

The Oxford NOTECHS System: reliability and validity of a tool for measuring teamwork behaviour in the operating theatre

A Mishra, K Catchpole, P McCulloch

Quality, Reliability, Safety and Teamwork Unit, Nuffield Department of Surgery, University of Oxford, Oxford, UK

Correspondence to: Mr A Mishra, Nuffield Department of Surgery, The John Radcliffe, Headington, Oxford OX3 9DU, UK; mishra007@hotmail.com

Accepted 28 March 2008

ABSTRACT

Introduction: The frequency of adverse events in the operating theatre has been linked to the quality of teamwork and communication. Developing suitable measures of teamwork may play a role in reducing errors in surgery. This study reports on the development and evaluation of a method for measuring operating-theatre teamwork quality.

Methods: The Oxford Non-Technical Skills (NOTECHS) scale was developed from an aviation instrument for assessment of non-technical skills. Consultation with experts and task analysis led to modifications reflecting the complexities of the theatre teamwork, particularly the coexistence of three subteams (surgeons, anaesthetists and nurses). The scale was then evaluated using teams performing laparoscopic cholecystectomy ($n = 65$) before and after teamwork training. Attitudes to teamwork and surgical error rates were assessed by questionnaire and direct observation methods, and used to assess the reliability and validity of the Oxford NOTECHS scale.

Results: The interobserver reliability was excellent in 24 operations independently assessed by two observers ($R_{wg} = 0.99$), confirmed by a third observer in 11 cases ($R_{wg} = 0.99$). Validity was demonstrated through improved scores after teamwork training ($t = -3.019$, $p = 0.005$), concurrent with improved attitudes to teamwork after training; inverse correlation between NOTECHS scores and surgical errors ($\rho = -0.267$, $p = 0.046$); strong inverse correlation between surgical subteam score and surgical errors ($\rho = -0.412$, $n = 65$, $p = 0.001$); and strong correlation with teamwork scores from an alternative system ($n = 5$, $r = 0.886$, $p = 0.046$).

Conclusion: The Oxford NOTECHS scale appears to be a reliable and valid instrument for assessing teamwork in the operating theatre, and is ready for further application.

Within hospitals, the operating theatre is reportedly the most common site for adverse events to occur,¹ probably because it represents a complex environment where technology, competence and resources require coordination under time pressure. This combination of factors has previously been identified in teams working in other complex, high-risk environments,² and analogies between healthcare and other industries have been frequently and plausibly made,³⁻⁵ supported by observations of the relationship in the operating theatre between potential adverse events and deficiencies in teamwork behaviour and coordination.⁶⁻¹² Thus, an ability to measure teamwork and communication performance is essential if we wish to investigate the role of these “non-technical” skills in influencing the quality and safety of healthcare.

Several methods have been developed for measuring, training or diagnosing teamwork and cognitive skills in the operating theatre.¹³⁻¹⁵ Our own research in this area has built upon extensive work in aviation by developing a scale to evaluate the behaviour of the operating team in relation to other intraoperative events, processes and outcomes.¹⁶⁻¹⁸ While taking advantage of a field where 30 years of research have been spent on understanding teamwork and communication, it is important to remember that concepts cannot be simply transplanted from aviation to surgery but have to be translated. In this paper, we describe the development of the Oxford Non-Technical Skills (NOTECHS) system for evaluating operating teams, seek to confirm that acceptable levels of reliability can be achieved and formally examine the validity of this scale, with the wider aim of establishing a clear evidential link between teamwork training programmes and improvements in surgical care.

METHOD

Development of Oxford NOTECHS system

The NOTECHS evaluation system used in aviation was developed in response to requirements for the training and assessment of teamwork and cognitive skills in the civil airline cockpit.¹⁹ It was structured along four behavioural dimensions: leadership and management; teamwork and cooperation; problem-solving and decision-making; and situation awareness. The extensive expertise and evaluation investments that validated this scale,¹⁹ combined with a recognised need for validated, performance-related behavioural markers in surgery,⁶ led to the adaptation of NOTECHS for operating theatre teams. Following usual practice,²⁰ a task analysis defined the domain in which the scale would be used, and consultation took place with content experts (two cardiac surgeons, one vascular surgeon, one orthopaedic surgeon, two anaesthetists, one human-factors expert and two aviation-crew resource-management trainers) to confirm the scoring system and translate skills from aviation to the operating-theatre context. The resultant NOTECHS scale for use in surgery (table 1), was found to be useful in early studies in paediatric cardiac surgery and orthopaedic surgery.^{16 17} In order to further examine the contribution of nursing, anaesthetic and surgical subteams to the functioning of the team, a refinement was then made which provided this extra layer of definition (table 2). This range of adapted markers

Table 1 Operating-theatre team Non-Technical Skills (NOTECHS) assessment tool

Leadership and management			
Leadership		Involves/reflects on suggestions/visible/accessible/inspires/motivates/coaches	
Maintenance of standards		Subscribes to standards/monitors compliance to standards/intervenes if deviation/deviates with team approval/demonstrates desire to achieve high standards	
Planning and preparation		Team participation in planning/plan is shared/understanding confirmed/projects/changes in consultation	
Workload management		Distributes tasks/monitors/reviews/tasks are prioritised/allots adequate time/responds to stress	
Authority and assertiveness		Advocates position/values team input/takes control/persistent/appropriate assertiveness	
Teamwork and cooperation			
Team building/maintaining		Relaxed/supportive/open/inclusive/polite/friendly/use of humour/does not compete	
Support of others		Helps others/offers assistance/gives feedback	
Understanding team needs		Listens to others/recognises ability of team/condition of others considered/gives personal feedback	
Conflict solving		Keeps calm in conflicts/suggests conflict solutions/concentrates on what is right	
Problem-solving and decision-making			
Definition and diagnosis		Uses all resources/analytical decision-making/reviews factors with team	
Option generation		Suggests alternative options/asks for options/reviews outcomes/confirms options	
Risk assessment		Estimates risks/considers risk in terms of team capabilities/estimates patient outcome	
Outcome review		Reviews outcomes/reviews new options/objective, constructive and timely reviews/makes time for review/seek feedback from others/conducts post-treatment review	
Situation awareness			
Notice		Considers all team elements/asks for or shares information/aware of available of resources/encourages vigilance/checks and reports changes in team/requests reports/updates	
Understand		Knows capabilities/cross-checks above/shares mental models/speaks up when unsure/updates other team members/discusses team constraints	
Think ahead		Identifies future problems/discusses contingencies/anticipates requirements	
Below standard = 1	Basic standard = 2	Standard = 3	Excellent = 4
Behaviour directly compromises patient safety and effective teamwork	Behaviour in other conditions could directly compromise patient safety and effective teamwork	Behaviour maintains an effective level of patient safety and teamwork	Behaviour enhances patient safety and teamwork; a model for all other teams

was used in conjunction with the existing skills sets to aid the observer to score accurately when studying these subteams. This produced the Oxford NOTECHS system for the evaluation of theatre teams.

A single observer examines behaviour on four dimensions: generic skills are defined in table 1, and subteam-specific modifiers described in table 2. Each subteam is scored on a scale of 1–4 for each dimension, with scores anchored to categories (below standard; basic standard; standard; excellent). These scores can be used individually (score of 1–4) and can be summed to provide a total score for each subteam (a score of 4–16) or for the whole team on each dimension (a score of 3–12), or used to form the total team score (score of 1–4×4 dimensions ×3 subteams=12–48). Since the scale had been developed from aviation principles in conjunction with surgeons, anaesthetists and nurses, was already demonstrably useful in assessing surgical teamwork and was similar in content to other instruments designed for the same type of task^{13–15}, face and content validity could be assumed. Evaluation therefore focused on reliability and other forms of validity.

Evaluation of the Oxford NOTECHS system

The properties of this system were examined as part of a larger study evaluating the effect of an aviation-style safety training intervention for operating teams performing laparoscopic cholecystectomy. The training intervention was devised and delivered in conjunction with two civil pilots with non-technical skills training experience. The course consisted of 9.5 h of lectures and interactive exercises, including threat and error management, personality, communication styles, conflict resolution and situation awareness. This was followed by 12 sessions of in-theatre coaching in preoperative briefings over

3 months. Laparoscopic Cholecystectomy was chosen, as it is performed frequently, requires both advanced technology and considerable teamwork, is moderately complex, and has recognisable complications which can be monitored. Operations were observed at the Oxford Radcliffe Hospitals Trust after obtaining Ethics Committee approval (LREC ref. no. 04/Q1603/35). Two observers were trained in the use of the Oxford NOTECHS scale. The principal observer (AM) was a surgical trainee who received training in assessing non-technical skills. The second observer (KC) was a human-factors practitioner with prior experience of observing theatre teams, who received additional technical and anatomical training. Observers kept free-form contemporary notes to provide contextual information when assigning NOTECHS scores, and three training operations were conducted to ensure baseline consistency in observations. After the teamwork training intervention had been completed, a third observer with prior experience in evaluating aviation non-technical skills used the Oxford NOTECHS scale in parallel with the other observers after initial instruction and three baseline operations. Theatre staff were aware that their teamwork and communication were being observed, and they quickly became used to the presence of the observers. Consent was obtained from them and the patients prior to commencement of observations.

Parallel independent scoring of operations with the two observers allowed assessment of inter-rater reliability, analysed with R_{wg} for overall NOTECHS scores and for each dimension in each subteam. The test–retest reliability could not be assessed directly, so observations before and after the training intervention were each divided into three temporally consecutive groups. One-way ANOVAs were used to test for significant variation in total team performance across these groups.

Table 2 Oxford Non-Technical Skills (NOTECHS) subteam modifiers

	Surgical subteam	Anaesthetic subteam	Nursing subteam
Leadership and management			
Positive modifiers	⇒Raises team morale	⇒Takes control when required	⇒Scrub provides clear instructions to circulating nurse(s)
	⇒Intervenes if deviation	⇒Demonstrates desire for high standard	⇒Senior nurse makes sure protocols are followed
	⇒Prioritises tasks	⇒Appropriately distributes tasks b/w rest of team	⇒Speaks up when unhappy
Negative modifiers	⇒Deflates or fails to motivate team	⇒Does not take control when required	⇒Senior nurse does not support juniors
	⇒Does not attempt to build cohesion	⇒Does not set standards	
		⇒Inappropriate task distribution	
Teamwork and cooperation			
Positive modifiers	⇒Open	⇒Supportive of other subteams	⇒Nurses cooperate and support each other well
	⇒Appropriate use of abilities within team	⇒Appreciates functions of other subteams	⇒Senior nurse covers for junior scrub
	⇒Supportive of other subteams when necessary		
Negative modifiers	⇒Aggressive in conflicts	⇒Remains idle when problems arise	⇒Poor coordination between equipment needs and those provided
	⇒Does not appreciate others' abilities	⇒Functions separately from other subteams	
Problem-solving and decision-making			
Positive modifiers	⇒Demonstrates generation of options	⇒Participates in solving problems	⇒Takes an active part in decision-making
	⇒Open discussion and agreement over anatomy	⇒Raises suggestions	⇒Suggests solutions to problems—eg, alternative equipment
	⇒Incorporates other subteam issues		
Negative modifiers	⇒Decisions made unsystematically	⇒Does not consider anaesthetic options when met with problem	⇒Blames the surgeons when faced with problems
	⇒Does not utilise team where it may benefit		
Situation awareness			
Specific to subteams			
Positive modifiers	⇒Periodically gathers awareness of surroundings	⇒Anticipates surgical and process needs	⇒Anticipates equipment needs
	Negative modifiers	⇒Is fixated on operative field	⇒Absent at stages when needed to provide service
For all subteams			
Positive modifiers	⇒Patient: Has awareness of patient condition/comorbidity		
	⇒Procedure: Appreciates stage of operation		
	⇒People: Who is present in theatre, what skills they have and what they are doing		

Validity was examined by comparing expectation with observed scale performance on a number of dimensions. More errors would be expected in teams with lower Oxford NOTECHS scores, with a stronger relationship between surgical errors and surgical subteam NOTECHS scores. Surgical errors were measured concurrently by the primary observer (AM) using the observational clinical human reliability analysis (OCHRA) technique.^{21, 22} A second assessment of validity took specific advantage of the training programme by examining the differences in Oxford NOTECHS scores before and after training. The Oxford NOTECHS system would be expected to measure any difference and was triangulated with results from the Safety Attitudes Questionnaire (SAQ),²³ which was also applied before and after the training programme, and would also be expected to change. Finally, we used the Oxford NOTECHS system in parallel with the Observational Teamwork Assessment for Surgery (OTAS)¹³ to examine the convergence of the two scales. For this component of the study, the primary observer recorded OTAS scores, which were examined with NOTECHS evaluations recorded by the second observer.

RESULTS

In total, 65 operations were observed, 26 before and 39 after the training intervention. Twenty-four cases were co-observed and independently scored by observers one and two, and 11 cases were also observed by observer 3. Agreement between observers 1 and 2 was generally excellent (table 3), with the exception of anaesthetic situation awareness.

The agreement between the third observer and the other two observers was also excellent ($R_{wg} = 0.99$). The test-retest reliability was found to be acceptable, with no differences in the mean NOTECHS scores during the three preintervention periods (ANOVA $F(2,1) = 1.341$, $p = 0.281$) or in the three postintervention periods (ANOVA $F(2,1) = 1.028$, $p = 0.368$). Thus, scoring of the scale appeared to be reliable across most dimensions and several observers, and over time.

The correlation between technical error and Oxford NOTECHS team score was negative and significant, though weak ($\rho = -0.267$, $n = 65$, $p = 0.045$). As expected, there was a stronger negative correlation between technical errors and the surgical subteam NOTECHS score ($\rho = -0.412$, $n = 65$, $p = 0.001$). Furthermore, the system was able to measure the effects of the training course, with a significant improvement ($t = -3.019$, $p = 0.005$) in scores after the team training programme (38.7 95% CI ± 0.9) compared with before (35.5 95% CI ± 1.9). This was in concurrent agreement with the SAQ score for teamwork climate, with a mean of 63.8 (95% CI ± 7.1) before training and 67.4 (95% CI ± 6.8) afterwards, though the difference was non-significant ($t = -1.81$, $p = 0.089$). Finally, the overall agreement between OTAS and NOTECHS was excellent ($r = 0.886$, $n = 5$, $p = 0.046$). The mean OTAS score for the five cases compared was 18.8 (range 14–22 out of a possible maximum of 30), and the mean Oxford NOTECHS score was 37.8 (range 33–45, out of a possible maximum of 48), suggesting that data on both scales covered a similar range in relation to the overall scale maxima and minima. The findings are summarised in table 4.

Table 3 Means (SD) for Non-Technical Skills (NOTECHS) scores obtained from 24 cases observed independently observers 1 and 2

	Leadership and management	Teamwork and cooperation	Problem-solving and decision-making	Situation awareness	Total
Surgeons					
Obs 1	3.54 (0.59)	3.42 (0.5)	3.58 (0.58)	3.25 (0.61)	13.8 (1.69)
Obs 2	3.54 (0.51)	3.38 (0.65)	3.54 (0.59)	2.92 (0.58)	13.4 (1.77)
R _{wg}	0.90	0.88	0.95	0.86	0.96
Anaesthetists					
Obs 1	2.92 (0.28)	3 (0.42)	2.92 (0.38)	2.88 (0.45)	11.7 (0.91)
Obs 2	2.92 (0.28)	3.04 (0.36)	2.96 (0.2)	2.96 (0.46)	11.9 (0.8)
R _{wg}	0.97	0.88	0.98	0.68	0.97
Nurses					
Obs 1	3.17 (0.76)	3.17 (0.64)	3.04 (0.75)	2.92 (0.88)	12.3 (2.46)
Obs 2	3.17 (0.7)	3.04 (0.69)	2.96 (0.36)	2.92 (0.78)	12.1 (1.84)
R _{wg}	0.87	0.88	0.83	0.93	0.95
Team					
Obs 1	9.63 (1.28)	9.58 (1.02)	9.54 (1.25)	9.04 (1.3)	37.8 (3.4)
Obs 2	9.63 (0.92)	9.46 (1.28)	9.46 (0.88)	8.79 (1.18)	37.3 (3.38)
R _{wg}	0.95	0.93	0.96	0.97	0.99

Inter-rater agreement per category is represented by R_{wg} results.

DISCUSSION

We have described the development of a scale for observing teamwork behaviour in an operating theatre, from its origin in aviation to its application to multidisciplinary aspects of theatre work. The Oxford NOTECHS system demonstrated excellent interobserver reliability, and has closely followed expectation in the aspects of validity examined here. It detected improvements in non-technical skills after specific training and was congruent with improvements in attitudes to teamwork, and scores were related in the expected manner to other measures of non-technical skills and technical performance skills. Reliability data from comparison with the third observer appear to refute the possibility of positive bias by observers 1 and 2 as a result of their involvement with development of both the scale and the training intervention.

The scale has certain advantages: it requires only one observer and can be used to evaluate the whole theatre team, and the performance of subteams separately. The scale can also be used by an observer from a variety of backgrounds, with a small provision for training.²⁴ Furthermore, as it captures non-technical skills independently of other operative events and can be used in several operative types,¹⁶⁻¹⁸ we believe the scale to be generally applicable across a wide range of operations. However, this study is not without limitations. The scale requires trained individuals with prior experience either in the

operating theatre or in non-technical skills. The concurrent use of NOTECHS with OCHRA by the same observer may have led to a greater agreement between scales than might otherwise be expected. Finally, even though attitudinal change is a prerequisite for behavioural change,²⁵ the concurrence between NOTECHS and SAQ results may not reflect a direct correlation. However, demonstrating a change in attitude at least provides the possibility for the change in behaviour measured by the scale. Thus, further work should focus on the development of the scale for ease of use, and on further independent concurrent validation.

Improvements to the scale may improve its value as a research tool. The reliability of the anaesthetic scores was disappointing, especially on the teamwork and cooperation dimension. This resulted from the limited involvement of the anaesthetist in this type of operation, leading to the minimal variation in these scores. Subsequent studies with carotid endarterectomy showed reliability to be good once sufficient variation was encountered. The tool also lacks scalability, due to the current limited understanding of teamwork skills in the OR and, by scoring all subteams equally, may not accurately reflect the contribution each team makes to the overall success of the operation. As our understanding of these complex relationships develops, it may be possible to enhance the system further. This study suggests that the Oxford NOTECHS scale is ready to help address these complex questions.

Table 4 Reliability and validity of the Oxford Non-Technical Skills (NOTECHS) tool

Reliability/validity	Measurement
Inter-rater reliability	R _{wg} statistics very high for total score, and most dimensions and subteam scores
Test-retest reliability	ANOVA analysis found no differences in the mean NOTECHS scores during the three preintervention periods or three postoperative periods
Face and content validity	Not quantitatively assessed, but qualitatively consistent with OTAS, ¹³ NOTSS ¹⁵ and ANTS ¹⁴
Predictive validity	Theoretical improvement in teamwork behaviour supported by increased NOTECHS performance
Concurrent validity	Correlation between overall NOTECHS and technical error, and between surgical subteam NOTECHS and technical error; increase in SAQ scores post-training concurrent with an increase in NOTECHS scores
Convergent validity	Agreement between OTAS and NOTECHS

ANTS, Anaesthetists' Non-Technical Skills; OTAS, Observational Teamwork Assessment for Surgery; SAQ, Safety Attitudes Questionnaire.

Acknowledgements: The QRST Unit wishes to thank the patients and staff at the Oxford Radcliffe Hospitals Trust for their participation; the Trust management for permission to conduct the study; P Smith for his help in observing; and T Dale and G Hirst, Atrainability, for their valued experience of team training in aviation, and their extensive assistance in developing both evaluation and training methods.

Funding: This study was funded by the BUPA Foundation.

Competing interests: None.

Ethics approval: Ethical approval was obtained from the Milton Keynes local research ethics committee (study no. 04/Q1603/35 amendment 2).

Patient consent: Obtained.

REFERENCES

1. **Leape LL.** Error in medicine. *JAMA* 1994;**272**:1851–7.
2. **Carthey J,** de Leval MR, Reason JT. The human factor in cardiac surgery: errors and near misses in a high technology medical domain. *Ann Thorac Surg* 2001;**72**:300–5.
3. **Sexton JB,** Thomas EJ, Helmreich RL. Error, stress and teamwork in medicine and aviation: cross sectional surveys. *BMJ* 2000;**320**:745–9.
4. **Thomas EJ,** Sexton JB, Helmreich RL. Translating teamwork behaviours from aviation to healthcare: development of behavioural markers for neonatal resuscitation. *Qual Saf Health Care* 2004;**13**(1 Suppl):i57–64S.
5. **Leonard M,** Graham S, Bonacum D. The human factor: the critical importance of effective teamwork and communication in providing safe care. *Qual Saf Health Care* 2004;**13**(1 Suppl):i85–90S.
6. **Carthey J,** de Leval MR, Wright DJ, *et al.* Behavioural markers of surgical excellence. *Saf Sci* 2003;**41**:409–25.
7. **Lingard L,** Espin S, Whyte S, *et al.* Communication failures in the operating room: an observational classification of recurrent types and effects. *Qual Saf Health Care* 2004;**13**:330–4.
8. **Rogers SO Jr,** Gawande AA, Kwaan M, *et al.* Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery* 2006;**140**:25–33.
9. **Helmreich RL.** On error management: lessons from aviation. *BMJ* 2000;**320**:781–5.
10. **Lingard L,** Reznick R, Espin S, *et al.* Team communications in the operating room: talk patterns, sites of tension, and implications for novices. *Acad Med* 2002;**77**:232–7.
11. **Way LW,** Stewart L, Gantert W, *et al.* Causes and prevention of laparoscopic bile duct injuries: analysis of 252 cases from a human factors and cognitive psychology perspective. *Ann Surg* 2003;**237**:460–9.
12. **Gawande AA,** Zinner MJ, Studdert DM, *et al.* Analysis of errors reported by surgeons at three teaching hospitals. *Surgery* 2003;**133**:614–21.
13. **Undre S,** Healey AN, Darzi A, *et al.* Observational assessment of surgical teamwork: a feasibility study. *World J Surg* 2006;**30**:1774–83.
14. **Fletcher GCL,** Flin RH, Glavin RJ, *et al.* Anaesthetists' Non-Technical Skills (ANTS): Evaluation of a behavioural marker system. *Br J Anaesth* 2003;**90**:580–8.
15. **Yule S,** Flin R, Paterson-Brown S, *et al.* Development of a rating system for surgeons' non-technical skills. *Med Educ* 2006;**40**:1098–104.
16. **Catchpole KR,** Giddings AE, Wilkinson M, *et al.* Improving patient safety by identifying latent failures in successful operations. *Surgery* 2007;**142**:102–10.
17. **Catchpole KR,** Giddings AE, Hirst G, *et al.* A method for measuring threats and errors in surgery. *Cogn Technol Work.* In press.
18. **Mishra A,** Catchpole K, Dale T, *et al.* The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surg Endosc.* In press.
19. **Flin RH,** Martin L, Goeters K-M, *et al.* Development of the NOTECHS (Non-Technical Skills) system for assessing pilots' CRM skills. *Hum Perf Extreme Environ* 2003;**3**:95–117.
20. **Stanton NA,** Salmon PM, Walker GH, *et al.* *Human factors methods: a practical guide for engineering and design.* Aldershot: Ashgate, 2005.
21. **Tang B,** Hanna GB, Joice P, *et al.* Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Arch Surg* 2004;**139**:1215–20.
22. **Joice P,** Hanna GB, Cuschieri A. Errors enacted during endoscopic surgery—a human reliability analysis. *Appl Ergon* 1998;**29**:409–14.
23. **Sexton JB,** Helmreich RL, Neilands TB, *et al.* The Safety Attitudes Questionnaire: psychometric properties, benchmarking data, and emerging research. *BMC Health Serv Res* 2006;**6**:44.
24. **Carthey J.** The role of structured observational research in health care. *Qual Saf Health Care* 2003;**12**:13ii–16.
25. **Kirkpatrick DL.** *Evaluating training programs: the four levels.* San Francisco: Berrett-Koehler, 1994.

Let us assist you in teaching the next generation

Figures from all articles on our website can be downloaded as a PowerPoint slide. This feature is ideal for teaching and saves you valuable time. Just click on the image you need and choose the "PowerPoint Slide for Teaching" option. Save the slide to your hard drive and it is ready to go. This innovative function is an important aid to any clinician, and is completely free to subscribers. (Usual copyright conditions apply.)