 = Never/Hardly, 1 = Seldom, 2 = About as often as not, 3 = Often, 4  
41  
42 = Always/Nearly always) and, finally, a Global rating of the team's overall performance on a  
43  
44 scale of 1 to 10. The original publication with the TEAM instrument had no descriptors for  
45  
46 the endpoints. For this study, we used 1 = Poor and 10 = Excellent.

47  
48  
49 The team instrument comprises three subscales: leadership (items 1–2), teamwork (items 3–9)  
50  
51 and task management (items 10–11). An index was constructed as a mean score for the  
52  
53 respective subscale and all 11 items in the instrument (mean score for TEAM). The global  
54  
55 rating of the overall performance was analysed separately.  
56  
57  
58  
59  
60

## 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,

1  
2  
3  
4 the questionnaire was translated into Swedish by the authors of this study as a basis for  
5  
6 developing scenario-specific items in Swedish. In accordance with SAGAT, targeted learning  
7  
8 objectives for the training scenarios were formulated and then the specific goals for each  
9  
10 simulation were set. An iterative process was used to reformulate the items in Swedish, using  
11  
12 a separate group of six professionals, all registered nurses, working in both a clinical context  
13  
14 and a teaching context. The final sets of SA items are shown in Tables 2A and 2B (author's  
15  
16 translation to English) with 11 items in each freeze. To determine whether an item is essential  
17  
18 in a specific context, Lawshe advocated the use of professional assessments such as a content  
19  
20 validity index (CVI) defined as the fraction of professionals who rate the item as  
21  
22 important.[32] In the present study, the relevance was reviewed by three professionals (nurse  
23  
24 n=1, physician n=2) before being used in the study. All professionals rated the questions as  
25  
26 relevant, i.e. the CVI was 1.0.  
27  
28  
29  
30  
31

32 The answers given by the participants in the SA questionnaire were classified as incorrect (0)  
33  
34 or correct (1). The classifications were discussed and agreed upon by the two authors (KJ and  
35  
36 MHu). For answers on a continuous scale (e.g. systolic blood pressure), a 10% range around  
37  
38 the intended correct answer was accepted as correct. One question was removed from the  
39  
40 questionnaire since it became apparent during the classification process that the question had  
41  
42 frequently been misinterpreted.  
43  
44  
45

46 In order to administer the SA questionnaires, the scenarios were frozen (i.e. paused) twice.  
47  
48 The first freeze of the scenario was five minutes into the scenario, unless there was an active  
49  
50 task activity or if the team was conducting a team re-evaluation, in which case, the freeze was  
51  
52 briefly postponed. During the freeze, the patient monitor was switched off and the participants  
53  
54 turned away from the patient simulator while individually answering the questions. The  
55  
56 second freeze took place according to the same principles after an additional five minutes. All  
57  
58 participants were allowed to complete the questionnaire before the scenario re-started.  
59  
60



1  
2  
3  
4  
5 The length of the freezes was measured from the video recordings (Table 3). The start of the  
6  
7 freeze was defined as the beginning of the sentence “Now we will pause the scenario so that  
8  
9 you can answer some questions about the patient case” and the end of the freeze was defined  
10  
11 as the end of the sentence “Is everyone in place and ready? Now the scenario will re-start”.

12  
13 Team SAGAT (TSAGAT) was developed by Crozier et al. [29] based on SAGAT as an  
14  
15 assessment tool for evaluating team performance. In Crozier’s study, each team comprised a  
16  
17 trauma leader, airway manager and nurse. Individual SA questionnaires were developed for  
18  
19 the three-team roles including both shared knowledge and complementary knowledge.  
20  
21 TSAGAT was calculated as the sum of individual SAGAT scores and the TSAGAT scores  
22  
23 had a high correlation to a traditional checklist (Pearson correlation,  $r=0.996$ ). However, Salas  
24  
25 defined team SA as a dynamic process defined as the team’s shared understanding of a  
26  
27 situation at a specific point in time [33] while Endsley argued that team SA involves unique  
28  
29 activities as information sharing and coordination.[20] Thus, in the present study, in order to  
30  
31 measure this efficiently in all team members, all participants in a case received identical SA  
32  
33 questionnaires. TSAGAT was calculated based on the mean scores of the team for each SA  
34  
35 item.  
36  
37  
38  
39  
40  
41

#### 42 Post-simulation questionnaire

43  
44 In order to measure whether the items in the SA questionnaire were considered relevant to the  
45  
46 scenarios, and whether pausing the scenarios affected the team training activity, a post-  
47  
48 simulation questionnaire (PSQ) was used.[22] The PSQ comprises 13 statements to be rated  
49  
50 on a four-point scale ranging from Strongly disagree (1) to Strongly agree (4), in which five  
51  
52 statements concern the SAGAT and the effect of freezing the scenario and eight statements  
53  
54 concern the simulation and the scenario *per se*. To the best of our knowledge, no data  
55  
56 regarding the reliability of PSQ were presented in the original study. In this study, we only  
57  
58 use the five questions relating to SAGAT and freezing the scenario. The PSQ was translated  
59  
60

1  
2  
3  
4 into Swedish by one of the authors and the translation was further refined based on iterative  
5 discussions within our research group. Next, the Swedish PSQ was sent to a professional  
6 translation agency together with the original PSQ for verification of the translation. The PSQ  
7 was answered in an anonymous web survey during the second week after the simulation  
8 training by 24 of the 55 participating students, i.e. the response rate was 44% (Table 4). In this  
9 study, the results of the four-point scale were dichotomised into disagree (Strongly disagree  
10 and Disagree) and agree (Agree and Strongly agree).  
11  
12  
13  
14  
15  
16  
17  
18  
19

### 20 21 Study size

22  
23 The aim of the study was to evaluate the feasibility of ABCDE, TEAM and SAGAT for use in  
24 further studies. In order to achieve this, the study sample must be large enough to allow for,  
25 with a fair precision, the calculation of descriptive statistics (means and standard deviations)  
26 and the calculation of reliabilities. In order to assess a large effect size (Cohen's d of 0.9) with  
27 t-test with a power of 80% at the 0.05 level, 16 participants per group would be needed as  
28 determined by G\*Power.[34] Thus, for the aims of the present study, the inclusion of 50  
29 individuals and 20 teams would suffice.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

### 40 Statistical analysis

41  
42 Statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM,  
43 Chicago, IL). Interrater reliability for the ABCDE checklist and the TEAM instrument was  
44 determined by intraclass correlation using a two-way random effects model (ICC (2,1) type  
45 absolute).[35,36] ICC is reported both as single measures and average measures since the ICC  
46 for single measures relates to the reliability of the individual ratings, while average measures  
47 relates to the reliability of the mean values. Cronbach's alpha was used to measure the  
48 internal consistency of the SAGAT and TSAGAT.[35] Internal consistency considered as the  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 extent to which all items measure the same latent variable was investigated by  $\chi^2$  and as  
6 suggested by Schweizer, a normed  $\chi^2$  below 2 was taken as an indication of a good fit.[37]  
7  
8

## 9 Ethics

10  
11 This study was approved by the Ethical Review Board of Northern Sweden (April 7, 2016,  
12 decision no. 2016-54-31M).  
13  
14

## 15 Patient and Public Involvement

16  
17 Neither patients nor the public were involved in the design or the data collection for this  
18 study.  
19  
20  
21  
22

