



Published in final edited form as:

Am J Surg. 2018 September ; 216(3): 630–635. doi:10.1016/j.amjsurg.2018.01.010.

Self-Assessment of Team Performance using T-NOTECHS in Simulated Pediatric Trauma Resuscitation is not Consistent with Expert Assessment

Minna M Wieck, MD^b, Cory McLaughlin, MD^b, Todd P Chang, MD MAcM^c, Alyssa Rake, MD^{d,e}, Caron Park, MS^f, Christianne Lane, PhD^f, Rita V Burke, PhD MPH^{a,b}, L Caulette Young, RN BSN CCRN^{d,e}, Elizabeth A Cleek, RN MS CPNP^{a,b}, Inge Morton, RN MSN CPN^c, Catherine J Goodhue, MN RN CPNP^{a,b}, Randall S Burd, MD PhD^g, Henri R Ford, MD MHA^{a,b}, Jeffrey S Upperman, MD^{a,b}, Aaron R Jensen, MD MEd^{a,b}

^aTrauma Program, Children's Hospital Los Angeles and the Keck School of Medicine - University of Southern California, Los Angeles, CA 90027

^bDivision of Pediatric Surgery, Children's Hospital Los Angeles and the Keck School of Medicine - University of Southern California, Los Angeles, CA 90027

^cDivision of Emergency and Transport Medicine, and Children's Hospital Los Angeles and the Keck School of Medicine - University of Southern California, Los Angeles, CA 90027

^dDivision of Anesthesiology and Critical Care Medicine, Children's Hospital Los Angeles and the Keck School of Medicine - University of Southern California, Los Angeles, CA 90027

^eLas Madrinas Simulation Research Laboratory, Children's Hospital Los Angeles and the Keck School of Medicine - University of Southern California, Los Angeles, CA 90027

^fSouthern California Clinical Translational Science Institute (SC-CTSI) and Department of Preventative Medicine, Keck School of Medicine - University of Southern California, Los Angeles, CA 90089

^gDivision of Trauma and Burn Surgery, Children's National Medical Center, Washington, DC 20010

1. Introduction

Poor teamwork and communication have been identified as important contributors to medical errors.^{1,2} The high stress, emergent, and time-dependent nature of acute trauma

Corresponding author: Aaron R Jensen, MD, MEd, Children's Hospital Los Angeles, 4650 Sunset Blvd, Department of Surgery, Mailstop 100, Los Angeles CA 90027, Phone: (323) 361-3341, Fax: (323) 361-3534, ajensen@chla.usc.edu.

Author Contributions

Study Design: Jensen, Burke, Upperman

Data Acquisition: Wieck, Chang, Rake, Young, Cleek, Morton, Goodhue, Upperman, Jensen

Statistical Analysis: Park, Lane

Data Interpretation: Chang, McLaughlin, Ford, Burd, Upperman, Jensen

Drafting of Manuscript: Wieck, McLaughlin, Chang, Park, Goodhue, Jensen

Revision of Manuscript: Rake, Lane, Burke, Young, Cleek, Morton, Ford, Burd, Upperman

Conflict of Interest

None of the authors have any conflicts of interest to disclose.

resuscitation can amplify problems with teamwork, which can potentially lead to medical error. Improving trauma team performance has become the focus of many studies, with simulation-based team training becoming increasingly popular to address deficiencies in team skills.³⁻⁶ Trauma team training using simulation is thought to improve leadership, teamwork skills, and resuscitation process efficiency.^{7,8} Accurately and reliably measuring teamwork skills to demonstrate the impact of team training, however, remains difficult due to a paucity of validated assessment tools that are easily deployable.

Expert assessment of team performance is the current gold standard for measuring teamwork, but the use of experts is subject to scheduling difficulties and is a strain on valuable faculty resources. These obstacles can be minimized by using video review by experts, but this methodology has an inherent delay and does not allow immediate assessment or feedback. Self-assessment of team performance by members of the resuscitation team is an attractive option because: 1) it does not incur the cost of expert evaluators, and 2) the assessment can be performed in real time. Self-assessment is also an integral component self-directed learning and can be as effective as instructor-led teaching in some settings.⁹ Physicians' self-assessments, however, are often inaccurate when compared to experts' assessments – particularly for non-technical skills.¹⁰⁻¹² Although self-assessment of team performance in a multidisciplinary trauma setting is appealing, its validity and reliability are not known.

