Title Page

Contest-style evaluation for the objective assessment of microsurgical techniques: an observational study

Running title: Contest-style evaluation of microsurgery

Authors: Yasuo Murai¹*, Eitaro Ishisaka¹, Atsushi Tsukiyama¹, Asami Kubota¹, Masahiro Yamaguchi², Fumihiro Matano¹, Tomonori Tamaki¹, Takayuki Mizunari², and Akio Morita¹

Affiliations:

¹ Department of Neurological Surgery, Nippon Medical School Hospital, Tokyo 113-8602, Japan

² Department of Neurosurgery, Nippon Medical School Chiba Hokusoh Hospital, Inzai, Chiba, 270-1694 Japan

Corresponding author: Yasuo Murai

Department of Neurological Surgery, Nippon Medical School Hospital

1-1-5, Sendagi, Bunkyo-ku, Tokyo 113-8603, Japan

Tel: 81-3-3822-2131 Fax: 81-3-5685-0986 E-mail: ymurai@nms.ac.jp
ABSTRACT

Background: Considering the lack of studies on the long-term evaluation of improvements in microsurgical techniques using simulation models, we determined whether technical improvements in surgical techniques could be assessed based on an increase in the score of contest-style continuous and objective evaluation systems involving the same microsurgical task.

Methods: Since 2014, neurosurgeons with 1–10 years of experience participated in a biannual competition-style test. The task involved suturing as many times as possible within 5 min after arteriotomy of 1-mm artificial vessels. A modified version of the Objective Structured Assessment of Technical Skills examination was created and used. Changes and differences in scoring results 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 obtained the highest score in each contest and those who obtained the lowest score. The linear regression equations for the highest and lowest scorer groups were $y=7.62x+81.56 \quad (R^2=0.628)$ and $y=1.94x+67.93 \quad (R^2=0.0433)$, respectively. The high-scoring group had high scores from the first time they participated, and their scores tended to increase further, while the low-scoring group did not show any tendency of score increases after gaining experience. No significant differences were found in scores according to four evaluators.

Conclusions: Our study showed the possibility of assessing technical improvements in surgery through long-term and continuous microsurgical technique evaluation. A surgical technique evaluation system was established, and its potential to contribute to surgical safety was demonstrated.

Keywords: microsurgery, objective assessment, skills, techniques
INTRODUCTION

Many surgical education studies\textsuperscript{1–3} have pointed out that effective surgical training can reduce technical errors. Various methods for evaluating surgical techniques have been reported\textsuperscript{1,4–7} in many fields such as gynecology, laparoscopy, and general surgery. In these reports,\textsuperscript{8,9} evaluation tables,\textsuperscript{3,8} differences between video evaluation and direct evaluation, and evaluation task contents\textsuperscript{10,11} were examined; however, there are no reports on the results of objective evaluations repeated over a number of years. The objectivity of the scoring method itself cannot be evaluated in a single technical evaluation because the tendencies of the participants and scorers at the time of the technical evaluation may affect the results.\textsuperscript{12} For this reason, a method evaluating video recordings of the same trial at intervals of 1–3 months has also been attempted.\textsuperscript{8,10,12} However, there are no previous reports on surgical technique evaluation by simulation that tracked the results of scoring for >6 months on the same participant. Thus, in this study, we examined the results of the same participants who took part multiple times with an interval of >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, wherein the evaluators were not changed for 4 years. This study aimed to investigate whether technical improvements in surgical techniques could be assessed based on an increase in the score through contest-style continuous and objective evaluation systems involving the same microsurgical task. Finally, the purpose of this study was to conduct a basic research to establish a technical evaluation system to qualify as a surgical operator and contribute to patient prognosis and surgical safety.

