Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.JournalofSurgicalResearch.com

Association for Academic Surgery

A comparative assessment and gap analysis of commonly used team rating scales

Carla M. Pugh, MD, PhD,[a,*] Elaine R. Cohen, MEd,[a] Calvin Kwan, BS,[a] and Janice A. Cannon-Bowers, PhD[b]

[a] Department of Surgery, University of Wisconsin, Madison, Wisconsin
[b] Institute for Simulation and Training, University of Central Florida, Orlando, Florida

ARTICLE INFO

Article history:
Received 5 January 2014
Received in revised form
4 April 2014
Accepted 22 April 2014
Available online 28 April 2014

Keywords:
Rating scale
Teamwork
Taskwork
Patient safety
Medical education
Performance assessment

ABSTRACT

Background: The purpose of this article was to conduct a gap analysis of important team constructs that may be absent in widely used team assessments.

Methods and materials: Two assessment tools with known validity evidence (1) Non-Technical Skills for Surgeons (NOTSS) and (2) the Cannon-Bowers Scale were used to evaluate 11 teams of surgical residents ($n = 33$) performing simulated laparoscopic hernia repairs. Faculty raters' scores were used to compare the surveys and assess validity and reliability. Raters' detailed observation notes were used to indicate important behavioral constructs that were missing from the team rating scales.

Results: When assessing inter-item correlations (reliability) four of five NOTSS' scale items had significant correlations ($r = 0.9–1.0$, $P < 0.05$) with the Cannon-Bowers items. While the correlations were only noted for three of six Cannon-Bowers items, in each instance the same four of five NOTSS items correlated with the three Cannon-Bowers items, thus providing further validity evidence for both scales. When evaluating the gap, key emerging themes included the need to focus on critical team errors, individual team member contributions, task performance, and overall team performance. These gaps, plus items from the NOTSS and Cannon-Bowers scales, were incorporated into a new rating scale.

Conclusions: Despite continued evidence of validity and reliability, there were several behavioral constructs that were not represented when using the NOTSS and Cannon-Bowers scales. Critical team errors, individual team member contributions, task performance, and overall team performance appear important in our ability to understand teams and teamwork.

1. Introduction

Teamwork has become a major topic of importance regarding health care delivery and patient safety in the medical field. However, despite an increase in the number of team training curricula, teamwork assessments are still being researched and developed. In 1987, Glickman *et al.* [1] published a highly cited article on teamwork skills. The authors hypothesize that teams progress through several stages of development during the course of training, and that the time and sequence of these stages are affected by the efficacy of the training program. In addition, the authors propose that there are two discrete developmental tracks: (1) a "taskwork" track and (2) a "teamwork" track. In order for team training to be successful, these

* *Corresponding author*. Department of Surgery, University of Wisconsin, 600 Highland Ave, H4 785B Clinical Science Center, Madison, WI 53792. Tel.: +1 (608) 262 2147; fax: (608) 252-0936.
E-mail address: pugh@surgery.wisc.edu (C.M. Pugh).
0022-4804/$ – see front matter © 2014 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jss.2014.04.034

two tracks must be separately developed and ultimately united [1]. Specifically, Glickman's study recommended that teamwork skills such as cooperation and coordination be trained in the early phases of team training. In subsequent training phases, taskwork competencies such as declarative knowledge, self-regulatory skills, and individual performance strategies serve as a foundation for continued development of teamwork competence.

In a review article on team-based performance assessments, the authors discussed the validity and reliability of three measurement tools currently used to evaluate surgical subteams: (1) Non-Technical Skills for Surgeons (NOTSS), (2) Non-Technical Skills, and (3) Observational Teamwork Assessment for Surgery [2]. A commonality among these three assessments is a focus on teamwork variables including (a) communication, (b) situation awareness, (c) leadership, (d) decision making, and (e) cooperation [3–5]. When evaluating these measurement tools, we found two versions of the NOTSS tool [2,3]. One version included a taskwork variable "task management." However, in the later versions, this variable was removed because of low inter-rater reliability and low sensitivity [3]. Although this may have been due to rater training, it may also relate to how the variable was defined.