The NON-TECHnical skills (NOTECHS) scale is a tool for evaluating team performance of non-technical skills in five categories: leadership, cooperation, communication, assessment and decision making, and situational awareness. It was initially adapted from the airline industry for assessing surgical teams and later modified for use with trauma teams as the T-NOTECHS scale.^{13,14} When used by expert reviewers, T-NOTECHS has been validated in simulated and live trauma resuscitation. Its clinical relevance has also been demonstrated by its correlation with improved clinical performance.⁸ While administration by the multidisciplinary resuscitation team members – rather than expert reviewers – may increase the practical utility of T-NOTECHS, its validity as a self-assessment tool is unknown. The purpose of this study was to: (1) determine the reliability of team-based self-assessment using T-NOTECHS during simulated pediatric trauma resuscitation, and (2) determine the concurrent validity of team-based self-assessment using T-NOTECHS compared to the gold standard expert assessment. We hypothesized that self-assessment using the T-NOTECHS instrument would be a valid and reliable method of measuring trauma team performance as compared to the gold standard of expert assessment.

2. Methods

An *in situ* simulated multidisciplinary pediatric trauma resuscitation training program was implemented to improve team performance at our urban, free-standing children's hospital that has been designated and verified as a level I pediatric trauma center. After Institutional Review Board approval, data was prospectively collected in conjunction with this training program from fifteen consecutive simulated pediatric trauma resuscitations. A panel of experts and all trauma team participants assessed team performance during each of the simulated resuscitations.

2.1. Participants

In situ high-fidelity simulated resuscitations were held in an emergency department (ED) trauma bay with a multidisciplinary team. All participants were providers who respond to level I trauma activations, which are the highest acuity trauma activations in our institution. Each trauma resuscitation team consisted of at least twelve providers according to hospital protocol: pediatric surgery fellow, junior surgical resident (PGY 2 or 3), pediatric emergency medicine (PEM) attending, three ED nurses, pediatric intensive care unit (PICU) nurse, respiratory care practitioner, radiology technologist, clinical pharmacist, ED technician, and nursing supervisor. Individual roles and responsibilities for each provider at a level I trauma activation were delineated in the hospital trauma resuscitation protocol. The pediatric surgery fellow functioned as the trauma team leader and the PEM attending was responsible for airway management.

2.2. Simulation-Based Training and Debriefing

Actual level I trauma activations were paged-out to the on-call trauma team with a ten-minute pre-arrival notification. Once respondents arrived in the trauma bay, they were informed it would be a simulation and were briefed on the simulation exercise. After a 30-second simulation briefing, teams were given about two minutes before patient 'arrival' to prepare for the resuscitation. Simulated scenarios included patients injured by several mechanisms, including (1) fall from a height with tension pneumothorax, (2) fall from height with traumatic brain injury, and (3) motor vehicle collision with facial trauma and hemorrhagic shock. We used high-fidelity pediatric simulators, each tailored to the respective educational scenario: SimBaby™ (Laerdal Medical, Stavanger, Norway), Gaumard HAL S3004 pediatric simulator (Gaumard® Scientific, Miami, FL), and TraumaChild® (Simulab Corporation, Seattle, WA). Each scenario was designed to progress rapidly to full traumatic arrest within minutes without intervention, and be completed within 10–15 minutes. All scenarios required procedural intervention, including intubation, chest tube placement, or surgical airway placement. To preserve fidelity, the team used actual supplies and medications from the trauma bay. Following completion of each scenario, all participants were immediately led through a 40-minute debriefing by either a PEM attending or pediatric surgery attending trained in debriefing methods. These sessions focused on teamwork, medical knowledge, and a review of hospital specific trauma protocols using the advocacy-inquiry debriefing method.^{15,16}

2.3. Outcomes Measurement

Team function was assessed with the T-NOTECHS instrument,⁸ a version of the NOTECHS tool revised for use in trauma resuscitation. T-NOTECHS assesses performance using a five-point scale with objective anchors for each of five domains: (1) leadership, (2) cooperation and resource management, (3) communication and interaction, (4) assessment and decision making, and (5) situational awareness/coping with stress. A score of 'five' represented ideal behavior within the domain, and a score of 'one' represented poor performance. Overall team performance was then quantified by total T-NOTECHS score, the additive sum of all five domain scores.

