

## PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

### ARTICLE DETAILS

<b>TITLE (PROVISIONAL)</b>	The reliability of instruments that measure situation awareness, team and task performance in a simulation setting with medical students
<b>AUTHORS</b>	Hultin, Magnus; Jonsson, Karin; Härgestam, Maria; Lindkvist, Marie; Brulin, Christine

### VERSION 1 – REVIEW

<b>REVIEWER</b>	Simon Cooper Federation University Australia Potential conflicts: Please note that I am the developer of the TEAM instrument and have published a number of papers related to the SAGAT, both described in this paper.
<b>REVIEW RETURNED</b>	07-Mar-2019

<b>GENERAL COMMENTS</b>	<p>Thank you for this interesting and well written paper. The work is thorough and considered. My main concern relates to the ‘so what question’ which I don’t think has been fully argued. Also I am left with the impression that I wanted to know more – especially related to student performance – which is not included. Perhaps include a short section including the performance outcomes of the groups? Further, at times the flow of the work is a little difficult to understand and requires some adaptations. Detailed comments for you below and again many thanks and well done on your work.</p> <p>Potential conflicts: Please note that I am the developer of the TEAM instrument and have published a number of papers related to the SAGAT.</p> <p>Abstract Is a brief introduction/background not required? Just note that there is no SAGAT instrument per se – SAGAT is a technique as in Situation Awareness Global Assessment Technique from which you develop a questionnaire. For clarity more details is needed in the ‘setting’ section and less is required in the primary/secondary outcomes section? Conclusions – perhaps a ‘bit light’ – I don’t think the ‘so what’ question has been answered. Also suggest you say ‘ABCDE’-concept as oppose to A-E – throughout the paper</p> <p>Article summary Suggest that you could have done a formal Content Validity Index (clarity/relevance etc) measurement with an expert team – as you would make the context/case clear to them</p>
-------------------------	--

	<p>Paper  With regard to TEAM development and testing the following paper is perhaps also highly relevant:  Cooper S. Cant R. Connell C, Sims L. Porter J. Symmons M. Nestel D. Liaw SY. (2016) Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. Resuscitation. 101; 97-101  And see <a href="http://medicalemergencyteam.com/">http://medicalemergencyteam.com/</a>  And as in point above the 't' in SAGAT is technique – note also that not all practitioners pause the simulation – as this tends to interrupt the flow – see reference below  Cooper S. Porter J. Peach L. (2014) Measuring situation awareness in emergency setting: a systematic review of tools and outcomes. Open Access Emergency Medicine. 6; 1-7.  Participants – how do you defend the inclusion of medical students only - this does not mirror real clinical inter-professional teams? Complicated by the fact that you used three professions to rate performance?  How was the English/Swedish translation performed – forward backward translation – details required.  Again point above about a formal measurement of content validity for SAGAT</p> <p>Results – suggest including the correlations immediately after each subscale</p> <p>End of review – no additional comments</p>
--	---

<b>REVIEWER</b>	A/Prof Robyn Cant School of Nursing and Healthcare Professoins, Federation University Australia.
<b>REVIEW RETURNED</b>	12-Mar-2019

<b>GENERAL COMMENTS</b>	<p>Thank you for this scholarly work. I suggest the paper is comprehensive, but a large number of changes are needed to present a final publication. I give my detailed suggestions below.  Title: suggest the title should refer to students and to non-technical skills?  Page 1, Keywords: suggest include use MeSH terms (seen at PubMed)</p> <p>Page 4 line 14-15: language problems- you suggest that this 'can be accomplished' however there is little evidence of simulation transfer into practice or preventing healthcare errors?? Your citations are lacking here?, or else you overstate the application.  Page 4 Line 16-18 I find your language a little non-specific for this type of technical paper. Consider a revision of every sentence to make clear what you refer to. Differentiate between 'technical skills' and 'non-technical skills'? as "task" covers both of these. Your saying "it is essential to evaluate task and team performance in order to develop strategies for improvements in clinical practice" Is unclear as both are 'tasks'. Technical skills and non-technical skills are the common terminology in this field?</p> <p>Page 4 Line 22: 'Handled' would be best stated as 'managed' as you don't 'handle' a patient.  Page 4 Line 24 -28: Consider having a native English speaker scholar edit the text throughout as there are better ways to express what you are saying and be clearer in this section. Line 20: Its not</p>
-------------------------	--

