Research Article

Anesthesia & Research

Anaesthetic Non-Technical Skills for Anaesthetic Practitioners (ANTS-AP) system: Behavioral rating system for assessing anaesthetic assistants' non-technical skills Disorders

J. S. Rutherford,[1*] R. Flin,[2] A. Irwin[3] and A. K. McFadyen[4]

1. Consultant, Department of Anaesthesia, Dumfries & Galloway Royal Infirmary, Dumfries, UK
2. Professor of Industrial Psychology,
3. Research Fellow, School of Psychology, University of Aberdeen, Aberdeen, UK
4. Statistical Consultant, AKM-STATS, Glasgow, UK

***Correspondence:**
J. S. Rutherford
Email: **john.rutherford@nhs.net**

Received: 7-Feb-2024; **Published:** 21-Feb- 2024;

Citation: J. S. Rutherford, Anaesthetic Non-Technical Skills for Anaesthetic Practitioners (ANTS-AP) system: Behavioral rating system for assessing anaesthetic assistants' non-technical skills Disorders

Abstract:

This study tested the reliability, validity and usability of a prototype behavioural rating system for the non-technical skills of assistants working with the anaesthetist. Anaesthetic nurses and operating department practitioners (n = 48) used the prototype Anaesthetic Non-technical Skills for Anaesthetic Practitioners (ANTS-AP) system to rate the non-technical skills of anaesthetic assistants in 12 videos of simulated theatre work. Test–retest reliability was assessed with a sub-sample (n = 12). The skill categories assessed were 'situation awareness', 'teamwork and communication' and 'task management'. The internal consistency for the ratings of elements in categories was acceptable (Cronbach's a of 0.78, 0.77 and 0.69, respectively), with more modest inter-rater reliability (intraclass correlations for categories 0.54, 0.70, 0.86), test–retest reliability (intraclass correlations 0.68, 0.58, 0.38) and accuracy (weighted kappa 0.39). Most participants considered the system complete (n = 42, 87%), the wording clear (n = 48, 100%) and the system useful for structuring observation (n = 48, 100%).

Introduction

Error and harm in the operating theatre have been recorded from surgery and anaesthesia for many years [1–3]. Many patient safety incidents in theatre have their origin in human error [4]. Assistants can minimise or ameliorate the development of such incidents, such as by reminding an anesthetist that he or she can stop trying to intubate the trachea of a patient with a difficult airway [5].

One approach to this problem has been to identify the non-technical skills that enhance safety and efficiency by utilizing techniques developed in aviation to study and rate human behaviour. These systems describe examples of good and poor behaviours within elements that are themselves grouped into skill categories.Within theatre, non-technical skills behaviour rating scales have been developed for anaesthetists (ANTS) [6], surgeons (NOTSS) [7], scrub practitioners (SPLINTS) [8] and nurse anaesthetists (NANTS) [9]. No non-technical skills taxonomy for staff assisting anaesthetists in theatre was found in the literature [10]. In the UK, this assistance is provided by anaesthetic nurses and operating department practitioners. For this article, we refer to anaesthetic nurses and operating department practitioners as 'anaesthetic practitioners'.

The task analysis to develop the prototype Anaesthetic Non-technical Skills for Anaesthetic Practitioners (ANTS-AP) system started with 33 interviews with anaesthetic practitioners (n = 22) and consultant anaesthetists (n = 11), asking about the skills required for effective anaesthetic assistance [5]. Following comments from both these groups regarding interactions with trainee anaesthetists, interviews were conducted with trainee anaesthetists (n = 12) (J.S. Rutherford, R. Flin, L. Mitchell, unpublished observations). Thematic analysis [11, 12] of the behaviours mentioned in the interview data identified the main non-technical skill categories for anaesthetic practitioners as being situation awareness, teamwork and task management.

In addition, critical incidents reported to the Australian Anaesthe-

-tic Incident Monitoring System from 2002 to 2008 were reviewed for evidence of the contribution of non-technical skills demonstrated by anaesthetic practitioners to anaesthetic critical incidents [13]. This analysis supported the findings of the interview study in that examples of situation awareness, teamwork and task management were all evident, and appeared much more frequently than examples of decision-making.