Each session was assessed by a different set of three expert raters who were selected from a group of eight expert raters based on availability. Expert raters included physicians and nurses with expertise in trauma care, and were all members of the study team (ARJ, JSU, TPC, EC, IM, MW, AR, LCY). Each expert evaluated between 1 to 11 different sessions (median 6 sessions). Expert raters received brief instruction in the use of the instrument, including specific review of the objective anchors on each scale as well as examples of team performance that demonstrate each anchor. Rating forms were then provided to each expert rater before the simulation began. Expert raters were not part of the resuscitation team during the simulation, but did participate in debriefing exercises after the simulation was completed.

At the conclusion of the debriefings, all members of the resuscitation team were given a T-NOTECHS assessment form and asked to assess the resuscitation team's performance. Resuscitation team members did not receive any instructions on how to use the T-NOTECHS instrument. No identifying information was collected on this form other than self-identified provider role in the resuscitation (i.e., 'nurse' or 'pharmacist'). All forms were collected from expert raters as well as team members prior to leaving the trauma bay.

2.4 Statistical Analysis

Boxplots and scatter plots with horizontal and vertical standard error bars were used to determine and understand the variability among the experts' and self-assessors' scores for the five T-NOTECHS domains (subscales) as well as the total composite T-NOTECHS score. Scatterplots with horizontal and vertical standard errors bars were produced to show the variability of the scores within experts (vertical bars) and non-experts (horizontal bars) together, which can be easily ignored when just looking at just the means or medians alone. Identities of expert raters were recorded and allowed analysis of variability between different expert raters, but team members performing self-assessment were anonymous and therefore treated as independent raters. Inter-rater reliability was assessed for the five domains with ordinal outcomes using a linear weighted kappa,¹⁷ and for total composite T-NOTECHS using intra-class correlations (ICC) obtained from a mixed model accounting for session clustering. The intra-class correlations were calculated from the covariance parameters from the mixed models. Using three expert raters and 8 to 15 randomly selected team self-assessment raters for each session would be equivalent to the type ICC (1,1) if using an ANOVA model. Agreement between every pairing of raters was also assessed by session using Bland Altman plots for the total score, which was treated as a continuous outcome. A mixed effects model accounting for session clustering was used to determine if expert ratings for the five domains and composite total were statistically different from the self-assessment team. All analyses were performed using SAS software v. 9.4 (SAS Institute Inc.), and the scatter plots were produced in R v. 3.3.1. Significance tests were two-tailed, with $\alpha=0.05$.

3. Results

One hundred sixty-three individual team members performed team self-assessment over 15 simulated scenarios (38 physicians, 57 nurses, 17 respiratory care practitioners, six

pharmacists, one radiology technologist, 10 ED technicians, and 34 did not indicate role). Eight to 15 (median 10) team members performed self-assessment per session. Since team members' ratings were anonymous, we were unable to determine if a rater rated more than one session.

Variability in the scores for the five T-NOTECHS domains and for the composite total score differed by session between experts and team self-assessment (Fig. 1). For most domains, expert raters demonstrated more variability in scores compared to the self-assessors. Evaluation of sessions with a wider range of variability indicated that among expert raters, one rater generally scored lower than others in most of the five domains. Another expert rater appeared to score much higher in the decision making domain; however, since this rater was assigned to rate only one session, we do not know if this was due to chance or if the rater tended to rate higher in general. With regards to total score, the Bland Altman plots showed much more disagreement across raters for the self-assessment group. Since the raters were anonymous, we were unable to determine if a specific rater was systematically scoring differently from other providers. For self-assessments that contained identifying information on provider role (i.e., nurse, physician, RCP, pharmacist) we assessed for systematic differences in ratings by role type. There was large disagreement across raters of specific roles, but this was random. Raters from any given role did not systematically score better or worse compared to the others' roles for the five domains or total composite score.

The weighted kappa's indicated slight to moderate agreement (0.01–0.20, slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement)¹⁸ ranging from 0.10 to 0.48 in experts and 0.10 to 0.31 in the self-assessment team for the five domains (Table 1). Experts tended to have better agreement than the self-assessors when scoring leadership, communication, and decision making domains. Team self-assessors agreed more in the domains of resource management and situational awareness. Total composite T-NOTECHS scores were similar between experts and self-assessors (ICC=0.57 and 0.56, respectively).