## 23 Results

### 24 Scores on ABCDE checklist, TEAM instrument and SA items

25  
26 Fifty-five participants participated in the study, combined into 23 teams with 3-4 participants  
27 in each team, running either case A or B. The scores on the ABCDE checklist, TEAM  
28 instrument and the SA items are shown in Table 5. The mean ABCDE score was 2.6 on a  
29 scale of 0–4, the mean TEAM item 1–11 score was 2.3 on a scale of 0–4, the mean TEAM  
30 global rating (item 12) was 4.8 on a scale of 0–10. The mean correct score per item ranged  
31 from 0.25 to 0.95 on a scale of 0 or 1, and the mean total SA item score per participant was  
32 13 with a maximal score of 22.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44

### 45 Interrater reliability for the ABCDE checklist and TEAM instrument

46  
47 Interrater reliability as measured by intraclass correlation was 0.55 (single measures)/0.83  
48 (average measures) for the ABCDE checklist and 0.54/0.83 for the TEAM scale. For the  
49 TEAM subscales of leadership, team work, task management and global rating, the intraclass  
50 correlations were 0.36/0.70, 0.45/0.77, 0.35/0.68 and 0.38 /0.72, respectively.  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## 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  $\chi^2$  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,

1  
2  
3  
4  
5 ABCDE, TEAM and SAGAT, could permit analysis of task performance, team performance  
6  
7 and the relationship to situation awareness in team training.  
8

9  
10 Scaling down the comprehensive TTET[15] to a smaller ABCDE checklist could introduce  
11  
12 unintended errors in measurements. In order to be sensitive to the intended use, the items  
13  
14 measured must be relevant to the task and the anchors of the scale used must be calibrated to  
15  
16 the setting in order to allow for significant changes to be reflected as the difference between  
17  
18 control and intervention.[35] In the present study, the developed ABCDE checklist was  
19  
20 perceived by professionals to be relevant to the case. Interrater variability was low, indicating  
21  
22 that the checklist could be reliably used for scoring task performance. The means of the scores  
23  
24 were in the middle of the scale, which indicates that the checklist might be sensitive to  
25  
26 differentiating between low and high performers.  
27  
28

29  
30 For the TEAM instrument, the interrater reliability was 0.55 for single measurements and 0.83  
31  
32 for average measurements, which are similar to the results reported by Cooper et al where the  
33  
34 interrater reliability for the TEAM instrument was 0.55 as measured by Cohen's kappa and  
35  
36 0.60 as measured by ICC [16], and also to the results reported by McKay et al where the ICC  
37  
38 were 0.59 – 0.88 for the different items.[38] According to Koo and Li, an ICC between 0.50  
39  
40 and 0.75 indicates moderate reliability, while an ICC between 0.75 and 0.90 indicates good  
41  
42 reliability.[36] Thus, the averaged measurements from four raters had a good reliability. In the  
43  
44 original TEAM publication by Cooper et al, the performance improved significantly from  
45  
46 novice learners to experts, and in this study the participants were a homogenous set of  
47  
48 medical students, and as such to be considered as novice learners.[16] The medical students in  
49  
50 the present study had an average TEAM item score of 2.3 of a maximum 4, which translates  
51  
52 to showing the desired behavior a bit more often than not. The national learning objectives  
53  
54 requires a systematic training in leadership and followership which might explain the rather  
55  
56 high score. [39]  
57  
58  
59  
60

1  
2  
3  
4  
5 Both the PSQ and the CVI indicate that SAGAT could be used to construct questions that  
6  
7 were considered relevant to the case. In both cases, the internal consistency was fair (0.80 and  
8  
9 0.61). When analysing the subscales, levels 1, 2 and 3, the internal consistency was low for  
10  
11 levels 1 and 2, and higher for level 3. This could indicate that the perception of the situation in  
12  
13 the groups was diverse and not related to the total score, while the ability to project the  
14  
15 direction in which the cases were heading was more homogeneously related to the total score.  
16  
17 TSAGAT had a higher overall homogeneity, as might be expected when analysing the means  
18  
19 of the group SAGAT for each question, instead of the specific SAGAT answers.  
20  
21

22  
23 The feasibility of using SAGAT to measure situation awareness was assessed by measuring  
24  
25 the length of the freeze (i.e. pause) needed to answer questions and by asking the participants  
26  
27 for their perception of the pauses in the scenario and how the pauses affected the training  
28  
29 session. According to Edsley, when measuring SA using SAGAT, the scenario is frozen while  
30  
31 the participants answer questions that probe the three levels of SA: perception, comprehension  
32  
33 and projection.[21] Each pause in this study lasted less than 3 minutes. The pauses and  
34  
35 SAGAT questions could influence SA in both a negative direction and a positive direction. In  
36  
37 the negative direction, the flow of the simulation training might be disrupted, stress be  
38  
39 induced by being forced to interrupt the case, and the commitment decrease. According to the  
40  
41 PSQ, the majority of the participants did not perceive that the pauses adversely affected  
42  
43 concentration and performance. Contrary, the pauses and questions might facilitate the  
44  
45 resolution of the clinical problem in the case by triggering and allowing time for reflection,  
46  
47 fully in line with Schön's reflection-on-action.[24]  
48  
49  
50  
51

#### 52 53 Limitations of this study

54  
55 The participants in the study comprised year four medical students for the video recordings of  
56  
57 simulation-based team-training and the rating of the video material was conducted by four  
58  
59 participants representing a wide spectrum of experience and training. It could be argued that it  
60

1  
2  
3  
4 is a limitation that this validation was not performed on a series of critical care teams, for  
5 example. However, for testing the reliability as well as the feasibility of the checklist,  
6 techniques and instruments, the simulation-based trainings with medical students was a  
7 readily available series of standardized simulations.  
8  
9  
10  
11  
12

13 The raters in this study included both experienced and inexperienced raters. The study was not  
14 designed to systematically test for differences in ratings between raters of different  
15 backgrounds. However, the inhomogeneity of the raters is a strength as it indicates that the  
16 checklists and instruments might be used by a group of raters with different experiences.  
17  
18 Difficulties were encountered in the interpretation of what constitutes partially performed  
19 versus performed completely in the ABCDE checklist. This could relate to the vast  
20 differences in the raters' level of education and previous experience. To resolve this, the raters  
21 were trained using a separate set of video recordings to improve the accuracy and minimize  
22 the variability. TEAM was not translated into Swedish in order to avoid inducing errors. This  
23 was possible as fluency in English is a prerequisite for academic studies in Sweden. The  
24 ratings based on the original TEAM instrument were consistent between the raters, indicating  
25 that this may have been a correct assumption. The developed set of SAGAT items contained  
26 only a few items for measuring each level of SA. The selected number of items were close to  
27 what was used by Gardner and Hogan.[22,25] Too few items increase measurement errors,  
28 while too many items result in extended freezes. However, in the present study, the probing of  
29 level 2 in particular would have benefited by 2–3 more items per measurement.  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49

50 In the present study, CVI for SAGAT was measured according to Lawshe [32] while the  
51 development of the ABCDE checklist and the two cases relied on the authors' experiences  
52 and iterative interactions with clinicians and experts in the field. Both the cases, the ABCDE  
53 checklists and the developed SAGAT questionnaires, could have benefitted from a full formal  
54 CVI by a review panel.  
55  
56  
57  
58  
59  
60

1  
2  
3  
4  
5 The PSQ was only translated from English and not translated back into Swedish in order to  
6  
7 formally check the identity of the items in the final sets. However, the translation was  
8  
9 adjusted by a professional translator before being used in the study. Thus, the results of the  
10  
11 PSQ can be used for probing the participants in the simulation with regards to their  
12  
13 experiences of SAGAT.  
14  
15

16 This study focused on the quantification of performance during a simulation based training.

17  
18 The transferability of the studied behaviours into a real world setting is an intriguing question  
19  
20 for further studies.  
21  
22

## 23 Conclusion

24  
25 In this setting with medical students, the developed ABCDE checklist and the TEAM  
26  
27 instrument had high interrater reliability, the process of using SAGAT questionnaires during  
28  
29 simulation-based training did not negatively affect the participants' evaluation of simulation-  
30  
31 based training and the developed SAGAT questionnaires had a fair internal consistency. Thus,  
32  
33 the measurement of task performance, team performance and situation awareness in future  
34  
35 studies in a Swedish simulation-based training setting may be conducted using these  
36  
37 techniques.  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



**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

For peer review only

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

correct answers.

