2023 SAGES POSTER

Using video-based assessment (VBA) to document fellow improvement in safely completing the jejunojejunostomy portion of laparoscopic Roux-en-Y gastric bypass (RYGB) surgery

Peter Nau[1] · Erin Worden[1] · Ryan Lehmann[1] · Kyle Kleppe[2] · Gregory J. Mancini[2] · Matt L. Mancini[2] · Bruce Ramshaw[3,4] · Michael S. Woods[4]

Abstract

Background Surgical assessment instruments are used for formative and summative trainee evaluations. To characterize the features of existing instruments and a novel 12-item objective, procedure-specific assessment tool for Roux-en-Y Gastric Bypass (RYGB-OPSA), we evaluated the progress of a single surgical fellow over 17 consecutive surgeries.

Methods Seventeen consecutive RYGB videos completed between 8/2021 and 1/2022 by an academic hospital surgical fellow were de-identified and assessed by four board-certified bariatric surgeons using Global Operative Assessment of Laparoscopic Skills (GOALS), General Assessment of Surgical Skill (GASS), and RYGB-OPSA which includes the reflection of transverse colon, identification of ligament of Treitz, biliopancreatic and Roux limbs orientation, jejunal division point selection, stapler use, mesentery division, bleeding control, jejunojejunostomy (JJ) anastomotic site selection, apposition of JJ anastomotic site, JJ creation, common enterotomy closure of JJ, and integrity of anastomosis. The GASS measured economy of motion, tissue handling, appreciating operative anatomy, bimanual dexterity, and achievement of hemostasis. RYGB-OPSA and GASS items were scored "poor—unsafe," "acceptable—safe," or "good—safe." Change in performance was measured by linear trendline slope.

Results Over the course of 17 procedures, significant improvement was demonstrated by three GOALS items, GOALS overall score, GASS bimanual dexterity, and three RYGB-OPSA tasks: JJ creation, jejunal division point selection, and stapler use. Achievement of hemostasis declined but never rated "poor—unsafe." Overall RYGB-OPSA and GOALS trendlines documented significant increase across the 17 procedures.

Conclusion This examination of a bariatric surgery fellow's operative training experience as measured by three surgical assessment instruments demonstrated anticipated improvements in general skills and safe completion of procedure-specific tasks. Effective surgical assessment instruments have enough sensitivity to show improvement to enable meaningful trainee feedback (low-stakes assessments) as well as the ability to determine safe surgical practice to enable promotion to greater autonomous practice.

✉ Peter Nau
  peter-nau@uiowa.edu

1 Section of Bariatric Surgery, Department of Surgery, University of Iowa Hospitals & Clinics, 200 Hawkins Drive, Iowa City, IA 52242, USA

2 Section of Foregut Surgery, Department of Surgery, University of Tennessee – Knoxville, 1924 Alcoa Highway, Knoxville, TN 37920, USA
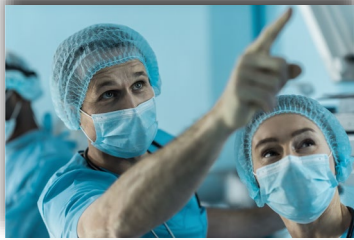
3 CQInsights PBC, Knoxville, TN, USA

4 Caresyntax Corporation, Boston, MA, USA

Graphical abstract

Video-Based Assessment for Bariatric Fellows: Measuring Improvement in Skill Acquisition

Real-world performance of the Objective, Procedure-Specific Assessment for Roux-en-Y Gastric Bypass (OPSA RYGB) to evaluate bariatric surgery fellow performance

Surgical trainee evaluation is transitioning to objective competency-focused assessments
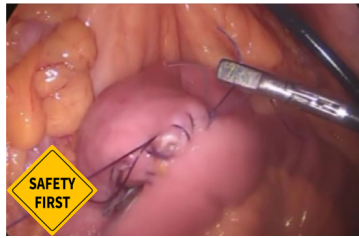
OPSA RYGB uses a simplified, competency-based safe vs. unsafe scoring framework

The OPSA RYGB demonstrated statistically significant improvement in skills over 17 procedures

Author Nau P, et al.