To create a behavioural rating system with categories and elements from the phrases identified by the interview studies, the phrases needed to be sorted into a structure that made sense to the anaesthetic practitioners. The 2205 behavioural phrases identified in the interviews by thematic analysis were discussed and debated in four focus groups [14] of anaesthetic practitioners (n = 6, 7, 3, 4) to form sets of behaviours. The headings of categorised behaviours developed by these focus groups were reviewed by a subject matter expert Delphi group [15] of anaesthetic practitioner lecturers (n = 6) to condense the headings and organise the resulting set into the structure required for a taxonomy of non-technical skills with categories, elements and behaviours.

The prototype ANTS-AP system and descriptive handbook were designed by JR and RF based on the results of the task analysis described above. The design criteria were that the system should fit on one side of a page to facilitate ease of use, with the minimum number of categories and elements consistent and adequate to describe non-technical skills frequently used by anaesthetic practitioners [16]. The prototype ANTS-AP taxonomy had three categories:'situation awareness'; 'teamwork and communication' and 'task management', with component elements as shown in Table 1.

The four-point rating scale (poor, marginal, acceptable, good), with the fifth 'not required' option used in ANTS, NOTSS and SPLINTS was adopted for the ANTS-AP assessment form, shown in Table 2.

The prototype ANTS-AP system required evaluation, and this study was designed to assess whether its reliability, validity and usability were sufficient for use. The research questions we posed for reliability included: whether the elements were consistent within their categories; could different users give similar scores; if seeing the same behaviours later, were their scores consistent over time; and how did the accuracy of novice raters compare with an expert panel. The research questions posed for validity were: was the system complete, and could the behaviours used to score the ANTS-AP system all be observed. The research questions posed for usability were: was the system.

Methods

The study was approved by the University of Aberdeen Psychology Ethics Committee. The West of Scotland Research Ethics Service confirmed that NHS Research Ethics approval was not required. Video clips (2–8 min) of simulated anaesthetic work in theatre (n = 12) were filmed at the Scottish Clinical Simulation Centre. The videos all portrayed an anaesthetic assistant working with one or two anaesthetists, as well as with other members of the surgical team. The scenario scripts were based on themes that had been described during the earlier interview study [5], and the aim was for the anaesthetic practitioner to demonstrate good, average and poor behaviour in each of the different non-technical skills categories. Twelve video clips were filmed to have enough material for inter-rater reliability testing, as described below.

Table 1 The prototype ANTS-AP system showing the categories, elements and rating scale, but omitting behavioural markers and definitions of categories and elements.

Category	Element
Situation awareness	Gathering information
	Recognising & understanding
	Anticipating
Teamwork & communication	Co-ordinating with team
	Supporting colleagues
	Asserting
Task management	Planning & preparing
	Prioritising & problem solving
	Coping with pressure

Table 2 The rating scale and definitions used for scoring the ANTS-AP categories and elements.

Score		Definition of standard of behaviour for the score
1	(Poor)	Performance was not acceptable and could potentially have endangered patient safety; remedial action required
2	(Marginal)	Performance indicated cause for concern; considerable improvement is needed
3	(Acceptable)	Performance was of a satisfactory standard but could be improved
4	(Good)	Performance was of a consistently high standard, enhancing patient safety; it could be used as a positive example for others
N/R	(Not required)	Not required; skill was not observed because it was not required in this case

Anaesthetic practitioners were eligible to participate if they had at least one year's experience working in NHS hospitals. For recruitment, senior theatre managers in 18 hospitals across the UK were asked to display a poster advertising the study on their theatre notice board. Anaesthetic practitioners who volunteered were invited to attend a workshop to rate the behaviour in the videos using the prototype ANTS-AP system. The flowchart for this study is shown in Fig. 1.

The participants who contacted the author before the workshop were given access to an online module introducing human factors (available on the North West Simulation Education Network) and asked to complete this before the workshop. The module took 1–2 h to complete. Participants were also sent the handbook of the ANTS-AP system for familiarisation.

