



# Validation study of a skill assessment tool for education and outcome prediction of laparoscopic distal gastrectomy

Yo Kurashima<sup>1,2</sup> · Hidehiko Kitagami<sup>3</sup> · Koichi Teramura<sup>4</sup> · Saseem Poudel<sup>1</sup> · Yuma Ebihara<sup>1</sup> · Noriyuki Inaki<sup>5</sup> · Fumitaka Nakamura<sup>6</sup> · Kazunari Misawa<sup>7</sup> · Kazunori Shibao<sup>8</sup> · Eishi Nagai<sup>9</sup> · Yoshimasa Akashi<sup>10</sup> · Shuji Takiguchi<sup>11</sup> · Shinichiro Kobayashi<sup>12</sup> · Satoshi Hirano<sup>1</sup>

Received: 10 January 2022 / Accepted: 23 April 2022 / Published online: 16 May 2022  
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

## Abstract

**Background** The Japanese operative-rating scale for laparoscopic distal gastrectomy (JORS-LDG) was developed through cognitive task analysis together with the Delphi method to measure intraoperative performance during laparoscopic distal gastrectomy. This study aimed to investigate the value of this rating scale as an educational tool and a surgical outcome predictor in laparoscopic distal gastrectomy.

**Methods** The surgical performance of laparoscopic distal gastrectomy was assessed by the first assistant, through self-evaluation in the operating room and by video raters blind to the case. We evaluated inter-rater reliability, internal consistency, and correlations between the JORS-LDG scores and the evaluation methods, patient characteristics, and surgical outcomes.

**Results** Fifty-four laparoscopic distal gastrectomy procedures performed by 40 surgeons at 16 institutions were evaluated in the operating room and with video recordings using the proposed rating scale. The video inter-rater reliability was  $> 0.8$ . Participating surgeons were divided into the low, intermediate, and high groups based on their total scores. The number of laparoscopic surgeries and laparoscopic gastrectomy procedures performed differed significantly among the groups according to laparoscopic distal gastrectomy skill levels. The low, intermediate, and high groups also differed in terms of median operating times (311, 266, and 229 min, respectively,  $P < 0.001$ ), intraoperative complication rates (27.8, 11.8, and 0%, respectively,  $P = 0.01$ ), and postoperative complication rates (22.2, 0, and 0%, respectively,  $P = 0.002$ ).

**Conclusions** The JORS-LDG is a reliable and valid measure for laparoscopic distal gastrectomy training and could be useful in predicting surgical outcomes.

**Keywords** Laparoscopy · Gastrectomy · Skill evaluation · Assessment scale · Surgical education · Outcome prediction

✉ Yo Kurashima  
yo.kurashima@huhp.hokudai.ac.jp

<sup>1</sup> Department of Gastroenterological Surgery II, Hokkaido University School of Medicine, Kita-15, Nishi-7, Kita-ku, Sapporo, Hokkaido 060-8638, Japan

<sup>2</sup> Clinical Simulation Center, Hokkaido University School of Medicine, Sapporo, Japan

<sup>3</sup> Department of Gastroenterological Surgery, Keiyukai Sapporo Hospital, Sapporo, Japan

<sup>4</sup> Department of Colorectal Surgery, National Cancer Center Hospital East, Kashiwa, Japan

<sup>5</sup> Department of Gastrointestinal Surgery, Kanazawa University Graduate School of Medical Sciences, Kanazawa, Japan

<sup>6</sup> Department of Gastroenterological Surgery, Teine Keijinkai Hospital, Sapporo, Japan

<sup>7</sup> Department of Gastroenterological Surgery, Aichi Cancer Center Hospital, Nagoya, Japan

<sup>8</sup> Department of Surgery I, School of Medicine, University of Occupational and Environmental Health, Kitakyushu, Japan

<sup>9</sup> Department of Gastrointestinal Surgery, Japanese Red Cross Fukuoka Hospital, Fukuoka, Japan

<sup>10</sup> Department of Gastrointestinal and Hepato-Biliary-Pancreatic Surgery, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan

<sup>11</sup> Department of Gastroenterological Surgery, Nagoya City University Graduate School of Medical Sciences, Nagoya, Japan

<sup>12</sup> Department of Surgery, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan

Performance evaluation plays a pivotal role in medical education. Reliable and valid global rating scales have been developed to evaluate surgical skill during surgical training, for program evaluation, and in surgical education research [1–4].

The development of endoscopic surgery has facilitated video recording of intraoperative surgical procedures, making it possible to analyze various aspects of technical performance during surgery, thereby enabling provision of detailed guidance and feedback. Skill evaluation scales were developed and investigated to validate educational achievements in specific procedures [5–8]. In the past decade, several studies reporting the correlation between surgical performance and postoperative outcomes have demonstrated the potential utility of intraoperative skill evaluation as a predictor of the surgical outcome [9–17]. However, most of these studies were limited to using global rating scales for skill evaluation of surgical procedures.