	<p>“the timing of the task” (when it was done) its the timing of the components of the task? (how long they took, etc). Line 24: suggest: as a pilot study, psychometric data such as reliability and validity were not reported.</p> <p>Line 26: What is “a case-specific A–E checklist”? The TTET is one, surely</p> <p>Line 30: I think you mean ‘nontechnical’ team performance?</p> <p>Line 38-9: “The TEAM instrument has not yet been validated for use outside the commonwealth nor has the instrument been tested at a subscale level: suggest this is unimportant, rephrase. See McKay et al. Team performance in resuscitation teams: comparison and critique of two recently developed scoring tools. Resuscitation 2012;83:1478–83 and Cooper, Cant. Measuring the non-technical skills of medical emergency teams: An update on validity and reliability of Team Emergency Assessment Measure (TEAM). Resuscitation, 2014; 85: 31-33. URL: <a href="http://dx.doi.org/10.1016/j.resuscitation.2013.08.276">http://dx.doi.org/10.1016/j.resuscitation.2013.08.276</a></p> <p>I am really not able to edit the whole of the Introduction for you so please attend to the issues I raise and improve the English grammar and specificity/accuracy of language in the whole section. Methods: Page 5 line 18: add ‘with medical students’ Participants should be described prior to the section on ‘setting’?? Give response rate.</p> <p>Page 5 Line 30: what are the roles that students played? Line 52: how can 55 students be divided into 23 teams? Please explain. Were the scenarios conducted in English ?</p> <p>The Methods section page 6 to page 9 is too long and tortuous. I suggest it would be beneficial to present a table of the ‘students’ experience’ from start to finish- showing each step of the simulation and testing in columns? Hence, you could reduce text</p> <p>Page 8 Line 37: The Methods language appears to be clearly written and considers relevant issues, thank you. The student sample appears to be too small.</p> <p>Page 9 line 5: we just need to see the code number of the ethics approval, not further details.</p> <p>Page 9 Results section: I would much prefer the headers under this section to reflect what you are testing rather than just given the instrument name?</p> <p>Page 9 line 15: The Results are presented as clinician rater’s consistency alone, and student teams performance results are missed. To be convincing, you should summarize the research in text (the student results) and then build on this by reporting instrument validity issues?</p> <p>Page 10, line 11. CVI? This is the first mention? Explain.</p> <p>Page 10 line 30: Limitations - please remove this text and redraft to address the limitations or shortcomings of your study, rather than what you might do. Design, sampling, transferability, what did you do to limit bias? Student rather than professional education?</p> <p>Page 10 line 58: suggest you redraft the conclusion- did you achieve the study aims? What was the outcome? What does the</p>
--	---

	<p>study contribute to knowledge? and to medical student education? professional education?  Abstract: Revise as applicable. The conclusion here should match the conclusion in the paper.</p> <p>Table 1 is incomplete, should present number and then percent. Some categories where there are n=2, can be collapsed.  Table 2: CV of professional raters should be removed and be summarized in text  Table 4a: I suggest place the response scale in the instrument to complete it?, place a footnote that a Swedish version was used for data collection? I suggest it would be more usual to state the header as Situation Awareness items for Case 1.</p> <p>Table 5: suggest alter header- this is the student agreement about usefulness and feasibility of prospective collection of situation awareness items.  I think this paper requires much revision-</p>
--	---

### VERSION 1 – AUTHOR RESPONSE

Reviewer 1	Response	Page in manuscript
<p>Thank you for this interesting and well written paper. The work is thorough and considered.</p> <p>My main concern relates to the ‘so what question’ which I don’t think has been fully argued.</p> <p>Also I am left with the impression that I wanted to know more – especially related to student performance – which is not included. Perhaps include a short section including the performance outcomes of the groups?</p> <p>Further, at times the flow of the work is a little difficult to understand and requires some adaptations. Detailed comments for you below and again many thanks and well done on your work.</p>	<p>Thank you for the review and the suggested adjustments of the manuscript.</p> <p>This is a methodological study focusing on evaluation of instruments that will be used in further intervention studies. Thus, the focus on the methodological issues is important in order to highlight trustworthiness of the techniques and instruments used in a Swedish context, therefore we have not the suggested additional comparisons.</p> <p>The manuscript has been thoroughly rewritten, sections has been added, and the structure revised to improve the flow.</p>	<p>Throughout the manuscript</p>
<p>Abstract  Is a brief introduction/background not required?</p>	<p>Unfortunately, the format of the journal does not allow a background section in the abstract.</p>	<p>P 3</p>

	We have added a sentence to the objectives section of the abstract to set the stage.	
Just note that there is no SAGAT instrument per se – SAGAT is a technique as in Situation Awareness Global Assessment Technique from which you develop a questionnaire.  For clarity more details is needed in the ‘setting’ section and less is required in the primary/secondary outcomes section?	This has been noted and the text has been clarified through the manuscript.	Throughout the manuscript
Conclusions – perhaps a ‘bit light’ – I don’t think the ‘so what’ question has been answered.	The conclusion has been revised.	P 4 and P 21
Also suggest you say ‘ABCDE’-concept as oppose to A-E – throughout the paper	This has been changed throughout the paper.	Throughout the manuscript
Suggest that you could have done a formal Content Validity Index (clarity/relevance etc) measurement with an expert team – as you would make the context/case clear to them	Thank you for the recommendation. The section on content validity index (CVI) has been rewritten and the CVI is also now discussed in limitations of the study.	P 13 and P 21
With regard to TEAM development and testing the following paper is perhaps also highly relevant: Cooper S. Cant R. Connell C, Sims L. Porter J. Symmons M. Nestel D. Liaw SY. (2016) Measuring teamwork performance: Validity testing of the Team Emergency Assessment Measure (TEAM) with clinical resuscitation teams. Resuscitation. 101; 97-101 And see <a href="http://medicalemergencyteam.com/">http://medicalemergencyteam.com/</a>	Thank you. This reference is now also cited in relevant sections.	P 34
And as in point above the ‘t’ in SAGAT is technique – note also that not all practitioners pause the simulation – as this tends to interrupt the flow – see reference below Cooper S. Porter J. Peach L. (2014) Measuring situation awareness in emergency setting: a systematic review of tools and outcomes. Open Access Emergency Medicine. 6; 1-7.	The text has been adjusted accordingly.	Throughout the manuscript