Freeze	Question	Response scale	Level	Correct
1			1	0.26
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	Free text	1	0.95
----	--	-----------	---	------

patient have?

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	e	3	0.82
----	----------------	---	---	------

10.	Saturation	Increase/Decrease	3	0.59
-----	------------	-------------------	---	------

11.	Respiratory rate	e	3	0.74
-----	------------------	---	---	------

Increase/Decrease

e

Increase/Decrease

e

---

Freeze	Question	Response scale	Level	Correct
2				
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	Question	Response scale	Level	Correct
1			1	0.28
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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

---

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	e	3	0.58
10.	Saturation	Increase/Decrease	3	0.78
11.	Respiratory rate	e	3	0.72
		Increase/Decrease		
		e		
		Increase/Decrease		
		e		

---



Freeze	Question	Response scale	Level	Correct
2				
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?  Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?	Yes/No	2	0.50
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)	1 <sup>st</sup> freeze m (sd)	2 <sup>nd</sup> freeze m (sd)	Total time m (sd)
<b>A</b>	12	14.1 (2.0)	2.7 (0.5)	2.5 (0.6)	19.3 (2.3)
<b>B</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 19 (79) 1 (4) 4 (17)

a tool for evaluating hands-on skills

during a trauma exercise.

---

Results from the post-simulation questionnaire

For peer review only

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

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

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

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

1  
2  
3  
4  
5 \*\*\*Mean score (sd). Max score 22.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

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

SAGAT/TSAGAT	Case A	Case B
	n/n	n/n
	39/12	36/11
<b>Level 1</b>	0.060/0.372	0.246/0.420
<b>Level 2</b>	0.321/-0.018	0.332/0.332
<b>Level 3</b>	0.891/0.891	0.659/0.786
<b>Total (Level 1–3)</b>	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](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

1  
2  
3  
4  
5  
6  
7 9 Brady PW, Muething S, Kotagal U, *et al*. Improving situation awareness to reduce unrecognized  
8 clinical deterioration and serious safety events. *Pediatrics* 2013;**131**:e298-308. doi:10.1542/peds.2012-  
9 1364  
10  
11  
12

13  
14  
15 10 Salas E, Cooke NJ, Rosen MA. On Teams, Teamwork, and Team Performance: Discoveries and  
16 Developments. *Hum Factors J Hum Factors Ergonomic Soc* 2008;**50**:540–7.  
17  
18  
19 doi:10.1518/001872008x288457  
20  
21  
22

23  
24 11 Kozlowski SW. Enhancing the Effectiveness of Work Groups and Teams: A Reflection. *Perspect*  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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.

1  
2  
3  
4  
5 doi:10.1016/j.resuscitation.2009.11.027  
6  
7

8  
9 17 Cooper SJ, Cant RP. Measuring non-technical skills of medical emergency teams: An update on the  
10 validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation*  
11 2014;**85**:31–3. doi:10.1016/j.resuscitation.2013.08.276  
12  
13  
14

15  
16  
17 18 Cooper S, Cant R, Connell C, *et al.* Measuring teamwork performance: Validity testing of the  
19 Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation*  
20 2016;**101**:97–101. doi:10.1016/j.resuscitation.2016.01.026  
21  
22  
23

24  
25 19 Schulz CM, Krautheim V, Hackemann A, *et al.* Situation awareness errors in anesthesia and critical  
26 care in 200 cases of a critical incident reporting system. *BMC Anesthesiol* 2016;**16**:4.  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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 measurement*. 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

1  
2  
3  
4 24 Schön DA. *The reflective practitioner: How professionals think in action*. The Perseus Books  
5  
6 Group 1984.  
7  
8

9  
10  
11 25 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the  
12  
13 SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc J Soc Simul*  
14  
15 *Healthc* 2017;**12**:17–21. doi:10.1097/sih.000000000000181  
16  
17

18  
19 26 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident's Situational  
20  
21 Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3.  
22  
23 doi:10.1097/BOT.000000000000003  
24  
25

26  
27 27 Lavoie P, Cossette S, Pepin J. Testing nursing students' clinical judgment in a patient deterioration  
28  
29 simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today*  
30  
31 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015  
32  
33  
34

35  
36 28 McKenna L, Missen K, Cooper S, *et al*. Situation awareness in undergraduate nursing students  
37  
38 managing simulated patient deterioration. *Nurs Educ Today* 2014;**34**:e27–31.  
39  
40 doi:10.1016/j.nedt.2013.12.013  
41  
42

43  
44 29 Crozier MS, Ting HY, Boone DC, *et al*. Use of human patient simulation and validation of the  
45  
46 Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team  
47  
48 assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009  
49  
50  
51

52  
53 30 Thim T, Krarup, Grove, *et al*. Initial assessment and treatment with the Airway, Breathing,  
54  
55 Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012;:117.  
56  
57 doi:10.2147/IJGM.S28478  
58  
59  
60

1  
2  
3  
4  
5 31 ATLS Advanced Trauma Life Support Student Course Manual. American College of Surgeons.  
6  
7 2018.

8  
9  
10  
11 32 Lawshe CH. A quantitative approach to content validity. *Pers Psychol* 1975;**28**:563–75.  
12  
13 doi:10.1111/j.1744-6570.1975.tb01393.x  
14

15  
16  
17 33 Salas E, Prince C, Baker DP, *et al.* Situation Awareness in Team Performance: Implications for  
18  
19 Measurement and Training. *Hum Factors J Hum Factors Ergonomics Soc* 1995;**37**:123–36.  
20  
21 doi:10.1518/001872095779049525  
22

23  
24  
25 34 Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. *Behavior research*  
26  
27 *methods, instruments, & computers* 1996;**28**:1–11. doi:10.3758/BF03203630  
28

29  
30  
31 35 Nunnally JC. *Psychometric Theory, 2nd edition.* McGraw-Hill 1978.  
32

33  
34  
35 36 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for  
36  
37 Reliability Research. *Journal of chiropractic medicine* 2016;**15**:155–63.  
38  
39 doi:10.1016/j.jcm.2016.02.012  
40  
41

42  
43  
44 37 Schweizer K. Some guidelines concerning the modeling of traits and abilities in test construction.  
45  
46 *European Journal of Psychological Assessment* 2010;**26**. doi:10.1027/1015-5759/a000001  
47  
48

49  
50  
51 38 McKay A, Walker ST, Brett SJ, *et al.* Team performance in resuscitation teams: comparison and  
52  
53 critique of two recently developed scoring tools. *Resuscitation* 2012;**83**:1478–83.  
54  
55 doi:10.1016/j.resuscitation.2012.04.015  
56  
57  
58  
59  
60

1  
2  
3  
4 39 Lindgren S, Brännström T, Hanse E, *et al.* Medical education in Sweden. *Medical teacher*;33:798–  
5  
6 803. doi:10.3109/0142159X.2011.570816  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

