

RESEARCH

Open Access



Evaluation of video review tools for assessing non-technical skills in emergency department resuscitation teams: a systematic review

Emily G Alexander¹, Fraser Denny^{1,2}, Malcom WG Gordon^{1,2}, Cieran McKiernan^{1,2} and David J Lowe^{1,2,3*}

Abstract

Background and importance Use of video review in medicine is established in contexts such as surgery. Although not widely used in the emergency department (ED), some centres use it to evaluate non-technical skills (NTS) to support teaching and quality improvement.

Objective There is no consensus on assessment of NTS using video review in the ED and the purpose of this review was to identify tools used in this context.

Design, setting and participants Studies were identified using Embase, Medline, CINAHL and Google Scholar. Inclusion criterion for the review was NTS of resuscitation teams working within the ED were assessed using video review. A systematic search method was used, and results were synthesised after search criteria was checked by two independent reviewers. Authors settled on the same 9 studies eligible for inclusion.

Outcome measures and analysis Reliability and validity of tools identified for use in this context. Due to the heterogeneity of studies, no meta-analysis occurred.

Main results There are 9 studies included in the review. The review was registered with PROSPERO (Ref No: CRD42022306129). Four unique tools were identified – 6 studies used T-NOTECHS, 1 used TTCA-24, 1 used CALM and 1 used the Communication tool. T-NOTECHS is validated in the literature for use in this context.

Conclusion T-NOTECHS is the tool of choice for assessing ED teams in this context.

Keywords Non-technical skills, Emergency department, Resuscitation team, Trauma, Video review, T-NOTECHS

*Correspondence:

David J Lowe

David.Lowe@ggc.scot.nhs.uk

¹University of Glasgow, Glasgow, UK

²Queen Elizabeth University Hospital, Glasgow, UK

³Institute of Health and Wellbeing, University of Glasgow, Glasgow, UK



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Introduction

Providing high quality resuscitation to patients presenting in the emergency department requires a coordinated performance of interventions to achieve resuscitation success and patient survival; [1] this requires non-technical skills (NTS). [2] NTS include skills such as leadership, communication, situational awareness, decision making and teamwork. [3] Leadership skills are correlated with increased quality of CPR and the International Liaison Committee on Resuscitation recommends that “specific teamwork training” should be taught on courses. [4] The importance of evaluating NTS within teams is increasing, as are the number of tools used to assess them. [1] Early examples are adapted from the aviation industry, where measuring NTS was already commonplace. [5, 6].

Assessing NTS of a resuscitation team in real time is challenging due to the emergency department (ED) environment. [7] One study showed that traditional review only detected 20% of errors that were seen in video review [8], highlighting the opportunity to enable forensic review of team performance. Clinical work must be examined in its natural setting to allow for inclusion of the nuances of real-life not accounted for in simulation. Introducing video review into the ED allows for critical review to gain insight from others. [9].

Video review in medicine is established in many contexts, including simulations and surgery. [10] Although not widely used in the ED yet, some hospitals use it to assess NTS to support teaching and quality improvement. [11] There is currently no consensus on assessment of NTS using video review in the ED. [12].

Aims

- To provide an overview of tools used to assess NTS in resuscitation teams within the ED using video review.
- To explore to evidence for the validity and usability of the tools.

Methods

This review is registered with PROSPERO (Ref No: CRD42022306129). Peer-reviewed studies were identified using electronic databases Medline, Embase and CINAHL. A grey literature search was completed using Google Scholar. A manual search of the reference list of relevant articles was conducted. The PRISMA diagram for review of NTS assessment tools is shown in Fig. 1. [13] The search strategy is further detailed in the [supplementary material](#).

The inclusion and exclusion criteria were informed by the authors’ experiences and familiarity with existing literature. We sought papers available in English and published between January 1995 and September 2023, which studied resuscitation teams within the ED. Terms

including other descriptors were included (e.g., trauma teams, resus team). Papers must also describe a tool used to assess at least one component of NTS where video review was utilised in a real clinical setting. Papers exclusively assessing simulation were excluded. Papers exclusively assessing resuscitation of paediatric patients were excluded due to the challenges presented by paediatric patients lying outside the scope of this paper.

The search criteria were checked by two independent reviewers. Papers for potential inclusion were checked for relevance by title and abstract (see Fig. 1 for PRISMA flowchart process). Relevant papers were retained for full review. Two papers did not have a full paper associated with their abstract, and one was not available in English. In the initial search, no papers required adjudication by the senior author as agreement between reviewers was achieved.

In the final analysis, three structured tools were found (T-NOTECHS, CALM and TTCA-24) and one tool assessing communication (Communication tool). Studies were analysed over three main domains: method of development, applicability and context use of tool, and evidence of validity. Data were collected and synthesised by one author and checked by another.

Risk of bias was considered throughout data analysis and interpretation. Potential bias includes study selection bias, language bias and anchoring bias. One author published a review in situational awareness, [14] a key component of NTS, which may lead to familiarity bias. Mitigations for these risks include review by three authors, use of a systematic search method, repeated re-examinations of papers in a random order, attempting to access pre-published papers from authors and an English translation. The latter were unsuccessful as shown in Fig. 1.

All reviewed articles were quality assessed using the Mixed Methods Appraisal Tool (MMAT) Version 2018 [15] by two authors. Discrepancies were discussed until agreement was reached. MMAT is a “critical appraisal tool designed for appraisal... of systematic mixed studies reviews”. Its validity and reliability meet accepted standards and it was pilot tested for reliability in systematic reviews. [16, 17].

Lack of homogeneity in design, definition, and study populations precluded the use of meta-analytic techniques. Findings were tabulated and summarised by detailed narrative analysis in accordance with the PRISMA checklist. [18].