We previously developed the Japanese operative-rating scale for laparoscopic distal gastrectomy (JORS-LDG), which is a procedure-specific assessment tool for skill evaluation and training in laparoscopic gastrectomy [7]. However, data on its validity assessment and its correlations with the patient outcome are lacking. This study aimed to investigate the validity of the JORS-LDG as an assessment tool and the correlation of the assessment score and clinical outcome.

## Methods

### Ethical considerations

All procedures conformed to the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Helsinki Declaration of 1964 and later versions. The Ethics Committee of Hokkaido University (investigator's institute; IRB No. 015-0245) and 15 other participating facilities approved the study. Written-informed consent was obtained from the participating patients and surgeons, and all personal information was protected.

### JORS-LDG: The assessment tool

JORS-LDG was developed through cognitive task analysis and expert consensus using the Delphi method [18]. It can be used to measure intraoperative performance during LDG using a scoring sheet (Table 1) [7]. The scale consists of the following tasks: (1) patient and energy device settings, (2) trocar placement, (3) investigation of distant intraperitoneal metastases, (4) lymph node dissection, (5) reconstruction, and (6) final check of the intraperitoneal space. The six tasks comprise 33 items describing the detailed steps

involved in LDG using two or three grade criteria. Simple steps are scored on the two-point scale—0 if a given step is not performed and 1 if performed. The other items are graded on a three-point scale—0, unable to perform due to lack of knowledge or skill; 1, needs moderate guidance due to insufficient knowledge or skill; 2, able to perform independently without guidance. If the first assistant took over as the operator for a short time during the procedure to ensure patient safety, the score for that section was marked as zero. The points were recorded in the event of takeover by the first assistant to provide the video raters with a reference during evaluation. The total score is the sum of all scores, with a maximum of 46 points for Billroth-I and Billroth-II reconstructions or 52 for Roux-en-Y reconstruction.

### Data collection in LDG cases

We evaluated the LDG performance of surgeons with various skill levels using the JORS-LDG. To minimize the differences in complexity among cases, patients with obesity (body mass index [BMI] > 30 kg/m<sup>2</sup>) or a history of upper abdominal laparotomy were excluded from the study. Intraoperative complications were defined as unintended organ injury or hemorrhage (> 100 mL) that might impact safe surgical progress. Short-term postoperative complications were defined as those with Clavien–Dindo classification grade III and above within 30 days following the initial LDG.

### LDG skill evaluation

#### Rating in the operating room

The operator's intraoperative performance during LDG was evaluated by the first assistant in the operating room, who was an expert in the procedure. Operator self-evaluations were also undertaken.

#### Video rating

The unedited LDG videos of all participating subjects were evaluated blindly. After proving the high reliability of the JORS-LDG as an assessment tool by comparing the scores of three raters (A, B, and C), the JORS-LDG score of one rater, i.e., video rater A, was adopted for comparison with the other factors.

#### Rater criteria

The investigator established the following criteria for the first assistant in the operating room and the video raters: (A) operated over 100 LG cases, or (B) qualified as a master of endoscopic surgery by the Endoscopic Surgical Skill Qualification System (ESSQS) of the Japan Society of Endoscopic

**Table 1** JORS-LDG scoring sheet [7]

Subtask	Score
<i>Set up</i>	
1. Place the patient in an appropriate position	0 1
2. Check and set up the surgical instruments	0 1
3. Check the operation of the instruments	0 1
<i>Port insertion</i>	
4. Safely insert the first camera port	0 1 2
5. Set the pneumoperitoneum to appropriate level	0 1
6. Insert operation ports under direct vision	0 1 2
<i>Check distant metastasis</i>	
7. Check for distant metastases, including peritoneal dissemination and liver metastasis	0 1
<i>Liver retraction</i>	
8. Ensure good visualization by gentle liver retraction	0 1 2
<i>Greater curvature LN dissection (#4d, 4Sb)</i>	
9. Resection of omentum by ensuring at least 3 cm margin from the gastroepiploic vessels	0 1 2
10. Visual confirmation to avoid injury to transverse colon	0 1
11. Resection of LGEA	0 1 2
<i>Subpyloric region LN dissection (#6)</i>	
12. Takedown of the transverse colon by fused tissue dissection	0 1 2
13. Confirm the confluence pattern of the gastrocolic trunk	0 1 2
14. Identify the ASPDV and determine the lower margin of the dissection of #6 LN	0 1 2
15. Resect the RGEV	0 1 2
16. Confirm the branching of the ASPDA, RGEA, and IPA from GDA	0 1 2
17. Resect the RGEA and IPA	0 1 2
<i>Duodenum resection</i>	
18. Determine the resection line and resect the duodenum after confirming the pyloric ring location	0 1 2
<i>Suprapyloric region LN dissection (#5)</i>	
19. Resect the RGA after confirming its root	0 1 2
<i>Upper pancreatic margin LN dissection (#8a, 9)</i>	
20. Ensure good visualization of upper pancreatic margin by gentle pancreatic retraction	0 1 2
21. Dissect the #8a LN by dissecting the outer layer of the CHA nerve plexus	0 1 2
22. Resection of the LGA and LGV after performing dissection around the vessels	0 1 2
<i>Lesser curvature LN dissection (#1, 3)</i>	
23. Dissect the #1–3 LN by dissecting the lesser curvature	0 1 2
<i>Stomach resection</i>	
24. Division of the stomach after confirming adequate margin from the lesion	0 1
<i>Select to evaluate A. Roux-en-Y or B. Bilroth-I reconstruction</i>	
<i>A. Roux-en-Y reconstruction</i>	
25. Construct a tension-free Roux limb	0 1
26. Gastro-jejunal anastomosis, considering the location and diameter of the anastomosis	0 1 2
27. Jejunum-jejunal anastomosis, considering the tension and intestinal fluid reflux to the gastro-jejunal anastomosis	0 1 2
28. Closure of the mesenteric defect between the mesentery of the limbs and the jejunum	0 1 2
29. Closure of the Peterson's defect	0 1 2
<i>B. Bilroth-I (II) reconstruction</i>	
30. Confirm the absence of excessive tension between the remnant stomach and duodenum (jejunum)	0 1
31. Gastroduodenal (jejunal) anastomosis, considering its diameter	0 1 2
<i>Check final appearance</i>	
32. Confirm that there is no bleeding in the entire operation field	0 1
33. Confirm the final appearance	0 1
/52 points (R-Y)	
/46 points (B-I, II)	