Participants – how do you defend the inclusion of medical students only - this does not mirror real clinical inter-professional teams? Complicated by the fact that you used three professions to rate performance?	It is difficult to make this kind of study in a clinical setting with personnel. Thus, we used an undergraduate setting for the methodological study. To test trustworthiness of instruments has been important in further intervention studies using simulation in situ at intensive care units at two different hospitals.  The selection of participants is now commented upon in the limitations section.	P 21
How was the English/Swedish translation performed – forward backward translation – details required.	Details have been added to the manuscript.	P11, P13 and P 115
Again point above about a formal measurement of content validity for SAGAT	See point above regarding formal vs informal measurement of CVI.	P 13 and P 21
Results – suggest including the correlations immediately after each subscale	See point above. This is interesting, however in this study we have focused upon the reliability and feasibility of using the instruments as this is a methodological study.	-

Reviewer 2	Response	Page in manuscript
Thank you for this scholarly work. I suggest the paper is comprehensive, but a large number of changes are needed to present a final publication. I give my detailed suggestions below.		
Title: suggest the title should refer to students and to non-technical skills?	The title has been revised.	P 1
Page 1, Keywords: suggest include use MeSH terms (seen at PubMed)	Keywords have been changed into MESH-terms, as suggested.	P 1
Page 4 line 14-15: language problems- you suggest that this 'can be accomplished' however there is little evidence of simulation transfer into practice or preventing healthcare errors?? Your citations are lacking here?, or else you overstate the application.	The section has been rephrased and the citations revised.	P 7
Page 4 Line 16-18 I find your language a little non-specific for this type of technical paper. Consider a revision of every sentence to make clear what you refer	The section has been rephrased according	P 7

to. Differentiate between 'technical skills' and 'non-technical skills'? as "task" covers both of these. Your saying "it is essential to evaluate task and team performance in order to develop strategies for improvements in clinical practice" Is unclear as both are 'tasks'. Technical skills and non-technical skills are the common terminology in this field?	to the definitions by Salas.	
Page 4 Line 22: 'Handled' would be best stated as 'managed' as you don't 'handle' a patient.	Thank you, corrected.	P 7
Page 4 Line 24 -28: Consider having a native English speaker scholar edit the text throughout as there are better ways to express what you are saying and be clearer in this section.	The section has been rephrased and the manuscript has undergone professional language editing an additional time by the same company.	P 7-8
Line 20: Its not "the timing of the task" (when it was done) its the timing of the components of the task? (how long they took, etc).	Thank you. Now rephrased to include also the timing of the components.	P 7
Line 24: suggest: as a pilot study, psychometric data such as reliability and validity were not reported.	Thank you. Changed accordingly.	P 7
Line 26: What is "a case-specific A-E checklist"? The TTET is one, surely	The section has been rephrased.	P 7
Line 30: I think you mean 'nontechnical' team performance?	We have now added a definition of team performance in the paragraph.	P 7-8
Line 38-9: "The TEAM instrument has not yet been validated for use outside the commonwealth nor has the instrument been tested at a subscale level: suggest this is unimportant, rephrase.	The sentence was deleted.	
See McKay et al. Team performance in resuscitation teams: comparison and critique of two recently developed scoring tools. Resuscitation 2012;83:1478-83 and Cooper, Cant. Measuring the non-technical skills of medical emergency teams: An update on validity and reliability of Team Emergency Assessment Measure (TEAM). Resuscitation, 2014; 85: 31-33. URL: <a href="http://dx.doi.org/10.1016/j.resuscitation.2013.08.276">http://dx.doi.org/10.1016/j.resuscitation.2013.08.276</a>	Both of the mentioned references have been cited in the rewritten introduction. Thank you for the suggestions.	P 7-8
I am really not able to edit the whole of the Introduction for you so please attend to the issues I raise and improve the English grammar and specificity/accuracy of language in the whole section.	The manuscript has been substantially rewritten and has also undergone a	Throughout the manuscript

