Video-based Assessment of Surgical Residents Performing Minimally Invasive Abdominal Surgery Offers Granular Feedback Using Multiple Assessment Tools

Ann DeBord Smith, MD, MPH, Director Bariatric Surgery[1]; Bruce Ramshaw, MD, CMIO[2]; [1]VA Boston Medical Center; [2]Caresyntax Corporation

Take home: Video-based assessments of residents' operative skills using the Caresyntax platform show increased skill among more senior residents. We can use this technique to monitor entrustable professional activities (EPAs).

Background:

- The purpose was to provide fair evaluation of operative performance to residents.
- Video-based assessment (using Caresyntax in this study) allows unbiased evaluation of anonymous residents doing laparoscopic or robotic cases.
- Faculty reviewing the videos were blinded to resident year, gender, and race freeing reviewers to give honest feedback based on the video before them without preconceived notions about the resident.
- Multiple assessment tools were available including SIMPL, GASS, and GCD among others as well as free text feedback.

Methods:

- Compared Global Case Difficulty (GCD) and Global Assessment of Surgical Skills (GASS)—both interpreted on a 1–3 scale (Poor, Adequate, Excellent)—and SIMPL global ratings across residents early (PGY-1), mid (PGY-4), and late (PGY-5) in training.
- These were evaluated by 1-4 attending surgeons at an academic medical center.
- Evaluations grouped by training year and compared with Kruskal-Wallis tests
- Data availability by metric: GCD (PGY-5 n=17, PGY-4 n=3, PGY-1 n=1), GASS (PGY-5 n=12, PGY-4 n=3, PGY-1 n=1), and SIMPL (PGY-5 n=7, PGY-4 n=3, PGY-1 n=1).

Assessment Tool	Description	PGY-1 (Early) n=1	PGY-4 (Mid) n=3	PGY-5 (Late) n=7–17	p-value	Trend
SIMPL	Global performance rating	2.30	2.63 ± 0.06	2.96 ± 0.45	0.14	↑ Increases with training
GASS	Global surgical skills	2.30	2.63 ± 0.06	2.68 ± 0.34	0.27	↑ Increases with training
GCD	Case complexity	2.00	2.33 ± 0.58	1.71 ± 0.59	0.25	↓* Chiefs take harder cases

*Scale: 1=Poor, 2=Adequate, 3=Excellent. All comparisons by Kruskal-Wallis test. *GCD lower score for PGY-5 reflects greater case complexity undertaken, not poorer performance.*

Figure 1: Qualitative Feedback Distribution (n=40 Comments)

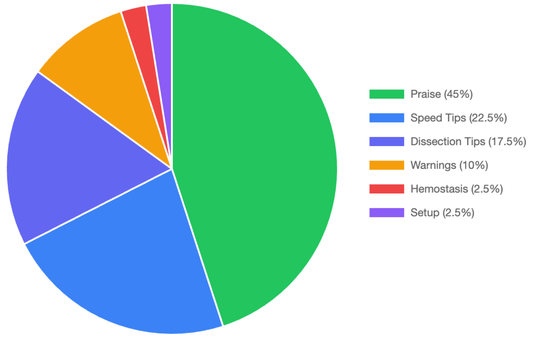

Figure 1. *Distribution of 40 free-text reviewer comments grouped into six thematic categories. Praise comprised the largest category (45%), followed by technical tips on operative speed (22.5%) and dissection technique (17.5%).*

Figure 2: Score Progression Across Training Levels

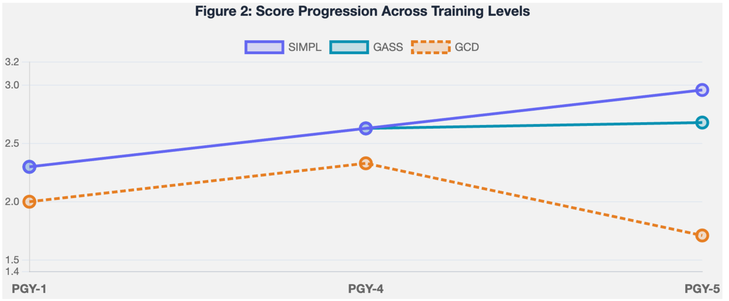

Figure 2. *Line chart showing score trends for SIMPL, GASS, and GCD across PGY-1, PGY-4, and PGY-5. GCD (dashed) trends inversely, reflecting increasing case complexity at the chief resident level.*

Conclusions: SIMPL and GASS scores rose with training year as anticipated. GCD showed the PGY-4's trended towards more complex cases. Further data needs to be collected so that assessment tool score can be adjusted for level of case complexity. Qualitative review of free text comments enabled reviewers to target feedback to residents based on growth opportunities noticed in videos with ample praise and encouragement offered in addition to useful technical tips.