—Original—

Use of a Contest Format for Objective Assessment of Microsurgical Technique: An Observational Study

Yasuo Murai[1], Eitaro Ishisaka[1], Atsushi Tsukiyama[1],
Asami Kubota[1], Masahiro Yamaguchi[2], Fumihiro Matano[1],
Tomonori Tamaki[3], Takayuki Mizunari[2] and Akio Morita[1]

[1]Department of Neurological Surgery, Nippon Medical School Hospital, Tokyo, Japan
[2]Department of Neurosurgery, Nippon Medical School Chiba Hokusoh Hospital, Chiba, Japan
[3]Department of Neurosurgery, Nippon Medical School Tama Nagayama Hospital, Tokyo, Japan

Background: Few studies have used simulation models to examine long-term improvement in microsurgical technique. We investigated whether improvement in surgical technique could be assessed by continuous, objective, contest-format evaluation of the same microsurgical task.

Methods: Since 2014, neurosurgeons with 1-10 years of experience participated in a biannual competition-format test. The task involved creating as many sutures as possible during the 5-minute interval after arteriotomy of a 1-mm artificial vessel. A modified version of the Objective Structured Assessment of Technical Skills examination was created and used. Changes and differences in scores over time were examined for each evaluator.

Results: Overall, 103 neurosurgeons participated in the study at least once, and those who participated more than once were divided into two groups: those who had the highest score in each contest and those who had the lowest score. The linear regression equations for the highest and lowest scorers were $y=7.62x+81.56$ ($R^2=0.628$) and $y=1.94x+67.93$ ($R^2=0.0433$), respectively. High scorers had high scores from the first time they participated, and their scores tended to increase further, while scores for low scorers tended not to increase with additional experience. Scores for the four evaluators did not significantly differ.

Conclusions: Our results suggest that technical improvement in surgery can be assessed by long-term, continuous evaluation of microsurgical technique and that the present evaluation system might help increase surgical safety. (J Nippon Med Sch 2022; 89: 405–411)

Key words: microsurgery, objective assessment, skills, techniques

Introduction

Many studies of surgical education[1–3] have found that effective surgical training reduces technical errors. Methods for evaluating surgical technique have been reported[1,4–7] in fields such as gynecology, laparoscopy, and general surgery. These reports[8,9] examined evaluation tables[3,8], differences between video evaluation and direct evaluation, and evaluation task contents[10,11]; however, no study assessed the results of objective evaluations repeated over a number of years. The objectivity of a scoring method cannot be evaluated in a single technical evaluation because the tendencies of the participants and scorers at the time of the technical evaluation might affect the results[12]. For this reason, a method evaluating video recordings of the same trial at intervals of 1-3 months has been attempted[8,10,12]. However, no previous study of surgical technique evaluation by simulation tracked the results of scoring for longer than 6 months for the same participant. Thus, in the present study, we examined the results for participants who took part multiple times during an

Correspondence to Yasuo Murai, Department of Neurological Surgery, Nippon Medical School Hospital, 1–1–5 Sendagi, Bunkyo-ku, Tokyo 113–8602, Japan
E-mail: ymurai@nms.ac.jp
https://doi.org/10.1272/jnms.JNMS.2022_89-407
Journal Website (https://www.nms.ac.jp/sh/jnms/)