Results

The screening process is shown in Fig. 1 as per PRISMA guidance. There were 378 discrete studies screened, 339 were eliminated based on title relevance and 28 were eliminated on abstract relevance. A total of 12 studies

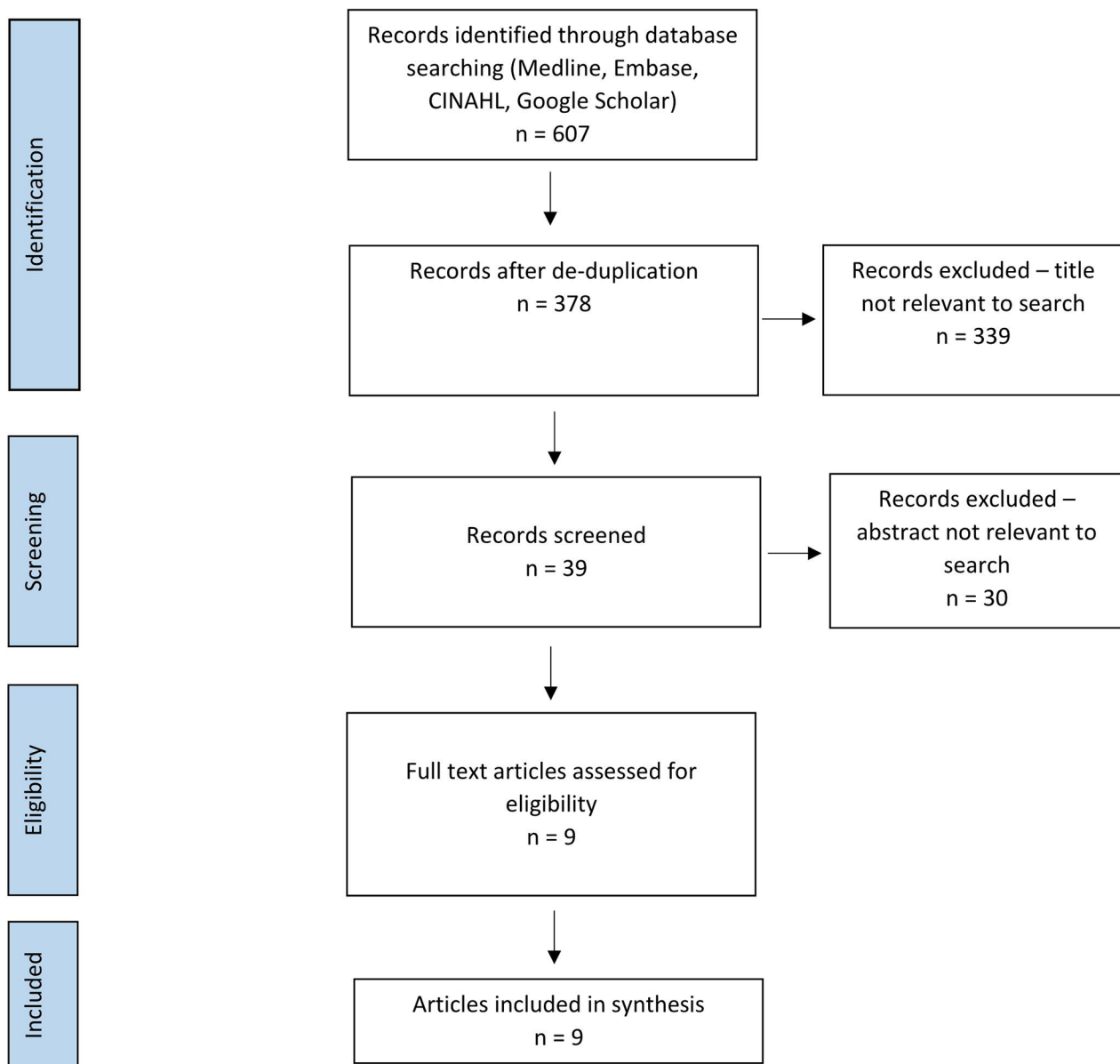


Fig. 1 PRISMA diagram for NTS assessment tools

were assessed for eligibility and 9 were included in the final study.

The summary of characteristics of studies is shown in Table 1. Six observational studies, two retrospective reviews, and one randomised controlled pilot study were included. The trials were conducted in the Netherlands [19, 20], USA [21–24, 27], Lithuania [25] and Canada [26]. Van Maarseveen et al [20] did not declare duration of time over which data was collected. The other studies were conducted over a mean of 6.94 months (range 2–24).

There was heterogeneity between studies in relation to patient groups, outcome measures and methodology. All

studies were single centre studies due to methodology. The key findings of the studies are highlighted in Table 2.

Four unique tools for assessing NTS in this setting were identified. The Communication tool was used to assess whether communication was audible or absent [19]. Three structured tools were identified: the Trauma Non-Technical Skills Assessment Tool (T-NOTECHS), the Concise Assessment of Leader Management (CALM) tool and the Trauma Team Communication Assessment (TTCA-24). The components of T-NOTECHS, CALM and TTCA-24 are shown in Figs. 2, 3 and 4 respectively [19, 23, 27].

