Assessment of non-technical skills: why aren't we there yet?

Adam P Johnson, Rajesh Aggarwal

Department of Surgery, Thomas Jefferson University Hospital, Philadelphia, Pennsylvania, USA

Correspondence to

Dr Adam P Johnson, Department of Surgery, Thomas Jefferson University Hospital, Philadelphia, PA 19107, USA; adam.johnson@jefferson.edu

Accepted 15 May 2019
Published Online First
25 May 2019

Knowledge and application of non-technical skills (NTS) may represent the greatest challenge facing medical education today. For centuries, medical education focused on developing individual clinical knowledge and technical skills. But, the modern complexities of healthcare delivery and rapid expansion of medical knowledge necessitate a high-functioning team approach, which requires human factors engineering and NTS to operate effectively.

Other complex high-risk industries—like aviation, oil drilling, nuclear power and the military—have aligned their educational systems to match.^{1–3} While certain healthcare disciplines have developed frameworks to ensure the acquisition and maintenance of clinical and technical skills, no standard framework for NTS exists. With the increasing computational support for clinical and technical skills—decision aids, predictive algorithms, robotic surgery and image interpretation—the true added value of human clinicians may lie with their mastery of NTS.

Failure of NTS has been linked to poor quality and safety of care.² A prospective observational study of 28 laparoscopic cholecystectomies found a strong correlation between surgical team situational awareness and fewer technical errors.⁴ In Japan, a 3-year retrospective review of fatal medical accidents submitted to a third-party safety organisation found roughly half to be due to failures of NTS, most often related to situational awareness, teamwork and decision-making.⁵ A review of malpractice claims identified >1 personnel involved in 83% of errors, but only 24% directly attributable to communication breakdown, which was the only NTS specifically studied.⁶ A review of trauma and orthopaedic-related adverse events from the National Reporting and Learning System found many to be

related to NTS—situational awareness (52%), communication/teamwork (21%), leadership (16%) and decision-making (12%).⁷ Prospective direct observation of 293 surgical procedures found a strong association between less effective teamwork behaviours—as measured by the Behavioural Marker Risk Index—and a higher risk of death or serious complication, even after controlling for American Society of Anesthesiologists risk category.⁸ A multi-institutional retrospective review of Veteran's Health Administration found interventions to increase NTS were associated with reductions in perioperative mortality.⁹ Although these studies demonstrate a link of NTS to quality and safety, quantifying the impact and comparing across studies is limited by a lack of standard definitions of NTS and relevant outcomes. Advancing our understanding of NTS requires a more thoughtful and standardised framework.

Boet *et al* review assessment focused on team performance in crisis situations¹⁰ and Higham *et al* review assessments of broader NTS domains across multiple healthcare disciplines.¹¹ These rigorous and thorough reviews of the literature provide a snapshot of the contemporary science underpinning the measurement of NTS in healthcare and provide researchers guidance for selecting the best tools available to measure NTS. However, the greatest lesson to take from these reviews may be the current gaps in the measurement of NTS and how to advance the field.

The first gap is the lack of a standardised definition of NTS and subdomains. The framework used by Boet *et al* for teamwork-related NTS included a total of 14 subdomains, whereas Higham *et al* include four domains for all NTS.^{10 11} Each tool included in these reviews measured a unique subset of overlapping domains and often apply

► <http://dx.doi.org/10.1136/bmjqs-2018-008565>

© Author(s) (or their employer(s)) 2019. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Johnson AP, Aggarwal R. *BMJ Qual Saf* 2019;**28**:606–608.

to narrow contexts.² In addition, each of these tools is based on external observations, which may not capture some domains effectively, such as situational awareness and decision-making. These may be better assessed through direct measurement via mental task load index, eye-gaze patterns to assess attention and written examinations.² Like the blind men and the elephant—each tool only measures a portion of NTS and so may instead interpret the elephant as a wall, spear, snake, tree, fan or rope.¹² Generalising a tool beyond the context in which it was developed, combined with Goodhart's Law where 'when a measure becomes a target, it ceases to be a good measure' leaves policy changes based on any of the existing measurement tools open to potentially damaging unintended consequences.¹³

The studies cited by both reviews also exhibit heterogeneity in the psychometric principles used to assess each tool. The Boet *et al* framework for reliability and validity include nine domains, whereas Higham and colleagues compare two domains of validity and treat reliability and usability as distinct single domains.^{10 11} Again, each of the tools is often assessed by only a subset of these psychometric test domains. Emphasising validity and reliability overlooks feasibility testing, which is critical for effective implementation. Assessing NTS at first may appear straightforward, but requires thorough training to assess properly.¹⁴