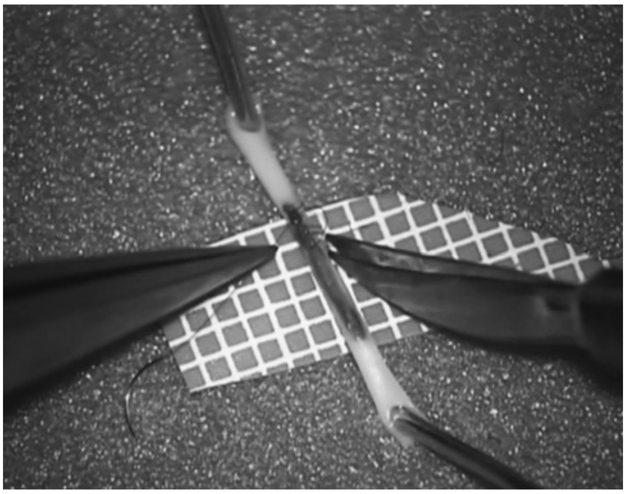

Fig. 1 Task involving suturing of an artificial vessel using 10-0 nylon sutures

interval longer than 6 months. We also investigated whether objective evaluation was performed, by comparing evaluation results among four evaluators and comparing the final comprehensive evaluation with the results of each evaluator, when the evaluators were unchanged for 4 years.

This study investigated whether technical improvements in surgical technique could be assessed by scores on a continuous, objective, contest-format evaluation of the same microsurgical task. Ultimately, the purpose of this study was to establish a technical evaluation system that would allow residents to qualify as a surgical operator and that would improve patient outcomes and surgical safety.

Materials and Methods

Modified Objective Structured Assessment of Technical Skills (OSATS) evaluation forms were created to assess individual technical improvement[12-14] and were used to evaluate participants in a contest format[15,16]. The results of long-term observations are discussed and reported.

Participants and Evaluators

Residents and neurosurgeons with 1-10 years of experience participated in the competitions, which have been held twice a year since 2014. A championship has been held nine times previously. All study participants provided informed consent, and the study design was approved by the institutional ethics review board (A-2019-009).

We chose evaluators with (a) experience as the first surgeon for more than 500 cerebrovascular-related surgeries, (b) experience as the main operator for more than 30 cerebrovascular reconstructive surgeries, (c) board certification by a neurosurgical society, and (d) qualification as a stroke surgeon technical mentor. Four evaluators remained throughout the competition, to measure improvement in the technical skills of each participant over time.

Competition Tasks

Microscopic procedures were performed by all participants and were scored; participants could observe the practical skills of their competitors during the practical tests. Each participant was given 5 minutes to complete the task; thus, the competition was limited to approximately 2 hours and did not interfere with daily duties. A model performance video was created to encourage advance practice by first-time participants.

The examination content was determined by considering reproducibility, time, cost, evaluation objectivity, and similarity to surgery. An actual surgical microscope (OPMI PENTERO 900, Carl Zeiss Meditec, Germany) was used. The examination included arteriotomy for a 1 mm-diameter artificial blood vessel linearly along the long axis and stitching the incision line with 10-0 nylon sutures (**Fig. 1**). The commercially available artificial blood vessel comprises one layer with a 1-mm diameter and simulates a blood vessel for anastomosis (Wetlab Co., Ltd., Shiga, Japan). To purchase consumables, we used the budget for educational equipment purchases provided by the hospital. Suturing was repeated for an interval of 5 minutes. To confirm the speed of the procedure, the incision area was sutured as many times as possible within 5 minutes. To ensure that forced use of unfamiliar tools did not affect the technical evaluation, the microsurgical instruments were not standardized; the participants chose the tools with which they were familiar. The contest organizers also provided forceps, microsurgical scissors, and 10-0 nylon sutures. A video of the entire procedure was recorded. During the week after the competition, a reflection meeting was held while watching the videos of all participants. The reflection meeting was intended for the participants to learn each other's strengths and weaknesses by viewing their peers' videos.

Scoring

The scoring chart (**Fig. 2**) was formatted as a score check only, so that the evaluator could complete it quickly. The evaluation table was created using the modified OSATS examination[5].

