Wong and colleagues report on the development of a classification system for surgical feedback. This study aimed to quantify intraoperative feedback during live surgical cases to create a reliable, deconstructed feedback framework. Using a mixed-methods analysis, the study examined audio and video recordings of robotic prostatectomies involving trainees, and an iterative coding process was performed using combined data to generate recurring themes. Faculty and resident interactions were ultimately coded into 3 types: triggers, feedback, and responses. This 3-category feedback framework was evaluated for its reliability, generalizability, and utility. In a series of 29 robotic prostatectomies involving 4 attending surgeons, 11 trainees, and 371 instances of feedback, 3 trained raters achieved a moderate to substantial interrater reliability (κ range, 0.6-1.0) in coding cases using different types of triggers, feedback, and responses. Furthermore, the authors found that significant differences in these domains reflected the surgical task being performed and experience level of the trainee. Combinations of certain forms of feedback were associated with different responses from trainees. The clear and rigorous methodology used by the authors in this study should be commended, and the large effort involved in development of this feedback framework and classification system acknowledged.

The authors sought to define the fundamental components of surgical feedback, which is ubiquitous but challenging to objectively measure and translate to intervention. This framework is compelling given that it underlines the importance of real-time feedback and acknowledges the complexity of surgery and intimate role that direct teaching plays in further development of surgical skills and problem-solving capabilities. The framework additionally acknowledges the critical role that communication and team relationships play within the operating room. This classification system may provide a uniform language for the analysis of surgical feedback, help bring standardization to feedback and coaching, and facilitate research into exploring, developing, and adopting more effective means of teaching surgical trainees.

This work suggests numerous implications and future directions. Full integration of this classification system across multiple training platforms, including simulation-based training, may allow for continuous improvement and refinement of interaction categories and additional testing of best combinations of triggers and feedback associated with appropriate learning responses. Multiplatform adoption and integration may also allow expanded solicitation and evaluation of nonverbal interactions, reflections from trainers and trainees on the quality of these interactions, and further subcategorization of different combinations of feedback. Moreover, with the increasing use of artificial intelligence (AI) in surgical education, especially in endoscopy and robotics, this feedback framework may serve as a foundation for developing a more streamlined, multifactorial, computer-based evaluation instrument. In addition, the proposed classification system may be used to train a deep-learning algorithm to perform automatic classification to deconstruct feedback into component parts. With additional data and further analysis, a future AI system may provide real-time feedback designed to generate high-impact learning responses to specific feedback triggers, ultimately associated with significantly improved training and patient outcomes. Timely facilitation of feedback and learning responses has great potential to democratize expertise and knowledge transfer for everyone involved in the process of surgical training.

Future work must also address limitations of this study. First, the generalizability of the framework needs to be confirmed given that it was developed using a limited number of cases of robotic prostatectomy performed in a single institution by a handful of personnel. Second, the
difficulty in quantifying confounding variables, such as preexisting mentoring relationships and interpersonal dynamics, needs to be acknowledged and empirically investigated to determine how this may be associated with classification system outcomes. Third, tracking the association of longitudinal feedback and feedback provided outside of the operating room with surgical education outcomes will likely enhance performance of this classification system and may aid in the creation of optimal feedback mechanisms.  

This development of a classification system is an important topic for surgical educators and represents a significant, foundational step forward in surgical education. The potential for standardization of surgical feedback using this classification system may open the door to further objective evaluation of surgical feedback, development of improved systems for feedback, and potential integration with other surgical technologies and approaches, including AI and coaching. We look forward to future work exploring this important and foundational topic in surgical education.

ARTICLE INFORMATION

Open Access: This is an open access article distributed under the terms of the CC-BY License. © 2023 Shkolyar E et al. JAMA Network Open.

Corresponding Author: Joseph C. Liao, MD, Department of Urology, Stanford University School of Medicine, 453 Quarry Rd, Mail Code 5656, Palo Alto, CA 94304 (jliao@stanford.edu).

Author Affiliations: Department of Urology, Stanford University School of Medicine, Stanford, California (Shkolyar, Liao); Department of Surgery, Stanford University School of Medicine, Stanford, California (Pugh).

Conflict of Interest Disclosures: Dr Shkolyar reported receiving personal fees from Johnson & Johnson outside the submitted work. No other disclosures were reported.

REFERENCES