The workshops were usually held at the hospital of the staff concerned, and lasted 3–4 h. Written consent was obtained, and

the lead author gave a short presentation on the ANTS-AP system as well as on rating methods and biases. After the ANTS-AP system was explained, the participants were then asked to use it to rate the behaviour of the anaesthetic practitioner assisting the anaesthetist in a practice video clip, and then to discuss their scores together. Emphasis was placed on only scoring the behaviours that could be observed. If the behaviour relating to an element was displayed in a scenario where that behaviour was not required, the participants were advised to score that element as 'not required', whereas if the behaviour was absent but should have been displayed it should be scored as poor or marginal. Once the training was completed, the participants were then shown the 12 video clips and asked to rate individually the non-technical skills of the anaesthetic practitioner from the behaviours observed using ANTS-AP. No feedback was given to the raters. On occasion, participants wanted to discuss the video they had just seen, but comments were not permitted until scoring was complete. The sheets used to record the ANTS-AP scores also had space for optional free-text comments by the raters.

Participants were asked to complete a questionnaire with background information at the end of the workshop. They were also asked about the design of the ANTS-AP system, and to comment on whether the quality of the videos was adequate to assess behaviours. The characteristics of the participants are outlined in Table 3.

A second workshop was held at least a week after the initial workshop and lasted less than an hour. Twelve participants who agreed to help with the test-retest reliability were re-shown three video clips and asked to rate them.

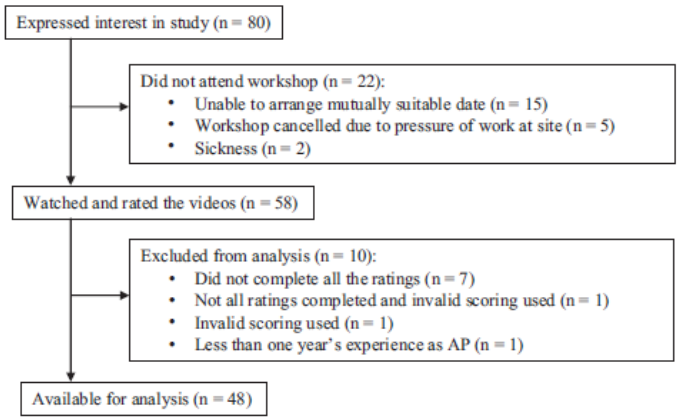

Figure 1 CONSORT flowchart. AP, anaesthetic practitioner.

The reliability of the ANTS-AP system was assessed for internal consistency, inter-rater reliability, test–retest reliability and accuracy.

The internal consistency of the elements within each category was tested by Cronbach's a (30 participants required as three elements per category and 10 participants per element required). A value $0.6 \leq a \leq 0.9$ was desirable (scores $a > 0.9$ may suggest that there

Table 3 Characteristics of participants meeting inclusion criteria for analysis in this study. Values are number (proportion) or median (IQR [range]).

Male/female	16 (33%)/32 (67%)
English first language	
Yes	45 (94%)
No	2 (4%)
Not answered	1 (2%)
Years of experience	10 (5.5–19.5 [1–30])
Previous training in assessment	28 (58%)
Involved with previous ANTS-AP studies	8 (17%)
Pre-workshop online human factors reading	
Completed	16 (33%)
Registered only	2 (4%)

are redundant elements) [17]. All values were used for analysis, with 'not required' included.

The inter-rater reliability was assessed by the scores of two raters chosen at random for all 12 videos, giving 80% power to detect an inter-class correlation of between 0.6 and 0.9 at the 5% level of significance. The 'not required' ratings were treated as missing values.

The test–retest reliability was assessed where the scores of 12 participants who rated three video clips on two occasions at least a week apart were correlated. This had 80% power to detect an intraclass correlation of between 0.6 and 0.9 with a 5% level of significance. The 'not required' values were treated as missing values.

Reference scores agreed by an 'expert panel' were compared with those of the participants by use of a weighted kappa. Again, 'not required' values were treated as missing values. An anaesthetist (JR), a psychologist (RF) and two experienced anaesthetist practitioners rated the scenarios, and their scores were used as the reference ratings to provide an 'expert' standard. If the difference in their scores was greater than 1 point, or one or more researchers had scored 1 or 2 while the others had scored 3 or 4, there was discussion to agree the expert rating. For example, if there were three 3 s and a 4, the modal value was chosen. If the rating had been two 3 s and two 4 s, the values of the clinicians were given priority.

Face validity and usability were assessed in the post-workshop questionnaire.

