The Generic Error Rating Tool: A Novel Approach to Assessment of Performance and Surgical Education in Gynecologic Laparoscopy ☆

Heinrich Husslein, MD,*,† Lindsay Shirreff, MD,* Eliane M. Shore, MD, MSc,*
Guylaine G. Lefebvre, MD,* and Teodor P. Grantcharov, MD, PhD‡

*Division of Obstetrics and Gynaecology, St. Michael's Hospital, University of Toronto, Toronto, Ontario, Canada; †Department of Obstetrics and Gynecology, Medical University Vienna, Vienna, Austria; and ‡Division of General Surgery, St. Michael's Hospital, University of Toronto, Toronto, Ontario, Canada

OBJECTIVE: Global rating scales are commonly used to rate surgeons' skill level. However, these tools lack granularity required for specific skill feedback. Recently, an alternative framework has been developed that is designed to measure technical errors during laparoscopy. The purpose of the present study was to gather validity evidence for the Generic Error Rating Tool (GERT) in gynecologic laparoscopy.

DESIGN: Video recordings of total laparoscopic hysterectomies were analyzed by 2 blinded reviewers using the GERT and the Objective Structured Assessment of Technical Skills (OSATS) scale. Several sources of validity were examined according to the unitary framework of validity. Main outcomes were interrater and intrarater reliability regarding total number of errors and events. Further, surgeons were grouped according to OSATS scores (OSATS ≥ 28 = high performers and OSATS < 28 = low performers), and the number of errors and events was compared between groups. Correlation analysis between GERT and OSATS scores was performed. Lastly, error distribution within procedure steps was explored and compared between high- and low-performing surgeons.

SETTING: University teaching hospital.

PARTICIPANTS: A total of 20 anonymized video recordings of total laparoscopic hysterectomies.

RESULTS: Interrater and intrarater reliability was high (intraclass correlation coefficient > 0.95) for total number of errors and events. Low performers made significantly more errors than high performers did (median = 49.5 [interquartile range: 34.5-66] vs median = 31 [interquartile range: 16.75-35.25], p = 0.002). There was a significant negative correlation between individual OSATS scores and total number of errors (Spearman $\rho = -0.76$, p < 0.001, and $\rho = -0.88$, p < 0.001, for raters 1 and 2, respectively). Error distribution varied between operative steps, and low performers made more errors in some steps, but not in others.

CONCLUSION: GERT allows for objective and reproducible assessment of technical errors during gynecologic laparoscopy and could be used for performance analysis and personalized surgical education and training. (J Surg 72:1259-1265. © 2015 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: error analysis, gynecology, laparoscopy, surgical education, technical error

COMPETENCIES: Practice-Based Learning and Improvement, Professionalism, Systems-Based Practice

INTRODUCTION

Laparoscopy requires psychomotor skills that may be difficult to learn and result in prolonged learning curves.[1-3] The increased complexity in specific skills acquisition has motivated refinements in surgical education and resulted in the development of several structured surgical courses and educational curricula over the last decades.[4-6]

To assess our progress in technical skill acquisition, we are now looking to objectify measures that document

performance. Video recordings offer a valuable opportunity for structured assessment and subsequent tailored coaching interventions. In-depth video analysis of technical skills is commonly used by athletes with the aim to enhance performance; however, this highly effective concept remains underused in medicine and surgery.[7,8] One reason for this may be that, to conduct meaningful feedback, we first need objective methods of performance analysis, which allow the detection of specific weaknesses that are to be addressed.

With an increased focus on surgical education, several global rating scales to assess surgical skill were developed.[9,10] Global rating scales have thereafter commonly been used to measure the effect of educational interventions and benchmark technical competency.[11] However, a common problem of most global rating scales is that they lack the granularity required for task- and skill-specific feedback. Therefore, there may also be a benefit in assessing technical surgical errors and injury mechanisms.[12] Studies investigating technical errors in laparoscopy are limited, and widespread implementation of error analysis as an educational tool or as a method of quality control is lacking.[13-15]

Recently, a new framework has been developed that was designed to measure technical errors during laparoscopy. The Generic Error Rating Tool (GERT) has been trialed in the context of laparoscopic Roux-en-Y gastric bypass procedures, satisfying multiple sources of validity.[12] The tool is intuitive to use and deemed applicable to any laparoscopic procedure. In contrast to global skill rating frameworks, analysis with the GERT assesses every technical error that occurs during a laparoscopic procedure and therefore allows an in-depth analysis of the surgeon's specific skills and respective weaknesses. It was suggested that GERT could be used for formative feedback in surgical education and may be a valuable tool for self-improvement. The aim of this study was to gather validity evidence for the GERT in gynecologic laparoscopic surgery.

METHODS

This study was performed at a tertiary teaching institution from May 2013 to January 2014 and received ethics approval by the local research ethics board (Number, 12-032).

Video Recordings

For the present analysis we used unedited and anonymized recordings of total laparoscopic hysterectomies. The recordings had been retained from a previous study conducted at the institution (not yet published) and had been collected in a prospective fashion. The entire procedures, starting with the introduction of the laparoscope and ending after completion of vaginal vault closure, had been recorded using the laparoscopic camera. A single standard operative technique had been used to perform all hysterectomies. The

procedures had been performed by staff surgeons ($n = 6$), minimally invasive surgical fellows ($n = 2$), or postgraduate year 4 and postgraduate year 5 residents ($n = 6$), all of whom had given written informed consent. Each procedure had been performed in its entirety by the same surgeon (i.e., the primary surgeon did not change within the procedure). As the videos were subsequently anonymized and reposited in an educational video archive, the identity and training level of the primary surgeon of each video was unknown during the present study. The reviewers in the current study were therefore blinded to the training level of the primary surgeon, avoiding introduction of bias at the time of error and skill rating. Skill level was determined by objective measurements, as obtained using the Objective Structured Analysis of Technical Skill (OSATS) scale.[9]

Error Analysis

The GERT is designed to capture and analyze technical errors and resulting events during laparoscopic procedures. Technical errors represent single actions (e.g., introduction of an instrument and overshooting, using energy without adequate visualization, and avulsing tissue when grasping) and not a sequence of actions, which might be specific to a single surgical procedure or surgical specialty. For the purpose of comparable analysis we defined technical error as "the failure of planned actions to achieve their desired goal"[16] and an event as "an action that may require additional measures to avoid an adverse outcome."[17]

Technical error analysis using GERT comprises 9 generic surgical tasks during which errors can occur: (1) abdominal access and removal of instruments or trocars; (2) use of retractors; (3) use of energy; (4) grasping and dissection; (5) cutting, transection, and stapling; (6) clipping; (7) suturing; (8) use of suction; and (9) other.[12] Each of these generic task groups is subdivided into 4 distinct error modes: (1) too much use of force or distance, (2) too little use of force or distance, (3) inadequate visualization, and (4) wrong orientation of instrument.[12] Technical errors and resulting events are marked in the GERT checklist (e-component) under the adequate surgical task group and error mode, according to the time they appear in the video material (time stamped). Every error is counted irrespective of perceived seriousness or the absence or presence of sequelae (e.g., bleeding and tissue damage). At the end of a procedure, the number of technical errors in each task group as well as the total number of errors are summed up.

Rater Training

Before video analysis, 2 reviewers were orientated to the tool by analyzing 5 laparoscopic hysterectomies together with an expert user of the GERT (trainer) as well as a faculty member of the department of minimally invasive gynecologic surgery. The procedures were watched as a group, and technical errors