

ACS 2024 Surgeons and Engineers: A Dialogue on Surgical Simulation Meeting

P-C-05

Research Abstracts

Instructive Feedback for Video-Based Surgical Coaching

Ryan Chou; Brittany Tsou; Jiawei Zhao; S. Swaroop Vedula; and Shameema Sikder

Whiting School of Engineering, Johns Hopkins University, Baltimore, MD; Johns Hopkins University School of Medicine, Baltimore, MD

Introduction: Machine learning methods for surgical skill assessment are typically trained to predict scores on structured rating scales. Currently-available structured rating scales are not specific in terms of the instructive feedback they provide to surgeons. Specific feedback minimizes inter-rater reliability and supports formative assessment and learning. Our objective was to develop a proof-of-concept feedback rubric to supplement an existing structured rating scale for the capsulorhexis step in cataract surgery.

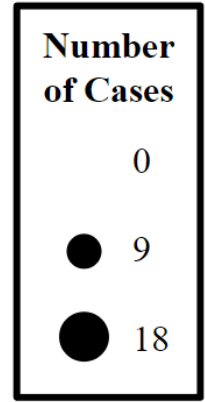
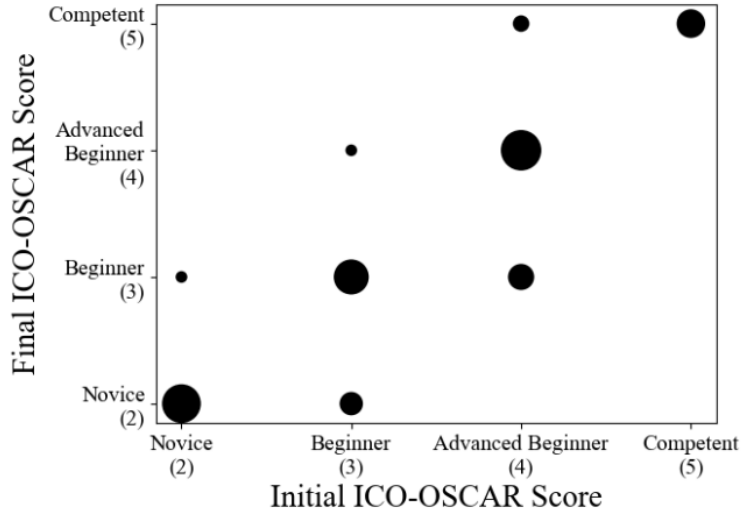
Methods: Expert surgeons used think-aloud narratives to develop a feedback rubric consisting of 27 behaviors observed during capsulorhexis. Three raters assessed videos of capsulorhexis performed by trainees using the feedback rubric. Before and after evaluation of videos using the rubric, the raters assessed skill using the same two items in an existing structured rating scale. We compared skill scores before and after the raters evaluated specific behaviors using the feedback rubric. We computed a weighted Cohen's kappa for inter-rater reliability. We also analyzed the learning curves of the trainees for the overall skill and each behavior.

Results: Our dataset included 51 videos by 10 residents. In many cases, expert surgeons chose to change their overall score given following the review of the feedback catalog (Figure). Inter-rater reliability of the skill assessments improved after raters evaluated the videos using the feedback rubric. Cohen's kappa increased from 0.2878 to 0.3872 for one item in the structured rating scale and from 0.3541 to 0.4623 for the other item. The learning curves for overall skill reflected potential trainee learning trajectories, though the learning curves for the behaviors are heterogeneous.

Conclusions: Our work shows a proof-of-concept feedback rubric to supplement a rating scale for use in formative assessment and for providing feedback to support skill acquisition. The rubric allows development of machine learning tools that go beyond summative assessment of surgical skill.

(a)

Commencement of Flap & Follow-through



(b)

Formation and Circular Completion