*JORS-LDG* Japanese operative-rating scale for laparoscopic distal gastrectomy, *LN* lymph node, *LGEA* left gastroepiploic artery, *ASPDV* anterior branch of the superior pancreaticoduodenal artery, *RGEV* right gastroepiploic vein, *RGEA* right gastroepiploic artery, *IPA* infrapyloric artery, *GDA* gastroduodenal artery, *RGA* right gastric artery, *CHA* common hepatic artery, *LGA* left gastric artery, *LGV* left gastric vein

Surgery. The ESSQS was established in 2004 [22] and performs the most rigorous examinations for Japanese endoscopic surgeons, with an LG pass rate of 21% in 2019.

### Guidance for raters

At the beginning of the study, the investigator guided the raters individually on the evaluation criteria through 20-min telephone conversations. The guidance confirmed the primary evaluation points: (A) the extent of understanding the anatomy and concept of each aspect, and (B) the level of autonomy and safety. No rater guidance was provided to the LDG surgeons about the self-evaluation.

### JORS-LDG scoring items and calculation adjustments

Six of the JORS-LDG evaluation items related to patient positioning, set up of surgical instruments, and first trocar placement were not recorded in the LDG videos; therefore, they were excluded from the evaluations in this study. The evaluation points for the Billroth-I and Billroth-II method (perfect, 3 points) were tripled to match the Roux-en-Y method (perfect, 9 points) to unify their total points for analysis. Following these adjustments, the maximum unified JORS-LDG score was 44 points.

### Setting of the raters for the investigation of JORS-LDG reliability and validity

Based on the sample size calculation, 17 LDG videos were chosen randomly from a total of 54 videos and evaluated by blinded raters (A, B, and C). After determining the high reliability of the JORS-LDG scores of three raters, rater A blindly evaluated the remaining 37 videos alone. Therefore, the JORS-LDG scores of 54 LDG videos, which were evaluated by rater A, were used for the comparison with the direct observation scores and surgical outcomes.

### Categorization into three groups based on the LDG skill level

The participating surgeons were categorized into three groups depending on the total JORS-LDG score for video evaluation. We arranged the total scores of the participants in a descending order and divided them evenly into three groups—high, intermediate, and low.

### Statistical analysis

The results are reported as medians and interquartile ranges (IQRs) or means and standard deviations (SDs). Statistical significance was set at  $P < 0.05$ . All analyses were performed using IBM SPSS Statistics for Macintosh, Version 26.0 (IBM Corp., Armonk, NY, USA) in consultation with a statistician. Intraclass correlation coefficients (ICCs) and 95% confidence intervals (CIs) were calculated by assessing the JORS-LDG scores of the three independent video raters to estimate the inter-rater reliability. The internal consistency of the JORS-LDG items was estimated using Cronbach's  $\alpha$ . The correlations between the evaluation methods and relevance of the association between the total score and each aspect of LDG were investigated using Spearman's correlation coefficient. The Kruskal–Wallis test compared the patient characteristics and surgical factors among the three LDG skill-level groups.

The sample size for the investigation of inter-rater reliability was calculated using the formula presented by Walter et al. [19]. The minimum acceptable ICC was set to 0.5, and the expected ICC was 0.8 in the study. The required sample size by three raters with the alpha error of 0.05 and power of 0.7 was 15.

Since this was the first study to examine the intraoperative performance of various surgeons during LDG using the JORS-LDG, it was difficult to hypothesize the relation between the assessment score and operative complications of LDG. Therefore, post-hoc power analysis was planned for operative outcomes.

### Results

We analyzed 54 LDG procedures performed by 40 surgeons at 16 institutions from January 2016 to December 2018. The operator characteristics and experience are shown in Table 2. Most surgeons were board certified with a median experience of 50 open and LG cases, and 40% were surgeons qualified by the ESSQS [20].