# BMJ Open

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

Journal:	<i>BMJ Open</i>
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; Umeå Universitet Medicinska fakulteten, Department of Nursing Lindkvist, Marie; Umeå Universitet Medicinska fakulteten, Department of Public Health and Clinical Medicine Brolin, Christine; Umeå Universitet Medicinska fakulteten, Department of Nursing
<b>Primary Subject Heading</b>:	Medical education and training
Secondary Subject Heading:	Communication
Keywords:	Patient care team, Resuscitation, Situation Awareness, Task performance and analysis, Teamwork

SCHOLARONE™  
Manuscripts

1  
2  
3  
4 The reliability of instruments that measure situation awareness, team and task performance in  
5 a simulation setting with medical students  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17

18 Magnus Hultin, MD Ph.D., Associate Professor, Department of Surgical and Perioperative  
19 Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå,  
20 Sweden. <https://orcid.org/0000-0003-2935-7161>  
21  
22

23 Karin Jonsson, RNT, Department of Nursing and Department of Surgical and Perioperative  
24 Sciences, Anesthesiology and Critical Care Medicine, Umeå University, S-901 87 Umeå,  
25 Sweden. <https://orcid.org/0000-0001-7974-6777>  
26  
27

28 Maria Härgestam, RNA Ph.D., Lecturer, Department of Nursing, Umeå University, S-901 87  
29 Umeå, Sweden. <https://orcid.org/0000-0003-0680-9962>  
30  
31

32 Marie Lindkvist, Ph.D., Associate Professor, Department of Epidemiology and Global Health,  
33 Umeå University, S-901 87 Umeå, Sweden  
34  
35

36 Christine Brulin, RNA Ph.D., Professor Emerita, Department of Nursing, Umeå University,  
37 S-901 87 Umeå, Sweden  
38  
39

40  
41  
42  
43  
44  
45  
46  
47  
48 **Abstract** 246 words  
49

50  
51 **Main article** 4736 words  
52  
53

54  
55 **Keywords**  
56

- 57 • Patient Care Team
- 58 • Resuscitation
- 59
- 60



- 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

For peer review only

## 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 4<sup>th</sup> 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  $\chi^2$  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.

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

## 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.

1  
2  
3  
4  
5 Team performance can be measured using the Team Emergency Assessment Measure  
6  
7 (TEAM), for example. It measures three dimensions of team performance: leadership,  
8  
9 teamwork and task management.[16,17] The instrument was initially developed and validated  
10  
11 for simulator-based team training and was recently validated for the collection of  
12  
13 observational ratings of non-technical skills during live resuscitations in emergency  
14  
15 departments.[18] Using instruments to score scenarios depend on reliable interpretations of  
16  
17 the instrument by the raters, and it might be even more important to ensure the reliability of  
18  
19 such interpretations when using an instrument in a non-native language.  
20  
21

22  
23 Moreover, and in addition to task and team performance, SA is a prerequisite for patient  
24  
25 safety and the prevention of errors, particularly in acute care situations.[9,19] SA includes  
26  
27 three levels of ability: (1) perception and attention (*What?*, what's going on), (2)  
28  
29 comprehension (*So what?*, the ability to understand what's going on), and (3) projection  
30  
31 (*What's next?*) in order to anticipate and plan for future events.[20] In order to measure SA in  
32  
33 a simulation setting, the Situation Awareness Global Assessment Tool (SAGAT) has been  
34  
35 developed and subsequently adapted for use in healthcare settings.[21,22] One feature of  
36  
37 SAGAT is that its use requires the simulation to pause[21], which might influence clinical  
38  
39 understanding, both by impeding the suspension of disbelief in the simulation setting [23] and  
40  
41 by facilitating reflection on action.[24] Gardner showed that it was feasible to use SAGAT to  
42  
43 measure SA in the team training of surgical trainees in advanced cardiac life support.[25]  
44  
45 Globally, SAGAT has been used to study, for example, the effect of sleep deprivation on SA  
46  
47 in trauma team training, how SA is associated with surgical trainee team performance, as well  
48  
49 as nurses' clinical assessment of patient deterioration.[26–28] SAGAT can be analysed on an  
50  
51 individual level and also on a team level. A specific application of SAGAT is team SAGAT  
52  
53 (TSAGAT) in which the different members of a team answer SAGAT questions specific to  
54  
55 their role. TSAGAT was constructed to account for the teams' shared situation awareness and  
56  
57  
58  
59  
60

1  
2  
3  
4 validated in trauma team training in a Canadian setting.[29] To our knowledge, SAGAT has  
5  
6 not previously been used in a Swedish context, and it is therefore important to evaluate both  
7  
8 its feasibility and its internal consistency.  
9

10  
11 To summarise, ABCDE checklists, the TEAM instrument and SAGAT have been developed  
12  
13 in order to evaluate different aspects of teamwork. However, these instruments and  
14  
15 questionnaires have neither been translated into Swedish nor tested for feasibility or  
16  
17 trustworthiness in a Swedish context. Such studies are necessary to enable the evaluation of  
18  
19 teamwork in Swedish acute care settings and simulation-based training. Thus, this study  
20  
21 aimed to test an ABCDE checklist that we developed for interrater reliability, to test the  
22  
23 TEAM instrument for interrater reliability, and to test two Swedish SAGAT questionnaires  
24  
25 for feasibility and internal consistency.  
26  
27  
28  
29  
30  
31

## 32 Method

33  
34  
35 This study is based on data collected during simulation-based team training sessions with  
36  
37 medical students.  
38

### 39 Participants and raters

40  
41 From March to October 2016, all medical students (n=68) in year 4 undertaking their  
42  
43 clerkship in anaesthesiology and critical care medicine were invited to participate in the study  
44  
45 while receiving mandatory simulator-based team training. In total, 55 students (81%)  
46  
47 participated in the study (Table 1) and 20 of them participated in both scenario A and scenario  
48  
49

### 50 B.

51  
52 All scenarios were video recorded and later scored by four raters to allow for calculation of  
53  
54 interrater reliability. First, a registered nurse with a Master's Degree (one year) in Nursing  
55  
56 (Critical Care Medicine), 20 years' working experiences at an ICU, nine years' experiences of  
57  
58  
59  
60



1  
2  
3  
4 human patient simulator team training and also a PhD student (author KJ). Second, a  
5  
6 physician with 20 years' working experience, associate professor and consultant in  
7  
8 Anesthesiology and Critical Care Medicine, with 14 years' experiences of human patient  
9  
10 simulator team training (author MHu). Third, a registered nurse with a Master's Degree (one  
11  
12 year) in Nursing (Acute Care Medicine), with six years' working experiences of prehospital  
13  
14 care and one year as a medical student. Fourth, a paramedic with limited experiences beyond  
15  
16 two years working as a paramedic in the Israeli Army and one year as a medical student.  
17  
18  
19

## 20 21 Setting

22  
23 Data collection in this study was carried out during simulation-based team training at the  
24  
25 Clinical Training Centre (CTC) of the Medical Faculty, Umeå University. The briefings,  
26  
27 scenarios and debriefings were conducted in Swedish. Two cameras mounted at an angle were  
28  
29 used to record videos in the simulation room, and one of the views included the patient  
30  
31 monitor.  
32  
33