Comparison of T-NOTECHS scores between self- and expert assessments demonstrated that leadership domain scores were not different. Self-assessment scores were significantly higher, however, for the other four domains (resource management, communication, decision making, situational awareness) and for total composite T-NOTECHS scores. (Table 2).

4. Discussion

In this study, several findings were observed regarding teamwork assessments. First, self-assessment T-NOTECHS scores were higher than expert scores in four of five domains. We also observed a higher level of inter-rater reliability among the self-assessment scores of team participants compared to expert scores in the domains of resource management and situational awareness. These results suggest that expert and self-assessments are different, and that self-assessment should not replace expert raters when evaluating resuscitation using the T-NOTECHS.

This study found team members assigned higher rating scores to their own team's performance than those scores assigned by expert independent observers, except in the domain of leadership. This difference is consistent with prior work by Steinemann showing that (1) trauma attendings and critical care nurses embedded in the resuscitation team rate team performance higher than expert observers in the total T-NOTECHS composite score and (2) nurses rate performance higher than expert observers in all domains except leadership.^{19,20} Self-assessments may be more accurate due to increased insight into nonverbal communication, decision making processes, and awareness of ongoing events. Expert reviewers completed their assessments in real time, and may have missed specific events or statements. Expert reviewers may also have been biased to search for areas of improvement to discuss in the subsequent debriefing and may have focused on poorly performed actions. Because most expert raters graded more than one simulation, they could have reflected on prior performances to introduce a bias into their scoring that the team members did not have.

Alternatively, self-assessment scores could be an inaccurate reflection of actual team performance due to lack of objectivity by team members. Team members provided self-assessments following the debriefing in which elements of the resuscitation process are summarized. They may be reporting post-debriefing impressions that were augmented only by the debriefing process and not necessarily reflective of their immediate perceptions during the actual resuscitation. We would expect that this effect would bias us toward the null hypothesis with self-assessment more closely resembling expert assessment. If this is the case, the true difference may be larger than we have demonstrated. Team members may have also overestimated the quality of team performance because all resuscitations ended in the patient "being saved". Participants may also have been too engaged in their tasks during the resuscitation to notice subtle errors and missed communication events.

Leadership, communication, and decision making often require more verbal communication skills, which may have made these domains much easier for expert reviewers to grade. In contrast, resource management and situational awareness may require more internalized thought processes, which may account for the discrepancy between self- and expert assessments. For example, team members may have been aware of resource utilization but no overt demonstration of this awareness was visible to raters. Expert reviewers may have observed chaos during the simulation and interpreted this as poor resource management and situational awareness. Although each individual team member may have had an individual understanding of the shared team mental models for resource management and situational awareness, these models may not have been consistent between individuals. Alternatively, the self-assessment may be overestimating team function in these domains and experts may more closely represent truth. This discrepancy highlights the challenge of verbalizing thoughts and communicating effectively in high stress, multidisciplinary trauma team resuscitations. Internalization of concerns, rather than verbalizing them, is potentially dangerous and a contributor to poor team performance. Because many procedures and assessments occur simultaneously during trauma resuscitation, all providers need to visibly and vocally demonstrate situational awareness so that all participants form a shared mental model.

The low to moderate kappa and ICC in both groups may indicate a lack of agreement within the raters. Because expert raters did not rate all sessions and self-assessment raters were anonymous, it is difficult to determine if the lack of agreement was due to using different raters. The validation study of T-NOTECHS showed highest expert inter-rater reliability when reviewing video recordings of trauma simulations.⁸ In contrast, our raters completed their assessments immediately after the simulation debriefing, in real time. Another T-NOTECHS study showed that inter-rater agreement among expert raters was lower when evaluating larger resuscitation teams (8–9 members) compared to smaller teams (5–7 members).¹⁹ At our institution, 12 providers respond to level 1 trauma activations. Therefore, it may have been unrealistic to have our expert raters evaluate such a large trauma resuscitation team in real time. Perhaps assessment at a delayed interval, after reviewing the video recording, and more training in the use of T-NOTECHS would yield higher inter-rater reliability.

Previous studies with other teamwork assessment tools have shown that greater similarity in experience level and medical specialty result in more consistent scores between raters.^{13,21} In the original T-NOTECHS study, expert raters were all critical care nurses with experience in scribing during trauma resuscitations but minimal patient contact.⁸ In contrast, our expert raters were physicians and nurses from different subspecialties and disciplines, but each with several years of experience in hands-on patient care during trauma resuscitations. Although these differences between our expert reviewers may have contributed to the lower degree of consistency, this group of reviewers is more representative of the diversity of roles on the trauma team. These results may therefore be a more accurate assessment of T-NOTECHS' validity when used by diverse expert reviewers.