#### Inter-rater reliability for the JORS-LDG

Three blinded raters (A, B, and C) evaluated 17 videos. The ICCs for the total JORS-LDG scores of the three raters were higher than 0.8, with excellent internal consistency of the JORS-LDG items, and a Cronbach's  $\alpha$  of 0.94 (Table 3).

#### Correlations between the evaluation scores of direct observations, videos, and self-evaluations

We observed good correlations between the video evaluations and first assistant evaluated JORS-LDG scores ( $R = 0.69$ ,  $P < 0.001$ ) and self-evaluations ( $R = 0.69$ ,  $P < 0.001$ ; Fig. 1). The correlation between the

**Table 2** Characteristics of surgeons

Characteristic	Number
Surgeons	40
Mean post-graduate year, n [SD]	15.5 [6.6]
Board Certified Surgeon, n (%)	37 (92.5)
Qualified surgeon of endoscopic surgical skill qualification system, n (%)	16 (40)
Surgical case number experience as an operator	Median (IQR)
Laparoscopic surgery	300 (120–500)
Open gastrectomy	50 (20–100)
Laparoscopic gastrectomy	50.0 (8.5–104)

IQR Interquartile range, SD Standard deviation

**Table 3** Inter-rater reliability for JORS-LDG

Rater	ICC	95% CI
Rater A—Rater B	0.85	0.61–0.94
Rater B—Rater C	0.84	0.60–0.94
Rater A—Rater C	0.82	0.56–0.93
Cronbach's $\alpha$	0.94	

JORS-LDG Japanese operative-rating scale for laparoscopic distal gastrectomy, ICC intraclass correlation coefficient, CI confidence interval

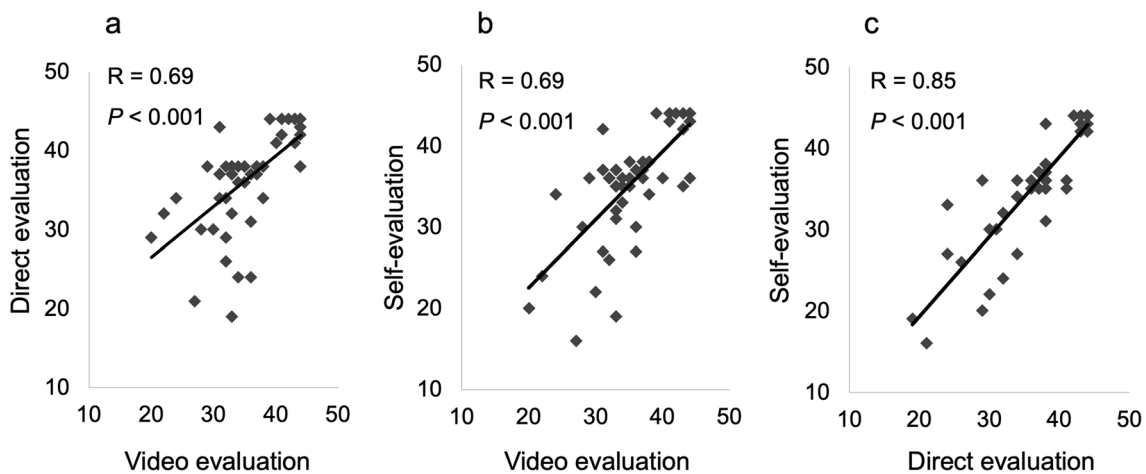
self-evaluation and the first assistant direct evaluation was excellent ( $R = 0.85$ ,  $P < 0.001$ ).

### Correlations between the total JORS-LDG score and each aspect score

The total JORS-LDG score had an excellent correlation with the scores for lymph node dissection in the infrapyloric (station no. 6;  $R = 0.81$ ,  $P < 0.001$ ) and upper pancreatic edge (station no. 8a + 9;  $R = 0.75$ ,  $P < 0.001$ ) regions (Table 4).

### Laparoscopic distal gastrectomy skill level groups

The participating surgeons were divided into three LDG skill-level groups: high (JORS-LDG score, 42–44;  $n = 19$ ), intermediate (JORS-LDG score, 39–41;  $n = 17$ ), and low (JORS-LDG score,  $\leq 38$ ;  $n = 18$ ) according to their total JORS-LDG scores. The first assistant took over the operation for a short time in five of the low-group cases to ensure patient safety.



**Fig. 1** Correlations between the scores for direct observation, video, and self-evaluation. Good correlation was observed between the JORS-LDG score for direct observation and video evaluation (a, b),

and excellent correlation existed between the self-evaluations and direct observations (c). JORS-LDG Japanese operative-rating scale for laparoscopic distal gastrectomy

**Table 4** Correlation between the total JORS-LDG score and the score for each aspect

JORS-LDG scores	Median (IQR)	R	P value
Total score (44)	40 (37–43)		
Lesser curvature and suprapyloric: #1, #3, #5 (4)	4 (3–4)	0.68	<0.001
Greater curvature: #4 (5)	5 (4–5)	0.69	<0.001
Infrapyloric: #6 (12)	10 (9–12)	0.81	<0.001
Upper pancreatic edge: #8a, #9 (6)	5 (4–6)	0.75	<0.001
Reconstruction (9)	9 (9–9)	0.49	<0.001

Numbers in parentheses after each scoring item indicate the perfect score

*JORS-LDG* Japanese operative-rating scale for laparoscopic distal gastrectomy, *IQR* interquartile range, *R* correlation coefficient

#1, 3, 4, 5, 6, 7, 8a, 9 Lymph node station number

## Patient characteristics

The patients' characteristics are summarized in Table 5. The characteristics, medical history, and pathological stages were similar among the three LDG skill level groups.