	second language revision.	
Methods: Page 5 line 18: add 'with medical students' Participants should be described prior to the section on 'setting'?? Give response rate.	Manuscript changed accordingly and the response rate is given in the manuscript.	P 9
Page 5 Line 30: what are the roles that students played?	All students played the roles of interns with the attending nurse presently unavailable. This has now been clarified in the manuscript.	P 10
Line 52: how can 55 students be divided into 23 teams? Please explain.  Were the scenarios conducted in English?	The students' participation in different scenarios is now clarified in the rephrased Participants section.  The scenarios were conducted in Swedish. This is now clarified in the manuscript.	P 9 and P 11
The Methods section page 6 to page 9 is too long and tortuous. I suggest it would be beneficial to present a table of the 'students' experience' from start to finish- showing each step of the simulation and testing in columns? Hence, you could reduce text	The text has been revised. We tried to make the suggested table. It became quite big and, in our view, it did not add enough to warrant publication. Thus, the table will not be added as the method section has been revised according to the suggestions.	
Page 8 Line 37: The Methods language appears to be clearly written and considers relevant issues, thank you.  The student sample appears to be too small.	The student sample was 55 and the aim was to allow for calculations of means and standard deviations with a fair precision.  We have added a short paragraph describing the power calculations.	P 15



Page 9 line 5: we just need to see the code number of the ethics approval, not further details.	We have kept the information regarding the approval and deleted three sentences with details to reduce the text.	P 16
Page 9 Results section: I would much prefer the headers under this section to reflect what you are testing rather than just given the instrument name?	The headers has been expanded to reflect what is tested	P 16 - 17
Page 9 line 15: The Results are presented as clinician rater's consistency alone, and student teams performance results are missed. To be convincing, you should summarize the research in text (the student results) and then build on this by reporting instrument validity issues?	The results section has been restructured, rewritten and expanded according to the suggestions.	P 16 - 17
Page 10, line 11. CVI? This is the first mention? Explain.	Content validity index (CVI) is now better introduced in the methods section.	P 13
Page 10 line 30: Limitations - please remove this text and redraft to address the limitations or shortcomings of your study, rather than what you might do. Design, sampling, transferability, what did you do to limit bias? Student rather than professional education?	Limitations has been completely revised.	P 20 - 21
Page 10 line 58: suggest you redraft the conclusion- did you achieve the study aims? What was the outcome? What does the study contribute to knowledge? and to medical student education? professional education?	The conclusions have been revised.	P 21
Abstract: Revise as applicable. The conclusion here should match the conclusion in the paper.	The abstract has been revised accordingly	P 4 - 5
Table 1 is incomplete, should present number and then percent. Some categories where there are n=2, can be collapsed.	Sorry about the missing numbers. Now the table is complete and some categories have been collapsed.	P 23
Table 2: CV of professional raters should be removed and be summarized in text	Table 2 has been removed and is now summarized in text under "Participants".	Deleted
Table 4a: I suggest place the response scale in the instrument to complete it?, place a footnote that a Swedish version was used for data collection? I suggest it would be more usual to state the header as Situation Awareness items for Case 1.	Tables 4a and 4b (Now tables 2A and 2B) has been improved according to the suggestions.	P 24 - 27

Table 5: suggest alter header- this is the student agreement about usefulness and feasibility of prospective collection of situation awareness items.	The title of the table (now table 4) has been adjusted accordingly.	P 29
I think this paper requires much revision	Yes, we have made substantial revisions according to the suggestions and the paper has undergone a renewed language revision by an accredited translation agency.	Throughout the manuscript

### VERSION 2 – REVIEW

<b>REVIEWER</b>	Simon Cooper Federation University Australia Declared: Note that I am one of the developers of the TEAM instrument used in this study
<b>REVIEW RETURNED</b>	24-May-2019

<b>GENERAL COMMENTS</b>	<p>The paper is much improved but there are still a few concerns – predominantly with the flow and a number of English grammar issues remain. Details as below:</p> <p>Abstract – the design .... You say methodological study..... but you use the observational cohort study checklist?</p> <p>I note that the correct title for SAGAT has not been corrected – it should read Situation Awareness Global Assessment Technique. See earlier comments in first review</p> <p>You indicate you measured performance ABCDE using a Likert scale – to be correct a Likert scale is used to measure attitudes – suggest therefore that you reword this as a “rating scale”.. or similar</p> <p>You have not referenced your CVI measure and it looks a little odd anyway.... CVI is used to measure relevance and clarity of each item ..... think Polit and Beck is the best reference for this</p> <p>Page 17 – why would you do ICC for one rater only? ...when your idea is to measure inter-rater reliability....not sure of your logic with this one?</p> <p>Page 18 – you don’t have to pause a scenario to use SAGAT you can use it at the end .....I think also you imply that the freeze did not affect their future performance but 29% thought it did?</p> <p>The limitations section is much too long with a repeat of the methods used as opposed to the limitations per se</p> <p>Also suggest that the stats/methods are checked by a statistician - unless you have one on your team - on a few occasions they are unclear</p>
-------------------------	--

<b>REVIEWER</b>	A/P Robyn Cant Federation University Australia, Victoria, Australia
<b>REVIEW RETURNED</b>	29-May-2019

**GENERAL COMMENTS**

Thank you for this well-written revision ( i.e., grammar) of an article in English based on a Swedish study. I find that in this revision 1 there are still issues that should be improved to inform the reader of your exact study details and justify your results - sorry!. I give my detailed suggestions for improvement below.