Variation in bariatric surgical outcomes is common, even among surgical centers of excellence [1]. Technical proficiency among bariatric surgeons varies widely and is an important predictor of clinical outcomes including surgical site infection, readmissions, hemorrhage, reoperation, length-of-stay, and mortality [2–5].

Increasingly, surgical training relies on the use of video-based assessment (VBAs) [6] and the evaluation of technical skill using validated instruments [7, 8]. The most common of these instruments are the Objective Structured Assessments of Technical Skills (OSATS) and the Global Operative Assessment of Laparoscopic Skills (GOALS) [9–11]. Both instruments generate a global rating of surgeon technique in select domains of surgical practice. However, evidence suggests that the instruments may not adequately assess surgeon technical skill during training [12] and more importantly, omit the specific skills required to safely complete a specific surgical procedure. A recent systematic literature review of tools to assess surgical technical skills concludes that most published and validated instruments have limited adoption [13].

More recently, there has been a trend toward development and validation of objective, procedure-specific assessments (OPSAs). The most cited and only procedure-specific scale focused on bariatric surgery, the Bariatric Objective Structured Assessment of Technical Skill (BOSATS) [14] is a checklist that rates 23 bariatric surgery tasks. Each task is rated on a unique scale with multiple criteria contributing to each level of response. Like many surgical assessments, the BOSATS has not enjoyed widespread adoption in part because its lengthy and complex scoring algorithm adds substantial administrative burden to already busy surgical training programs.

Nonetheless, this type of procedure-specific assessment has the advantage of providing surgical residents and fellows with necessary feedback required for safe surgical practice. In alignment with competency-based medical education and the evolving focus on entrustable professional activities (EPAs) in training programs and professional societies, [15–17] OPSAs provide a logical, data-driven methodology to ensure trainees are prepared for safe autonomous practice [18].

In support of developing and implementing EPAs focused on safe completion of surgical procedures, the objective of this research was to evaluate changes in general surgical skill acquisition as well as changes in the safe completion of 12 consecutive tasks required for the jejunojejunostomy (JJ) segment of the Roux-en-Y Gastric Bypass (RYGB) procedure.

Methods

This was a prospective cohort study of consecutive case series assessments of 17 RYGB procedures performed by a single post-doctoral surgical fellow from the University of Iowa Hospital & Clinics, Department of Gastroenterology, Bariatric Surgery Section. The bariatric surgery fellow's laparoscopic surgical skill and competence was assessed by four board-certified, fellowship-trained bariatric surgeons using GOALS and two novel instruments: an RYGB OPSA and a General Assessment of Surgical Skill (GASS).

Instrument development and selection

The assessment included independent completion of three rating scales related to surgical technical performance: Global Operative Assessment of Laparoscopic Skills (GOALS) the RYGB OPSA (Supplemental Fig. 1), and the General Assessment of Surgical Skill (GASS) (Supplemental Fig. 2). GOALS is a standardized and validated instrument for grading overall technical proficiency for laparoscopic surgery, using a 5-point scale with a 3-level response option for assessing depth perception, bimanual dexterity, efficiency, and tissue handling. The second rating scale, the RYGB OPSA, was designed and developed by five study authors (PN, RL, EW, BR, MSW). The RYGB OPSA was designed to assess surgeon competence in completing discrete tasks of the JJ, pouch, and gastrojejunostomy. This study used only the JJ portion of the assessment. The following 12 tasks comprised the OPSA for the JJ portion of gastric bypass surgery: adequate reflection of the transverse colon, clear identification of the ligament of Treitz, maintenance of appropriate orientation of biliopancreatic and Roux limb, selection of the jejunal division point, stapler use, mesentery division, bleeding control, selection of JJ anastomotic site, apposition of JJ anastomotic site, creation of JJ, common enterotomy closure of JJ, and finally, evaluation of integrity of anastomosis. Each task was rated as poor—unsafe, acceptable—safe, or good—safe (scored numerically as 1, 2, and 3, respectively). Finally, raters assessed each surgery for global case difficulty (GCD) as "easy," "average," or "hard." Similar to the GOALS, the GASS measures aspects of overall surgical technical performance including economy of motion, tissue handling, appreciating operative anatomy, bimanual dexterity, achievement of hemostasis, and overall performance with scoring rubric of poor—safe (1), adequate—safe (2), and good—safe (3).