Using this table (**Fig. 2**), the tasks were evaluated based on multiple factors, such as preparation (posture of the operator and instrument layout), microscope operation (focusing and positioning), motion (hand tremor), instrument knowledge (linear incision using scissors), nee-

8th 5min championship Grade Slip

□ [redacted] □ [redacted] □ [redacted] □ [redacted]

Grader A □ B □ C □ D □

	Failing	Bad	Poor: below average	Fair: average	Good: Fairly good	Excellent: Almost expert
	0	1	2	3	4	5
1. Surgeons Positioning / Instrument layout (preparation)						
2. Microscope operation / Focusing, Positioning, Hand movement (knowledge of microscope)						
3. Hand tremor (motion)						
4. Draw a line, linear incision, how to use scissors (knowledge of microinstrument)						
5. Handring of Needle / Insettion Angle / Vessel Treatment (handring of needle and thread)						
6. Pull and Stop the thread (respect for vessel wall)						
7. Ligation/ make a loop, pull direction (flow of suture)						
8. Final ligature/ Slack, direction, length (knowledge of completion)						
Handicap for young neurosurgeon	2 point (5-6 year)	4 point (3-4 year)			Overall Points	
Grader Comments						

Fig. 2 Scoring chart used by the evaluators

dle handling (insertion angle and thread treatment), attention to the vessel wall, suture flow (ligation and pull direction), and knowledge of the completeness of the task. Because the OSATS examination is designed for endoscopic surgery[5,8], assistant usage is usually included as an evaluation item. However, we excluded this item because microsurgery is often performed alone. In addition, we excluded the time item because the time allotted for the procedure was part of the competition design. The final ranking was determined by summing the scores of the four evaluators. A perfect score for each evaluator was 40 points; thus, the highest possible score was 160 points and the lowest score was 0 points. First-year doctors were allotted 4 handicap points, while second-year doctors received 2 handicap points. Handicap points were allotted to motivate beginners to practice in order to win. Therefore, the highest potential score was 164 points. To maintain the independence of the evaluation results, graders were prohibited from speaking until all participants had completed the practical skills and the grader had completed the grade sheet. It was forbidden to collect or modify the scoring table after each participant had finished. The reflection meeting was held a week after the contest. Final scores were announced before the reflection meeting.

Statistical Analyses

JMP 14.0 software (SAS Institute Inc., Cary, NC, USA) was used to perform all statistical analyses. All data were presented as mean ± SD, but the median and standard error were also added to the table. A p-value of <0.05 was considered significant. To observe improvements in microsurgical technique, only physicians who participated in the competition more than once were included in the statistical analyses.

We used Student's t-test to compare a competitor's scores from the first and final competitions. For doctors who participated only twice, the second round was considered the final round. Additionally, we divided participants into two groups based on their scores. Participants with the highest score in any competition in which they participated were assigned to group H, while those with the lowest score in any competition were assigned to group L. Scores at the first time of participation, number of years of experience as a physician, and increase in scores from the first to the final participation were compared between groups L and H, using the Wilcoxon test. A linear analysis was performed by plotting the scores. For both groups, all scores were plotted in relation to the number of participations, and linear regression analysis was performed. The independent variable was number of participations, and the dependent variable was total number of points. Coefficients of determination (R^2) were obtained based on linear regression analysis. Furthermore, number of years of experience at the time of the

Table 1 Participant scores for their first, second, and final rounds of competition

Competition Round	Mean	Median	SE	SD	Time from First Round (Months)	p-value*
First	84.52	83	4.56	19.37		
Second	92.89	96	5.84	24.81	6.4	0.266
Final	96.63	94	6	25.46	20.6	0.11

SE, standard error; SD, standard deviation. *Student's t-test

Table 2 Participant scores from group H and group L

	Group H	Group L	p-value*
Participants	6	8	
First round score	111.75±14.31	69.83±11.29	0.0142
Doctor's years of experience in the first round	2.0±1.73	1.42±0.66	0.559
Score increase from the first to the final round	24.95±10.11	9.095±19.86	0.164
Score for all rounds	124.75±14.40	80.205±19.17	0.0001
Doctor's years of experience in all rounds	5.67±1.50	6.33±2.06	0.549
R^2 in linear regression	0.3486	0.0111	
R^2 corrected for doctor's years of experience	0.6284	0.0433	

Data are presented as mean ± standard deviation. *Wilcoxon test

competition was adjusted for, and the evaluation results were compared using linear regression analysis. In this case, the independent variable was years of experience at the time of participation, and the dependent variable was total score. R^2 values were again obtained based on linear regression analysis.