Page 7 line 12: "Scoring scenarios ..." - sentence is unclear please revise.

I am unclear about your sample: For example, "in total, 55 students (81%) participated in the study (Table 1) and 20 of them participated in both scenario A and scenario B". ... ("In each scenario, three students were active and 1–2 were observers"). Therefore, only three participated in the interactive role play in each scenario (the actual simulation). Please make this clear.

Page 8 line 50: this next section is no longer about 'participants' – I suggest you create a section on verification of the instruments and particularly, on photo-elicitation?

Page 10 line 38: Please justify how a valid checklist of 58 items can be reduced to ten items for students' clinical performance? – what was omitted?

Page 11 line 37: I suggest the heading here refer to the post- test ratings of student performance using assessment instruments? (not clear in your language at the start). How many raters?

Page 12 line 48: please state how many items in each freeze. With your tracked changes in the tables it is impossible to see the final version in a pdf. I suggest in future just present the final tables- I have viewed them in Word but they are not properly formatted. Your tables: you present the questions asked in the SAGAT (all well and good). However there is a need to report the results in these tables as the responses are just dichotomous- either yes or no? how many were correct, how many were incorrect - totals/range.

Page 14 line 12: again, give the sample number- all three participants in each scenario reported the SAGAT? 23 teams by three in each?

Page 15 line 45: the name of the committee chairperson is not required.

Page 15 line 57 Results: please state the student sample numbers included in analysis. I am still not clear whether this the three person teams x 23 scenarios.

Page 16 line 50: please summarize data (add number) in this sentence.

Table 5 should become Table 3 as you present the main findings earlier in Results? Also specify in the table, the sample you report on. Mention total possible score for each instrument? Note that the TEAM instrument score should represent items 1-11 in total (a score out of a possible 44)? as each section differs. It would be helpful to report the range of scores seen.

Discussion page 17: line 23: You seem to have missed stating that this study explored students' performance and this may be a major limitation of current results? and may or may not reflect more

	<p>experienced staff's performance. Your results do provide an indication- however, professionals may rate differently? (see TEAM papers around professionals' scores). The discussion section should be shorter.</p> <p>Page 17 line 58: and the impact of student (novice learners) scores on instrument reliability? How do your scores compare with published data? Suggest place this earlier</p> <p>Page 19 limitations- this section should relate to efforts to preclude bias- selection and representativeness of students? Prior simulation experience? The section is too long and unnecessarily repeats earlier stated issues about instruments/statistics.</p> <p>Conclusion: include the fact this this refers to a student study. Strengths and limitations section: include the fact this this refers to student study. Abstract: include the fact this this refers to student study.</p> <p>Title: : include the fact this this refers to a student study. I think refer to reliability first? The reliability of instruments that measure situation awareness, team and task performance in a simulation setting with medical students in Sweden.</p> <p>Table 4: suggest correct the header to read Agree/strongly agree and disagree/strongly disagree? Table 6: please explain the header n/n And the data for the reader to be able to understand the content without referring to the text of the paper. Thank you.</p>
--	--

### VERSION 2 – AUTHOR RESPONSE

<p>Reviewer: 1</p> <p>The paper is much improved but there are still a few concerns – predominantly with the flow and a number of English grammar issues remain.</p> <p>Details as below:</p>	<p>The manuscript has been language edited twice by a professional agency. According to reviewer 2 “Thank you for this well-written revision (ie., grammar) of an article in English based on a Swedish study.” However, if needed, this process can be repeated.</p>	
<p>Abstract – the design .... You say methodological study..... but you use the observational cohort study checklist?</p>	<p>The STROBE checklist is used for a number of different study designs as Cohort, Case-control and Cross-sectional studies. It was the best fit as we could not find a more appropriate checklist for this kind of methodological approach, e.g. from <a href="https://www.equator-network.org/">https://www.equator-network.org/</a>. The GRRAS checklist would be too limited</p>	