Face validity was assessed for completeness, where the participants were asked if there were any non-technical skills categories or elements that were missing or redundant, and observability, where participants were asked if they could observe behaviours demonstrating non-technical skills as described by the ANTS-AP system.

For usability, the participants were asked about the acceptability of the ANTS-AP system for training and assessing staff, and how easy or difficult the system was to use.

The ANTS-AP scores were analysed with SPSS for Windows (Version 22; SPSS, IBM Corp., Armonk, New York, USA), whereas the questionnaire responses were analysed with NVivo (Version 9; QSR, Doncaster, Australia).

Results

The questionnaire data indicated that 40 (83%) respondents said that the quality of the videos was adequate to rate the behaviours. The quality of the sound was suboptimal on a number of the videos, as one participant wrote in the questionnaire, "a bit poor, but I was still able to hear all the dialogue". This was compounded at one hospital by a loud air conditioning system. Most participants (n = 34, 71%) thought it would be easier to observe the behaviours on the videos rather than being in the operating theatre. Most (n = 44, 92%) thought they had enough training to be able to use the system effectively.

The internal consistency of the category ratings was respectable, with Cronbach's a = 0.78 for 'situation awareness', 0.77 for 'teamwork and communication', but 0.69 (minimally acceptable) for 'task management' [17]. Video 5 appeared to be unusual, with much lower scores than any other video, but it was not clear why (it illustrated a rapid sequence induction with poor preparation). If the Cronbach's a was recalculated with video 5's scores omitted, the a scores increased to very good for 'situation awareness' (0.82) and 'teamwork and communication' (0.81) and respectable for 'task management' (0.71).

Table 4 Inter-rater reliability scores for categories and elements of two randomly chosen participants for all 12 videos. The values are intraclass correlations. Values are number (95% CI).

Category or element	Intraclass correlations
Situation awareness	0.54 (0.01–0.84)
Gathering information	0.47 (0.00–0.81)
Recognising & understanding	0.52 (0.01–0.83)
Anticipating	0.62 (0.13–0.87)
Teamwork and communication	0.70 (0.27–0.90)
Co-ordinating with team	0.62 (0.11–0.87)
Supporting colleagues	0.61 (0.12–0.87)
Asserting	0.58 (0.08–0.86)
Task management	0.86 (0.59–0.96)
Planning & preparing	0.76 (0.20–0.93)
Prioritising & problem solving	0.43 (0.00–0.79)
Coping with pressure	0 (0.00–0.51)

The inter-rater reliability results are shown in Table 4 and the test–retest reliability in Table 5. The weighted kappa as a marker of the participants' accuracy compared with the researchers was fair [18], with mean (SD) j = 0.39 (0.15) for categories and 0.30 (0.07) for elements. Table 6 shows the responses to the validity and usability of the ANTS-AP system.

Discussion

The ANTS-AP system appears to be a usable prototype behavioural rating system (with the exception of the 'coping with pressure' element). The internal consistency of the elements within the categories was generally good. This could be a reflection of the process of development – the elements having been identified in the task analysis and organised by subject matter experts into groups that formed the categories. Alternatively, if the elements were scored first and the categories scored as an average of the elements, it would be surprising not to find high internal consistency.

The inter-rater reliability was disappointing. 'Situation awareness' as an intraclass correlation > 0.6 is usually considered the minimum for development of a scale [19], but in view of the very limited training given this is hardly surprising [20–22]. Reliable calibration of assessment of non-technical skills is normally considered to take at least two days, but this was not feasible for this study as we had difficulty recruiting staff for even a half-day workshop. That 'teamwork and communication' and 'task management' were acceptable is reassuring. The ANTS, NOTSS and SPLINTS used within-group agreement (rwg) instead of intraclass correlation. None of the four categories in ANTS met the acceptable agreement level of rwg > 0.7, whereas the social categories of NOTSS and all three categories in SPLINTS met acceptable levels of agreement.

The element 'coping with pressure' stands out as having the worst intraclass correlation for inter-rater reliability. If an anaesthetic practitioner is failing to cope, this will result in poor behaviour in the other non-technical skills elements. In development of the ANTS system, it was decided to omit 'coping with pressure' as it overlapped with the other elements so much (personal communication, Glavin R, Maran N). 'Coping with pressure' is an element in both NOTSS and SPLINTS, but had lower levels of agreement than other elements in both systems. It would seem reasonable to omit 'coping with pressure' from the ANTS-AP system as its inter-rater reliability is so much worse than any other element, and the behaviours overlap with the other elements.