Finally, the included studies provide no benchmark for adequate performance in NTS. While NTS have been associated with improved quality and safety, the dose-response relationship remains poorly understood. Providing actionable feedback to clinicians requires comparison to a standard, therefore establishing 'good' or 'poor' performance. Currently, each individual institution or researcher is left to interpret the results on their own. Without standard definitions for NTS, psychometric assessment inclusive of feasibility and benchmarks for determining competence, any system-wide intervention to increase assessment and training in NTS will likely fail to reach critical mass and acceptance.

An important initial next step is to exponentially increase the data collected regarding the current state of NTS and their impact on patient outcomes. Automated real-time data collection of clinician interactions may provide better insight into the association between NTS and decrease the burden of assessment. In aviation, the cockpit black box records all communications between pilot and team, which are then rigorously analysed after adverse events. In healthcare, the medical record and clinician recollection of events are used as a proxy for this black box, which under-represents the nuances of decisions and clinical communications.¹⁵ Some surgeons are routinely capturing video to assess for technical skill.^{16 17} A system to automatically collect intraoperative events,

dubbed the OR Black Box, has been used to assess for intraoperative distractions and adverse events but may also provide insight into NTS.¹⁸ Ongoing fears of loss of prestige and litigation stand in the way of routine recording of clinical interactions.² Legal protection against the subpoena of clinical recordings is necessary.¹⁹

Once this robust data set is collected, we must establish a parsimonious set of domains—prioritised according to impact on patient outcomes, transferable across specialties and clinical domains and benchmarked for adequate performance. We must design assessment tools that prioritise domains of NTS most closely associated with poor outcomes. The emphasis should be on a small set of generalisable domains. Opportunities may then arise to automate data analysis through natural language processing and machine learning algorithms.^{20 21}

The final step will be to link incentives to performance in NTS and an ability to surpass the benchmark. In the UK and the USA, the General Medical Council and Accreditation Council for Graduate Medical Education outline general principles, but no specific standards for certification.^{22 23} There exists a curriculum through the American College of Surgeons geared towards surgical residents.²⁴ Guidebooks are available but lack a broader systematic approach with external accountability for organisations and clinicians.²⁵ Some countries have attempted to more systematically incorporate human factors and ergonomics systematically into healthcare.²⁶ Any system for accountability must provide a blame-free framework to support and remediate poor performers—both trainees and active clinicians—and thus prevent a failure to fail.²⁷ Mature team simulation for NTS has already been initiated by malpractice insurers.²⁸ We can look to Crew Resource Management (CRM) framework for licensing and competence assurance from civil aviation, nuclear power, offshore oil drilling, mining, rail and emergency services.^{29 30} However, we must not forget the importance to align systems design with any training paradigm to foster the application of CRM in the workplace. Training alone without system redesign with human factors in mind will be insufficient to ensure appropriate quality and safety.^{3 31}

The authors of both of these systematic reviews^{10 11} should be commended for their rigorous work to aggregate and compare a wide range of disparate instruments. They provide guidance on how to navigate the current literature on assessing NTS in healthcare. However, the strongest take away from their work may be recognising the vast amount of work yet to do to quantify the impact of NTS in healthcare and standardise assessment. We need more robust data, a parsimonious set of NTS and a set of benchmarks and incentives to guide adoption among clinicians.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests RA is a consultant for Applied Medical

Patient consent for publication Not required.

Provenance and peer review Commissioned; internally peer reviewed.

REFERENCES

- 1 Flin R, O'Conner P. *Safety at the sharp end: a guide to non-technical skills*. Burlington, VT: CRC Press, 2008.
- 2 Flin R, Youngson G, Yule S. *Enhancing surgical performance: a primer in non-technical skill*. Boca Raton, FL: CRC Press, 2016.
- 3 Catchpole K. Spreading human factors expertise in healthcare: untangling the knots in people and systems. *BMJ Qual Saf* 2013;22:793–7.
- 4 Mishra A, Catchpole K, Dale T, *et al*. The influence of non-technical performance on technical outcome in laparoscopic cholecystectomy. *Surg Endosc* 2008;22:68–73.
- 5 Uramatsu M, Fujisawa Y, Mizuno S, *et al*. Do failures in non-technical skills contribute to fatal medical accidents in Japan? A review of the 2010–2013 national accident reports. *BMJ Open* 2017;7:e013678–7.
- 6 Rogers SO, Gawande AA, Kwaan M, *et al*. Analysis of surgical errors in closed malpractice claims at 4 liability insurers. *Surgery* 2006;140:25–33.
- 7 Panesar SS, Carson-Stevens A, Mann BS, *et al*. Mortality as an indicator of patient safety in orthopaedics: lessons from qualitative analysis of a database of medical errors. *BMC Musculoskelet Disord* 2012;13.
- 8 Mazzocco K, Petitti DB, Fong KT, *et al*. Surgical team behaviors and patient outcomes. *Am J Surg* 2009;197:678–85.
- 9 Neily J, Mills PD, Young-Xu Y, *et al*. Association between implementation of a medical team training program and surgical mortality. *JAMA* 2010;304:1693–700.
- 10 Boet S, Etherington N, Larrigan S, *et al*. Measuring the teamwork performance of teams in crisis situations: a systematic review of assessment tools and their measurement properties. *BMJ Qual Saf* 2019;28:327–37.
- 11 Higham H. Observer-Based tools for the assessment of non-technical skills in simulated or real clinical environments in healthcare.
- 12 Saxe JG. *The blind men and the elephant*. McGraw-Hill, 1963.
- 13 Strathern M. 'Improving ratings': audit in the British University system. *Eur.rev.* 1997;5:305–21.
- 14 Hull L, Arora S, Symons NRA, *et al*. Training faculty in nontechnical skill assessment: national guidelines on program requirements. *Ann Surg* 2013;258:370–5.
- 15 Wong BM, Dyal S, Etchells EE, *et al*. Application of a trigger tool in near real time to inform quality improvement activities: a prospective study in a general medicine ward. *BMJ Qual Saf* 2015;24:272–81.
- 16 Birkmeyer JD, Finks JF, O'Reilly A, *et al*. Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 2013;369:1434–42.
- 17 Bonrath EM, Gordon LE, Grantcharov TP. Characterising 'near miss' events in complex laparoscopic surgery through video analysis. *BMJ Qual Saf* 2015;24:516–21.
- 18 Jung JJ, Jüni P, Lebovic G, *et al*. First-year analysis of the operating room black box study. *Ann Surg* 2018.
- 19 Lloyd A, Dewar A, Edgar S, *et al*. How to implement live video recording in the clinical environment: a practical guide for clinical services. *Int J Clin Pract* 2017;71:e12951–7.
- 20 Deo RC. Machine learning in medicine. *Circulation* 2015;132:1920–30.
- 21 Hart Y, Czerniak E, Karnieli-Miller O, *et al*. Automated video analysis of Non-verbal communication in a medical setting. *Front Psychol* 2016;7.
- 22 General Medical Council. Generic professional capabilities framework guidance on implementation for colleges and faculties 2017.
- 23 Accreditation Council for Graduate Medical Education, American Board of Surgery. The general surgery milestone project 2015.
- 24 American College of Surgeons. ACS/APDS Surgery Resident Skills Curriculum – Phase 3, 2018. Available: <https://learning.facs.org/content/acsapds-surgery-resident-skills-curriculum-phase-3>
- 25 Carthey J, Clarke J. Implementing human factors in healthcare, 2010. Available: <http://www.patientsafetyfirst.nhs.uk/>
- 26 National Quality Board. Human factors in healthcare a Concordat from the National quality board. *Natl Qual Board* 2013;1–22.
- 27 Dudek NL, Marks MB, Regehr G. Failure to fail: the perspectives of clinical supervisors. *Acad Med* 2005;80(10 Suppl):S84–S87.
- 28 Arriaga AF, Gawande AA, Raemer DB, *et al*. Pilot testing of a model for insurer-driven, large-scale multicenter simulation training for operating room teams. *Ann Surg* 2014;259:403–10.
- 29 Flin R. Safe in their hands? Non-technical skills and competence assessment, 2015. Available: https://ncsbn.org/2015_AM_RFlin.pdf
- 30 Okray R, Lubnau T. Crew resource management for the fire service, 2004. Available: <https://www.fireengineering.com/articles/print/volume-154/issue-8/features/crew-resource-management-for-the-fire-service.html>
- 31 Russ AL, Fairbanks RJ, Karsh B-T, *et al*. The science of human factors: separating fact from fiction. *BMJ Qual Saf* 2013;22:802–8.