Developing Metrics for Simulated Surgical Training by Proficiency-Based Progression

P. Guy, MD*; Matt L. Graves, MD; Kevin Inkpen, MSc; Michael R. Baumgaertner; Anthony G. Gallagher, PhD*
University of British Columbia, Vancouver, British Columbia, CANADA

Purpose: Traditional assessment of surgical teaching based on years of training is not reflective of individual learner proficiency. With increased demand to quantify learning and skill level (proficiency), the interest for standardized measures has grown. Likert-type assessment tools are prone to subjectivity. We propose to use a method that is based on a binomial (yes/no) rating of simulated surgical tasks based on Proficiency-Based Progression (PBP). The poster describes the metric development and validation.

Methods: In PBP, the development of objective metrics of surgical task performance involves: (1) procedure identification, (2) tasks analysis, (3) operational metrics definition and verification, (4) metric validation by a Delphi panel (face validity), and (5) measurement of construct validity, inter-rater reliability (IRR), and responsiveness of task metrics. We added (6) measurement of baseline demographic, clinical experience, and visuospatial ability. We developed metrics for the performance of surgical tasks at 4 skill stations and 1 clinical simulation laboratory (fibula fracture fixation in an ankle fracture model) used for training of orthopaedic surgery residents attending an AO Basic Principles Course. The investigator team completed steps 1, 2, and 3. A Delphi panel of experienced surgeon-educators approved the measures, (step 4). Steps 5 and 6 were completed by asking faculty members (Experienced [E], n = 20), and postgraduate year (PGY)1-2 participants (Novice [N], n = 23). Tasks were recorded on video and scored by 2 independent raters. Validity, reliability, and responsiveness were measured and group results were compared by IRR, and chi-squared and analysis of variance, respectively.

Results: No significant difference in visuospatial testing was identified between E and N participants. Significant differences were observed between E and N in their ability to complete tasks within the set measured parameter tolerances or absolute measures: drill aiming and plunging tasks: E = 73%, N = 58% ($P < 0.05$); screw torque task: E = 93%, N = 67% ($P < 0.05$); hardware removal time ($P = 0.001$); hardware removal errors ($P = 0.004$); ankle fixation time ($P < 0.05$); and ankle fixation errors ($P < 0.05$). IRR for occurrence of an error for hardware removal and ankle fixation were 0.7 and 0.75.

Conclusion: The development of discrete binomial metrics for assessment and training of tasks in a PBP model was successfully extended to surgical fracture care training, showing good face and construct validity and good ability to discriminate between novice and expert performers of the tasks.