MATERIALS AND METHODS

In this study, modified Objective Structured Assessment of Technical Skills (OSATS) evaluation forms were created to assess individual technical improvements.\textsuperscript{12–14} They are used to evaluate each participant in a contest format,\textsuperscript{15,16} and the results of long-term observations are discussed and reported.
Participants and evaluators

Residents and neurosurgeons with 1–10 years of experience participated in each competition, which has been held twice a year since 2014. A championship has been held nine times in the past. 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 of neurosurgical society qualifications, and (d) stroke surgeon technical mentor qualifications. There were four evaluators who remained throughout the competition to measure improvements 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 min to complete the task; thus, the competition was limited to approximately 2 h and did not interfere with daily work. A model performance video was created to encourage practices in advance for first-time participants.

The examination content was determined by considering reproducibility, time, cost, evaluation objectivity, and similarity to surgery. An actual surgical microscope (OMI PENTERO 900, Carl Zeiss Meditec, Germany) was used. The examination included performing arteriotomy on a 1-mm-diameter artificial blood vessel linearly in the long-axis direction and stitching the incision line with 10-0 nylon sutures (Figure 1). The commercially available artificial blood vessel comprises one layer with a 1-mm diameter, simulating the blood vessel for anastomosis (Wetlab Co., Ltd., Shiga, Japan). For purchasing consumables, we used the educational equipment purchase cost provided by the hospital. The suturing was repeated within
To confirm the speed of the procedure, the incision area was sutured as many times as possible within 5 min. To ensure that the forced use of unfamiliar tools does not affect the technical evaluation, the microsurgical instruments were not standardized, and 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 each other’s videos.

[Figure 1]

**Scoring**

The scoring chart (Figure 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.\(^5\)

[Figure 2]

Using this table (Figure 2), the tasks were evaluated based on multiple factors such as the preparation (posture of the operator and instrument layout), microscope operation (focusing and positioning), motion (hand tremor), knowledge of the instrument (linear incision using scissors), handling of the needle (insertion angle and thread treatment), attention to the vessel wall, flow of the suture (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 one of the evaluation items. 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 limited based on the design of the competition. The final ranking was determined by the total score of the four evaluators. A perfect score for each evaluator was 40 points, indicating that the highest possible score was 160 points, while the lowest score was 0 points. First-year doctors were given 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 independent evaluation results, the graders were prohibited from speaking until all participants had completed the practical skills and the grader had filled in the grade sheet. It was forbidden to collect and modify the scoring table after each participant had finished. The reflection meeting was held a week after the contest. The scoring results were announced before the reflection meeting.

Statistical analyses

JMP 14.0 software (SAS Institute Inc., Cary, NC) was used to perform all statistical analyses. All data were presented as mean ± standard deviation, 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 techniques, only medical doctors who participated in the competition more than once were included in the statistical analyses.

We compared each competitor’s scores from the first and final competitions using Student’s t-test. For doctors who participated only twice, the second round was considered the final round. Additionally, we divided the participants into two groups based on their scores. Participants gaining the highest score at least once in all the competitions they participated in were assigned to group H, while those obtaining the lowest score at least once were assigned to group L. The scores at the first time of participation, number of years of experience of the physician, and increase in scores from the first to the last time of 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 by the number of participations were plotted, and linear regression analysis was performed. The independent variable was the number of participations, and the dependent variable was the total number of points. Coefficients of determination (R²) were obtained based on linear regression analysis. Furthermore, the number of years of experience at the time of the competition was adjusted, and the evaluation results were compared using linear regression analysis. In this case, the independent variable was the
doctor’s years of experience at the time of participation, and the dependent variable was the total score. \(R^2\) values were again obtained based on linear regression analysis.

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

**RESULTS**

In total, the contests were held nine times from 2014 to 2019. Among all participants, the ones who participated at least twice were included in this study. Finally, 21 neurosurgeons participated in the contest 103 times. The scores for the first (mean ± standard deviation), 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 the scores during the first and second rounds \((p = 0.27)\) or the first and last rounds \((p = 0.11)\) (Table 1).

**[Table 1]**

Participants who gained the highest score at least once in all the competitions they participated in were assigned to group H, while those who obtained the 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 the scores according to the Wilcoxon test \((p = 0.0142)\). The scores (Table 1) in groups H and L for all rounds were 124.75 ± 14.40 and 80.20 ± 19.17, respectively, and significant differences were noted between these scores according to the Wilcoxon test \((p < 0.0001)\).