Table 1 Characteristics and findings of included studies

Paper, Year, Country	Type of Study	Length of Study (months)	Patient population	Sample Size	Tool used	Methodology
Bergs et al. [19], 2005, Netherlands	Observational	4	Patients RTS* <12 assessed by major trauma team, > 12 assessed by minor trauma team	193	Communication tool	Resuscitations consecutively enrolled. Information transfer evaluated for all ABCDE's†. Observer trained and first 30 videos excluded.
Van Maarseveen et al. [20], 2020, Netherlands	Retrospective review	None stated	Patients whose condition is severe enough to activate trauma team	18	T-NOTECHS	Pre-power calculation determined 18 videos needed for analysis using 3 reviewers and T-NOTECHS. Investigated interrater reliability.
Dumas et al. [21], 2020, USA	Observational	24	Patients who underwent an emergency department thoracotomy (EDT)	61	T-NOTECHS	Used T-NOTECHS to evaluate NTS. Modified to 3-point Likert scale.
Nagaraj et al. [22], 2021, USA	Retrospective review	3	Patients with level 1 trauma or an injury severity score (ISS) > 15	99	T-NOTECHS	Compared standard of handover from paramedics to T-NOTECHS score.
Steinemann et al. [23], 2012, USA	Observational	6.5	Blunt, multisystem trauma in nonpregnant patients > 6 years	69	T-NOTECHS	Developed T-NOTECHS. Interrater reliability and correlation assessed.
Kava et al. [24], 2019, USA	Prospective randomised controlled pilot study	2	Patients triaged as a resuscitation	20	CALM	Postgraduate year 2 and 3 residents acted as team leader for resuscitation. Intervention group reviewed performance before 2nd attempt. Used CALM tool.
Aukstakalnis et al. [25], 2020, Lithuania	Observational	8	All emergency department resuscitation patients > 18 years	143	T-NOTECHS	Audio / video review process to evaluate technical and non-technical skills performance. Used T-NOTECHS.
Bhangu et al. [26], 2022, Canada	Observational	2	Patients for whom a trauma code was activated	55	T-NOTECHS	2 independent reviewers used T-NOTECHS with 5 point Likert scale to assess NTS using video review.
DeMoor et al. [27], 2017, USA	Observational	6	Patients recorded as trauma resuscitations (1:1 ratio of stable:unstable patients)	70	TTCA-24	Compared TTCA-24 to T-NOTECHS and TEAM tool scores for same videos to assess concurrent validity

*RTS=Revised Trauma Score; †ABCDE's=Airway, Breathing, Circulation, Disability, Exposure

T-NOTECHS, CALM and TTCA-24 measure NTS, however, they score components differently. T-NOTECHS splits NTS into 5 distinct categories with a 5-point Likert scale for each heading, whereas CALM and TTCA-24 utilise a 4-point Likert scale for each individual element under its headings. They all assess leadership, communication, and general team performance; however, they adopt individual approaches.

The Mixed Methods Appraisal Tool (MMAT) was used to assess quality of papers; however, this was limited by the variability in provision of evidence [15]. The highest quality papers by MMAT standards was Bergs et al [19] and DeMoor et al [27] with 100% quality review. The T-NOTECHS papers scored a mean of 87.5% (range 75–100%) [20–23, 25, 26]. Kava et al [24] scored 80%, however, there were five sections to score this paper due to its methodology compared to four sections for the others.

Reliability was assessed within studies using inter-class correlation coefficient (ICC). Rater reliability represents the extent to which the data collected in the study correctly represents the variables measured. [28] The

T-NOTECHS papers which used ICC were Steinemann et al [23] who had an ICC score of 0.48 for real-life resuscitations and van Maarseveen [20] et al. had an ICC of 0.94 (0.87–0.98). Steinemann et al [23] report poor reliability and van Maarseveen et al [20] report excellent reliability. [29] Bhangu et al [26] had an ICC score of 0.52 but did not comment on the reliability.

The CALM paper, Kava et al [24], used weighted Kappa between two experts to ensure agreement and it was 0.45 (CI 0.35–0.56, $p < 0.0001$). This is a weak level of agreement. [28] Both ICC and weighted Kappa can be used to assess inter-rater reliability. The other papers did not demonstrate assessment of reliability.

The TTCA-24 paper, DeMoor et al [27], assessed the ICC in both stable and unstable patient activations and reported 0.87 and 0.78, respectively, which demonstrates excellent reliability. [29].

T-NOTECHS is suggested to be a more reliable assessment of NTS than CALM, as inter-rater reliability is higher across the studies that assess it. Both T-NOTECHS and the CALM tool are previously validated in the literature. [23, 30]. The TTCA-24

Table 2 Main findings of included studies

Paper, Year	Main Aim	Main Outcome Measure	Reviewers, interrater reliability	Key findings	Limitations	Conclusions
Bergs et al., 2005 [19]	Document information transfer during multi-disciplinary team resuscitation	Presence of audible communication (16–74%)	1, n/a	Better communication in major trauma team than minor trauma team. Better communication during exposure of severely injured patient ($p=0.06$)	No formal taxonomy used	Lack of feedback in current training system. Communication is sub-optimal. Other industries have better guidelines
Van Maarseveen et al., 2020 [20]	Assess reliability of T-NOTECHS	Interrater reliability using ICC*	3, ICC = 0.94 (0.87–0.98)	Used T-NOTECHS. Average ICC excellent across all 5 domains	Unable to assess intra-observer variability. Used non-experts to assess.	High ICC in all domains. T-NOTECHS is reliable, especially with mean of 3 assessors
Dumas et al., 2020 [21]	Evaluate association between T-NOTECHS scores and ROSC in patients undergoing an EDT	Return of Spontaneous Circulation (ROSC)	3, ICC not calculated	Association between T-NOTECHS scores and ROSC did not reach statistical significance. Assessment and decision-making high scores were 5.3x more likely to lead to ROSC	Small sample size. Did not calculate ICC. Knowledge of outcome may bias evaluation. Not all treatment-specific outcome captured.	Demographic and injury data not associated with ROSC in univariate analysis
Nagaraj et al., 2021 [22]	Characterise EMS handoff experience and effects of non-technical performance	T-NOTECHS, median 10/15 (3-point Likert scale)	1, n/a	Significant difference in score based on experience of team leader. There is a relationship between quality of handover and T-NOTECHS score.	Cessation of videos cause some videos to be excluded. Inability to adjust audio based on noise level.	No significant difference on whether team leader was present, only level of experience
Steinmann et al., 2012 [23]	Develop a tool on 5 essential behavioural domains	T-NOTECHS, mean 16.3–17.7/25 (5-point Likert scale)	2, ICC = 0.48	Significant correlation between T-NOTECHS ratings and standard outcomes such as mortality and length of stay, as well as higher T-NOTECHS correlating with faster resuscitation	Moderate inter-rater reliability and inexperienced raters	Preparedness and resilience associated with situational awareness enabled teams to cope
Kava et al., 2019 [24]	Assess utility of video review compared with self-reflection alone	Assessing gain scores in residents with and without self-reflection using video review	2, weighted Kappa = 0.45	Positive effect of incorporating video review feedback into leadership training for EM	Convenience sampling. Variability in time to complete self-reflection	Potential use is direct coaching from mentor while watching together
Aukstakalnis et al., 2020 [25]	Create framework for trauma team performance analysis in ED	T-NOTECHS, mean 11.9/25 (5-point Likert scale used)	1, n/a	ISS > 15 in 16% of cases. 5.6% trauma patients died. Mean rate of completion of primary survey was 68.5%. Poor assessment and decision making. Low levels of ATLS guideline compliance	One video recording reviewer – fine for timings, bias for T-NOTECHS. Poor sound quality during severely injured resuscitation due to overcrowding and shouting	Larger variability supports loss of standardisation. Performs worse when compared to other studies – lack of briefing, room crowding and poor non-technical skills. Trauma team unfamiliar with each other