	as it focusses only on reliability studies. As the checklist is not optimal for the present study it has been removed from the submission.	
I note that the correct title for SAGAT has not been corrected – it should read Situation Awareness Global Assessment Technique. See earlier comments in first review	Thank you, we have corrected this in the manuscript by consistently rephrasing SAGAT instrument to SAGAT.	Consistently in the manuscript
You indicate you measured performance ABCDE using a Likert scale – to be correct a Likert scale is used to measure attitudes – suggest therefore that you reword this as a “rating scale’.. or similar	The scale used for rating ABCDE has been renamed to “rating scale”.	Consistently in the manuscript
You have not referenced your CVI measure and it looks a little odd anyway.... CVI is used to measure relevance and clarity of each item ..... think Polit and Beck is the best reference for this	Sorry, flow of the manuscript was unfortunately interrupted and the reference (32) was not in immediate proximity with the sentence describing CVI. The sentence has been rephrased.	P 13
Page 17 – why would you do ICC for one rater only? ...when your idea is to measure inter-rater reliability....not sure of your logic with this one?	The idea was to show the effect of using several raters on the reliability of the instrument. The paragraph relating to ICC and the TEAM instrument has been rephrased slightly in the discussion.	P18
Page 18 – you don’t have to pause a scenario to use SAGAT you can use it at the end .....I think also you imply that the freeze did not affect their future performance but 29% thought it did?	Both sentences have been rephrased.	P18
The limitations section is much too long with a repeat of the methods used as opposed to the limitations per se	Agree. The section has now been shortened.	P19
Also suggest that the stats/methods are checked by a statistician - unless you	One of the coauthors are statistician.	

have one on your team - on a few occasions they are unclear		
Reviewer: 2  Thank you for this well-written revision (ie., grammar) of an article in English based on a Swedish study. I find that in this revision 1 there are still issues that should be improved to inform the reader of your exact study details and justify your results - sorry!. I give my detailed suggestions for improvement below.	Thank you, we will follow the suggestions.	
Page 7 line 12: "Scoring scenarios ..." - sentence is unclear please revise.  I am unclear about your sample: For example, "in total, 55 students (81%) participated in the study (Table 1) and 20 of them participated in both scenario A and scenario B". ... ("In each scenario, three students were active and 1-2 were observers"). Therefore, only three participated in the interactive role play in each scenario (the actual simulation). Please make this clear.	Thank you, rephrased.  This has been rephrased: <i>In each scenario, 3-4 students were active and 1-2 were observers. Therefore, only 3-4 participated in the interactive role play in each scenario (the actual simulation). In all, the 55 unique students made up a total of 23 teams with 3-4 participants in each team and 20 students participated in both case A and case B.</i>	P7  P9
Page 8 line 50: this next section is no longer about 'participants' – I suggest you create a section on verification of the instruments and particularly, on photo-elicitation?	In order to clarify the content, we have changed the heading to <i>Participants and raters</i> .	P8
Page 10 line 38: Please justify how a valid checklist of 58 items can be reduced to ten items for students' clinical performance? – what was omitted?	The reduction of the checklist into ten items is now described in the methods section.	P10
Page 11 line 37: I suggest the heading here refer to the post- test ratings of student performance using assessment instruments? (not clear in your language at the start).	The heading for participants has been changed to participants and raters.	P8

How many raters?	The sentence has been rephrased to better clarify the number of raters.	
Page 12 line 48: please state how many items in each freeze. With your tracked changes in the tables it is impossible to see the final version in a pdf. I suggest in future just present the final tables- I have viewed them in Word but they are not properly formatted.	The sentence has been rephrased to include the number of items in each freeze.	P12
Your tables: you present the questions asked in the SAGAT (all well and good). However there is a need to report the results in these tables as the responses are just dichotomous- either yes or no? how many were correct, how many were incorrect - totals/range.	The fraction of correct answers has been added to Table 2 for each item in the SA questionnaire.	Table 2
Page 14 line 12: again, give the sample number- all three participants in each scenario reported the SAGAT? 23 teams by three in each?	Yes, all participants answered the SAGAT questionnaires and there were 23 teams with 3 participants in each team. How 55 unique students were combined into 23 teams with 3 participants in each team is now stated under Setting.	P 9
Page 15 line 45: the name of the committee chairperson is not required.	The name has been removed.	P 15
Page 15 line 57 Results: please state the student sample numbers included in analysis. I am still not clear whether this the three person teams x 23 scenarios.	The paragraph now starts with a new sentence stating the number of participants and the number of unique teams.	P16
Page 16 line 50: please summarize data (add number) in this sentence.  Table 5 should become Table 3 as you present the main findings earlier in Results? Also specify in the table, the sample you report on. Mention total possible score for each instrument? Note that the TEAM instrument score should represent items 1-11 in total (a score out of a possible 44)? as each section differs.	The data has been added to the sentence.  The number of the tables follows from their first reference in the manuscript. Table 3, 4 and 5 are first mentioned in the methods section.  For the TEAM instrument we report the average score per item as it was	P17  No change