Operative video selection and data collection

The first 17 consecutive gastric bypass procedures completed by a single fellow from the Bariatrics Section of the Fellowship Training Program at the University of Iowa Hospitals and Clinics were video recorded. The cases were completed between August 2021 and January 2022. All videos were de-identified and uploaded to a proprietary online software-as-a-service (SaaS) platform. Each case's JJ portion was isolated and clipped using the platform's video editing function. The clipped JJ video was downloaded and labeled with a unique number generated using a random number generator with a lower limit setting of 1 and an upper limit setting of 5000, generating 100 numbers. The videos were then uploaded to a secure, password-protected cloud that the reviewers accessed to complete the video reviews. Two raters completed the GOALS assessment first, followed by the OPSA and GASS. Two raters completed the OPSA assessment first, followed by the GOALS and GASS. Raters were provided scoring sheets and received no specific training in the completion of assessments. All four reviewers were board-certified general surgeons trained in bariatric surgery, ranging in experience from 4 to 28 years in practice and a mean of 17.8 years. Each reviewer used a pre-formatted Excel spreadsheet to rate each video. Assessments were completed between January 10th and February 11th, 2022. Finally, each rater scored each operative video for case difficulty on a scale of easy (score: 1), medium (score: 2), or hard (score: 3).

Data analysis

The mean GOALS score was calculated for each case based upon the sum of the five individual item scores divided by five. Additionally, we calculated percent agreement between the categorical (safe vs. unsafe) measure for each of the 12 tasks in the OPSA and for the GCD. To calculate percent agreement across the raters, both the OPSA and GASS scale scores were categorized as either safe or unsafe. For the OPSA, unsafe was assigned to a score of 1 and safe to scores of 2 or 3. For the GOALS, unsafe was assigned for a score less than 3 and safe for scores of 3 or higher. Measurement of performance was assessed for all cases, cases with an average GCD less than 2, and cases with an average GCD 2 or greater. Evidence of a learning curve was measured by change in score over time, as measured by the slope of the linear trend line fit to the chronologically order average rater score for each component and the overall average of both the GOALS and OPSA scales.

Results

A total of 17 gastric bypass procedures were performed by the Fellow between August 2021 and January 2022. As assessed by the raters, the average case difficulty score was 1.8 (SD 0.5), with only one surgical case receiving a GCD score of "hard" from one of the four raters.

GOALS

On the five-point GOALS rating scale, the average score for all cases was highest for tissue handling (3.76, SD 0.76) and for depth perception (3.74; SD 0.89). The lowest average rating was for efficiency (3.46; SD 0.82). Scores stratified by GCD averaged mildy worse performance for harder cases. (Table 1) The slope of the linear trendline across the 17 consecutive operative videos was positive and significant for the total GOALS score (Table 2; Supplemental Fig. 3) and for the bimanual dexterity, depth perception, and efficiency items (Table 2). Only tissue handling did not show a positive trend toward improvement.
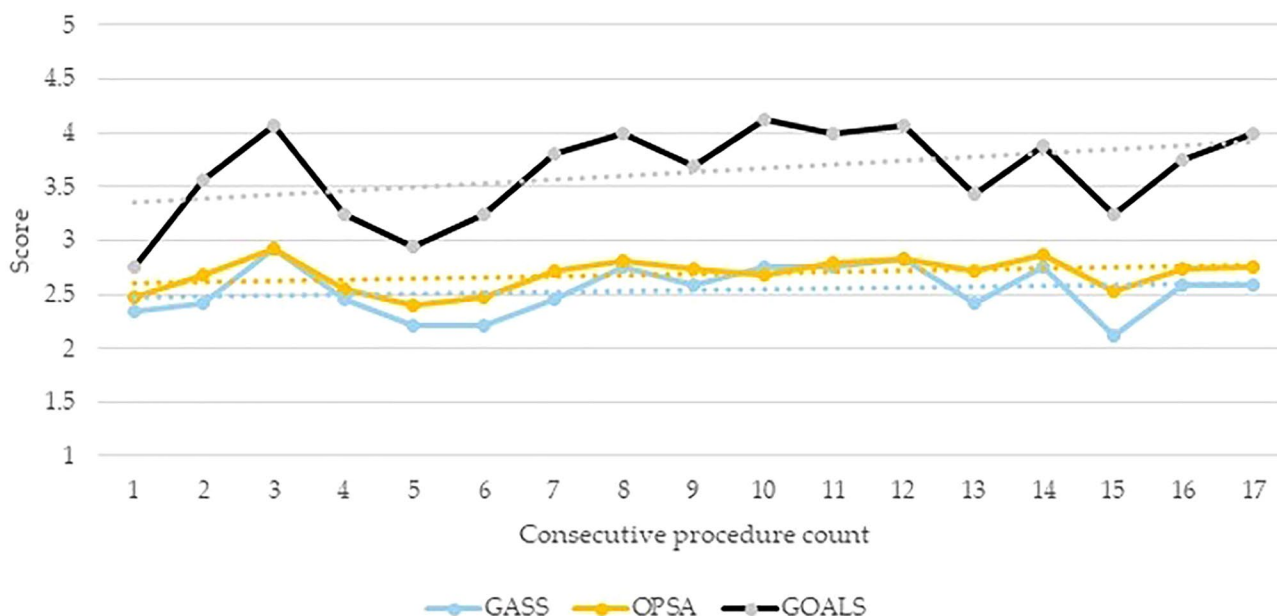
GASS