**[Table 2]**

Doctors’ 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 according to the Wilcoxon test \((p = 0.559)\). Doctors’ 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.
according to the Wilcoxon test \((p = 0.549)\). The increases in scores from the first participation to the final participation in groups H and L were 24.95 ± 10.11 and 9.0 ± 19.86, respectively, and the difference was not significant according to the Wilcoxon test \((p = 0.164)\) (Table 2).

Results of the regression analysis for groups H and L are shown in Figure 3. Among the rounds of competition, higher mean scores were recorded in group H than in 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 specific pattern was observed \((y = 1.47x + 69.44, R^2 = 0.011)\) for the scores among successive competitions in group L. These findings revealed that high scores were recorded in group H during the first rounds, which tended to increase in subsequent rounds of successive competitions.

[Figure 3]

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

Reliability was estimated using the intraclass correlation for each of the evaluators. The intraclass correlation coefficients \((2,1)\) was 0.6905. The results indicate a strong positive correlation between the score of each pair of evaluators.

DISCUSSION

In this study, the top performers had higher scores than others during the first rounds of the competitions, which tended to increase over time. However, the lower scores did not increase considerably over time. These results indicate that our method has the potential to objectively assess changes in individual skills over time. The results also showed that our
method is an objective assessment, with no differences in scoring among scorers.

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

A systematic review\textsuperscript{19} of the literature highlights the need for a structured classification of surgical skills and the OSATS\textsuperscript{5,9} to accurately transfer skills from the training room to the operating room. Recently, actual surgical images were evaluated using scales such as OSATS and Generic Error Rating Tool,\textsuperscript{2} and their relationship with prognosis was reported. Furthermore, a study reported that a low score on the OSATS assessment was an independent predictor of short-term primary outcome.\textsuperscript{2} Thus, since the beginning of the competition, we classified the technologies structurally, set evaluation standards, and used these evaluation standards continuously. Time is included as an evaluation item on the OSATS,\textsuperscript{5} but we decided\textsuperscript{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 daily work. Furthermore, by limiting the time, the competition could be repeated over a long period.

Various issues have been noted in the methods of evaluating surgical skills. Studies have reported\textsuperscript{7,9} that evaluations made based on direct practical observations place a heavy burden on the evaluator. Thus, the recommendation was to conduct simulation task training in an environment with other participants and instructors.\textsuperscript{18,19} This is because it has the effect of performing practical skills in an environment that has the same tension and stress\textsuperscript{20} as those of
actual surgery. Nickel et al. reported the usefulness of assessing surgical procedure videos, rather than actual observations in the operating room, to reduce the burden placed on senior physicians in charge of assessments. The study also suggested that in the case of video evaluations, more direct evaluations can be made by considering the operative time. However, the authors believe that it is not appropriate to use the time of the actual surgical video as an evaluation item because in actual surgery, the preparation of the equipment and the ability of the assistant affect the surgical time.

Because contests by direct observation involves a sense of tension, 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 min. 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 judging results could be announced in a short time.

When considering the content of the contest actual technique, we took into account the following aspects: high reproducibility, close to that of actual surgical techniques, low cost (approximately 30 US dollar per person at a time), content that can be evaluated in 5 min, safety, and a procedure not limited to neurosurgery. After more than 4 years of continuous contests, we concluded that no problems were associated with this method. The results of this study revealed that some doctors have not improved their techniques even after years of teaching. We believe that we need to create another educational program for those doctors whose objective and continuous evaluation results do not show improvement in their skills. As a way to motivate those with low scores, it has been reported that creating an awareness of the existence of doctors younger than them but have mastered the technique can motivate to recognize the need for training. We have been conducting the reflection meeting with this expectation, but the results of this study have not shown effects. At the current time, daily practice is left to the
autonomy of individual doctors. However, for those with low scores, it is necessary to increase opportunities for direct guidance from the supervising physician. 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 had some limitations that should be addressed. First, the study included a small sample size. Second, some evaluators were in their 60s; thus, in future rounds of competition, new evaluators would have to be chosen to replace them. Therefore, future studies should analyze whether similar results are obtainable even if the evaluators are changed. Third, some operations could not be performed in a short time, which limited the assessor’s ability to evaluate the entire procedure. In such cases, evaluation is considered possible by extracting scenes and procedures that are likely to cause complications during surgery.