Table 2 (continued)

Paper, Year	Main Aim	Main Outcome Measure	Reviewers, interrater reliability	Key findings	Limitations	Conclusions
Bhangu et al., 2022 [26]	Determine whether T-NOTECHS can be used to identify communication gaps in the trauma bay	T-NOTECHS, median 22/25 (5-point Likert scale used)	2, ICC=0.52	ISS > 16 in 37% of cases. Communication and interaction significantly lower median score ($p < 0.0001$) when compared to other domains. More completed closed loop communications in severe cases.	Moderate inter-rater reliability. Limitation on team members due to COVID-19 protocols.	Video review provides an opportunity to identify areas for improvement, such as communication, which can be identified by the T-NOTECHS tool. Video review can further be used to assess callouts and closed loop communication.
DeMoor et al. [27]	Demonstrate TTCA-24 is a valid and reliable measurement of communication	TTCA-24	No. of reviewers not stated, ICC=0.78 (unstable) and 0.87 (stable)	Spearman rank correlation coefficient between TTCA-24 and T-NOTECHS was $r = 0.261$. No correlation with TEAM score. High interrater reliability of TTCA-24.	TTCA-24 places heavy focus on communication and lacks evaluation of other aspects of NTS	TTCA-24 provides more in depth analysis to communication as a NTS but forgoes a broader impression of NTS as a whole.

*ICC = inter-class correlation coefficient; †ISS = injury severity score

demonstrates excellent reliability however, there has only been one study published so far in this context by the author of the tool. The T-NOTECHS reliability is more variable across studies, however, has been better studied.

Discussion

The first tool for assessing NTS in healthcare was developed by Gaba et al [31] in 1998. This was an adaptation of an instrument called NOTECHS where performance was assessed using video recordings from simulated resuscitations [32] in the context of anaesthetic practice. They found high levels of team variability and concluded that the rating system needs refinement before effectively assessing clinical competence. [31] A number of tools have been validated in clinical context, and although Gaba et al [31] is a different context than this review, it demonstrates validity of using video review to assess NTS in simulated resuscitations.

Bergs et al [19] used the Communication tool to assess presence of audible information transfer from physician to team members. The tool focused on a single element, communication, an important NTS and function of leadership and teamwork. They assessed 204 recordings in a single centre. There was a trend towards better communication during care of the severely injured patient ($p = 0.06$). Some information may not have been picked up due to background noise, a confounder which is not corrected for. Bergs et al [19] concluded communication was sub-optimal.

T-NOTECHS was adapted from NOTECHS, a tool previously used in aviation [5]. which had to be validated for clinical application using several steps. [33] Firstly, a draft tool must be developed. This was done for use in the trauma context by Steinemann et al. [23]. Then, a tool must be adapted based on findings of pilot data. Adaptations of T-NOTECHS between papers in this review are the variation in the number of points in the Likert scale used. Five papers used the original 5-point Likert scale. [20, 23, 25–27] The other two papers [21, 22] utilised the same headings, but reduced the respective scales to a 3-point Likert scale. No study has been identified to validate this contraction. The 5-point scale is more accepted in practice due to increased reliability and validity, alongside its ability to identify extreme attitudes. [34] One paper argued that 3-point Likert scales introduce rounding error but they are quicker to complete which increases the usability. [35] Finally, a tool becomes validated when “researcher has come to the opinion that the instrument measures what it was supposed to measure”. [20, 33] In the context of measuring NTS in a trauma setting, the application of T-NOTECHS by more studies shows that authors of further studies agree with the findings of Steinemann et al., [23] and applied the tool to

TRAUMA NOTECHS

Leadership

5	4	3	2	1
Clearly defined team leader. Good time management, all tasks completed, non-hierarchical		Individual defined, but some tasks not completed.		Identity of team leader not clear

Cooperation and Resource Management

5	4	3	2	1
All team members clearly fill a role and perform all designated tasks		Identity of all members not clear, some do not perform assigned tasks		Unable to discern role identity of team members

Communication and Interaction

5	4	3	2	1
Clear communication with Team Leader as a hub, relayed to scribe		Communication not always through Team Leader, or not relayed rapidly to scribe		Unorganized or incoherent communication on many different levels

Assessment and Decision Making

5	4	3	2	1
Orderly and complete 1° (ABCDE) and 2° surveys. Plan communicated to team.		Assessment somewhat out of order, all major tasks complete		1° and 2° surveys disorderly and/or incomplete. Plan not clear.

Situation Awareness / Coping with Stress

5	4	3	2	1
Untoward findings and distractions did not upset systematic and orderly flow. Team is calm and plans ahead.		Untoward findings caused disruption but did not preclude task completion.		Untoward findings or interruptions completely upset orderly assessment and task completion Not anticipatory.