Scores from each of the four evaluators were compared. We used Pearson correlation coefficients to analyze correlations between each evaluator and the total score for each participant.

Results

The contest was held nine times from 2014 to 2019. Among all participants, those who participated at least twice were included in this study. Ultimately, 21 neurosurgeons participated in the contest a total of 103 times. Scores for the first (mean ± SD), second, and final rounds for all participants were 84.52 ± 19.37, 92.89 ± 24.81, and 96.63 ± 25.46, respectively. Student's t-test showed no significant differences between scores during the first and second rounds (p = 0.27) and between the first and last rounds (p = 0.11) (**Table 1**).

Participants who earned a highest score at least once in a competition were assigned to group H, while those who earned a lowest score at least once were assigned to group L. Group H comprised six participants, and group L comprised eight participants. The scores (**Table 2**) in groups H and L for the first round were 111.75 ± 14.31 and 69.83 ± 11.29, respectively, and significant differences were found between scores (p = 0.0142, Wilcoxon test). The scores (**Table 1**) in groups H and L for all rounds were 124.75 ± 14.40 and 80.20 ± 19.17, respectively (p < 0.0001, Wilcoxon test).

Physicians' years of experience in groups H and L for the first round were 2.0 ± 1.73 and 1.42 ± 0.66 years, respectively; however, this difference was not significant (p = 0.559, Wilcoxon test). Physicians' years of experience in groups H and L for all rounds were 5.67 ± 1.50 and 6.33 ± 2.06 years, respectively, which was also not significantly different (p = 0.549, Wilcoxon test). The increase in scores from the first participation to the final participation in groups H and L was 24.95 ± 10.11 and 9.0 ± 19.86, respectively, and the difference was not significant (p = 0.164, Wilcoxon test) (**Table 2**).

Results of regression analysis of groups H and L are shown in **Figure 3**. Among competition rounds, mean scores were higher for group H than for group L. The scores in each round for group H improved in each successive competition (y = 3.73x + 108.73, R^2 = 0.25), whereas no obvious pattern was observed (y = 1.47x + 69.44, R^2 = 0.011) for scores among successive competitions in group L. The findings revealed that high scores were recorded in group H during the first rounds and that scores tended to increase in subsequent rounds of successive competitions.

Comparisons were made between the groups with the

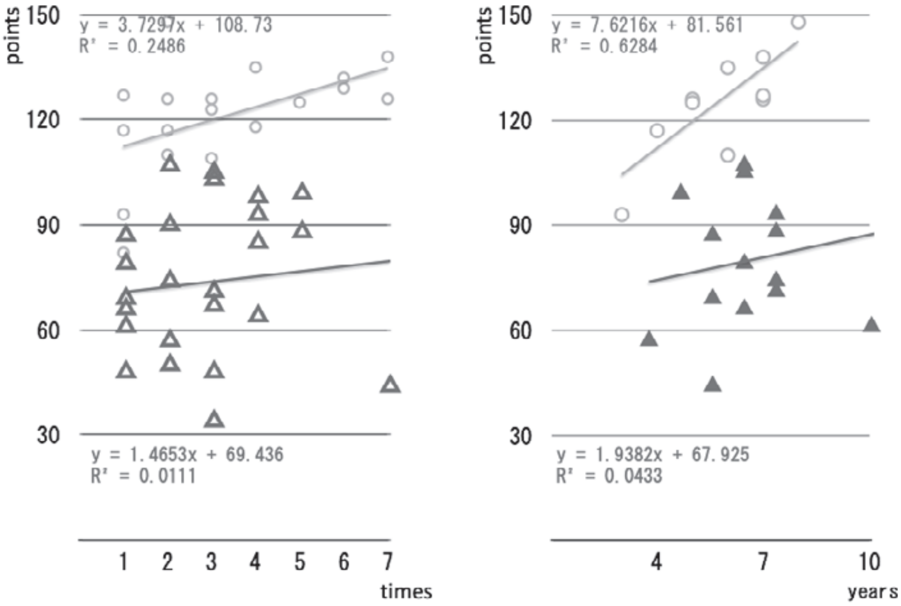

Fig. 3
left: Linear regression model for groups H (blue) and L (red). right: Linear regression model for groups H (blue) and L (red) after adjusting for years of experience after receiving a medical license