To obtain a more global view of operative performance for surgical subteams, assessment tools must be explicit in separating teamwork variables (i.e., communication, leadership, decision making, cooperation, and coordination) from taskwork variables (i.e., task knowledge, task performance, motor skills, and self-awareness) [1]. Moreover, there needs to be some consensus on how these variables are operationalized. Task management theory can help in developing assessments that include both "teamwork" and "taskwork" items. In 1991, Funk *et al.* [6] described cockpit task management as a separate and measurable entity contrasted with cockpit resource management. Task management was described as the process that flight crews use to initiate, monitor, prioritize, execute, and terminate multiple concurrent tasks.

The aim of this study was to conduct an in-depth analysis of surgical subteam behaviors and actions using two previously developed surveys with documented validity evidence: (1) NOTSS [7] and (2) Cannon-Bowers [8]. The Cannon-Bowers survey was chosen because this work introduced variables that were not covered by the other surveys. These variables include (a) affect and attitude management, (b) motivation building, and (c) adaptability. Our goal was to evaluate the gap between behavioral constructs represented in the two previously developed surveys and the constructs noted by trained observers reviewing operative performance. Team performance was evaluated during a decision-based simulation for surgical residents.

2. Methods

2.1. Study design

This study used a comparative assessment and gap analysis design [10]. Data for the analysis were captured using qualitative research methods (structured observations [11,12] and deductive reasoning [12]), and quantitative measures (survey-based data). The purpose of this study was to (1) compare how two previously developed assessment tools categorize team performance and (2) document the gaps in currently used assessments when evaluating both team and taskwork.

2.2. Setting and procedure

Data were collected during once-a-year formative assessment, over a 3-y period from September 2008 to January 2012 at an academic training program. In the first year, five three-member teams participated. In the third year, six three-member teams participated. The teams were stratified in such a way that junior (postgraduate year 1–3) and senior (postgraduate year 4–5) participants were placed into groups in a randomized fashion. Final team composition included two junior residents and one senior resident. This activity was a one-time team assessment for all groups. The final study group included 11 three-member teams of surgery residents ($n = 33$) performing a simulated laparoscopic hernia repair in a skills laboratory setting. Each team was tasked with repairing a 10 × 10-cm right upper quadrant ventral hernia within 30 min. This task required team members to work together to execute important intraoperative decisions, including surgical planning and technical approach. There were five steps in the laparoscopic ventral hernia (LVH) operation: (1) port insertion, (2) reduction of incarcerated omentum, (3) hernia defect measurement, (4) mesh preparation, and (5) mesh attachment. All simulation sessions were videotaped for further review. Both internal views from the laparoscope and external views from an outside camera were captured. The external camera also captured audio, which was used as part of the analysis.

2.3. Approach

Two previously developed assessment tools currently being used the operating room environment, (1) NOTSS [7] and (2) the Cannon-Bowers Scale [8], were used to evaluate the LVH team performance videos from the decision-based simulation sessions [13]. In preparation for this study, we reviewed several theoretical propositions regarding teamwork and task performance [1–9,14,15].

The NOTSS behavior rating system (tool) was developed to assess intraoperative team behaviors. The tool is in surgical language for suitably trained surgeons to observe, rate, and give feedback on nontechnical skills in a structured manner. One version of the skills assessed includes communication and teamwork, decision making, situation awareness, and leadership [3]. Another version used task management instead of leadership. For this study, we used both leadership and task management [2]. The NOTSS system uses a four-point rating scale: 4 = good, 3 = acceptable, 2 = marginal, 1 = poor, and N/A = not applicable (skill not required or expected for given clinical situation).

The Cannon-Bowers scale was developed based on a newly described theory of shared mental models [8]. The Cannon-Bowers scale contains six global performance items based on the shared mental models framework. Performance items include planning, monitoring, affect management, motivation building, adaptability, and shared mental models. This team assessment tool is rated on a five-point scale: 1 = not