34  
35 One week before team training, the students were asked to watch a twelve-minute video  
36  
37 available at the learning platform that introduced both the ABCDE concept when caring for a  
38  
39 patient and the room and equipment to be used during simulation-based training.[30,31] At  
40  
41 the start of the training session, the students participated in a 15-minute live introduction to  
42  
43 the simulation laboratory presented by the operator and the instructor. Each student group (4–  
44  
45 5 students/group) trained on four scenarios that focused on assessment and treatment of  
46  
47 severely ill patients in an emergency room. In each scenario, 3-4 students were active, and 1–  
48  
49 2 were observers. Therefore, only 3-4 participated in the interactive role-play in each scenario  
50  
51 (the actual simulation). In all, the 55 unique students made up a total of 23 teams with 3-4  
52  
53 participants in each team and 20 students participated in both case A and case B. The students  
54  
55 all played the role of interns, i.e. physicians who are in training to become licensed, with the  
56  
57 attending nurse currently unavailable. The assigned task involved conducting a primary  
58  
59  
60

1  
2  
3  
4  
5 survey and stabilising the patient until more senior staff arrived in 15 minutes. After each  
6  
7 scenario, a 10–15-minute debriefing session permitted reflection and shared learning. The first  
8  
9 scenario was a warm-up, and the last included a training summary. Case A and B, or in the  
10  
11 reverse order, as determined by the toss of a coin immediately before the second scenario,  
12  
13 were used as the second and third scenarios.  
14  
15

16 All scenarios were conducted in Swedish, were designed to last 10–15 minutes and were pre-  
17  
18 programmed into a Laerdal SimMan simulator in order to support the standardisation of the  
19  
20 simulation. In essence, the patient cases used were slightly modified versions of the scenarios  
21  
22 used by Hogan.[22] Case A was hypovolemic shock following a traffic incident. Case B was a  
23  
24 pneumothorax with affected vital signs following a traffic incident.  
25  
26  
27

#### 28 Background questionnaire

29  
30 The background questionnaire included informed consent and was answered immediately  
31  
32 before team training started. The questionnaire included questions such as year of birth,  
33  
34 male/female, previous medical training, previous experiences of team training, previous  
35  
36 experiences of human patient simulator-based training, previous experiences of CRM and  
37  
38 previous experiences of live trauma care.  
39  
40  
41

#### 42 ABCDE checklist

43  
44 In order to measure the completeness of critical tasks in acute care scenarios, an ABCDE  
45  
46 checklist was used. The original TTET comprised 58 items derived from the ATLS  
47  
48 protocol.[15] Each item in the TTET was scrutinised by the authors of this study and the  
49  
50 number of items reduced to those that reflected the measures expected to be carried out by  
51  
52 year four medical students in the acute care scenarios in this study. Some items were also  
53  
54 slightly modified to reflect the current ATLS standards. The final list contained ten items  
55  
56 reflecting the key elements of the ATLS primary survey, i.e. the core management of  
57  
58  
59  
60

1  
2  
3  
4  
5 ABCDE: airway assessed, airway secured, saturation assessed, oxygen applied, ventilation  
6  
7 assessed, ventilation optimised, pulses checked (radial – femoral – carotid), venous access and  
8  
9 infusions established, neurological disabilities checked (consciousness, pupils), and  
10  
11 full/complete exposure. The items are well described in major textbooks and were translated  
12  
13 into Swedish in line with the nomenclature being used. Compared to the original scoring  
14  
15 system, an additional scoring option was added: performed after a reminder from the  
16  
17 instructor. Each item was rated on a rating scale of 0 to 4 (0 = Not initiated, 1= Performed  
18  
19 after a reminder from the instructor, 2 = Partially performed, 3 = Performed completely before  
20  
21 the end of the simulation, and 4 = Performed consistently during the whole simulation, NA =  
22  
23 not applicable). Based on all items in the ABCDE checklist an index was constructed as a  
24  
25 mean score ranging from 0.0 to 4.0.  
26  
27  
28  
29

### 30 TEAM instrument

31  
32 In order to measure team leadership, teamwork and task management, the TEAM instrument  
33  
34 was used unmodified, i.e. in English, as developed by Cooper et al.[16,17] The published  
35  
36 internal consistency (Cronbach's alpha) was 0.97, and the interrater reliability was 0.55, as  
37  
38 measured by Cohen's Kappa (adjusted for chance) with a mean intraclass correlation  
39  
40 coefficient of 0.60 for the 11 items. The instrument comprises 11 critical behaviours rated on  
41  
42 a Likert scale of 0 to 4 (0 = Never/Hardly, 1 = Seldom, 2 = About as often as not, 3 = Often, 4  
43  
44 = Always/Nearly always), which are summed into a TEAM score ranging from 0-44, and  
45  
46 finally, a global rating of the team's overall performance on a scale of 1 to 10. The original  
47  
48 publication with the TEAM instrument had no descriptors for the endpoints. For this study,  
49  
50 we used 1 = Poor and 10 = Excellent.  
51  
52  
53

54  
55 The TEAM instrument comprises three subscales: leadership (items 1–2), teamwork (items 3–  
56  
57 9) and task management (items 10–11). Indexes were constructed based on the items in the  
58  
59 subscales and on all items in the instrument, ranging from 0.0 to 4.0.  
60

#### 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,

1  
2  
3  
4 the questionnaire was translated into Swedish by the authors of this study as a basis for  
5  
6 developing scenario-specific items in Swedish. In accordance with SAGAT, targeted learning  
7  
8 objectives for the training scenarios were formulated, and then the specific goals for each  
9  
10 simulation were set. An iterative process was used to reformulate the items in Swedish, using  
11  
12 a separate group of six professionals, all registered nurses, working in both a clinical context  
13  
14 and a teaching context. The final sets of SA items are shown in Tables 2A and 2B (author's  
15  
16 translation to English) with 11 items in each freeze. To determine whether an item is essential  
17  
18 in a specific context, Lawshe advocated the use of professional assessments such as a content  
19  
20 validity index (CVI) defined as the fraction of professionals who rate the item as  
21  
22 important.[32] In the present study, the relevance was reviewed by three professionals (nurse  
23  
24 n=1, physician n=2) before being used in the study. All professionals rated the questions as  
25  
26 relevant, i.e. the CVI was 1.0.  
27  
28  
29  
30  
31

32 The answers given by the participants in the SA questionnaire were classified as incorrect (0)  
33  
34 or correct (1). The classifications were discussed and agreed upon by the two authors (KJ and  
35  
36 MHu). For answers on a continuous scale (e.g. systolic blood pressure), a 10% range around  
37  
38 the intended correct answer was accepted as correct. One question was removed from the  
39  
40 questionnaire since it became apparent during the classification process that the question had  
41  
42 frequently been misinterpreted.  
43  
44  
45

46 In order to administer the SA questionnaires, the scenarios were frozen (i.e. paused) twice.

47  
48 The first freeze of the scenario was five minutes into the scenario, unless there was an active  
49  
50 task activity or if the team was conducting a team re-evaluation, in which case, the freeze was  
51  
52 briefly postponed. During the freeze, the patient monitor was switched off, and the  
53  
54 participants turned away from the patient simulator while individually answering the  
55  
56 questions. The second freeze took place according to the same principles after an additional  
57  
58  
59  
60

1  
2  
3  
4 five minutes. All participants were allowed to complete the questionnaire before the scenario  
5  
6 re-started.  
7  
8

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

13  
14 Team SAGAT (TSAGAT) was developed by Crozier et al. [29] based on SAGAT as an  
15 assessment tool for evaluating team performance. In Crozier’s study, each team comprised a  
16 trauma leader, airway manager and nurse. Individual SA questionnaires were developed for  
17 the three-team roles, including both shared knowledge and complementary knowledge.  
18

