The American Journal of Surgery (2017) 214, 365-372

The American Journal of Surgery®

Strategies for increasing the feasibility of performance assessments during competency-based education: Subjective and objective evaluations correlate in the operating room

Peter Szasz, M.D.[a,*], Marisa Louridas, M.D.[a], Kenneth A. Harris, M.D., F.R.C.S.C., F.A.C.S.[b], Teodor P. Grantcharov, M.D., Ph.D., F.A.C.S.[a]

[a]Department of Surgery, St. Michael's Hospital, Toronto, Canada; [b]Royal College Physicians and Surgeons of Canada, Ottawa, Canada

KEYWORDS:
General surgery;
Residents;
Formative assessment;
Summative assessment;
Competency-based education;
Competency-based assessment

Abstract

BACKGROUND: Competency-based education necessitates assessments that determine whether trainees have acquired specific competencies. The evidence on the ability of internal raters (staff surgeons) to provide accurate assessments is mixed; however, this has not yet been directly explored in the operating room. This study's objective is to compare the ratings given by internal raters vs an expert external rater (independent to the training process) in the operating room.

METHODS: Raters assessed general surgery residents during a laparoscopic cholecystectomy for their technical and nontechnical performance.

RESULTS: Fifteen cases were observed. There was a moderately positive correlation (r_s = .618, P = .014) for technical performance and a strong positive correlation (r_s = .731, P = .002) for nontechnical performance. The internal raters were less stringent for technical (mean rank 3.33 vs 8.64, P = .007) and nontechnical (mean rank 3.83 vs 8.50, P = .01) performances.

CONCLUSIONS: This study provides evidence to help operationalize competency-based assessments.

This work was supported by the Royal College of Physicians and Surgeons of Canada (RCPSC) through the H.S. Morton Exchange Fellowship Fund.

Dr. Grantcharov is a Canadian Institutes of Health Research (CIHR) Tier II Canada Research Chair in Simulation and Surgical Safety. He also holds research grants and receives consultancy fees from Covidien Ltd and Ethicon Inc. Dr. Harris is employed by the RCPSC as the Director of Specialty Education. No other authors have financial interests to disclose. None of the authors have potential conflicts of interest.

* Corresponding author. Tel.: +1-647-207-9660; fax: +1-416-864-5343.
E-mail address: peter.szasz@utoronto.ca

Manuscript received April 26, 2016; revised manuscript June 23, 2016

0002-9610/$ - see front matter © 2016 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.amjsurg.2016.07.017

Competency-based education, an outcome-focused training paradigm, is significantly altering the way in which surgical residents are being trained; with less of a focus on the duration of training and more of a focus on the acquisition and demonstration of specific competencies.[1,2] Although relatively new, it has permeated every medical specialty worldwide, and it is changing not only the way surgical residents are being educated but also assessed.[1–4]

Although formative assessments, which are used regularly for trainee learning, development, and feedback, are at the center of competency-based education, periodic assessments that determine whether residents have acquired specific competencies are also required.[4,5] These summative assessments are used to assess trainee learning at specific intervals to make decisions about that trainee (ie, promotion, remediation).[4] Given that summative assessments can have significant implications for residents, training programs, and licensing authorities, they must be developed, organized, and implemented in a thoughtful manner.[6–8] One area of focus in relation to assessment practices, particularly important for summative assessments, has been who the assessors ought to be.[4,9]

There is an emerging area of literature that evaluates the utility of staff physician assessments (deemed internal assessments for the remainder of the manuscript) for both technical and nontechnical resident performance.[10,11] In particular, the literature is mixed as to the accuracy and reliability of internal assessments, compared with other more "objective" measures of trainee performance (ie, standardized assessments, licensing examinations).[11–16] Despite comparisons to other such "objective" measures of performance, a comparison between internal and external assessors (those individuals who are independent to the training process of the resident being assessed)—while often if only anecdotally presumed superior, is lacking. Potential reasons against using internal assessors for summative-type assessments include their knowledge of a trainee and previous trainee encounters, and the natural conflict of interest that exists between an internal assessor and trainee (the internal assessor is intimately tied to and responsible for training a resident—therefore they are not only assessing trainee performance but also their ability to teach).[17] In contrast, potential reasons for using internal assessors for summative type assessments within competency-based education includes their ability to address and possibly minimize some of the logistical issues inherent to such assessment frameworks (ie, cost).[18,19]

A study that directly compares internal and external raters in the operating room has yet to be completed. Therefore, the objective of this study is to compare the ratings given by internal raters vs an external rater in terms of the performance scores they attribute to trainees in the operating room during a laparoscopic cholecystectomy. This will be accomplished by evaluating (1) total score correlations; and (2) mean rank differences, between internal and external rater attributed scores for both technical and nontechnical performance.

Methods

Ethics

The research ethics boards at the University of Toronto, and St. Michael's Hospital approved this study. Informed consent was obtained from all participating residents.

Procedure and assessment instruments

The procedure chosen for this study was the laparoscopic cholecystectomy, based on our earlier research ascertaining procedures for milestone assessments.[20] Resident technical performance was evaluated using the Objective Structured Assessment of Technical Skills (OSATS)[21] global rating instrument, and resident nontechnical performance was evaluated using the Objective Structured Assessment of Nontechnical Skills (OSANTS)[22] global rating instrument. Both of these rating instruments have 7 categories, each ranked on a 5-point likert scale with a maximum score of 35.[21,22] In a traditional sense, the OSATS[21] instrument has previously demonstrated construct validity, inter-rater reliability, and internal consistency and in a more contemporary sense meets many of the criteria within Messick's conceptual framework of validity.[21,23–26] Similarly, the OSANTS[22] instrument has also previously demonstrated construct validity, concurrent validity, inter-rater reliability, and internal consistency and in a more contemporary sense fits many of the criteria within Messick's conceptual framework of validity.[22,24–26]

Participants

The resident participants were general surgery trainees at the University of Toronto, completing a rotation at St. Michael's Hospital.

The internal raters were all board certified General Surgeons at St. Michael's Hospital with experience in minimally invasive surgery. They are deemed internal as they are intimately tied to the training of the residents they are assessing, often having previous knowledge of and experience with these residents. These surgeons had some previous informal experience assessing trainee technical performance and little to no experience assessing trainee nontechnical performance. Prior to any assessments taking place, each internal rater underwent formalized training. A combination of the Performance Dimension Training and Rater Error Training strategies were used.[27] During a 1-hour tutorial, the specific constructs under study for each rating scale item on the OSATS[21] and OSANTS[22] rating instruments were explained and discussed, and examples of good and bad markers of performance were described for each.[27] Furthermore, the internal raters attention was drawn to rating errors and again this was discussed.[27] Finally, after the