The expert raters here used different methods of training to expertise than a previously used approach.⁸ Reported training protocols consisted of pre-reading, a one-hour workshop on T-NOTECHS, practice evaluations of videotaped simulations, and then further additional training sessions over the six-month trial period. Expert reviewers in our study participated in a shorter session with only verbal discussion of the T-NOTECHS instrument, specific criteria, and examples of representative team behavior for each point on the scale. This abbreviated protocol is one of the significant limitations of this study, but it is a more pragmatic implementation of T-NOTECHS. More extensive training is not practical when available external expert reviewers are integrated into a quality improvement-targeted training program or when used for large groups as a self-assessment tool.

The self-assessment scores in this study should be interpreted in light of key limitations. Specifically, one compromise of the study design was to not incorporate team members into the T-NOTECHS training. As outlined above, only expert raters underwent the training – a relatively exhaustive process which is most practical for a small group of individuals. A significant resource investment would have been required to train all 190 potential providers, although this may be possible in future follow-up studies. An additional limitation of the self-assessment scores is the potential relationship with provider specialty. We attempted to account for this potential confounder by performing subgroup analyses based on specialty, which showed no significant differences in T-NOTECHS scores. However, 21% of team members did not report their specialty, so this analysis was likely biased due to missing data.

A further limitation of this study is the lack of ‘low-performing’ teams. We noted a clustering of teams toward the middle and high-performing range of the scale with no teams performing at the bottom of the scale. This is not surprising as this *in situ* study was performed with on-duty clinician resuscitation teams and not exclusively with trainees. ‘Low functioning teams’ would not be permitted to function as trauma resuscitation teams in our institution. The inclusion of an extremely low- or high-performing team would have increased the spread of the scale and perhaps increased the expert rater reliability of the assessment and the ability to correlate across a wider range of scores. Future studies that aim to study discriminant validity of this assessment will need to intentionally include novice teams to adequately assess validity.

In conclusion, assessment of team function continues to be a challenge. Expert assessment and self-assessment cannot be used interchangeably with the T-NOTECHS instrument for pediatric trauma resuscitation. It is unclear from these results which assessment closer reflects the ‘truth’. Certain aspects of team performance may be better assessed by members within the team, while others may be better assessed by independent observers. The T-NOTECHS instrument may benefit from further refinement – potentially using a mixture of self- and expert assessments for differing domains to increase the accuracy. Criterion-related validation studies are necessary to investigate if this hybrid approach would improve accuracy.

Acknowledgments

Funding

This work was supported by the 2014 Children’s Hospital Los Angeles Barbara M. Korsch Award for Research in Medical Education and SC CTSI (NIH/NCRR/NCATS) Grants KL2TR001854 and UL1TR000130.

References

1. Kohn L. To err is human: an interview with the Institute of Medicine’s Linda Kohn. *Jt Comm J Qual Improv* 2000;26(4):227–34. [PubMed: 10749007]
2. Gawande AA, Zinner MJ, Studdert DM, Brennan TA. Analysis of errors reported by surgeons at three teaching hospitals. *Surgery* 2003;133(6):614–21. [PubMed: 12796727]
3. Scott DJ, Dunnington GL. The new ACS/APDS Skills Curriculum: moving the learning curve out of the operating room. *J Gastrointest Surg* 2008;12(2):213–21. [PubMed: 17926105]
4. Knudson MM, Khaw L, Bullard MK, et al. Trauma training in simulation: translating skills from SIM time to real time. *J Trauma* 2008;64(2):255–63; discussion 63–4. [PubMed: 18301184]
5. Holcomb JB, Dumire RD, Crommett JW, et al. Evaluation of trauma team performance using an advanced human patient simulator for resuscitation training. *J Trauma* 2002;52(6):1078–85; discussion 85–6. [PubMed: 12045633]
6. Falcone RA Jr., Daugherty M, Schweer L, et al. Multidisciplinary pediatric trauma team training using high-fidelity trauma simulation. *J Pediatric Surg* 2008;43(6):1065–71.
7. Capella J, Smith S, Philp A, et al. Teamwork Training Improves the Clinical Care of Trauma Patients. *J Surg Educ* 2010;67(6):439–43. [PubMed: 21156305]
8. Steinemann S, Berg B, DiTullio A, et al. Assessing teamwork in the trauma bay: introduction of a modified “NOTECHS” scale for trauma. *Am J Surg* 2012;203(1):69–75. [PubMed: 22172484]
9. Devine LA, Donkers J, Brydges R, et al. An Equivalence Trial Comparing Instructor-Regulated With Directed Self-Regulated Mastery Learning of Advanced Cardiac Life Support Skills. *Simul Healthc* 2015;10(4):202–9. [PubMed: 26154249]