Fig. 2 T-NOTECHS tool [23]

their own studies. [20–23, 25] The T-NOTECHS scale is shown in Fig. 2.

The CALM tool was developed by Nadkarni et al [30] in 2018 and validated in paediatric simulations to assess team leader performance. It was applied to adult real-life resuscitations by Kava et al [24] to assess individual resident performance as team leader. The CALM tool is shown in Fig. 3. It assessed 15 NTS components which is more than the 5 components assessed in T-NOTECHS, providing a greater scope of assessment. T-NOTECHS may be able to give a greater insight into smaller range of NTS assessed.

The TTCA-24 tool was designed by DeMoor et al [27] as they commented on the use of T-NOTECHS and the Team Emergency Assessment Measure (TEAM) developed by Cooper et al [38]. The senior author felt that these tools lacked scope to adequately assess communication as a NTS so developed the TTCA-24 tool to be

used live or during video review. DeMoor et al. assessed concurrent validity between TTCA-24 and T-NOTECHS and TTCA-24 and TEAM. The Spearman rank correlation coefficient between TTCA-24 and T-NOTECHS is $r=0.261$, demonstrating positive correlation that was statistically significant ($p=0.029$). There was no statistically significant correlation between TTCA-24 and TEAM. As T-NOTECHS contains a distinct communication category, it is understandable how these tools would be correlated.

The T-NOTECHS, CALM and TTCA-24 tools both assess leadership, communication and team management. T-NOTECHS emphasises decision making and situational awareness, CALM focuses on medical management and knowledge and TTCA-24 focuses on team communication. These are not distinct categories and demonstrate overlap in some areas. T-NOTECHS recognises the response to “untoward findings”, a useful

Concise Assessment of Leader Management

VIDEO: _____
 DATE: _____ TRAINEE: _____ PGY: _____ ASSESSOR: _____ CASE: _____

I. LEADERSHIP

A. ROLE

1. Announced role as leader no yes
 2. Clear role as leader throughout case rarely sometimes mostly always

B. STYLE

1. Style appropriate and effective for situation rarely sometimes mostly always

Specific examples/comments: _____

II. COMMUNICATION

- A. Voice is appropriately loud and clear rarely sometimes mostly always
 B. Addresses people explicitly (e.g. by name) rarely sometimes mostly always
 C. Reinforces closed-loop communication rarely sometimes mostly always

Specific examples/comments: _____

III. TEAM MANAGEMENT

- A. Assigns or acknowledges roles rarely sometimes mostly always
 B. Directs team effectively / assigns tasks rarely sometimes mostly always
 C. Balances work load of team rarely sometimes mostly always
 D. Engages team members in decision making rarely sometimes mostly always
 E. Summarizes case status periodically rarely sometimes mostly always

Specific examples/comments: _____

IV. MEDICAL MANAGEMENT

- A. Prioritizes task order rarely sometimes mostly always
 B. Maintains global view (avoids fixation bias) rarely sometimes mostly always
 C. Periodically reassesses patient rarely sometimes mostly always
 D. States next step(s) in patient care rarely sometimes mostly always
 E. Aware of limitations and seeks help as needed rarely sometimes mostly always

Specific examples/comments: _____

V. MEDICAL KNOWLEDGE

Prescribe an action plan to address any knowledge gaps identified from today's scenario: _____

VI. GLOBAL ASSESSMENT

How did the leader perform in comparison to peers?

- below expected for level as expected for level above expectations for level top 5%

Fig. 3 CALM tool

Trauma Team Communication Assessment (TTCA-24)						
Team Flow						
<i>The team's efforts to be...</i>	Poor	Fair		Good	Excellent	Observation Notes
Emotionally controlled	1	2		3	4	
Collaborative/organized/structured	1	2		3	4	
Focused and alert	1	2		3	4	
Global Perception	1	2		3	4	
Team Relationships						
<i>Team members demonstrating...</i>	Poor	Fair		Good	Excellent	Observation Notes
Assertive/responsive behaviors	1	2		3	4	
Appropriate task/social talk behaviors	1	2		3	4	
Competent, caring, trustworthy behaviors	1	2		3	4	
Global Perception	1	2		3	4	
Team Space Negotiation						
<i>The team's efforts to...</i>	Poor	Fair		Good	Excellent	Observation Notes
Yield to each other when necessary	1	2		3	4	
Remain at bedside only when necessary	1	2		3	4	
Get-in/get-out	1	2		3	4	
Global Perception	1	2		3	4	
Team Noise Management						
<i>The team's efforts to...</i>	Poor	Fair		Good	Excellent	Observation Notes
Manage environmental noise	1	2		3	4	
Manage team member noise	1	2		3	4	
Global Perception	1	2		3	4	
Team Listening						
<i>The team's efforts to...</i>	Poor	Fair		Good	Excellent	
Pay attention to each other	1	2		3	4	
Understand each other	1	2		3	4	
Respond to each other	1	2		3	4	
Global Perception	1	2		3	4	
Team Emergent Leadership						
<i>The team's efforts to...</i>	Poor	Fair		Good	Excellent	Observation Notes
Instruct each other	1	2		3	4	
Delegate roles and duties	1	2		3	4	
Compensate for others	1	2		3	4	
Ask insightful questions to close gap	1	2		3	4	
Global Perception	1	2		3	4	
Scale Range = 24 to 96 Midpoint = 60 > 60 = Effective Team Communication < 60 = Ineffective Team Communication	1. Team Flow total score:					
	2. Team Relationships total score:					
	3. Team Space Negotiation total score:					
	4. Team Noise Management total score:					
	5. Team Listening total score:					
	6. Team Emergent Leadership total score:					
Team TTCA-24 Total Score:						

Fig. 4 Similarities and differences between the tools identified

inclusion that helps to validate its use in real-life resuscitations, as this is common in the ED. [36] T-NOTECHS and TTCA-24 are designed to assess team performance, whereas CALM is better suited to assessing individual performance.