On the 3-point GASS rating scale, the highest average score for all cases was achievement of hemostasis (2.82; SD 0.38) followed by appreciating operative anatomy (2.62; SD 0.52). The lowest average score was reported for economy of motion (2.32; SD 0.53). Using the categorized

safe vs. unsafe rating scale, both achievement of hemostasis and overall performance were rated as safe across all 17 procedures, with tissue handling and economy of motion having the highest percent rated as poor/unsafe (4.4% and 3.9%, respectively) (Table 1). Scores stratified by GCD were mildly reduced for harder cases. Despite having no ratings of poor (unsafe), the slope of the linear trendline for achievement of hemostasis was significant and negative, while the slope of the trendline for bimanual dexterity was significant and positive (Table 2). No other items showed a significant change across the 17 operative videos (Table 2; Fig. 1).

RYGB OPSA

Overall, the highest average rated items of the RYGB OPSA were maintenance of orientation of biliopancreatic and Roux limb (3.0; S.D. 0), and evaluation of integrity of anastomosis (2.94; S.D. 0.34). The lowest average ratings were for common enterotomy closure of the JJ (2.37; S.D. 0.54) and for creation of the JJ (2.54; S.D. 0.5) (Table 1). Scores stratified by GCD showed no significant changes. Of interest, no raters scored creation of the JJ, apposition of JJ anastomotic site, creation of the JJ, maintenance of orientation of the biliopancreatic and Roux limb, or stapler use as unsafe for any of the operative videos. The tasks with the highest percent of unsafe ratings were clear identification of the ligament of Treitz (13.2%), followed by adequate reflection of the transverse colon (11.8%) and selection of the JJ anastomotic site (4.4%). In addition to the average score across all 12 items (Fig. 1), three RYGB OPSA items showed a statistically

Fig. 1 Average assessment scores with linear trendline for 17 consecutive procedures