<p>It would be helpful to report the range of scores seen.</p>	<p>reported by Cooper et al in Resuscitation 2010. The max score for Global rating has been added as a footnote to the table and are also described in methods.</p>	<p>Table 5</p>
<p>Discussion page 17: line 23: You seem to have missed stating that this study explored students' performance and this may be a major limitation of current results? and may or may not reflect more experienced staff's performance. Your results do provide an indication- however, professionals may rate differently? (see TEAM papers around professionals' scores).</p>	<p>The sentence (P17 L 23 in old submission) has been changed to reflect that professionals most likely scores higher on a scale measuring performance.</p>	<p>P17</p>
<p>The discussion section should be shorter.</p>	<p>We do believe that the different instruments needs to be discussed. Thus it does take some space. Hopefully, the discussion will advance our understanding of the field.</p>	
<p>Page 17 line 58: and the impact of student (novice learners) scores on instrument reliability? How do your scores compare with published data? Suggest place this earlier</p>	<p>Our scores are now discussed in relation to published data</p> <p>The impact of novice learners are now discussed in the revised limitations.</p>	<p>P18</p> <p>P20</p>
<p>Page 19 limitations- this section should relate to efforts to preclude bias-selection and representativeness of students? Prior simulation experience? The section is too long and unnecessarily repeats earlier stated issues about instruments/statistics.</p>	<p>The limitations section has been shortened and refocused.</p>	<p>P 19</p>
<p>Conclusion: include the fact this this refers to a student study.</p> <p>Strengths and limitations section: include the fact this this refers to student study.</p> <p>Abstract: include the fact this this refers to student study.</p>	<p>This issue is now addressed at all three locations.</p>	
<p>Title: include the fact this this refers to a student study. I think refer to reliability</p>	<p>Thank you for the suggested title.</p>	<p>Title</p>



first? The reliability of instruments that measure situation awareness, team and task performance in a simulation setting with medical students in Sweden.		
Table 4: suggest correct the header to read Agree/strongly agree and disagree/strongly disagree?	Corrected.	Table 4
Table 6: please explain the header n/n And the data for the reader to be able to understand the content without referring to the text of the paper.	A footnote has been added to explain the table.	Table 6

### VERSION 3 – REVIEW

<b>REVIEWER</b>	SIMON Cooper Federation University Australia Developer of TEAM tool cited in this paper
<b>REVIEW RETURNED</b>	11-Jul-2019

<b>GENERAL COMMENTS</b>	Many thanks for this re-submission. Just some minor typos throughout and I note that in table 2B etc you have added a header with Correct - I assume this is the mean score listed below so suggest indicate this in the column heading?
-------------------------	--

<b>REVIEWER</b>	A/P Robyn Cant Fedartion University Australia
<b>REVIEW RETURNED</b>	22-Jul-2019

<b>GENERAL COMMENTS</b>	<p>Thank you for this revision, taking into account reviewers' previous suggestions.</p> <p>The paper is improved with better justification of methodology and clearer (more specific) language,. However there are still several amendments that need to be made to your results- perhaps as the result of how you use English.</p> <p>Page 83 line 32: As before: the TEAM instrument is a valid scale therefore you should present the scores as a total of a possible 44 points. Then, if you wish you could present the subscore means? In addition, when discussing the TEAM scores in your discussion section (p64) I suggest comparison be made with the performance scores of students' data as seen in the original TEAM development paper- how were these higher or lower than the later interprofessional team scores in clinical practice?</p>
-------------------------	---

	<p>The word limit usually allowed for papers is 4000 words and you are over the word limit by 800 words, making the paper less readable and your main points are less well argued because they become lost.</p> <p>Please note the BMJ Open Journal's guidance on this point: 'We also recommend, but do not insist, that the discussion section is no longer than five paragraphs and follows this overall structure (you do not need to use these as subheadings): a statement of the principal findings; strengths and weaknesses of the study; strengths and weaknesses in relation to other studies, discussing important differences in results; the meaning of the study: possible explanations and implications for clinicians and policymakers; and unanswered questions and future research.'</p> <p>The discussion is too lengthy (is it 1800 words?) and meanders off topic, suggest revise and shorten (as requested R1). In research articles there is no way that a whole project can be fully explained in a single paper so there is a skill in being selective as an author. In many places you can reduce words to be more concise- by not restating the full descriptive terminology when we have already had it described to us -you need telegraphic language. There are areas that can be removed: eg., page 66, line 35 -47. In the discussion we are not asking you for a rehash of what you did or did not do, we are asking you to compare your results with current literature and say how the results are supported by the literature or not? page 67 line 14-23 is not essential information and should be omitted. I suggest re-edit the paper to a maximum 4000 words plus 5% if needed. This is the longest paper I have ever reviewed (of many).</p> <p>Conclusion Lines 28-40 A five-line sentence is too long and the actual meaning may be lost to readers. Please revise to shorten the sentence to be more meaningful. Suggest revise the last sentence to read: '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.' (better clarity)</p> <p>Table 2a, table 2b- please state in the table header how many participants are rated in these data</p> <p>Please present these tables as smaller font they are too lengthy at present</p> <p>Please be selective with tables some are not essential in the actual paper? And can be seen in supplementary data? The results tables need to be a priority in the text and any that describe an instrument are numbered as supplementary data?</p> <p>The tables are almost unreadable as they are not presented in final version.</p> <p>Did the students achieve a pass score?</p>
--	--