In conclusion, although it cannot be stated that the contest contributes to skill improvement, this study revealed that objective changes in individual skills of young surgeons could be determined using a contest-style continuous evaluation system involving the same microsurgical task. The results also suggest that the time course is not a predictor of low-performance group scores. A system in which only surgeons with a certain number of points are qualified to perform the actual surgery was established, and this system was shown to have the potential to contribute to patient prognosis.

Acknowledgments: None

Conflicts of interest/Competing interests: The authors declare that they have no conflicting or competing interests.

Funding: No financial support was provided for this study.

Ethics approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional research committee and with the 1964
Helsinki Declaration and its later amendments or comparable ethical standards.

**Consent to participate:** Informed consent was obtained from all individual participants who were included in the study.

**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):115-120. 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):1639-1644. doi:10.1002/lary.24023

4. Cokluk C. Using of the Chicken Wing’s Bone in the Microneurosurgical Training Model for Microdrilling. *Asian J Neurosurg.* 2018;13(4):990-994. 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):1856-1865. doi:10.1016/j.ajog.2005.07.057

6. Javid P, Aydin A, Mohanna P-N, Dasgupta P, Ahmed K. Current status of simulation and training models in microsurgery: A systematic review. *Microsurgery.* 2019;39(7):655-668. 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):279-283.
8. Chang OH, King LP, Modest AM, Hur H-C. Developing an Objective Structured Assessment of Technical Skills for Laparoscopic Suturing and Intracorporeal Knot Tying. *J Surg Educ*. 2016;73(2):258-263. 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):454-458. doi:10.1007/s00266-012-0005-0

11. Hirche C, Megerle K, Heitmann C, et al. [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*. October 2019. 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):216-221. 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):273-278. 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):1077-1084. 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):1052-1058. 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):4711-4716. 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):146-157. 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):114-119. 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*. 2018;7(2):166-170. http://www.ncbi.nlm.nih.gov/pubmed/30083498. Accessed February 11, 2021.

22. Tsutsumi K. The Education for Young Neurosurgerons. *Surg Cereb Stroke*(Jpn). 2007;35(3): 361-363 (Japanese)
Figures and figure legends

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

Fig. 2 Scoring chart used by the evaluators

Fig. 3
a: Linear regression model for groups H (blue) and L (red). b: Linear regression model for groups H (blue) and L (red) after adjusting for doctor’s years of experience after receiving a licentiate degree
| 8th 5min championship Grade Slip | Grader A | B | C | D |
|----------------------------------|----------|---|---|---|
| Lucas | Ray | Luke | Kubota |
| **Grading Criteria** | Failing | Bad | Poor: below average | Fair: average | Good: Fairly good | Excellent: Almost expert |
| 1. Surgeons Positioning / Instrument layout (preparation) | 0 | 1 | 2 | 3 | 4 | 5 |
| 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 / Insetion 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. 3

---

Left graph:
- Equation: $y = 3.729x + 108.73$
- $R^2 = 0.2486$

Right graph:
- Equation: $y = 7.6216x + 81.561$
- $R^2 = 0.6284$

---

Left graph:
- Equation: $y = 1.4653x + 69.436$
- $R^2 = 0.0111$

Right graph:
- Equation: $y = 1.9382x + 67.925$
- $R^2 = 0.0433$
**Tables and footnotes**

**Table 1.** Participant scores for the 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    |
| Increased score from the first to 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    |
| R² in linear regression| 0.3486           | 0.0111           |          |
| R² corrected for doctor’s years of experience | 0.6284     | 0.0433           |          |

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