The test–retest reliability and inter-rater reliability scores had similar issues. In contrast with the first workshop, the average intraclass correlation increased with each video. Unlike the initial workshop, there was no practice video on the retest workshop, so the poor intraclass correlation scores on the first video may have been learning effect. If we were to repeat this study, we would use practice video clips before the retest material.

The accuracy ratings were only fair [18], but the participants had no feedback to enable calibration. Some of the disagreements were due to limitations of the simulations used in the video. The actors in a video showing blood transfusion did not have the full paperwork normally available, resulting in lower scores for the 'situation awareness' element 'gathering information' by some participants. We had not scripted this as an example of poor behaviour, and Weber et al. recently reported a similar finding where pilots rated a pilot's behaviour differently depending on whether they noticed an unscripted hazard on the video recording [23]. Hospitals have their own routine ways of working, and the anaesthetic practitioners filmed for the videos performed cricoid pressure with one hand and used their free hand to pass equipment to the anaesthetist. Some participants commented adversely on this, as the routine in their hospital was to have a second person trained to perform cricoid pressure, thus freeing both hands of the anaesthetic practitioner to assist the anaesthetist.

Table 5 Test–retest reliability as assessed by the intraclass correlation of the scores of the categories and elements of the three test–retest videos. The values are intraclass correlations. Values are number (95% CI).

Category or element	Video 4	Video 6	Video 8	Average
Situation awareness	0.48 (0–0.82)	0.84 (0.54–0.95)	0.72 (0.27–0.91)	0.68
Gathering information	0.30 (0–0.73)	0.61 (0.09–0.87)	0.35 (0–0.75)	0.42
Recognising & understanding	0.27 (0–0.72)	0.88 (0.66–0.97)	0.65 (0.15–0.89)	0.60
Anticipating	0.65 (0.16–0.88)	0.76 (0.35–0.93)	0.85 (0.55–0.95)	0.75
Teamwork and communication	0.44 (0–0.80)	0.48 (0–0.82)	0.82 (0.50–0.94)	0.58
Co-ordinating with team	0.56 (0.03–0.85)	0.41 (0–0.79)	0.66 (0.21–0.89)	0.54
Supporting colleagues	0.40 (0–0.78)	0.31 (0–0.75)	0.75 (0.36–0.92)	0.49
Asserting	0.44 (0–0.79)	0.74 (0.24–0.93)	0.78 (0.29–0.95)	0.65
Task management	0.27 (0–0.72)	0.18 (0–0.68)	0.69 (0.20–0.90)	0.38
Planning & preparing	0.69 (0.24–0.98)	0.66 (0.10–0.90)	0.72 (0.22–0.92)	0.68
Prioritising & problem solving	0.36 (0–0.76)	0.28 (0–0.72)	0.71 (0.22–0.91)	0.45
Coping with pressure	0.42 (0–0.80)	0.52 (0–0.86)	0.42 (0–0.79)	0.45
Average	0.44	0.56	0.68	

Table 6 Participants' responses to questions about the usability and validity of ANTS-AP system in the end-of-workshop questionnaire. Values are number (proportion).

Construct	Abbreviated version of question	Yes	No	Unanswered		
Usability						
Acceptability	ANTS-AP useful for structuring observation?	48 (100%)	0	0		
	ANTS-AP helpful for training trainee AP?	47 (98%)	0	1 (2%)		
	ANTS-AP helpful for assessing trainee AP?	47 (98%)	0	1 (2%)		
	ANTS-AP helpful in developing the skills to be a good AP	47 (98%)	0	1 (2%)		
	ANTS-AP helpful in supporting theatre teaching?	47 (98%)	0	1 (2%)		
Ease of use	Wording meaningful for categories & elements	48 (100%)	0	0		
	Were the wording for categories and elements clear?	48 (100%)	0	0		
	Were examples of good behaviours helpful in identifying NTS?	48 (100%)	0	0		
	Were examples of poor behaviours helpful in identifying NTS?	46 (96%)	2 (4%)	0		
Validity						
Completeness	Did ANTS-AP address the key NTS displayed?	47 (98%)	1 (2%)	0		
	Do you think any categories or elements missing?	6 (12%)	41 (85%)	1 (2%)		
	Do you think any categories or elements unnecessary?	0	48 (100%)	0		
		Very difficult	Difficult	Average	Easy	Very easy
Observability	How easy to associate observed behaviours with ANTS-AP elements?	1 (2%)	3 (6%)	16 (33%)	23 (48%)	5 (10%)
	How easy was it to associate observed behaviours with ANTS-AP categories?	0	7 (15%)	12 (25%)	23 (48%)	6 (13%)