Table 1 Performance measures for the RYGB OPSA, GOALS, and GASS

Instrument	Item	All videos $n=17$			GCD score <2 (easier) $n=8$			GCD score $2+$ (harder) $n=9$		
		Mean	SD	% Safe	Mean	SD	% Safe	Mean	SD	% Safe
GASS	Achievement of hemostasis	2.82	0.38	100	2.84	0.37	100	2.81	0.40	100
	Appreciating operative anatomy	2.62	0.52	98.5	2.75	0.44	100	2.50	0.56	97.2
	Bimanual dexterity	2.43	0.53	98.5	2.50	0.57	96.9	2.36	0.49	100
	Economy of motion	2.32	0.53	97.1	2.38	0.61	93.8	2.28	0.45	100
	Overall performance	2.49	0.5	100	2.59	0.50	100	2.39	0.49	100
	Tissue handling	2.54	0.58	95.6	2.66	0.55	96.9	2.44	0.61	94.4
	Average score	2.54	0.38	95.6	2.62	0.37	96.9	2.46	0.38	94.4
GOALS	Bimanual dexterity	3.59	0.85		3.62	0.87		3.56	0.84	
	Depth perception	3.74	0.89		3.78	0.87		3.69	0.92	
	Efficiency	3.46	0.82		3.53	0.84		3.39	0.80	
	Tissue handling	3.76	0.76		3.88	0.75		3.67	0.76	
	Total GOALS score	14.54	3.01		14.81	3.09		14.31	2.96	
RYGB OPSA	Adequate reflection of the transverse colon	2.76	0.65	88.2	2.88	0.49	93.8	2.67	0.76	83.3
	Apposition of jejunojejunostomy anastomotic site	2.63	0.49	100	2.69	0.47	100	2.58	0.50	100
	Bleeding control	2.69	0.5	98.5	2.69	0.54	96.9	2.69	0.47	100
	Clear identification of the ligament of Treitz	2.74	0.68	86.8	2.81	0.59	90.6	2.67	0.76	83.3
	Common enterotomy closure of the JJ[a]	2.37	0.54	97.1	2.41	0.56	96.9	2.33	0.53	97.2
	Creation of the JJ	2.54	0.5	100	2.59	0.50	100	2.50	0.51	100
	Evaluation of integrity of anastomosis	2.94	0.34	97.1	2.88	0.49	93.8	3.00	0.00	100
	Maintenance of orientation of biliopancreatic and Roux limb	3.00	0	100	3.00	0.00	100	3.00	0.00	100
	Mesentery division	2.71	0.49	98.5	2.66	0.55	96.9	2.75	0.44	100
	Selection of jejunal division point	2.59	0.53	98.5	2.53	0.57	96.9	2.64	0.49	100
	Selection of JJ anastomotic site	2.53	0.59	95.6	2.62	0.49	100	2.44	0.65	91.7
	Stapler use	2.72	0.45	100	2.66	0.48	100	2.78	0.42	100
	Average score	2.69	0.23	100	2.70	0.25	100	2.67	0.22	100

[a]Jejunojejunostomy

significant improvement in scores across the 17 operative videos: creation of the JJ, selection of jejunal division point, and stapler use (Table 2).

Discussion

To create a laparoscopic bariatric surgery assessment instrument aligned with the needs of competency-based medical education and EPAs (enabling trainee micro-assessments in routine clinical practice), a novel 12-item RYGB OPSA focused upon safe completion of narrowly defined surgical tasks was developed and field tested for initial performance against a single surgical fellow over a series of 17 procedures. Preliminary results indicate the instrument was able to measure meaningful changes in surgical performance, both for general assessments of surgical technical skill and procedure-specific assessments of the specific tasks required to complete a procedure. OPSA RYGB performance scores

demonstrated did not vary substantially based upon case difficulty.

The improvement in performance documented in this study is consistent with one of the few prior studies evaluating surgical fellows' performance using objective assessments. In a study including 98 assessments among 31 surgical fellows, Hogle et al. [19] reported that GOALS scores for overall performance, bimanual dexterity, efficiency, and autonomy significantly improved throughout the fellowship year and that depth perception and tissue handling improved but didn't reach statistical significance.

The novel instrument introduced in this study—the 12-item RYGB OPSA—was deliberately designed with a consistent "poor-unsafe," "adequate—safe" vs. "good—safe" scoring rubric to enable low-stakes feedback to surgical trainees. Though the average score across all 12 items was high (2.69 out of 3), all but four of the tasks had at least one unsafe rating and two tasks had a substantially higher number of unsafe ratings. Providing unambiguous