same number of years of medical experience (**Fig. 3**). The results of regression analysis for groups H and L are shown in **Figure 3**. Regarding number of years of medical experience, the mean score for group H improved in each successive competition (y = 7.62x + 81.56, R^2 = 0.63), whereas no specific pattern was observed for scores across competitions in group L (y = 1.94x + 67.93, R^2 = 0.043). R^2 values indicated that these trends were stronger for the number of years of medical experience than for the subsequent rounds of competition.

Reliability was estimated by using intraclass correlation for each of the evaluators. The intraclass correlation coefficient (2, 1) was 0.6905, indicating a strong positive correlation between the scores from each pair of evaluators.

Discussion

In this study, top performers had the highest scores during the first rounds of the competitions, and these scores tended to increase over time. However, lower scores did not substantially increase over time. These results indicate that our method has the potential to objectively assess changes in individual skills over time and that it is an objective assessment, with no significant scoring differences among evaluators.

To our knowledge, no study has reported repeated contest-format assessments of microsurgical skills with the same evaluators over several years. Small randomized controlled trials[1,15,17] have shown that competition can help improve laparoscopic surgical techniques. One study[15] included short-term results and skills obtained by watching surgical videos. Another study showed[1] that introducing the principle of competition through full participation-type simulation training is useful for improving willingness to participate and decision-making abilities. However, many studies have reported[8,15,18] that most participants in simulation training are young surgeons who received short-term training; therefore, it is difficult to clarify the effects of simulation training on preventing complications[15]. For this reason, we reported the results of our long-term observations.

A systematic review[19] of the literature highlighted the need for structured classification of surgical skills and use of the OSATS[5,9] to accurately transfer skills from the training room to the operating room. Recently, actual surgical images were evaluated by using scales such as the OSATS and Generic Error Rating Tool[2], and their relationship with outcomes was reported. Another study reported that a low score on the OSATS assessment was an independent predictor of short-term primary outcome[2]. Thus, since the beginning of the competition, we classified the techniques structurally, set evaluation standards, and used these evaluation standards continuously. Time is included as an evaluation item on the OSATS[5], but we decided[7] to limit the time required for each person and exclude time as an evaluation item. This made it possible for many doctors to participate in the competition during their busy day. Furthermore, by limiting the time, the competition could be repeated over a long period.

Various issues have been noted when evaluating surgical skills. Evaluations based on direct practical observations place a heavy burden on the evaluator[7,9]. Thus, one recommendation is to conduct simulation task training in an environment with other participants and instructors[18,19]. This has the effect of performing practical skills in an environment that has the same tension and stress[20] as actual surgery[15,18]. Nickel et al.[9] reported that assessing surgical procedure videos was more useful than actual observations in the operating room in reducing the burden placed on senior physicians in charge of assessment. They also suggested[9] that in the case of video evaluations, more direct evaluations can be made by considering operative time. However, the authors felt that it was not appropriate to use the time of the actual surgical video as an evaluation item because, in actual surgery, equipment preparation and the ability of the assistant affect surgical time.

Because contests utilizing direct observation involve a sense of tension[15,16,18,21], we emphasized the effectiveness of direct face-to-face evaluation rather than surgical video observation. By contrast, we attempted to reduce the time burden on the evaluator and the participant by limiting the time per participant to 5 minutes. To further shorten the duration of the entire competition, the organizers prepared the following two points. First, while a practitioner is performing, the next person should be prepared to start as soon as their turn arrives. Second, a scoring sheet that can be filled up in a short time should be prepared so that the assessment results can be announced in a short time.