## Correlation between the total JORS-LDG scores and surgical outcomes

Comparison of the surgical factors revealed that the number of laparoscopic surgery ( $P < 0.001$ ) and LG ( $P < 0.001$ ) cases differed significantly among the three LDG skill-level groups (Table 6). The high group performed more

Roux-en-Y reconstructions than the other two groups (low, 22.2%; intermediate, 23.5%; high, 57.9%;  $P = 0.01$ ). The low, intermediate, and high groups differed significantly in terms of the median operating time (311, 266, and 229 min, respectively;  $P < 0.001$ ), rate of unintended intraoperative organ injury or hemorrhage  $> 100$  mL (27.8, 11.8, and 0%, respectively;  $P = 0.01$ ), and postoperative complication rate (22.2, 0, and 0%, respectively;  $P = 0.002$ ). There were seven cases with hemorrhage  $> 100$  mL, two with splenic injury, and one with duodenal injury. These include multiple complications in the same case. Postoperative complications included anastomotic leakage in two cases, pancreatic fistula in two, intraperitoneal abscess in one, and surgical site infection in one. More than one complication occurred in a single case. The post-hoc power of this study was 0.71 with an alpha error of 0.05.

## Discussion

This was the first study to demonstrate the reliability and validity of an LG-specific performance scale (JORS-LDG) and the correlation between intraoperative performance and clinical outcomes for LG.

Performance evaluation, which plays an essential role in surgical education, requires reliability among raters because proper evaluation and feedback on trainees' performance cannot be provided without a reliable measure [21, 22]. ICCs  $> 0.8$  were demonstrated for the three raters during the video evaluation of LDG performance using

**Table 5** Patient characteristics

Characteristic	Total ( <i>n</i> = 54)	Three LDG skill level groups			<i>P</i> value
		Low ( <i>n</i> = 18)	Intermediate ( <i>n</i> = 17)	High ( <i>n</i> = 19)	
Mean age, years [SD]	69.3 [10.1]	68.8 [7.8]	69.1 [11.5]	69.9 [11.2]	0.79
Male sex, <i>n</i> (%)	36 (67)	12 (66.7)	14 (82.4)	10 (52.6)	0.17
Mean BMI [SD]	22.0 [3.1]	21.7 [2.6]	22.7 [2.9]	21.6 [3.6]	0.50
<i>Medical history, n (%) (multiple answers allowed)</i>					
Diabetes	8 (14.8)	3 (16.7)	4 (23.5)	1 (5.3)	0.10
Cardiovascular disease	12 (22.2)	4 (22.2)	1 (5.9)	7 (36.8)	0.27
Cerebrovascular disease	1 (1.9)	0 (0)	0 (0)	1 (5.3)	0.54
Respiratory disease	1 (1.9)	0 (0)	1 (5.9)	0 (0)	0.54
Digestive disease	5 (9.3)	1 (5.6)	1 (5.9)	3 (15.8)	0.21
Other malignant disease	4 (7.4)	0 (0)	2 (11.8)	2 (10.5)	0.36
Others	5 (9.3)	1 (5.6)	2 (11.8)	2 (10.5)	0.21
<i>Pathological stage, n (%)</i>					
IA	38 (70.4)	12 (66.7)	13 (76.5)	13 (68.4)	0.79
IB	8 (14.8)	4 (22.2)	2 (11.8)	2 (10.5)	
IIA	3 (5.6)	1 (5.6)	1 (5.9)	1 (5.3)	
IIB	5 (9.3)	1 (5.6)	1 (5.9)	3 (15.8)	

*LDG* laparoscopic distal gastrectomy, *SD* standard deviation, *BMI* body mass index (kg/m<sup>2</sup>)