19  
20 TSAGAT was calculated as the sum of individual SAGAT scores, and the TSAGAT scores  
21 had a high correlation to a traditional checklist (Pearson correlation,  $r=0.996$ ). However, Salas  
22 defined team SA as a dynamic process defined as the team’s shared understanding of a  
23 situation at a specific point in time [33] while Endsley argued that team SA involves unique  
24 activities as information sharing and coordination.[20] Thus, in the present study, in order to  
25 measure this efficiently in all team members, all participants in a case received identical SA  
26 questionnaires. In this study, to account for slight differences in number of team members  
27 between the teams, TSAGAT was calculated as the mean SA score in each team.  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48

#### 49 Post-simulation questionnaire

50 In order to measure whether the items in the SA questionnaire were considered relevant to the  
51 scenarios, and whether pausing the scenarios affected the team training activity, a post-  
52 simulation questionnaire (PSQ) was used.[22] The PSQ comprises 13 statements to be rated  
53 on a four-point scale ranging from Strongly disagree (1) to Strongly agree (4), in which five  
54 statements concern the SAGAT and the effect of freezing the scenario and eight statements  
55  
56  
57  
58  
59  
60

1  
2  
3  
4 concern the simulation and the scenario *per se*. To the best of our knowledge, no data  
5  
6 regarding the reliability of PSQ were presented in the original study. In this study, we only  
7  
8 use the five questions relating to SAGAT and freezing the scenario. The PSQ was translated  
9  
10 into Swedish by one of the authors, and the translation was further refined based on iterative  
11  
12 discussions within our research group. Next, the Swedish PSQ was sent to a professional  
13  
14 translation agency together with the original PSQ for verification of the translation. The PSQ  
15  
16 was answered in an anonymous web survey during the second week after the simulation  
17  
18 training by 24 of the 55 participating students, i.e. the response rate was 44% (Table 4). In this  
19  
20 study, the results of the four-point scale were dichotomised into Disagree (Strongly disagree  
21  
22 and Disagree) and Agree (Agree and Strongly agree).  
23  
24  
25  
26  
27

#### 28 Study size

29  
30 The aim of the study was to evaluate the feasibility of ABCDE, TEAM and SAGAT for use in  
31  
32 further studies. In order to achieve this, the study sample must be large enough to allow for,  
33  
34 with fair precision, the calculation of descriptive statistics (means and standard deviations)  
35  
36 and the calculation of reliabilities. In order to assess a large effect size (Cohen's *d* of 0.9) with  
37  
38 t-test with a power of 80% at the 0.05 level, 16 participants per group would be needed as  
39  
40 determined by G\*Power.[34] Thus, for the aims of the present study, the inclusion of 50  
41  
42 individuals and 20 teams would suffice.  
43  
44  
45

#### 46 Statistical analysis

47  
48 Statistical analysis was performed using IBM SPSS Statistics for Windows, version 24 (IBM,  
49  
50 Chicago, IL). Interrater reliability for the ABCDE checklist and the TEAM instrument was  
51  
52 determined by intraclass correlation using a two-way random-effects model (ICC (2,1) type  
53  
54 absolute).[35,36] ICC is reported both as single measures and average measures since the ICC  
55  
56 for single measures relates to the reliability of the individual ratings, while average measures  
57  
58  
59  
60

1  
2  
3  
4  
5 relate to the reliability of the mean values. Cronbach's alpha was used to measure the internal  
6 consistency of the SAGAT and TSAGAT.[35] Internal consistency considered as the extent to  
7 which all items measure the same latent variable was investigated by  $\chi^2$ , and as suggested by  
8 Schweizer, a normed  $\chi^2$  below 2 was taken as an indication of a good fit.[37]  
9  
10  
11  
12

## 13 14 Ethics

15  
16 This study was approved by the Ethical Review Board of Northern Sweden (April 7, 2016,  
17 decision no. 2016-54-31M).  
18  
19

## 20 21 Patient and Public Involvement

22  
23 Neither patients nor the public was involved in the design or the data collection for this study.  
24  
25

## 26 27 Results

### 28 29 Descriptives of ABCDE checklist, TEAM instrument and SA items

30  
31 Fifty-five participants participated in the study, combined into 23 teams with 3-4 participants  
32 in each team, running either case A or B. The ABCDE mean item score was 2.6. The mean  
33 TEAM score was 25.3, and the mean TEAM global rating was 4.8. The mean SA score per  
34 item ranged from 0.25 to 0.95, and the mean SA score per participant was 13 (Table 5).  
35  
36  
37  
38  
39

### 40 41 Interrater reliability for the ABCDE checklist and TEAM instrument

42  
43 Interrater reliability, as measured by intraclass correlation, was 0.55 (single measures)/0.83  
44 (average measures) for the ABCDE checklist and 0.54/0.83 for the TEAM scale. For the  
45 TEAM subscales of leadership, team work, task management and global rating, the intraclass  
46 correlations were 0.36/0.70, 0.45/0.77, 0.35/0.68 and 0.38 /0.72, respectively.  
47  
48  
49  
50  
51

### 52 53 Feasibility and internal consistency of SAGAT

54  
55 The PSQ showed that 96% of the participants considered the SA items to be relevant to the  
56 case, and 96% considered the questions to be easy to understand (Table 4). About three out of  
57  
58  
59  
60



1  
2  
3  
4  
5 four (72%) participants stated that the freeze did not negatively impact their concentration or  
6  
7 performance during the simulation session.  
8

9  
10 The internal consistency of SAGAT measured as Cronbach's alpha was 0.80 for case A and  
11  
12 0.62 for case B (Table 5), and normed  $\chi^2$  was 1.72 vs. 1.62. For level 1 (perception),  
13  
14 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  
15  
16 and 0.66, respectively.  
17

18  
19 For TSAGAT, the internal consistency as measured by Cronbach's alpha was good for the  
20  
21 SAGAT questionnaire (All levels) with 0.83 and 0.76 for case A and B, respectively, and for  
22  
23 level 3 it was 0.89 and 0.79, respectively (Table 6).  
24  
25  
26  
27

## 28 Discussion

29  
30 Research on how to maximise team performance depends on the availability of sensitive and  
31  
32 reliable tools for measuring the impact of an intervention. This study aimed to test the  
33  
34 usability of three instruments and techniques developed in English-speaking contexts for  
35  
36 rating in a Swedish setting with teams performing in their native language and cultural  
37  
38 context. The main finding of this study was that the adapted ABCDE checklist and the TEAM  
39  
40 instrument could be used with acceptable interrater reliability and that it was feasible to use  
41  
42 SAGAT to measure SA. The scores were in the middle of the scales indicating that the scales  
43  
44 could purposefully be used in a future effect study providing that the test group also score in  
45  
46 the sensitive range of these instruments. The combination of these three measurements,  
47  
48 ABCDE, TEAM and SAGAT, could permit analysis of task performance, team performance  
49  
50 and the relationship to situation awareness in team training.  
51  
52

53  
54 Scaling down the comprehensive TTET[15] to a smaller ABCDE checklist could introduce  
55  
56 unintended errors in measurements. In order be sensitive to the intended use, the items  
57  
58  
59  
60

1  
2  
3  
4  
5 measured must be relevant to the task and the anchors of the scale used must be calibrated to  
6  
7 the setting in order to allow for significant changes to be reflected as the difference between  
8  
9 control and intervention.[35] In the present study, the developed ABCDE checklist was  
10  
11 perceived by professionals to be relevant to the case. Interrater variability was low, indicating  
12  
13 that the checklist could be reliably used for scoring task performance. The means of the scores  
14  
15 were in the middle of the scale, which indicates that the checklist might be sensitive for  
16  
17 differences between low and high performers.  
18  
19

20  
21 For the TEAM instrument, the interrater reliability was 0.55 for single measurements and 0.83  
22  
23 for average measurements, which are similar to the results reported by Cooper et al where the  
24  
25 interrater reliability for the TEAM instrument was 0.55 as measured by Cohen's kappa and  
26  
27 0.60 as measured by ICC [16], and also to the results reported by McKay et al where the ICC  
28  
29 were 0.59 – 0.88 for the different items.[38] According to Koo and Li, an ICC between 0.50  
30  
31 and 0.75 indicates moderate reliability, while an ICC between 0.75 and 0.90 indicates good  
32  
33 reliability.[36] Thus, the averaged measurements from four raters had good reliability. In the  
34  
35 original TEAM publication by Cooper et al., the performance improved significantly from  
36  
37 novice learners to experts, and in this study, the participants were a homogenous set of  
38  
39 medical students, and as such to be considered as novice learners.[16] The medical students in  
40  
41 the present study had an average TEAM item 1-11 score of 2.3 of a maximum 4, which  
42  
43 translates to showing the desired behaviour a bit more often than not. This is in agreement  
44  
45 with the 2.49 score reported by Cooper et al. for a group of second-year medical and nursing  
46  
47 students rated with TEAM during an interprofessional one-day resuscitation course.[16] The  
48  
49 national learning objectives requires systematic training in leadership and followership, which  
50  
51 might explain the rather high score. [39]  
52  
53  
54  
55

56  
57 Both the PSQ and the CVI indicate that SAGAT could be used to construct questions that  
58  
59 were considered relevant to the case. In both cases, the internal consistency was fair (0.80 and  
60

1  
2  
3  
4  
5 0.61). When analysing the subscales, levels 1, 2 and 3, the internal consistency was low for  
6  
7 levels 1 and 2, and higher for level 3. This could indicate that the perception of the situation in  
8  
9 the groups was diverse and not related to the total score, while the ability to project the  
10  
11 direction in which the cases were heading was more homogeneously related to the total score.  
12  
13 TSAGAT had a higher overall homogeneity, as might be expected when analysing the means  
14  
15 of the group SAGAT for each question, instead of the specific SAGAT answers.  
16  
17

18  
19 The feasibility of using SAGAT to measure situation awareness was assessed by measuring  
20  
21 the length of the freeze (i.e. pause) needed to answer questions and by asking the participants  
22  
23 for their perception of the pauses in the scenario and how the pauses affected the training  
24  
25 session. According to Endsley, when measuring SA using SAGAT, the scenario is frozen  
26  
27 while the participants answer questions that probe the three levels of SA: perception,  
28  
29 comprehension and projection.[21] Each pause in this study lasted less than 3 minutes. The  
30  
31 pauses and SAGAT questions could influence SA in both a negative direction and a positive  
32  
33 direction. In the negative direction, the flow of the simulation training might be disrupted,  
34  
35 stress be induced by being forced to interrupt the case, and the commitment decrease.  
36  
37

38  
39 According to the PSQ, the majority of the participants did not perceive that the pauses  
40  
41 adversely affected concentration and performance. Contrary, the pauses and questions might  
42  
43 facilitate the resolution of the clinical problem in the case by triggering and allowing time for  
44  
45 reflection, fully in line with Schön's reflection-on-action.[24]  
46  
47

#### 48 49 Limitations of this study

50  
51 The participants in the study comprised year four medical students for the video recordings of  
52  
53 simulation-based team-training and the rating of the video material was conducted by four  
54  
55 participants representing a wide spectrum of experience and training. It could be argued that it  
56  
57 is a limitation that this validation was not performed on a series of critical care teams, for  
58  
59 example. However, for testing the reliability as well as the feasibility of the checklist,  
60

1  
2  
3  
4  
5 techniques and instruments, the simulation-based training with medical students was a readily  
6  
7 available series of standardized simulations.

8  
9 Difficulties were encountered in the interpretation of what constitutes partially performed  
10  
11 versus performed completely in the ABCDE checklist. This could relate to the vast  
12  
13 differences in the raters' level of education and previous experience. To improve the accuracy  
14  
15 and minimize the variability, the raters were trained using a separate set of video recordings.  
16  
17 TEAM was not translated into Swedish in order to avoid inducing errors. This was possible as  
18  
19 fluency in English is a prerequisite for academic studies in Sweden. The ratings based on the  
20  
21 original TEAM instrument were consistent between the raters, indicating that this may have  
22  
23 been a correct assumption.  
24  
25

26  
27 In the present study, CVI for SAGAT was measured according to Lawshe [32] while the  
28  
29 development of the ABCDE checklist and the two cases relied on the authors' experiences  
30  
31 and iterative interactions with clinicians and experts in the field. Both the cases, the ABCDE  
32  
33 checklists and the developed SAGAT questionnaires, could have benefitted from a full formal  
34  
35 CVI by a review panel.  
36  
37

38  
39 The PSQ was only translated from English and not translated back into Swedish in order to  
40  
41 formally check the identity of the items in the final sets. However, the translation was  
42  
43 adjusted by a professional translator before being used in the study. Thus, the results of the  
44  
45 PSQ can be used for probing the participants in the simulation with regards to their  
46  
47 experiences of SAGAT.  
48  
49

50  
51 This study focused on the quantification of performance during simulation-based training.

52  
53 The transferability of the studied behaviours into a real-world setting is an intriguing question  
54  
55 for further studies.  
56  
57  
58  
59  
60

## 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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

**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?  Presuming an ordinary course of development, how will vital signs change in	Free text	1	95%



1  
2  
3  
4  
5 the next 10 minutes?  
6

7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17

---

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%

---

18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

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?  Presuming an ordinary course of development, how will vital signs change in the next 10 minutes?	Free text	2	21%
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%

1  
2  
3  
4  
5 The questions are shown translated into English in the table, and the original questions in Swedish are  
6  
7  
8 available upon request from the authors of the study.  
9

10 \*Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

**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?  Presuming an ordinary course of development, how will vital signs	Yes/No	1	69%

---

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%

---

For peer review only

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.

1  
2  
3  
4  
5 Levels of situation awareness: 1 – Perception, 2 – Comprehension, 3 – Projection  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only

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

Case	No. of teams	Scenario	1 <sup>st</sup> freeze	2 <sup>nd</sup> freeze	Total time
	n	m (sd)	m (sd)	m (sd)	m (sd)
<b>A</b>	12	14.1 (2.0)	2.7 (0.5)	2.5 (0.6)	19.3 (2.3)
<b>B</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 (%)
<b>Concerning the SAGAT questionnaire.</b>			
<b>In my opinion:</b>			
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	Case B
	m (sd)	m (sd)
<b>ABCDE checklist<sup>1</sup></b>	2.70 (0.19)	2.58 (0.27)
<b>TEAM instrument</b>		
<b>Sum of items 1-11<sup>2</sup></b>	25.5 (5.9)	25.0 (5.1)
<b>Mean item score 1-11<sup>1</sup></b>	2.32 (0.54)	2.27 (0.46)
<b>Mean subscore Leadership (items 1-2)<sup>1</sup></b>	2.27 (0.77)	2.32 (0.65)
<b>Mean subscore Team Work (items 3-9)<sup>1</sup></b>	2.28 (0.55)	2.23 (0.51)
<b>Mean subscore Task Management (items 10-11)<sup>1</sup></b>	2.49 (0.56)	2.38 (0.50)
<b>Global rating<sup>3</sup></b>	4.90 (1.13)	4.75 (1.04)
<b>SAGAT questionnaire<sup>4</sup></b>	13.95 (4.26)	12.60 (3.35)

<sup>1</sup>Max score per item 4.

<sup>2</sup>Max score 44.

<sup>3</sup>Max score 11.

<sup>4</sup>Max score 22.

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

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

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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

For peer review only

## 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](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

1  
2  
3  
4  
5  
6  
7 9 Brady PW, Muething S, Kotagal U, *et al.* Improving situation awareness to reduce unrecognized  
8 clinical deterioration and serious safety events. *Pediatrics* 2013;**131**:e298-308. doi:10.1542/peds.2012-  
9 1364  
10  
11  
12

13  
14  
15 10 Salas E, Cooke NJ, Rosen MA. On Teams, Teamwork, and Team Performance: Discoveries and  
16 Developments. *Hum Factors J Hum Factors Ergonomic Soc* 2008;**50**:540–7.  
17  
18  
19 doi:10.1518/001872008x288457  
20  
21  
22

23  
24 11 Kozlowski SW. Enhancing the Effectiveness of Work Groups and Teams: A Reflection. *Perspect*  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

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.

1  
2  
3  
4  
5 doi:10.1016/j.resuscitation.2009.11.027  
6  
7

8  
9 17 Cooper SJ, Cant RP. Measuring non-technical skills of medical emergency teams: An update on the  
10 validity and reliability of the Team Emergency Assessment Measure (TEAM). *Resuscitation*  
11 2014;**85**:31–3. doi:10.1016/j.resuscitation.2013.08.276  
12  
13  
14

15  
16  
17 18 Cooper S, Cant R, Connell C, *et al.* Measuring teamwork performance: Validity testing of the  
19 Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. *Resuscitation*  
20 2016;**101**:97–101. doi:10.1016/j.resuscitation.2016.01.026  
21  
22  
23

24  
25  
26 19 Schulz CM, Krautheim V, Hackemann A, *et al.* Situation awareness errors in anesthesia and critical  
27 care in 200 cases of a critical incident reporting system. *BMC Anesthesiol* 2016;**16**:4.  
28  
29  
30 doi:10.1186/s12871-016-0172-7  
31

32  
33  
34 20 Endsley MR. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors*  
35 1995;**37**:32–64. doi:10.1518/001872095779049543  
36  
37  
38

39  
40 21 Endsley MR. Direct measurement of situation awareness: validity and use of SAGAT. In: *Situation*  
41 *awareness analysis and measurement*. Lawrence Erlbaum Associates 2000. 1–21.  
42  
43  
44

45  
46  
47 22 Hogan MP, Pace DE, Hapgood J, *et al.* Use of human patient simulation and the situation  
48 awareness global assessment technique in practical trauma skills assessment. *J Trauma* 2006;**61**:1047–  
49 52. doi:10.1097/01.ta.0000238687.23622.89  
50  
51  
52

53  
54  
55 23 Hamstra SJ, Brydges R, Hatala R, *et al.* Reconsidering fidelity in simulation-based training. *Acad*  
56 *Med* 2014;**89**:387–92. doi:10.1097/ACM.0000000000000130  
57  
58  
59  
60

1  
2  
3  
4  
5 24 Schön DA. *The reflective practitioner: How professionals think in action*. The Perseus Books  
6  
7 Group 1984.  
8  
9

10  
11 25 Gardner AK, Kosemund M, Martinez J. Examining the Feasibility and Predictive Validity of the  
12  
13 SAGAT Tool to Assess Situation Awareness Among Medical Trainees. *Simul Healthc J Soc Simul*  
14  
15 *Healthc* 2017;**12**:17–21. doi:10.1097/sih.000000000000181  
16  
17

18  
19 26 Stratton A, Furey A, Hogan M. The Effect of Sleep Deprivation on a Resident's Situational  
20  
21 Awareness in a Trauma Scenario. *J Orthop Trauma* 2014;**28**:e160–3.  
22  
23 doi:10.1097/BOT.000000000000003  
24  
25

26  
27 27 Lavoie P, Cossette S, Pepin J. Testing nursing students' clinical judgment in a patient deterioration  
28  
29 simulation scenario: Development of a situation awareness instrument. *Nurse Educ Today*  
30  
31 2016;**38**:61–7. doi:10.1016/j.nedt.2015.12.015  
32  
33  
34

35  
36 28 McKenna L, Missen K, Cooper S, *et al*. Situation awareness in undergraduate nursing students  
37  
38 managing simulated patient deterioration. *Nurs Educ Today* 2014;**34**:e27–31.  
39  
40 doi:10.1016/j.nedt.2013.12.013  
41  
42

43  
44 29 Crozier MS, Ting HY, Boone DC, *et al*. Use of human patient simulation and validation of the  
45  
46 Team Situation Awareness Global Assessment Technique (TSAGAT): a multidisciplinary team  
47  
48 assessment tool in trauma education. *J Surg Educ* 2015;**72**:156–63. doi:10.1016/j.jsurg.2014.07.009  
49  
50  
51

52  
53 30 Thim T, Krarup, Grove, *et al*. Initial assessment and treatment with the Airway, Breathing,  
54  
55 Circulation, Disability, Exposure (ABCDE) approach. *Int J Gen Med* 2012;:117.  
56  
57 doi:10.2147/IJGM.S28478  
58  
59  
60



1  
2  
3  
4  
5 31 ATLS Advanced Trauma Life Support Student Course Manual. American College of Surgeons.  
6  
7 2018.

8  
9  
10  
11 32 Lawshe CH. A quantitative approach to content validity. *Pers Psychol* 1975;**28**:563–75.  
12  
13 doi:10.1111/j.1744-6570.1975.tb01393.x  
14

15  
16  
17 33 Salas E, Prince C, Baker DP, *et al.* Situation Awareness in Team Performance: Implications for  
18  
19 Measurement and Training. *Hum Factors J Hum Factors Ergonomics Soc* 1995;**37**:123–36.  
20  
21 doi:10.1518/001872095779049525  
22

23  
24  
25 34 Erdfelder E, Faul F, Buchner A. GPOWER: A general power analysis program. *Behavior research*  
26  
27 *methods, instruments, & computers* 1996;**28**:1–11. doi:10.3758/BF03203630  
28

29  
30  
31 35 Nunnally JC. *Psychometric Theory, 2nd edition*. McGraw-Hill 1978.  
32

33  
34  
35 36 Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for  
36  
37 Reliability Research. *Journal of chiropractic medicine* 2016;**15**:155–63.  
38  
39 doi:10.1016/j.jcm.2016.02.012  
40  
41

42  
43  
44 37 Schweizer K. Some guidelines concerning the modeling of traits and abilities in test construction.  
45  
46 *European Journal of Psychological Assessment* 2010;**26**. doi:10.1027/1015-5759/a000001  
47  
48

49  
50  
51 38 McKay A, Walker ST, Brett SJ, *et al.* Team performance in resuscitation teams: comparison and  
52  
53 critique of two recently developed scoring tools. *Resuscitation* 2012;**83**:1478–83.  
54  
55 doi:10.1016/j.resuscitation.2012.04.015  
56  
57  
58  
59  
60

1  
2  
3  
4 39 Lindgren S, Brännström T, Hanse E, *et al.* Medical education in Sweden. *Medical teacher*;33:798–  
5  
6 803. doi:10.3109/0142159X.2011.570816  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

For peer review only