Table 2 Test for trend for each item assessed for the fellow surgeon

Instrument	Item	Trend	p value
GASS	Achievement of hemostasis	− 0.003	< 0.05
	Appreciating operative anatomy	0.002	NS[a]
	Bimanual dexterity	0.004	< 0.05
	Economy of motion	0.003	NS
	Overall case difficulty level	0.003	NS
	Overall performance	0.002	NS
	Tissue handling	0.001	NS
	Average score	0.001	NS
GOALS	Bimanual dexterity	0.008	< 0.05
	Depth perception	0.008	< 0.05
	Efficiency	0.008	< 0.01
	Tissue handling	0.004	NS
	Total GOALS	0.026	< 0.05
OPSA	Adequate reflection of the transverse colon	− 0.001	NS
	Apposition of jejunojejunostomy anastomotic site	0.003	NS
	Bleeding control	− 0.002	NS
	Clear identification of the ligament of Treitz	0.001	NS
	Common enterotomy closure of the jejunojejunostomy	0.003	NS
	Creation of the jejunojejunostomy	0.004	< 0.05
	Evaluation of integrity of anastomosis	− 0.001	NS
	Maintenance of appropriate orientation of biliopancreatic and Roux limb	0	NS
	Mesentery division	0.002	NS
	Selection of jejunal division point	0.005	< 0.05
	Selection of jejunojejunostomy anastomotic site	0.001	NS
	Stapler use	0.005	< 0.001
	Average score	0.002	0.05

[a]Not significant

assessment of safe task-specific performance coupled with VBAs where trainees can visualize their technique provides a rich context for providing feedback that surgical residents and fellows [20] desire as part of their training and has been demonstrated to improve operative performance [21, 22].

The approach used to assess surgical fellow performance in this study is consistent with a recent review of the use of VBA in surgical education which summarized results from 199 peer-reviewed manuscripts [7]. The authors report on numerous benefits of VBA in the educational process, concluding with potentially the most relevant, that VBA may help decrease the assessment demands of medical education. The two novel instruments included in this study, the GASS and the RYGB OPSA, were designed to work seamlessly with VBA and to explicitly prioritize the most important aspect of surgical performance, namely the safe completion of a procedure, while also focusing on ease of use and administration. Efficiency of use and ease of administration may be particularly important as it is estimated that the learning curve range for RYGB is anywhere from 30 to 500 procedures [23].

Advancing the real-world adoption of competency-based medical education and entrustable professional activities

While the vision and value of competency-based medical education and EPAs is clear, the pathway to widespread adoption is less certain. The core operating components of EPAs include reliable assessments of trainee performance as part of routine surgical practice and the provision of robust feedback for both formative (low-stakes) assessments and summative (high-stakes) assessments. One of the critical enabling factors of EPAs will be assessment instruments that are: (a) reliable and bias-free, (b) easy to use as part of routine practice, and (c) valid in that they improve trainee performance and enable promotion of safe, competent surgeons. The design approach for the RYGB OPSA included iterative engagement with clinical experts, narrowly defining surgical tasks and the use of unambiguous assessment scales, testing of instrument constructs against a variety of surgical videos, focus on safety and brevity, and statistical validation within the context of intended use. One of the core drivers of the

adoption of EPAs is the willingness of busy surgical mentors to actually perform assessments of their trainees [24]. While sound methodologies such as those used in the development of the 23 item BOSATS including hierarchical task analysis and Delphi questionnaires may deliver theoretically consistent instruments, the lack of broad adoption forces one to question the ultimate effectiveness at delivering instruments that can achieve intended goals established by EPAs. This study provides preliminary evidence that a 12-item RYGB OPSA is sensitive enough to measure improvements in novice surgical performance, and provides evidence regarding the value of rapid, real-world approach to developing and testing instruments focused upon safe surgical practice.

Future research

This research investigation represents a proof-of-concept, namely, that it is practical to assess procedure-specific surgical skill at the level of definable tasks and that measurement can document improvement in a surgical fellow's performance. The current laparoscopic gastric bypass surgery OPSA represents an initial step toward the development of objective, measurable EPAs. However, addressing known challenges of existing instruments including reliability, usability, and validity for high-stakes assessments will require additional investigation [11, 25]. First, continued evaluation and validation of the GASS and RYGB OPSA with investigations focused on test–retest reliability, evaluation of trainer and trainee qualitative feedback on the utility and value of the instruments, and on the association of performance on clinical and financial outcomes following surgery. In support of this work, the RYGB OPSA will continue to be used in fellowship training and will be used in a national surgical collaborative to collect additional validation data.

Strengths and limitations

The strength of this study is reflected in the uniqueness of the data. VBA is a practical method for allowing the assessment of surgical performance and safety. Further, this is the first time that the RYGB procedure has been deconstructed into its major steps with a simple, easy-to-rate scoring rubric focused on the safe completion of each task. The focus on safe completion of each surgical task may support more focused training and feedback on the specific areas of improvement required to grant autonomy.