When selecting the technical content of the contest, we sought high reproducibility (close to that of actual surgical techniques), low cost (approximately 30 US dollars per person per participation), short evaluation time (5 min), safety, and a procedure not limited to neurosurgery. After more than 4 years of continuous contests, we have found no problems with our method. The present results revealed that some surgeons have not improved their technique even after years of teaching, which indicates that we need to create another educational program for physicians whose objective and continuous evaluation results do not show skill improvement. To motivate those with low scores, a previous study reported[22] that making such surgeons aware that doctors younger than them have mastered the technique can motivate the surgeons to recognize the need for training. We have been conducting our reflection meetings with this expectation, but the present results have shown no effects. Currently, daily practice is at the discretion of individual physicians. However, those with low scores need additional direct guidance from supervising physicians. For neurosurgeons who do not intend to acquire skills in this field, we encourage them to improve their skills in other fields, such as endovascular treatment and stereotactic surgery, after fully confirming their intentions.

This study has limitations that warrant mention. First, the sample size was small. Second, some evaluators were in their 60s; thus, new evaluators will need to be selected for future competitions. Therefore, future studies should analyze whether similar results are obtainable even when evaluators are replaced. Third, some operations could not be performed in a short time, which limited the assessors' ability to evaluate the entire procedure. In such cases, evaluation might be facilitated by extracting scenes and procedures that are likely to cause complications during surgery.

In conclusion, although the contest did not appear to contribute to skill improvement, objective changes in the skills of young surgeons could be assessed with a contest-format continuous evaluation system involving the same microsurgical task. The results also suggest that scores for low performers do not improve with time. A system in which only surgeons with a certain number of points are qualified to perform the actual surgery was established, and this system might improve patient outcomes.

Conflict of Interest: None declared.

References

1. Graafland M, Schraagen JMC, Boermeester MA, Bemelman WA, Schijven MP. Training situational awareness to reduce surgical errors in the operating room. Br J Surg. 2015;102(1):16–23. doi: 10.1002/bjs.9643
2. Fecso AB, Bhatti JA, Stotland PK, Quereshy FA, Grantcharov TP. Technical performance as a predictor of clinical outcomes in laparoscopic gastric cancer Surgery. Ann Surg. 2019 ; 270 (1) : 115 – 20. doi : 10.1097 / SLA.0000000000002741
3. Ahmed A, Ishman SL, Laeeq K, Bhatti NI. Assessment of improvement of trainee surgical skills in the operating room for tonsillectomy. Laryngoscope. 2013;123(7):1639–44. doi: 10.1002/lary.24023
4. Cokluk C. Using of the chicken wing's bone in the microneusurgical training model for microdrilling. Asian J Neurosurg. 2018;13(4):990–4. doi: 10.4103/ajns.AJNS_372_16
5. VanBlaricom AL, Goff BA, Chinn M, Icasiano MM,