**Table 6** Surgical factors

Factor	Total (n = 54)	Three LDG skill level groups			P value
		Low (n = 18)	Intermediate (n = 17)	High (n = 19)	
JORS-LDG score	40 (37–43)	35 (33–37)	40 (39–40)	44 (43–44)	<0.001
<i>Number of surgical cases performed by operator</i>					
Laparoscopic surgery	300 (120–500)	100 (90–300)	250 (145–425)	600 (300–1,050)	<0.001
Open gastrectomy	50 (20–100)	40(17–50)	100 (37–180)	100 (20–100)	0.06
Laparoscopic gastrectomy	50 (8.5–104)	9.5 (3–25)	60 (50–117)	51 (30–225)	<0.001
<i>Extent of lymphadenectomy, n (%)</i>					
D1 +	42 (77.8%)	14 (77.8%)	14 (82.4%)	14 (73.7%)	0.83
D2	12 (22.2%)	4(22.2%)	3(17.6%)	5(26.3%)	
<i>Type of reconstruction, n (%)</i>					
Billroth -I	34 (63.0%)	14 (77.8%)	14 (82.4%)	14 (73.7%)	0.01
Billroth -II	1 (1.9%)	0 (0%)	0(0%)	1 (5.3%)	
Roux-en-Y	19 (35.2%)	4 (22.2%)	4 (23.5%)	11 (57.9%)	
<i>Surgical outcome</i>					
Operation time (min)	281 (227–310)	311 (290–331)	266 (206–298)	229 (205–272)	<0.001
Number of harvested lymph node	D1 + lymphadenectomy				
	33.5 (27–45)	28.5 (26–41)	34 (30–38)	39 (24–49)	0.66
Hemorrhage volume (mL)	D2 lymphadenectomy				
	42 (31–54)	35.5 (28–45)	37 (28–43)	69.5 (50–85)	0.15
Intraoperative complications, n (%) <sup>a</sup>	10 (1–47.5)	25 (0–105)	25 (7.5–50)	10 (5–15)	0.21
Postoperative surgical complication, n (%) <sup>b</sup>	7 (13.0)	5 (27.8)	2 (11.8)	0(0)	0.01
Postoperative stay (day)	4(7.4)	4 (22.2)	0 (0)	0(0)	0.002
	11 (7–14)	11 (9–19)	8 (6–12)	11 (10–14)	0.16

JORS-LDG score, experienced surgical case number; operation time, harvested lymph node number, hemorrhage volume, and postoperative stay are reported as median (IQR)

*JORS-LDG* Japanese operative-rating scale for laparoscopic distal gastrectomy, *LDG* laparoscopic distal gastrectomy, *IQR* interquartile range

<sup>a</sup>Intraoperative complications: case number (%) of unintended organ injury or hemorrhage > 100 mL

<sup>b</sup>Postoperative surgical complication: ≥ grade III of the Clavien–Dindo classification

the JORS-LDG. This result revealed that the JORS-LDG, which was developed by us [7] using the Delphi method [18], was easy to understand and apply by different evaluators. Further, the LDG experts of the participating facilities were asked to offer 20-min telephonic guidance and serve as the first assistant in every case to maintain the quality of surgery and conduct direct evaluation. In Japan, general surgeons usually acquire board certification after completing a 5-year surgical residency program. However, most board

certificated surgeons cannot perform LDG independently without advanced training in minimally invasive surgery. Therefore, an expert in minimally invasive surgery usually acts as the first assistant during an LDG case to teach and control the quality of the procedure, similar to the setting of this study. The excellent inter-rater reliability observed in this study clearly implies that telephonic guidance contributed toward raters' understanding of the definitive criteria of the JORS-LDG, which includes some subjective

criteria. Therefore, a user manual that defines each step and aspect of the evaluation criteria of the JORS-LDG is required to ensure widespread adoption of the JORS-LDG in various educational facilities in the future. Furthermore, we observed excellent correlation between direct and self-evaluations ( $R=0.85$ ,  $P<0.001$ ). Even after consideration of bias in the direct and self-evaluations, the results suggested that the operators could perform LDG based on proper communication with the first assistants, mutually confirming what they understood and what they could manage at each step of the operation. This result demonstrated that the JORS-LDG, initially developed as a formative assessment tool, could be suitable for this very purpose.

The high, intermediate, and low skill-level groups stratified according to the JORS-LDG scores exhibited significant differences in the total number of performed laparoscopic surgeries and LGs. This demonstrated the construct validity with correlation between the JORS-LDG score and experience in performing laparoscopic surgery and LG. Birkmeyer et al. [11], Fecso et al. [14], and Curtis et al. [15] argued that the surgical outcomes were affected by intraoperative performance and not by the duration of training and history. Based on the abovementioned factors, it is important to provide surgical trainees with abundant case experience in addition to competency-based training focusing on the performance of each procedure on advanced procedures such as LDG.

It is well known that lymph node dissection in the infrapyloric and upper pancreatic edge regions is a technically difficult aspect of LG [23–25], and unsafe execution of these steps could cause intraoperative and postoperative complications such as hemorrhage and pancreatic fistula [26–28]. It was interesting to note that the scores of these two steps correlated with the total JORS-LDG score more closely than the scores for the other steps. From an educational perspective, deliberate training focusing on such aspects that would significantly impact the total score could aid in efficient acquisition of LDG skills. Specific analysis of the JORS-LDG scores and errors at each step could be useful for predicting the onset of complications caused by the individual LDG steps and prognosticating the clinical outcomes of the surgery in future research.