Several limitations should be considered when interpreting these results. First, raters received no training in the interpretation and scoring of any of the instruments and no follow-up interviews were conducted among the raters to assess why a score—especially an unsafe rating—was assigned. Additionally, surgical performance was only

assessed for a single surgical fellow, a fellow who entered the program with substantial surgical experience. This fellow's experience and scores may not reflect those of surgical residents in training or with less experience. As a result, improvement in performance across all the instruments may be demonstrated in a surgical trainee population that is earlier in their surgical training residency.

Conclusion

This study demonstrated improved skill acquisition by a minimally invasive bariatric surgery fellow in the first six months of their fellowship training. The study also demonstrated that improvement occurs even in the latest stage of surgical training and that improvement is not just measurable at the general skill level but also at the level of procedure-specific tasks. With further validation, the new scales may assist in documenting competency and support the evaluation of EPAs during surgical training programs. Finally, with the goal of optimizing surgical technique, associated outcomes and patient safety, surgical training may benefit significantly from more systematic targeted feedback, coaching, and guidance provided for each task in a procedure.

Declarations

Disclosures Nau reported receipt of support for this manuscript and payment or honoraria from Caresyntax. Worden reported receipt of support for this manuscript from Caresyntax. Lehmann reported receipt of support for this manuscript, consulting fees and payment or honoraria from Caresyntax. Kleppe reported receipt of support for this manuscript from Caresyntax. G. Mancini reported receipt of support for this manuscript, consulting fees and payment or honoraria from Caresyntax. M. Mancini reported receipt of support for this manuscript, consulting fees and payment or honoraria from Caresyntax. Dr. Ramshaw reported receipt of support for this manuscript and consulting fees from Caresyntax. Dr. Woods is an employee of Caresyntax.

Ethical approval This study was approved by the University of Iowa Hospitals & Clinics Institutional Review Board (Approval No. 202109638).

References

1. Ibrahim AM, Ghaferi AA, Thumma JR, Dimick JB (2017) Variation in outcomes at bariatric surgery centers of excellence. JAMA Surg. https://doi.org/10.1001/jamasurg.2017.0542

2. Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, Dimick J, Banerjee M, Birkmeyer NJO, Collaborative MBS (2013) Surgical skill and complication rates after bariatric surgery. N Engl J Med. https://doi.org/10.1056/nejmsa1300625

3. Varban OA, Thumma JR, Finks JF, Carlin AM, Ghaferi AA, Dimick JB (2021) Evaluating the effect of surgical skill on outcomes for laparoscopic sleeve gastrectomy: a video-based study. Ann Surg. https://doi.org/10.1097/sla.0000000000003385

4. Stulberg JJ, Huang R, Kreutzer L, Ban K, Champagne BJ, Steele SR, Johnson JK, Holl JL, Greenberg CC, Bilimoria KY (2020) Association between surgeon technical skills and patient outcomes. JAMA Surg. https://doi.org/10.1001/jamasurg.2020.3007

5. Woods MS, Liberman JN, Rui P, Wiggins E, White J, Ramshaw B, Stulberg JJ (2023) Association between surgical technical skills and clinical outcomes; a systematic literature review and meta—analysis. JSLS. https://doi.org/10.4293/JSLS.2022.00076

6. Ahmet A, Gamze K, Rustem M, Sezen KA (2018) Is video-based education an effective method in surgical education? A systematic review. J Surg Educ. https://doi.org/10.1016/j.jsurg.2018.01.014

7. McQueen S, McKinnon V, VanderBeek L, McCarthy C, Sonnadara R (2019) Video-based assessment in surgical education: a scoping review. J Surg Educ. https://doi.org/10.1016/j.jsurg.2019.05.013

8. Huang RJ, Limsui D, Triadafilopoulos G (2018) Video-based performance assessment in endoscopy: moving beyond "see one, do one, teach one"? Gastrointest Endosc. https://doi.org/10.1016/j.gie.2017.09.014

9. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. https://doi.org/10.1046/j.1365-2168.1997.02502.x

