External Validity and Reliability of Task-Specific Metrics for Assessment of Performance in Laparoscopic Crural Repair

Sofia Garces Palacios MD, Alexis Desir MD, Carla Holcomb MD, Daniel J. Scott MD, Ganesh Sankaranarayanan, PhD

Background

Laparoscopic hiatal hernia repair (LHHR) is a widely performed foregut procedure requiring advanced surgical training due to its complexity. Proficiency in LHHR is critical, as recurrence rates can reach up to 50%, particularly in paraesophageal hernias.

The reported learning curve ranges from 50 to 200 cases, highlighting the need for optimized training methods to accelerate skill acquisition.

To address this, our group has developed a simulator and validated Task-Specific Metrics (TSM) for an objective performance assessment of the LHHR.

Objective

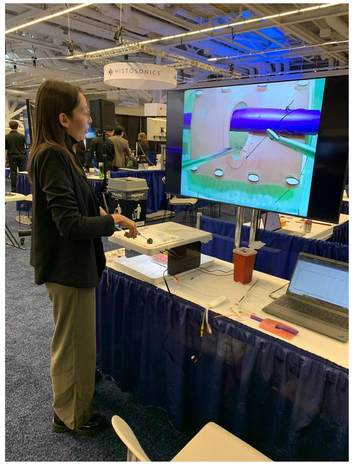
This study aimed to evaluate the external validity of task-specific metrics for assessing the laparoscopic crural repair portion of the LHHR.

Methods

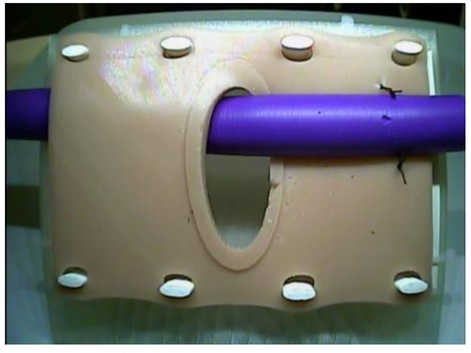
In this IRB-approved study, participants performed a simulated laparoscopic crural repair on a silicone model within a laparoscopic trainer box at the SAGES 2024 Education and Innovation Center.

After consent, a pre-survey collected demographic data, training level, LHHR experience, and geographic location (U.S. or international). Each procedure was videotaped and assessed using OSATS and TSM by three trained graders.

Participants were categorized as novices (PGY 1-2) or experienced (PGY 3-5, fellows, attendings). Inter-rater reliability was measured using the Intraclass Correlation Coefficient (ICC). Group performance was compared via the Wilcoxon-Mann-Whitney test, and Spearman Rank correlation assessed the relationship between OSATS and TSM.



Laparoscopic CR simulator setup at the SAGES EI center

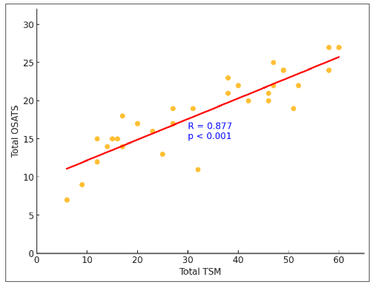


Crural repair silicone model

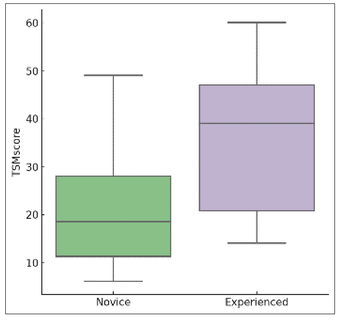
Results

Thirty participants were included: attendings (n=9), fellows (n=3), PGY-5 (n=5), PGY-4 (n=3), PGY-3 (n=3), PGY-2 (n=4), and PGY-1 (n=3). Twenty-two trained in the U.S., while eight were from Asia, Europe, and South America.

Inter-rater agreement was high (ICC = 0.85, $p < 0.001$), with a strong correlation between TSM and OSATS scores ($R = 0.88$, $p < 0.001$). Experienced participants outperformed novices in OSATS (14 vs 20, $W = 43.5$, $p = 0.03$) and TSM (18.5 vs 39, $W = 39.5$, $p = 0.02$).



Correlation plot



TSM score for the two groups

Conclusion

This study validated the TSM for assessing laparoscopic crural repair across diverse local and international participants. Global inclusiveness supports metric standardization across training programs.

Artificial Intelligence and Medical Simulation (AIMS) Lab

UT Southwestern Medical Center