10. Arora S, Miskovic D, Hull L, et al. Self vs expert assessment of technical and non-technical skills in high fidelity simulation. *Am J Surg* 2011;202(4):500–6. [PubMed: 21943950]
11. Moorthy K, Munz Y, Adams S, et al. Self-assessment of performance among surgical trainees during simulated procedures in a simulated operating theater. *Am J Surg* 2006;192(1):114–8. [PubMed: 16769287]
12. Davis DA, Mazmanian PE, Fordis M, et al. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA* 2006;296(9):1094–102. [PubMed: 16954489]
13. Sevdalis N, Lyons M, Healey AN, et al. Observational teamwork assessment for surgery: construct validation with expert versus novice raters. *Ann Surg* 2009;249(6):1047–51. [PubMed: 19474694]
14. Nick S, Davis R, Koutantji M, et al. Reliability of a revised NOTECHS scale for use in surgical teams. *Am J Surg* 2008;196(2):184–90. [PubMed: 18558392]
15. Rudolph JW, Simon R, Raemer DB, Eppich WJ. Debriefing as formative assessment: closing performance gaps in medical education. *Acad Emerg Med* 2008;15(11):1010–6. [PubMed: 18945231]
16. Rudolph JW, Simon R, Dufresne RL, Raemer DB. There's no such thing as "nonjudgmental" debriefing: a theory and method for debriefing with good judgment. *Simul Healthc* 2006;1(1):49–55. [PubMed: 19088574]
17. Cicchetti D, Allison T. A new procedure for assessing reliability of scoring eeg sleep recordings. *Am J EEG Technol* 1971;11:101–9.
18. Landis J, Koch G. The measurement of observer agreement for categorical data. *Biometrics* 1977;33(1):159–74. [PubMed: 843571]
19. Lim YS, Steinemann S, Berg BW. Team size impact on assessment of teamwork in simulation-based trauma team training. *Hawaii J Med Public Health* 2014;73(11):358–61. [PubMed: 25414806]
20. Steinemann S, Kurosawa G, Wei A, et al. Role confusion and self-assessment in interprofessional trauma teams. *Am J Surg* 2016;211(2):482–8. [PubMed: 26801092]
21. Robertson ER, Hadi M, Morgan LJ, et al. Oxford NOTECHS II: a modified theatre team non-technical skills scoring system. *PLOS ONE* 2014;9(3):e90320.