### VERSION 3 – AUTHOR RESPONSE

Comment	Response	
<p>Reviewer 1</p> <p>Many thanks for this re-submission. Just some minor typos throughout and I note that in table 2B etc you have added a header with Correct - I assume this is the mean score listed below so suggest indicate this in the column heading?</p>	<p>Thank you!</p> <p>We have tried to find the minor typos and correct them.</p> <p>To clarify the meaning of correct, the headers has been changed to Situation awareness items for Case X. Response scale, level and proportion correct (%) (n=xx).</p>	<p>Tables 2A and 2B</p>

Full Manuscript Page 2-45

Full Manuscript with revisions shown Page 46-91

Comment	Response	
<p>Reviewer 2</p> <p>Thank you for this revision, taking into account reviewers' previous suggestions.</p> <p>The paper is improved with better justification of methodology and clearer (more specific) language,. However there are still several amendments that need to be made to your results- perhaps as the result of how you use English.</p>	<p>Thank you! The process does indeed improve the manuscript.</p>	
<p>Page 83 line 32: As before: the TEAM instrument is a valid scale therefore you should present the scores as a total of a possible 44 points. Then, if you wish you could present the subscore means? In addition, when discussing the TEAM scores in your discussion section (p64) I suggest comparison be made with the performance scores of students' data as seen in the original TEAM development</p>	<p>Table 5 has been changed according to the suggestions to show the sum of the instruments items and then showing the mean item score as a complement.</p> <p>The results have been updated to show TEAM sum instead of TEAM average for items 1-11.</p> <p>The discussion has been changed to include a brief comparison to the students'</p>	<p>Table 5</p> <p>Page 16</p> <p>Page 18</p>

<p>paper- how were these higher or lower than the later interprofessional team scores in clinical practice?</p>	<p>performance in the original paper.</p>	
<p>The word limit usually allowed for papers is 4000 words and you are over the word limit by 800 words, making the paper less readable and your main points are less well argued because they become lost. Please note the BMJ Open Journal's guidance on this point:          'We also recommend, but do not insist, that the discussion section is no longer than five paragraphs and follows this overall structure (you do not need to use these as subheadings): a statement of the principal findings; strengths and weaknesses of the study; strengths and weaknesses in relation to other studies, discussing important differences in results; the meaning of the study: possible explanations and implications for clinicians and policymakers; and unanswered questions and future research.'</p>	<p>Yes, this paper includes a report on several instruments/questionnaires and thus this makes the paper a bit wordy. However, we believe that benefits of reporting the details of the different measurements in a combined paper are more than the drawbacks with a lengthy paper.</p>	<p>-</p>
<p>The discussion is too lengthy (is it 1800 words?) and meanders off topic, suggest revise and shorten (as requested R1). In research articles there is no way that a whole project can be fully explained in a single paper so there is a skill in being selective as an author. In many places you can reduce words to be more concise- by not restating the full descriptive terminology when we have already had it described to us - you need telegraphic language. There are areas that can be removed: eg., page 66,</p>	<p>The discussion per se is 810 words and limitations of the study is an additional 482 words, a total of 1292 words. By deleting the suggested paragraphs, and making some adjustments to the text, the discussion is now down to 1159 words and the main text is 4736 words.</p>	<p>Page 20 + reduction of the text at several places.</p>

<p>line 35 -47. In the discussion we are not asking you for a rehash of what you did or did not do, we are asking you to compare your results with current literature and say how the results are supported by the literature or not? page 67 line 14-23 is not essential information and should be omitted. I suggest re-edit the paper to a maximum 4000 words plus 5% if needed. This is the longest paper I have ever reviewed (of many).</p>		
<p>Please present these tables as smaller font they are too lengthy at present Please be selective with tables some are not essential in the actual paper? And can be seen in supplementary data? The results tables need to be a priority in the text and any that describe an instrument are numbered as supplementary data?</p> <p>The tables are almost unreadable as they are not presented in final version.</p> <p>Did the students achieve a pass score?</p>	<p>The final design of the tables is in the hands of the journal, but we will in the future try to make the tables more comprehensible.</p> <p>As requested by the journal we have submitted two versions of the manuscript and both are shown in the pdf – first a final version as submitted and then a version with all changes indicated. The page numbers in the review includes that you might be reading the version with all changes indicated. We agree that the version with all changes shown with revisions shown is almost unreadable.</p> <p>No, there were no pass or fail decisions made using either of the tools/instruments.</p>	<p>No change except for adjusted font size in tables</p> <p>No change</p> <p>No change</p>

<p>Conclusion Lines 28-40 A five-line sentence is too long and the actual meaning may be lost to readers. Please revise to shorten the sentence to be more meaningful.</p> <p>Suggest revise the last sentence to read: '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.' (better clarity)</p> <p>Table 2a, table 2b- please state in the table header how many participants are rated in these data</p>	<p>The conclusion has slightly been rephrased to increase the readability.</p> <p>Thank you for the suggestion. Changed accordingly.</p> <p>Added as requested</p>	<p>Conclusion</p> <p>Conclusion</p> <p>Table 2</p>
--	--	--