All tools demonstrate a high level of usability. T-NOTECHS provides an explanation for the lowest, highest and middle score to guide the user. CALM uses a simple scoring system which enables the user to assess the frequency at which each NTS is exhibited.

T-NOTECHS is potentially easier to complete as limited number of components to rate. When paired with video review, reviewers can pause or rewind the video for a more accurate assessment of NTS. [23] The TTCA-24 tool was designed for interprofessional use and comes with a codebook. The high inter-rater reliability suggests that the raters utilised the tool in the same way. The inter-rater reliability of TTCA-24 is highest of the three, however, it has both TTCA-24 and CALM have only been utilised in this context in one paper so more research is needed. T-NOTECHS has more variable ICC across studies, so more research would be beneficial in getting a truer picture of ICC across a larger sample size. [20–27].

Higham et al [1] evaluated tools used for assessment of NTS in healthcare. Due to broader inclusion criteria, this study identified 76 distinct tools, including T-NOTECHS, for assessment of NTS. They noted a large amount of variation between methodology of design of tools, extent of their validity and usability. This was also evident in the comparison of our three assessed tools. They suggest that there is a “need for rationalisation and standardisation in the way we assess non-technical skills in healthcare”. This study was published in 2019 and included Steinemann et al. [23], and 6 out of 7 of the studies we reported that used T-NOTECHS were published later. The inclusion of the newer studies in our review furthers the research into the standardisation of assessment of NTS.

Bhangu et al [37] also published a scoping review in 2022 evaluating tools used to assess NTS in both real world and simulated settings. They identified the T-NOTECHS and TEAM tool as the most reliable for use in this context. The TEAM tool was used in studies utilising simulation which means they do not fit the inclusion criteria for this review. This tool was adapted from a paper by Cooper et al. in 2010 [38] and further validated in 2016 [39] in both simulated and real-life settings, without video review. No studies included in this review utilised the TEAM tool.

The aim of this review was to provide an overview of tools used to assess NTS in resuscitation teams within the ED using video review and to explore the evidence for the validity and usability of the tools. This review has answered the stated aims despite having a limited number of papers included. We found T-NOTECHS to be the most valid tool and has been shown to be a reliable tool to assess NTS during resuscitation in the emergency department using video review. The TTCA-24 tool showed early signs of good reliability but will need to be further validated. The TTCA-24 provides more insight into communication as a NTS than T-NOTECHS, but when assessing NTS more holistically, T-NOTECHS demonstrates usability, reliability and validity. The authors are aware of the difficulty of excluding bias and can hope that the techniques utilised minimised bias.

Due to the heterogeneity of studies, there was limited application of statistical approaches to compare tools. A similar review identifies a need to benchmark outcomes between studies, thus enabling a potential future meta-analysis. [40] The findings of our review provide more clarity on the use of T-NOTECHS as a standardised tool which would enable use of video review as a tool in education and quality improvement. [41] One study translated T-NOTECHS into Finnish to assess translatability and validity and found that it can still be used to assess efficacy of trauma team resuscitations. This study used simulated trauma resuscitations, which was an exclusion criteria for our review. [34].

Steinemann et al [23] also assessed use of T-NOTECHS in the context of simulated resuscitations using video review. Rater agreement was higher in simulated resuscitations than in real-life resuscitations (ICC=0.71). There was a significant correlation found between the number of completed resuscitation tasks ($r=0.50$, $P<0.01$) and faster time to completion of the 3 common resuscitation tasks ($r=-0.38$, $P<0.05$). [23] Simulated resuscitations are a useful tool to assess NTS of staff as there are less ethical considerations when filming patients. However, the nature of the simulated environment does not provide assessors with a true picture of how teams would perform in a real life clinical setting, hence the exclusion from our review.

This review highlights the tools used in this setting and recommends use of T-NOTECHS to assess NTS in resuscitation teams within the ED using video review. In terms of future study, using T-NOTECHS with larger sample sizes, such as in a multi-centre study may greatly establish utility of this tool. TTCA-24 may have uses in departments where communication is identified as a weakness by the use of T-NOTECHS or other means. Both tools can be used to identify areas where further clinician education is indicated. Furthermore, there is scope to formally compare NTS with TS using video review within the ED.

Limitations

One of the limitations of this review is the small sample size. There is a breadth of tools available that assess NTS across all domains of healthcare, however, use of video review in the ED is a growing field and excluding studies without video review reduced the number available. Due to the infrastructure and resource demands to review video creation and validation of a new tool and demonstrating generalisability will be challenging. Use of tools developed and validated in the simulation context requires demonstration of their utility in real-world clinical care.

Many institutions lack audio-visual recording access due to financial and ethical restraints, therefore there

is limited generalisability for these findings. Researchers may be faced with a reluctance to be filmed due to privacy concerns from staff regarding patients and themselves. There should be strict measures in place to ensure recordings are only accessed by appropriate personnel to ensure privacy and security.

Conclusion

The aim of this review was to provide overview of tools used to assess NTS in resuscitation teams within the ED using video review and to explore the evidence for the validity and usability of the tools. T-NOTECHS was first validated in Steinemann et al [23] and therefore was the tool of choice for the majority of future papers assessing NTS in the ED using video review. This review found T-NOTECHS to be valid and reliable. The conclusion that T-NOTECHS is the best tool of those used in this context is suggested, but not able to be proven fully due to small sample sizes.

Acknowledgements:

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12873-023-00895-7>.

Supplementary Material 1

Acknowledgements

I would like to express my deepest appreciation to Mark and Valerie Alexander, without whom I would never have been given the opportunity to pursue my career. I would like to thank my supervisors Professor David J Lowe and Dr Fraser Denny for their help and guidance.