10. Balvardi S, Semsar-Kazerooni K, Kaneva P, Mueller C, Vassiliou M, Al Mahroos M, Fiore JF Jr, Schwartzman K, Feldman LS (2023) Validity of video-based general and procedure-specific self-assessment tools for surgical trainees in laparoscopic cholecystectomy. Surg Endosc. https://doi.org/10.1007/s00464-022-09466-6

11. Kramp KH, van Det MJ, Hoff C, Lamme B, Veeger NJGM, Pierie JPEN (2015) Validity and reliability of global operative assessment of laparoscopic skills (GOALS) in novice trainees performing a laparoscopic cholecystectomy. J Surg Educ. https://doi.org/10.1016/j.jsurg.2014.08.006

12. Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD (2016) Objective structured assessments of technical skills (OSATS) does not assess the quality of the surgical result effectively. Clin Orthop Relat Res. https://doi.org/10.1007/s11999-015-4603-4

13. Vaidya A, Aydin A, Ridgely J, Raison N, Dasgupta P, Ahmed K (2020) Current stats of technical skills assessment in surgery: a systematic review. J Surg Res. https://doi.org/10.1016/j.jss.2019.09.006

14. Zevin B, Bonrath EM, Aggarwal R, Dedy NJ, Ahmed N, Grantcharov TP, ATLAS Group (2013) Development, feasibility, validity, and reliability of a scale for objective assessment of operative performance in laparoscopic gastric bypass surgery. J Am Coll Surg. https://doi.org/10.1016/j.jamcollsurg.2013.01.003

15. Greenberg JA, Minter RM (2019) Entrustable professional activities: the future of competency-based education in surgery may already be here. Ann Surg. https://doi.org/10.1097/sla.0000000000003153

16. Wagner JP, Lewis CE, Tillou A, Agopian VG, Quach C, Donahue TR, Hines OJ (2018) Use of entrustable professional activities in the assessment of surgical resident competency. JAMA Surg. https://doi.org/10.1001/jamasurg.2017.4547

17. American Board of Surgery (2022) Entrustable professional activities. https://www.absurgery.org/default.jsp?epahome. Accessed 1 Mar 2023

18. White A, Moran HRM, Ryan J, Mador B, Campbell S, Turner SR (2022) Validity evidence for procedure-specific competency assessment tools in cardiovascular and thoracic surgery: a scoping review. J Surg Educ 79:1016–1023

19. Hogle NJ, Liu Y, Ogden RT, Fowler DL (2014) Evaluation of surgical fellows' laparoscopic performance using global operative assessment of laparoscopic skills (GOALS). Surg Endosc. https://doi.org/10.1007/s00464-013-3324-6

20. Bello RJ, Sarmiento S, Meyer ML, Rosson GD, Cooney DS, Cooney DS, Lifchez SD, Cooney CM (2018) Understanding surgical resident and fellow perspectives on their operative performance feedback needs: a qualitative study. J Surg Educ. https://doi.org/10.1016/j.jsurg.2018.04.002

21. Ducournau F, Meyer N, Xavier F, Facca S, Liverneaux P (2021) Learning a MIPO technique for distal radius fractures: mentoring versus simple experience versus deliberate practice. Orthop Traumatol Surg Res. https://doi.org/10.1016/j.otsr.2021.102939

22. Bonrath EM, Dedy NJ, Gordon LE, Grantcharov TP (2015) Comprehensive surgical coaching enhances surgical skill in the operating room: a randomized controlled trial. Ann Surg. https://doi.org/10.1097/sla.0000000000001214

23. Wehrtmann FS, de la Garza JR, Kowalewski KF, Schmidt MW, Müller K, Tapking C, Probst P, Diener MK, Fischer L, Müeller-Stich BP, Nickel F (2020) Learning curves of laparascopic Roux-en-Y gastric bypass and sleeve gastrectomy in bariatric surgery: a systematic review and introduction of a standardization. Obes Surg. https://doi.org/10.1007/s11695-019-04230-7

24. Champagne BJ (2013) Effective teaching and feedback strategies in the OR and beyond. Clin Colon Rectal Surg. https://doi.org/10.1055/s-0033-1356725

25. Groenier M, Brummer L, Bunting BP, Gallagher AG (2020) Reliability of observational assessment methods for outcome-based assessment of surgical skill: systematic review and meta-analysis. J Surg Educ. https://doi.org/10.1016/j.jsurg.2019.07.007