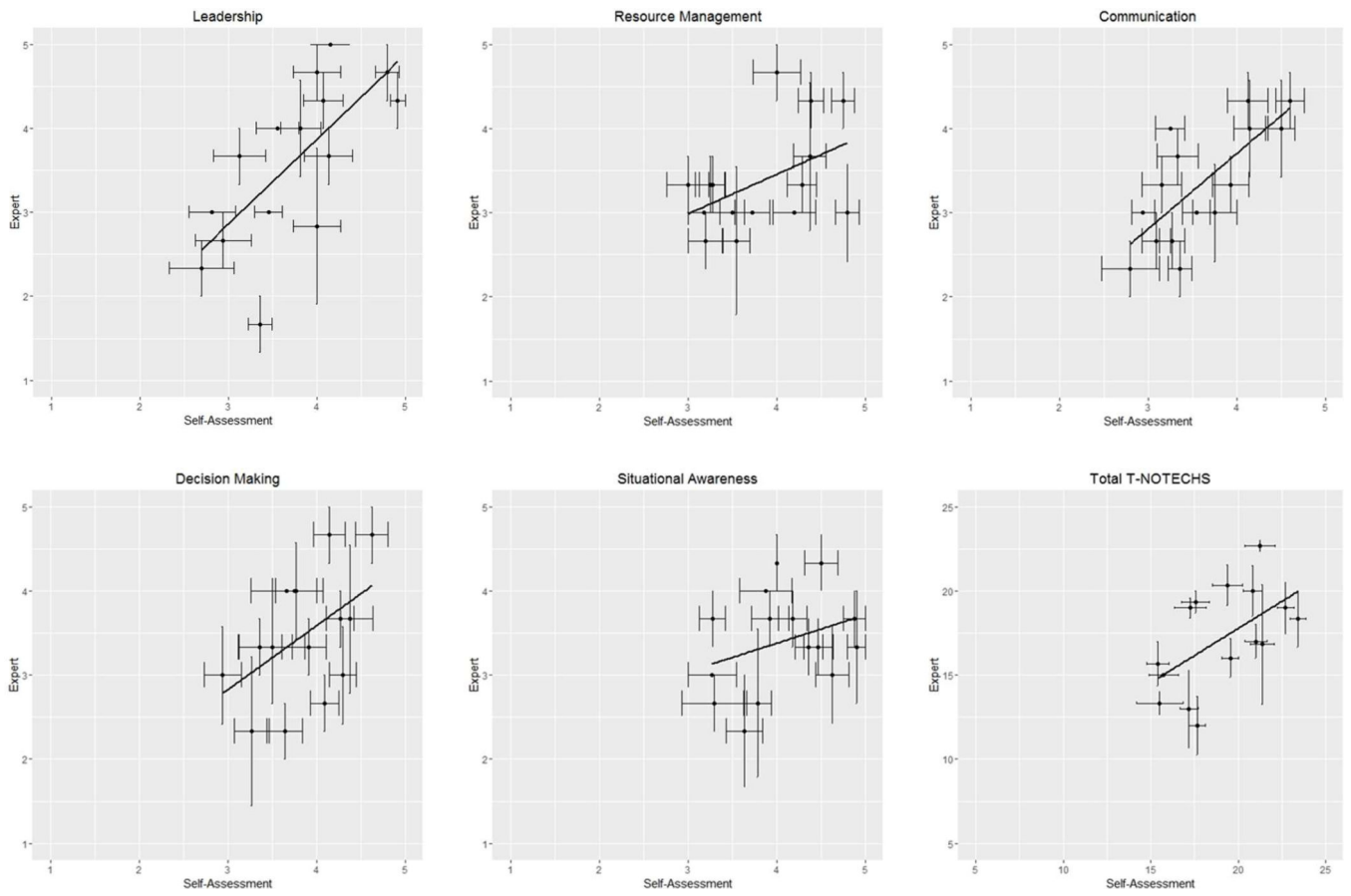


Figure 1: Scatterplot of mean scores comparing expert and team self-assessment for each of the five T-NOTECHS domains and total composite score (N=15 teams assessed). Error bars represent standard error for expert raters (vertical) and team member self-assessment (horizontal).

Table 1:

Weighted kappa and intra-class correlation for T-NOTECHS assessment within expert and team self-assessment. Weighted kappa reported as kappa score (95% confidence interval).

	Expert	Self-Assessment
Domains	Weighted Kappa	Weighted Kappa
Leadership	0.48 (0.33, 0.64)	0.26 (0.21, 0.30)
Resource Management	0.20 (0.03, 0.37)	0.31 (0.27, 0.36)
Communication	0.27 (0.10, 0.44)	0.24 (0.20, 0.29)
Decision Making	0.17 (-0.02, 0.35)	0.10 (0.05, 0.14)
Situational Awareness	0.10 (-0.05, 0.26)	0.25 (0.20, 0.30)
	Intra-Class Correlation	Intra-Class Correlation
Total Composite T-NOTECHS	0.57	0.56

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Comparison of T-NOTECHS rating between expert and team self-assessment. LS Mean = Least squares mean, SE = Standard Error.

Domains	Expert		Self-Assessment		p-value
	LS Mean	SE	LS Mean	SE	
Leadership	3.59	0.21	3.73	0.18	0.28
Resource Management	3.38	0.17	3.85	0.14	<0.01
Communication	3.33	0.17	3.60	0.15	0.02
Decision Making	3.47	0.16	3.85	0.12	<0.01
Situational Awareness	3.40	0.16	4.07	0.13	<0.01
Total Composite T-NOTECHS	17.17	0.72	19.10	0.65	<0.01

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript