Observational Teamwork Assessment for Surgery: Content Validation and Tool Refinement

Louise Hull, MSc, Sonal Arora, PhD, Eva Kassab, MSc, Roger Kneebone, PhD, Nick Sevdalis, PhD

BACKGROUND: Effective teamwork is crucial for safe surgery. Failures in nontechnical and teamwork skills are frequently implicated in adverse events. The Observational Teamwork Assessment for Surgery (OTAS) tool assesses teamwork of the entire team in the operating room. Empirical testing of OTAS has yet to explore the content validity of the tool.

STUDY DESIGN: This was a cross-sectional observational study. Data were collected in 30 procedures by 2 trained researchers. Five teamwork behaviors were scored (ie, communication, leadership, cooperation, coordination, and monitoring) and behavior exemplar completion was recorded (phase 1). Expert operating room personnel (5 surgeons, 5 anesthesiologists, and 5 scrub nurses) assessed the content validity of the OTAS exemplar behaviors. Finally, a panel of operating room patient-safety experts refined the exemplars (phase 2).

RESULTS: In total, the observability (presence/absence) of 130 exemplars was assessed by 2 blinded observers in 30 general surgical cases. Observer agreement was high (Cohen's $\kappa \geq 0.41$) for 83.85% (109 of 130) of exemplar behaviors; 60.77% (79 of 130) of exemplar behaviors were observed frequently with high observer agreement. The majority of the exemplars were rated by expert operating room practitioners and an expert panel as substantial contributors to teamwork and patient safety. Based on expert consensus, 21 behavior exemplars were removed from OTAS and an additional 23 were modified.

CONCLUSIONS: The exemplars of OTAS demonstrated very good content validity. Taken together with recent evidence on the construct validity of the tool, these findings demonstrate that OTAS is psychometrically robust for capturing teamwork in the operating room. (J Am Coll Surg 2011;212: 234–243. © 2010 by the American College of Surgeons)

Analyses of adverse events in health care have altered the way in which surgical performance is understood and analyzed.[1,2] Empirical evidence supports the notion that although technical expertise is essential, it is not sufficient to consistently achieve high levels of surgical performance.[1] Nontechnical skills reflect the interpersonal (eg, communication, teamwork, and leadership) and cognitive skills (ie, decision-making, situational awareness, and mental readiness) that complement clinician's technical skills. Failures in nontechnical skills in the operating room (OR) have been frequently implicated in adverse events in surgical patients.[3-6] In contrast, superior teamwork is associated with fewer errors in the OR.[7,8]

Despite the impact of teamwork on technical performance and clinical outcomes,[8-10] OR team training has yet to be explicitly addressed in the surgical curriculum. This is especially important because nontechnical skills do require training.[11-14] The high frequency of teamwork failures in the OR, which have been documented in the literature,[9] emphasize that the informal and unstructured manner in which these skills are assumed to be acquired is ineffective.

To effectively provide surgical teams with the nontechnical skills required to achieve and maintain an optimal environment for surgical safety, it is essential to have a robust tool to measure these skills in the OR. Such a tool is necessary to benchmark good teamwork skills, and to guide formative feedback and debriefing in clinical practice. It can also form an integral element of simulation-based training, such as the training required by the new American College of Surgeons/Association of Program Directors in

Disclosure Information: Nothing to disclose.

This research was funded by the UK's National Institute for Health Research (NIHR), and a joint Economic and Social Research Council/Medical Research Council grant.

Received September 10, 2010; Revised November 1, 2010; Accepted November 2, 2010.

From the Department of Surgery and Cancer, Imperial College London, UK. Correspondence address: Louise Hull, MSc, Department of Surgery and Cancer, Imperial College London, 5th Floor, Wright Fleming Building, Norfolk Place, London, W2 1PG UK. email: l.hull@imperial.ac.uk

ISSN 1072-7515/11/$36.00
doi:10.1016/j.jamcollsurg.2010.11.001

Abbreviations and Acronyms
CVM = content validity metric
OR = operating room
OTAS = Observational Teamwork Assessment for Surgery

Surgery phase 3 curriculum.[15] In addition, with an increasing number of interventions being developed to improve teamwork and patient safety in the OR,[16-19] it is essential that such a tool be available to robustly evaluate the effectiveness of such interventions.

The Observational Teamwork Assessment for Surgery (OTAS) aims to capture teamwork in the OR comprehensively, in a reliable and valid manner (ie, psychometrically robust).[20,21] OTAS distinguishes between different subteams within the OR (ie, surgeons, anesthesiologists, nurses) and different stages of a procedure (ie, pre-, intra-, and postoperative). Quality of teamwork is assessed through direct real-time observation for each subteam separately across the 3 operative stages. Five behaviors are assessed in the OTAS tool on 0 to 6-point scales: communication, leadership, cooperation, coordination, and team monitoring. To guide the ratings, exemplar behaviors are used ("exemplars" hereafter). These are key observable behaviors that indicate exemplary teamwork and are associated with effective, safe surgical practice. In addition, descriptive scale anchors are provided to observers to assist scoring. Behavior ratings can be summed to provide team performance scores. Existing evidence shows that OTAS is feasible for use in general and urological surgery,[20,21] the scoring is reliable between expert raters[21] and construct valid.[13]

The present study focuses on the content validity of the exemplars used to guide the scoring of the 5 teamwork behaviors assessed in the OTAS tool. These exemplars act as behavioral markers developed to guide the assessor in what to look for to give a specific rating. OTAS contains 130 such exemplars. These exemplars were originally derived from ethnographic field notes after OTAS use in ORs,[20,21] existing literature and best practice guidelines, and documentation for ORs,[20] as well as input from expert OR personnel who took part in the tool development process.[22] An example of a communication exemplar for surgeons is, "asks team if all prepared to begin the operation" (Fig. 1). Consistent presence of these exemplar behaviors allows an observer to rate highly the quality of communication in an OR team. In contrast, if these exemplars are not observed or are carried out in an inconsistent manner, the rating for the quality of surgical team communication is likely to be lower. Although the provenance of these exemplars would suggest that they are indeed valid for team and nontechnical skill assessment in the OR, to date this remains to be empirically proven.

The aim of the present study was to assess quantitatively the content validity of these exemplars through a phased, multimethod investigation. For OTAS to be content-valid, the exemplars available to observers should be observable in the OR and blinded observers should agree that these exemplars have or have not occurred during a case. In addition, the exemplars should be empirically demonstrated to capture comprehensively the range of behaviors likely to occur within an OR.[23] In the present study, observability of the exemplars was first assessed through quantitative real-time OR observation. Then, inter-rater agreement between 2 blinded observers in capturing these exemplar behaviors was assessed. Finally, exemplars in need of additional refinement were submitted to a process of refinement through expert consensus. The following hypotheses were tested:

Hypothesis 1: Blinded observers will demonstrate sizeable and substantial interobserver agreement as to whether an exemplar is present/absent during a case.

Hypothesis 2: Exemplars will exhibit sizeable content validity as assessed through expert consensus (Fig. 1).

METHODS

A 2-phase approach was used. Phase 1 was an observational study that assessed exemplar observability and interobserver agreement (hypothesis 1). Phase 2 involved an expert consensus process to refine and validate exemplars (hypothesis 2).

Phase 1: exemplar observability in the OR

Design and case sample

This was a prospective, cross-sectional, observational study. Two blinded observers with a background in psychology (LH, EK) recorded the presence/absence of exemplars during surgery in real time and scored the 5 OTAS behaviors (ie, communication, leadership, cooperation, coordination, and team monitoring). Both observers were previously trained and calibrated in using OTAS (having carried out 70 case observations each in real and simulated ORs at the time of the study). Observations were carried out in real time in ORs of a London teaching hospital. Data were collected from 30 general surgical procedures (14 open, 16 laparoscopic) that lasted between 30 and 240 minutes. Typical procedures included laparoscopic cholecystectomies and open/laparoscopic hernia repairs. Ethical approval and informed consent were obtained.

Procedure

The 2 observers were provided with 130 OTAS exemplars in the form of a list. These covered the 5 OTAS behaviors,