Authors' contributions

E.A. wrote the main manuscript text and prepared the figures and tables. The project was supervised by F.D. and D.J.L. The manuscript was reviewed by E.A., F.D., M.W.G.G., C.M. and D.J.L. The search strategy was reviewed by C.M. and E.A. and where agreement was not reached, D.J.L. further reviewed the papers and decided on inclusion. All authors read and approved the final manuscript.

Funding

Not applicable.

Data availability

All of the material is owned by the authors and no permissions are required. For access to raw data analysed, contact Emily G Alexander at 237165a@student.gla.ac.uk.

Declarations

Ethics approval and consent to participate

All methods were carried out in accordance with relevant guidelines and regulations. As this paper is a scoping review, there were no experiments on humans performed as part of this research. No experimental protocols or consent was required.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 June 2023 / Accepted: 17 October 2023

Published online: 29 November 2023

References

1. Higham H, Greig PR, Rutherford J, Vincent L, Young D, Vincent C. Observer-based tools for non-technical skills assessment in simulated and real clinical environments in healthcare: a systematic review. *BMJ Qual Saf.* 2019;28:672–86. <https://doi.org/10.1136/bmjqs-2018-008565>.
2. Krage R, Zwaan L, Len LTS, Kolenbrander MW, van Groeningen D, Loer SA, Wagner C, et al. Relationship between non-technical skills and technical performance during cardiopulmonary resuscitation. *Emerg Med J.* 2017;34(11):728–33. <https://doi.org/10.1136/emered-2016-205754>.
3. Prineas S, Mosier K, Mirko C, Guiccardi S. Non-technical skills in healthcare. In: Donaldson L, Ricciardi W, Sheridan S, Tartaglia R, editors. *Textbook of Patient Safety and Clinical Risk Management.* Springer; 2021. pp. 413–34.
4. Soar JB, Bili F, et al. Part 12: education, implementation and teams: 2010 international consensus on cardiopulmonary resuscitation and emergency cardiovascular care with treatment recommendations. *Resuscitation.* 2010;81:e288–e330.
5. Flin R, Martin L, Goeters K, Hoermann J, Amalberti R, Valot C, et al. Development of the NOTECHS (non-technical skills) system for assessing Pilots' CRM skills. *Hum Factors Aerosp Saf.* 2003;3(2):95–117.
6. Flight Safety Foundation. *Assessment and Feedback of non-technical skills (OGHFA BN).* EUROCONTROL; 2022.
7. Kovacs G, Croskerry P. Clinical decision making: an emergency medicine perspective. *Acad Emerg Med.* 1999;6:947–52. <https://doi.org/10.1111/j.1553-2712.1999.tb01246.x>.
8. Oakley E, Stocker S, Staubli G, Young S. Using video recording to identify management errors in paediatric trauma resuscitation. *Paediatrics.* 2006;117:658–64. <https://doi.org/10.1542/peds.2004-1803>.
9. Wears RL, Schubert CC. Visualizing expertise in Context. *Ann Emerg Med.* 2016;67:752–4. <https://doi.org/10.1016/j.annemergmed.2015.11.027>.
10. Williams KS, Pace C, Milia D, Juern J, Rubin J. Development of a Video Recording and Review process for Trauma Resuscitation Quality and Education. *West J Emerg Med.* 2019;20(2):228–31. <https://doi.org/10.5811/westjem.2018.12.40951>.
11. Lowe DJ, Dewar A, Lloyd A, Edgar S, Clegg GR. Optimising clinical performance during resuscitation using video evaluation. *Postgrad Med J.* 2017;93:449–53. <https://doi.org/10.1136/postgradmedj-2016-134357>.
12. Cooper S, Endacott R, Cant R. Measuring non-technical skills in medical emergency care: a review of assessment measures. *Open Access Emergency Medicine.* 2010;2:7–16. <https://doi.org/10.2147/oaem.s6693>.
13. Page MJ, McKenzie JE, Bossuyt PM, Bourton I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ.* 2021;372:n71.
14. Lowe DJ, Ireland A, Ross AI, Ker J. Exploring situational awareness in emergency medicine: developing a shared mental model to enhance training and assessment. *Postgrad Med J.* 2016;92(1093):653–8. <https://doi.org/10.1136/postgradmedj-2015-133772>.
15. Hong QN, Fàbregues S, Bartlett G, Boardman F, Cargo M, Dagenais P, et al. Mixed methods Appraisal Tool (MMAT) Version 2018 for information professionals and researchers. *Educ Inf.* 2018;34(4):285–91.
16. Pluye P, Cargo M, Robert E, Bartlett G, O' Cathain A, Griffiths F, et al. *A pilot mixed methods Appraisal Tool (MMAT) for systematic mixed studies reviews.* In: abstracts of the 19th Cochrane Colloquium. John Wiley & Sons; 2011.
17. Pace R, Pluye P, Bartlett G, Macaulay AC, Salsberg J, Jagosh J, et al. Testing the reliability and efficiency of the pilot mixed methods Appraisal Tool (MMAT) for systematic mixed studies review. *Int J Nurs Stud.* 2012;49:47–53. <https://doi.org/10.1016/j.ijnurstu.2011.07.002>.
18. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Med.* 2009;6(7). <https://doi.org/10.1371/journal.pmed.1000097>.
19. Bergs EA, Rutten FLPA, Tadors T, Krijnen P, Schipper IB. Communication during trauma resuscitation: do we know what is happening? *Injury.* 2005;36(8):905–11. <https://doi.org/10.1016/j.injury.2004.12.047>.
20. van Maarseveen OEC, Ham WHW, Huijsmans RLN, Dolmans RGF, Leened LPH. Reliability of the assessment of non-technical skills by using video-recorded trauma resuscitations. *Eur J Trauma Emerg Surg.* 2020;48(1):441–7. <https://doi.org/10.1007/s00068-020-01401-5>.