Nielsen P, Mandel L. A new curriculum for hysteroscopy training as demonstrated by an objective structured assessment of technical skills (OSATS). Am J Obstet Gynecol. 2005;193(5):1856–65. doi: 10.1016/j.ajog.2005.07.057
6. Javid P, Aydın A, Mohanna PN, Dasgupta P, Ahmed K. Current status of simulation and training models in microsurgery: A systematic review. Microsurgery. 2019;39(7): 655–68. doi: 10.1002/micr.30513
7. Hu Y, Tiemann D, Michael Brunt L. Video self-assessment of basic suturing and knot tying skills by novice trainees. J Surg Educ. 2013;70(2):279–83. doi: 10.1016/j.jsurg.2012.10.003
8. Chang OH, King LP, Modest AM, Hur HC. Developing an objective structured assessment of technical skills for laparoscopic suturing and intracorporeal knot tying. J Surg Educ. 2016;73(2):258–63. doi: 10.1016/j.jsurg.2015.10.006
9. Nickel F, Hendrie JD, Stock C, et al. Direct observation versus endoscopic video recording-based rating with the objective structured assessment of technical skills for training of laparoscopic cholecystectomy. Eur Surg Res. 2016;57(1-2):1–9. doi: 10.1159/000444449
10. Leclère FMP, Trelles M, Lewbart GA, Vögelin E. Is there good simulation basic training for end-to-side vascular microanastomoses? Aesthetic Plast Surg. 2013;37(2):454–8. doi: 10.1007/s00266-012-0005-0
11. Hirche C, Megerle K, Heitmann C, et al. Konsensus der Deutschsprachigen Arbeitsgemeinschaft für Mikrochirurgie der peripheren Nerven und Gefäße (DAM) für Mindeststandards mikrochirurgischer Trainings- und Übungskurse zur Vergabe eines Qualitätssiegels [Consensus of the German-Speaking Society for Microsurgery of Peripheral Nerves and Vessels (DAM) on minimum standards for microsurgical training courses and accreditation - Minimum Standards for Microsurgical Training Courses and Accreditation]. Handchir Mikrochir Plast Chir. 2020 Apr;52(2):135–9. German. doi: 10.1055/a-1017-3688
12. Zhang Z, Zhou M, Liu K, et al. Development of a new valid and reliable microsurgical skill assessment scale for ophthalmology residents. BMC Ophthalmol. 2018;18(1):1–7. doi: 10.1186/s12886-018-0736-z
13. Rajan S, Sathyan R, Sreelesh LS, et al. Objective assessment of microsurgery competency-in search of a validated tool. Indian J Plast Surg. 2019;52(2):216–21. doi: 10.1055/s-0039-1695658
14. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997;84(2):273–8. doi: 10.1046/j.1365-2168.1997.02502.x
15. Hashimoto DA, Gomez ED, Beyer-Berjot L, et al. A randomized controlled trial to assess the effects of competition on the development of laparoscopic surgical skills. J Surg Educ. 2015;72(6):1077–84. doi: 10.1016/j.jsurg.2015.06.005
16. Mokadam NA, Lee R, Vaporciyan AA, et al. Gamification in thoracic surgical education: Using competition to fuel performance. J Thorac Cardiovasc Surg. 2015;150(5):1052–8. doi: 10.1016/j.jtcvs.2015.07.064
17. Sharma M, Macafee D, Horgan AF. Basic laparoscopic skills training using fresh frozen cadaver: a randomized controlled trial. Am J Surg. 2013;206(1):23–31. doi: 10.1016/j.amjsurg.2012.10.037
18. McCreery GL, El-Beheiry M, Schlachta CM. Local and national laparoscopic skill competitions: residents' opinions and impact on adoption of simulation-based training. Surg Endosc. 2017;31(11):4711–6. doi: 10.1007/s00464-017-5546-5
19. Buckley CE, Kavanagh DO, Traynor O, Neary PC. Is the skillset obtained in surgical simulation transferable to the operating theatre? Am J Surg. 2014;207(1):146–57. doi: 10.1016/j.amjsurg.2013.06.017
20. LeBlanc V, Woodrow SI, Sidhu R, Dubrowski A. Examination stress leads to improvements on fundamental technical skills for surgery. Am J Surg. 2008;196(1):114–9. doi: 10.1016/j.amjsurg.2007.08.059
21. Ramos-Gallardo G, Mecott-Rivera GÁ, Limon-Cervantes R, García-Pérez M, Rodríguez-Olivares E. How to improve speaking skills in plastic surgery training? Survey in residents participants in Pecha Kucha Contest. World J Plast Surg [Internet]. 2018 [cited 2021 Feb 11];7(2):166–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/30083498
22. Tsutsumi K. Watashi no shujutsu kyoiku [The Education for Young Neurosurgerons]. Nousocchuu no Geka [Surg Cereb Stroke]. 2007;35(5):361–3. Japanese.

(Received, November 16, 2021)
(Accepted, December 24, 2021)
(J-STAGE Advance Publication, January 25, 2022)