There were no superfluous categories or elements identified by the participants, but some participants considered there were missing elements. The issues that were raised were of wanting an element of 'supporting patients' and being unsure where to score 'communication'. As looking after the patient is the job of the anaesthetic practitioner, we felt this was the primary activity and hence not a non-technical skills element. Communication is involved in all the nontechnical skills, and is nominally described in 'teamwork and communication'. The wording of the ANTSAP system was generally accepted as clear.

Only a third (n = 16, 33%) of participants completed the online human factors material, so the participants did not have the same minimal knowledge of human factors. Some of the hospitals already had some education about human factors, but the concepts were novel to others. Only 28 (58%) of the participants said they were trained in assessment, but the question failed to confirm that this referred to the use of behavioural rating systems. The assessment of staff with behavioural rating systems appears deceptively simple, but their use takes both time and calibration to achieve reliable assessments [22]. That we got a positive evaluation of the ANTS-AP system in such circumstances is encouraging.

The participants had to have at least a year's experience. This might not be long enough to have gained expertise as an anaesthetic practitioner. However, the training to become a nurse or operating department practitioner is usually three years, and so they would not have been completely new to healthcare and should have seen good and poor non-technical skills while training. In a previous interview study, anaesthetic trainees (even with only 1–2 years experience) were able to describe good and poor non-technical skills demonstrated by anaesthetic practitioners who had assisted them (J.S. Rutherford, R. Flin, L. Mitchell, unpublished observations). We did not have a means of assessing expertise in the anaesthetic practitioners, and hence used an arbitrary criterion of at least a year's experience in their role as an anaesthetic practitioner.

The questionnaire we used to gather the participants' opinions on the ANTS-AP system was distributed at the end of the workshop, and while this was practical, asking the anaesthetic practitioners for their views immediately after completing the scoring may not have been long enough for considered reflection.

The scenarios in the 12 video clips were written to illustrate specific problems with 'situation awareness', 'teamwork and communication' or 'task management' with good, average or poor behaviours portrayed. To be able to show a practice clip and 12 rating videos in the limited time available, we omitted the filming of the anaesthetic machine check at the start of each video, even though this is a key task for the anaesthetic practitioner. The absence of seeing the anaesthetic machine check was adversely commented upon in the feedback after videos, despite the instructions to score only what was shown, and to assume the check had been done unless otherwise evident. The video clips had the simulated patient's vital signs displayed as a box within a corner of the screen, and the size of this box was a compromise between a risk of obscuring action in the video, and the vital signs' being too small to read comfortably.

The participants were asked if it would have been easier watching the scenario in a real operating theatre or on the video. The staff who thought it was easier in theatre cited the ability to gain the context of what was going on, whereas the participants who preferred the video said the angle of view was better than they could get in theatre, and also that they would have felt the need to intervene when poor behaviour was demonstrated. The limited training to use the system might have led to a learning effect's being evident in the scoring. However, the ratings of the main workshop did not appear to alter in a systematic fashion, although a learning effect cannot be excluded in the test–retest workshop. In conclusion, we have demonstrated that the ANTS-AP system has acceptable internal consistency, and with the omission of 'coping with pressure', is of sufficient reliability, validity and usability to warrant further investigation. The ANTS-AP system is available on www.abdn.ac.uk/iprc/ANTS-AP.

Acknowledgements

The authors acknowledge the help of Scottish Clinical Simulation centre and colleagues who helped in the filming of the scenarios used in this project, the North West Simulation Education Network for use of their Human Factors online teaching material, and Mark Hellaby and David Caldwell for their advice and help with expert ratings.