21. Dumas RP, Vella MA, Chreiman KC, Smith BP, Subramanian M, Maher Z, et al. Team Assessment and decision making is Associated with outcomes: a trauma. *Video Rev Anal*. 2020;246:544–49. <https://doi.org/10.1016/j.jss.2019.09.033>.
22. Nagaraj MB, Lowe JE, Marinica AL, Morshedi BB, Isaacs SM, Miller BL. Using Trauma Video Review to assess EMS Handoff and Trauma Team Non-technical Skills. *Prehosp Emerg Care*. 2021;1–12. <https://doi.org/10.1080/10903127.2021.2000684>.
23. Steinemann S, Berg B, DiTullio A, Skinner A, Terada K, Anzelon K, Ho HC. Assessing teamwork in the trauma bay: introduction of a modified NOTECHS scale for trauma. *Am J Surg*. 2012;203(1):69–75. <https://doi.org/10.1016/j.amjsurg.2011.08.004>.
24. Kava L, Jones K, Ehrman R, Smylie L, McRae M, Dubey E, et al. Video-assisted self-reflection of resuscitations for resident education and improvement of leadership skills: a pilot study. *Perspect Med Educ*. 2022;11(2):80–5. <https://doi.org/10.1007/s40037-021-00690-9>.
25. Aukstakalnis V, Dambrauskas Z, Stasaitis K, Darginavicius L, Doboziuskas P, Jasinskis N, Vaitkaitis D. What happens in the shock room stays in the shock room? A time-based audio/video audit framework for trauma team performance analysis. *Eur J Emerg Med*. 2020;27(2):121–4. <https://doi.org/10.1097/MEJ.0000000000000627>.
26. Bhangu A, Notario L, Pinto RL, Pannell D, Thomas-Boaz W, Freedman C, Tien H, Nathens AR, da Luz L. Closed loop communication in the trauma bay: identifying opportunities for team performance improvement through a video review analysis. *Can J Emerg Med*. 2022;24:419–25. <https://doi.org/10.1007/s43678-022-00295-z>.
27. DeMoor S, Shady A, Olmsted R, Myers JG, Parker-Raley J. Evaluating trauma team performance in a Level I trauma center validation of the trauma team communication assessment (TTCA-24). *J Trauma Acute Surg*. 2017;83(1):159–64. <https://doi.org/10.1097/TA.0000000000001526>.
28. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)*. 2012;22(3):276–82.
29. Bobak CA, Barr PJ, O'Malley AJ. Estimation of an inter-rater intra-class correlation coefficient that overcomes common assumption violations in the assessment of health measurement scales. *BMC Med Res Methodol*. 2018;93. <https://doi.org/10.1186/s12874-018-0550-6>.
30. Nadkarni LD, Roskind CG, Auerbach MA, Calhoun AW, Adler MD, Kessler DO. The Development and Validation of a Concise Instrument for Formative Assessment of Team Leader Performance during simulated Pediatric resuscitations. *Simul Healthc*. 2018;13(2):77–82. <https://doi.org/10.1097/SIH.0000000000000267>.
31. Gaba DM, Howard SK, Flanagan B, Smith BE, Fish KJ, Botney R. Assessment of clinical performance during simulated crisis using both technical and behavioral ratings. *Anesthesiology*. 1998;89(1):8–18. <https://doi.org/10.1097/0000542-199807000-00005>.
32. van Avermaete JAG. *NOTECHS: non-technical skill evaluation in JAR-FCL*. Natl Aerosp Libr NLR. 1998.1.1.178.5648&rep = rep1&type = pdf
33. Alyusuf RH, Prasad K, Satir AMA, Abalkhail AA, Arora RK. Development and validation of a tool to evaluate the quality of medical education websites in pathology. *J Pathol Inform*. 2013;4:29. <https://doi.org/10.4103/2153-3539.120729>.
34. Repo JP, Rosqvist E, Lauritsalo S, Paloneva J. Translatability and validation of non-technical skills scale for trauma (T-NOTECHS) for assessing simulated multi-professional trauma team resuscitations. *BMC Med Educ*. 2019;19(40). <https://doi.org/10.1186/s12909-019-1474-5>.
35. Lehmann DR, Hulbert J. Are three-point scales always good Enough? *J Mark Res*. 1972;9(4):444–6. <https://doi.org/10.1177/002224377200900416>.
36. Platts-Mills TF, Nagurney JM, Melnick ER. Tolerance and uncertainty and the practice of Emergency Medicine. *Ann Emerg Med*. 2020;75(6):715–20. <https://doi.org/10.1016/j.annemergmed.2019.10.015>.
37. Bhangu A, Stevenson C, Szulewski A, MacDonald A, Nolan B. A scoping review of nontechnical skill assessment tools to evaluate trauma team performance. *J Trauma Acute Care Surg*. 2022;92(5):e81–e91. <https://doi.org/10.1097/TA.0000000000003492>.
38. Cooper S, Cant R, Porter J, Sellick K, Somers G, Kinsman L, Nestel D. Rating medical emergency teamwork performance: development of the Team Emergency Assessment measure (TEAM). *Resuscitation*. 2010;81(4):446–52. <https://doi.org/10.1016/j.resuscitation.2009.11.027>.
39. Cooper S, Cant R, Connell C, Sims L, Porter JE, Symmons M, Nestel D, Liaw SY. Measuring teamwork performance: validity testing of the Team Emergency Assessment measure (TEAM) with clinical resuscitation teams. *Resuscitation*. 2016;101:97–101. <https://doi.org/10.1016/j.resuscitation.2016.01.026>.
40. Brogaard L, Ulldberg N. Filming for auditing of real-life emergency teams: a systematic review. *BMJ Open Qual*. 2019;8(4). <https://doi.org/10.1136/bmjopen-2018-000588>.
41. Quirion A, Nikouline A, Jung J, Nolan B. Contemporary uses of trauma video review: a scoping review. *CJEM*. 2021;23(6):787–96. <https://doi.org/10.1007/s43678-021-00178-9>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.