Over the past decade, studies have demonstrated a correlation between intraoperative performance and short-term outcomes, particularly the postoperative complication rates [11–17]. Fecso et al. [14] examined 61 LG procedures for patients with gastric cancer performed by three surgeons at three institutions. The researchers used the Objective-Structured Assessments of Technical Skills [1] and Generic Error-Rating Tool [29] and demonstrated a relationship between intraoperative LG performance and postoperative complications. Our study demonstrated that the intraoperative performances of 40 surgeons in 54 LDGs, scored by the JORS-LDG, were correlated with intraoperative and

postoperative complications. Our study is valuable since it demonstrated the utility of the JORS-LDG for evaluating surgeons at various skill levels at 16 institutions.

Curtis et al. [15] first demonstrated the correlation between intraoperative performance and short-term clinical outcomes using a procedure-specific skill measure for laparoscopic total mesorectal excision. As the only procedure-specific rating scale for LG skill, the JORS-LDG proved reliable, valid, and demonstrated a correlation with the short-term clinical outcomes. Although few rating scales exist for procedure-specific skills because of the extensive labor required for their development [30], they possess tremendous potential to provide specific educational feedback for surgery and detailed analysis of the causes for surgical complications. Moreover, the robust correlation between the total JORS-LDG scores and surgical outcomes suggested that the scale could play a role as a prognostic factor, operator's criterion in surgical training, and quality control of surgical intervention in clinical trials.

Few studies have investigated the correlation between intraoperative performance and long-term surgical outcomes [15–17]. Nevertheless, the correlation between intraoperative performance and long-term outcomes has not been proven in surgeries for malignant diseases [15]. Future research should examine the relationship between procedure-specific performance and postoperative complications by type and site and the short- to long-term outcomes related to the complications associated with malignant diseases.

This study has several limitations. First, since participating surgeons were volunteers asked to present LDG videos for this research, video selection for evaluation could have been biased. To minimize bias, the LDG cases chosen for evaluation should have been randomly selected from various cases of participating surgeons. However, this was not possible due to the limited number of LDG cases available for each participating surgeon. Moreover, our data analysis revealed correlations between LDG performance and short-term clinical outcomes. These correlations suggest that the quality of the participating surgeon's performance in each LDG reflected the clinical outcomes of the case.

Second, the quality of LDG performance by non-expert surgeons may have been influenced by the team's ability, including the assistants' support and advice, and the laparoscope intra-abdominal view in this complicated procedure. However, there was no prior confirmation regarding the extent and type of support and advice provided by the assistants during LDG in the operating room. The relationship between team performance, including assistants' support, and the overall quality of surgery should be evaluated in future research. Moreover, the expert assistants decided that patient safety was of utmost importance and took over the operation for a short time in five low-group cases. Although all operator alternation times were recorded and accurately



reflected in the direct and video evaluations, the possible effects of operator alternations on the intraoperative and postoperative outcomes were not investigated in this study. Appropriate and timely operator alternations in challenging situations could prevent unintended intraoperative complications. Investigation into the occurrence of complications due to lack of experience and the appropriate skill level in each step will require detailed analysis using a procedure-specific assessment scale.

## Conclusion

This study demonstrated the reliability and validity of the JORS-LDG, exhibiting excellent correlation between its scores and short-term surgical outcomes. This newly developed JORS-LDG could be useful in surgical training and surgical outcome prediction, potentially improving patient outcomes.

**Acknowledgements** The authors would like to thank the surgeons who participated in this study. We would have been unable to complete this study without their support.

**Funding** This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Declarations

**Disclosures** Drs. Yo Kurashima, Hidehiko Kitagami, Koichi Teramura, Saseem Poudel, Yuma Ebihara, Noriyuki Inaki, Fumitaka Nakamura, Kazunari Misawa, Kazunori Shibao, Eishi Nagai, Yoshimasa Akashi, Shuji Takiguchi, Shinichiro Kobayashi, and Satoshi Hirano have no conflicts of interest or financial ties to disclose.

## References

- Martin JA, Regehr G, Reznick R, Macdrae H, Murnaghan J, Hutchison C, Brown M (1997) Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg* 84:273–278
- Doyle JD, Webber EM, Sidhu RS (2007) A universal global rating scale for the evaluation of technical skills in the operating room. *Am J Surg* 193:551–555 (**discussion 555**)
- Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ (2012) The Ottawa surgical competency operating room evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med* 87:1401–1407
- Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 190:107–113
- Kurashima Y, Feldman LS, Al-Sabah S, Kaneva PA, Fried GM, Vassiliou MC (2011) A tool for training and evaluation of laparoscopic inguinal hernia repair: the global operative assessment of laparoscopic skills-groin hernia (GOALS-GH). *Am J Surg* 201:54–61
- Palter VN, MacRae HM, Grantcharov TP (2011) Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg* 201:251–259
- Kurashima Y, Watanabe Y, Hiki N, Poudel S, Kitagami H, Ebihara Y, Murakami S, Shichinohe T, Hirano S (2019) Development of a novel tool to assess skills in laparoscopic gastrectomy using the Delphi method: the Japanese operative rating scale for laparoscopic distal gastrectomy (JORS-LDG). *Surg Endosc* 33:3945–3952
- Poudel S, Kurashima Y, Kawarada Y, Watanabe Y, Murakami Y, Matsumura Y, Kato H, Miyazaki K, Shichinohe T, Hirano S (2016) Development and validation of a checklist for assessing recorded performance of laparoscopic inguinal hernia repair. *Am J Surg* 212:468–474
- Fecso AB, Szasz P, Kerezov G, Grantcharov TP (2017) The effect of technical performance on patient outcomes in surgery: a systematic review. *Ann Surg* 265:492–501
- Arvidsson D, Berndsen FH, Larsson LG, Leijonmarck CE, Rimback G, Rudberg C, Smedberg S, Spangen L, Montgomery A (2005) Randomized clinical trial comparing 5-year recurrence rate after laparoscopic versus Shouldice repair of primary inguinal hernia. *Br J Surg* 92:1085–1091
- Birkmeyer JD, Finks JF, O'Reilly A, Oerline M, Carlin AM, Nunn AR, Dimick J, Banerjee M, Birkmeyer NJO (2013) Surgical skill and complication rates after bariatric surgery. *N Engl J Med* 369:1434–1442
- Mackenzie H, Ni M, Miskovic D, Motson RW, Gudgeon M, Khan Z, Longman R, Coleman MG, Hanna GB (2015) Clinical validity of consultant technical skills assessment in the english national training programme for laparoscopic colorectal surgery. *Br J Surg* 102:991–997
- Hogg ME, Zenati M, Novak S, Chen Y, Jun Y, Steve J, Kowalsky SJ, Bartlett DL, Zureikat AH, Zeh HJ (2016) Grading of surgeon technical performance predicts postoperative pancreatic fistula for pancreaticoduodenectomy independent of patient-related variables. *Ann Surg* 264:482–491
- Fecso AB, Bhatti JA, Stotland PK, Quereshey FA, Grantcharov TP (2019) Technical performance as a predictor of clinical outcomes in laparoscopic gastric cancer surgery. *Ann Surg* 270:115–120
- Curtis NJ, Foster JD, Miskovic D, Brown CSB, Hewett PJ, Abbott S, Hanna GB, Stevenson ARL, Francis NK (2020) Association of surgical skill assessment with clinical outcomes in cancer surgery. *JAMA Surg* 155:590–598
- 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* 155:960–968
- 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* 273:766–771
- Diamond IR, Grant RC, Feldman BM, Pencharz PB, Ling SC, Moore AM, Wales PW (2014) Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol* 67:401–409
- Walter SD, Eliasziw M, Donner A (1998) Sample size and optimal designs for reliability studies. *Stat Med* 17:101–110
- Mori T, Kimura T, Kitajima M (2010) Skill accreditation system for laparoscopic gastroenterologic surgeons in Japan. *Minim Invasive Ther Allied Technol* 19:18–23
- Szasz P, Louridas M, Harris KA, Aggarwal R, Grantcharov TP (2015) Assessing technical competence in surgical trainees: a systematic review. *Ann Surg* 261:1046–1055

22. 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-analyses. *J Surg Educ* 77:189–201
23. Shibasaki S, Suda K, Nakauchi M, Nakamura T, Kadoya S, Kikuchi K, Inaba K, Uyama I (2018) Outermost layer-oriented medial approach for infrapyloric nodal dissection in laparoscopic distal gastrectomy. *Surg Endosc* 32:2137–2148
24. Kitano S (2009) What technique is suitable for laparoscopic suprapancreatic lymph node dissection? *Gastric Cancer* 12:67–68
25. Fukunaga T, Hiki N, Tokunaga M, Nohara K, Akashi Y, Katayama H, Yoshida H, Yamada K, Ohyama S, Yamaguchi T (2009) Left-sided approach for suprapancreatic lymph node dissection in laparoscopy-assisted distal gastrectomy without duodenal transection. *Gastric Cancer* 12:106–112
26. Guerra F, Giuliani G, Iacobone M, Bianchi PP, Coratti A (2017) Pancreas-related complications following gastrectomy: systematic review and meta-analysis of open versus minimally invasive surgery. *Surg Endosc* 31:4346–4356
27. Washio M, Yamashita K, Niihara M, Hosoda K, Hiki N (2020) Postoperative pancreatic fistula after gastrectomy for gastric cancer. *Ann Gastroenterol Surg* 4:618–627
28. Hiki N, Honda M, Etoh T, Yoshida K, Kodera Y, Kakeji Y, Kumamaru H, Miyata H, Yamashita Y, Inomata M, Konno H, Seto Y, Kitano S (2018) Higher incidence of pancreatic fistula in laparoscopic gastrectomy. Real-world evidence from a nationwide prospective cohort study. *Gastric Cancer* 21:162–170
29. Bonrath EM, Zevin B, Dedy NJ, Grantcharov TP (2013) Error rating tool to identify and analyse technical errors and events in laparoscopic surgery. *Br J Surg* 100:1080–1088
30. Ghaderi I, Manji F, Park YS, Juul D, Ott M, Harris I, Farrell T (2015) Technical skills assessment toolbox: a review using the unitary framework of validity. *Ann Surg* 261:251–262

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.