Competing interests

No external funding and no competing interests declared.

References

1. Commission on Anaesthetics. Report of the Lancet Commission appointed to investigate the subject of the administration of chloroform and other anaesthetics from a clinical standpoint. Lancet 1893; 141: 693–708.
2. Buck N, Devlin HB, Lunn JN, Vickers MD. The Report of a Confidential Enquiry into Perioperative Deaths. London: King's Fund Publishing Office, 1987.
3. Fenwick D. The history of anaesthetic mortality reporting. Anaesthesia and Intensive Care 2007; 35: 21–5.
4. Cooper JB, Newbower RS, Long CD, McPeek B. Preventable anesthesia mishaps: a study of human factors. Anesthesiology 1978; 49: 399–406.
5. Rutherford JS, Flin R, Mitchell L. Teamwork, communication, and anaesthetic assistance in Scotland. British Journal of Anaesthesia 2012; 109: 21–6.
6. Fletcher G, Flin R, McGeorge P, Glavin R, Maran N, Patey R. Anaesthetists' Non-Technical Skills (ANTS): evaluation of a behavioural marker system. British Journal of Anaesthesia 2003; 90: 580–8.
7. Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behaviour rating system. World Journal of Surgery 2008; 32: 548–56.
8. Mitchell L, Flin R, Yule S, Mitchell J, Coutts K, Youngson G. Development of a behavioural marker system for scrub practitioners' non-technical skills (SPLINTS system). Journal of Evaluation in Clinical Practice 2013; 19: 317–23.

9. Lyk-Jensen HT, Jepsen RMHG, Spanager L, Dieckmann P, Østergaard D. Assessing Nurse Anaesthetists' Non-Technical Skills in the operating room. Acta Anaesthesiologica Scandinavica 2014; 58: 794–801.
10. Rutherford JS, Flin R, Mitchell L. Non-technical skills of anaesthetic assistants in the perioperative period: a literature review. British Journal of Anaesthesia 2012; 109: 27–31.
11. Miles MB, Huberman AM, Saldana JM. Qualitative Data Analysis: A Methods Sourcebook, 3rd edn. Thousand Oaks, CA: Sage Publications Inc., 2013.
12. Braun V, Clarke V. Using thematic analysis in psychology. Qualitative Research in Psychology 2006; 3: 77–101.
13. Rutherford JS, Flin R, Irwin A. The non-technical skills used by Anaesthetic Technicians in critical incidents reported to the Australian Incident Monitoring System between 2002 and 2008. Anaesthesia and Intensive Care 2015; 43: 512–7.
14. Krueger RA, Casey MA. Focus Groups: A Practical Guide for Applied Research, 3rd edn. Thousand Oakes, CA: Sage Publications Inc., 2000.
15. Hasson F, Keeney S, McKenna H. Research guidelines for the Delphi survey technique. Journal of Advanced Nursing 2000; 32: 1008–15.
16. Flin R, O'Connor P, Crichton M. Safety at the Sharp End: A Guide to Non-Technical Skills. Aldershot: Ashgate, 2008.
17. DeVellis RF. Scale Development: Theory and Applications, 1st edn. Newbury Park: Sage, 1991.
18. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977; 33: 159–74.
19. Evers A. The revised Dutch rating system for test quality. International Journal of Testing 2001; 1: 155–82.
20. Klampfer B, Flin R, Helmreich RL, et al. Group Interaction in High Risk Environments: Enhancing Performance in High Risk Environments, Recommendations for the Use of Behavioural Markers. Berlin: GIHRE, 2001.
21. Graham J, Hocking G, Giles E. Anaesthesia non-technical skills: can anaesthetists be trained to reliably use this behavioural marker system in 1 day? British Journal of Anaesthesia 2010; 104: 440–5.
22. Hull L, Arora S, Symons NR, et al. Training faculty in nontechnical skill assessment: national guidelines on program requirements. Annals of Surgery 2013; 258: 370–5.
23. Weber DE, Mavin TJ, Roth W-M, Henriqson E, Dekker SWA. Exploring the use of categories in the assessment of airline pilots' performance as a potential source of examiners' disagreement. Journal of Cognitive Engineering and Decision Making 2014; 8: 248–64.