



Universiteit
Leiden
The Netherlands

3D Learning in anatomical and surgical education in relation to visual-spatial abilities

Bogomolova, K.

Citation

Bogomolova, K. (2022, February 3). *3D Learning in anatomical and surgical education in relation to visual-spatial abilities*. Retrieved from <https://hdl.handle.net/1887/3274191>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3274191>

Note: To cite this publication please use the final published version (if applicable).

6

Global versus task-specific post-operative feedback in surgical procedure learning: a randomized controlled trial

Tahmina Nazari* & Katerina Bogomolova*, Marlike Ridderbos,
Mary E.W. Dankbaar, Jeroen J.G. van Merriënboer, Johan F. Lange,
Theo Wiggers, Jos A. van der Hage

* Shared first authorship

Surgery 2021; 170:81-87



ABSTRACT

Background

Task-specific checklists and global rating scales are both recommended assessment tools to provide constructive feedback on surgical performance. This study evaluated the most effective feedback tool by comparing the effects of the Observational Clinical Human Reliability Analysis (OCHRA) and the Objective Structured Assessment of Technical Skills (OSATS) on surgical performance in relation to the visual-spatial abilities (VSA) of the learners.

Methods

In a randomized controlled trial, medical students were allocated to either the OCHRA (n = 25) or OSATS (n = 25) feedback group. VSA was measured by a Mental Rotation Test (MRT). Participants performed an open inguinal hernia repair procedure on a simulation model twice. Feedback was provided after the first procedure. Improvement in performance was evaluated blindly using a global rating scale (performance score) and hand-motion analysis (time and path length).

Results

Mean improvement in performance score was not significantly different between the OCHRA and OSATS feedback groups ($p = .100$). However, mean improvement in time (371.0 ± 223.4 vs 274.6 ± 341.6 ; $p = .027$) and path length (53.5 ± 42.4 vs 34.7 ± 39.0 ; $p = .046$) was significantly greater in the OCHRA feedback group. When stratified by MRT scores, the greater improvement in time ($p = .032$) and path length ($p = .053$) was observed only among individuals with low VSA.

Conclusions

A task-specific (OCHRA) feedback is more effective in improving surgical skills in terms of time and path length in novices compared to a global rating scale (OSATS). The effects of a task-specific feedback are present mostly in individuals with lower VSA.

INTRODUCTION

Feedback has long been recognized for its positive effect in surgical knowledge and skills training.¹ It has been shown to be crucial in technical skill development because it increases motivation, prevents incorrect actions, and reinforces correct actions.^{2,3} Feedback can be provided based on direct observation of technical skills.⁴ Within the surgical field, different observational assessment tools are available.⁵ Assessment tools assess surgical performance on competences, skills, or surgical-specific items on a checklist. These tools can be used as a medium for feedback to provide information regarding a trainee's performance to improve on specific items that are being assessed.^{1,5} Two main types of assessment tools can be recognized: global rating scales, which rate general surgical skills and are applicable to all surgical procedures, or procedure-specific checklists.⁵ In both categories, many tools have been developed and validated.^{4,5}

A commonly used and generally accepted as "gold standard" assessment tool is Objective Structured Assessment of Technical Skills (OSATS), a global rating scale introduced by Martin et al for assessing technical skills of an entire surgical procedure.^{5,6} OSATS is a reliable, validated tool that assesses 7 competencies on a 5-point Likert scale.⁶ It is feasible and effective in assessment of surgical skills of trainees in the operating room.⁷

Although global rating scales such as the OSATS are easy in use, these scales can be imprecise.⁴ A task-specific method may provide more concise and precise feedback.⁴ A task-specific technical skills assessment method is the Observational Clinical Human Reliability Analysis (OCHRA).⁸ An OCHRA checklist assesses in a stepwise manner whether a substep was correct or incorrect.⁸ Both OSATS and OCHRA assessment tools have shown to be valid for providing constructive feedback.^{4,7} However, according to constructive alignment theory, the OCHRA feedback might be more effective when the surgical procedure is also learned in a stepwise manner.⁹

Although the validity of OSATS and OCHRA is demonstrated, these assessment tools are still based on individual judgments, which are inevitably associated with subjectivity.¹⁰ Quantifying measures of technical skills may potentially mitigate this subjectivity. For open surgery, different motion tracking devices are described to measure either hand or instrument movements.¹¹⁻¹⁴ The outcomes of time to complete a task and total path length can differentiate between novices and experts.¹³⁻¹⁵

Additionally, the effect of feedback in relation to visual-spatial abilities (VSA), as another determining factor for technical skills development, is unrecognized. VSA are defined as the ability that allows individuals to construct visual-spatial (i.e., 3-dimensional) mental

representations of 2D images and to mentally manipulate these representations.^{16,17} This ability determines how well individuals are able to translate the acquired anatomical knowledge into clinical and surgical practice. Consequently, VSA determine how well surgical residents can understand and perform spatially complex procedures. The positive association between VSA and acquisition of surgical skills, including quality of hand motion, has been observed especially in the early phases of surgical training.^{15,18-20} Moreover, VSA can have a modifying effect on outcomes. Individuals with lower VSA tend to perform worse than individuals with high VSA on acquisition of anatomical knowledge and surgical skills. However, with supportive instructional methods and deliberate practice and feedback they are able to achieve a comparable level of competency.^{15,21-23}

The aim of this study was to investigate whether a task-specific, stepwise feedback checklist (OCHRA) leads to a greater improvement in performance of a surgical procedure compared to a global rating scale method (OSATS) in terms of improvement of overall performance score, time to complete task, and total path length. These outcomes were also evaluated in relation to learners' VSA.

METHODS

Study design and population

A randomized controlled trial was conducted at the Leiden University Medical Center, The Netherlands. Participants were medical students and novices to almost any type of surgical procedures. Only right-handed students were included because left-handed novice students may have difficulties with the surgical instruments.²⁴ Participation was voluntary, and written consent was obtained from all participants. The study protocol was approved by the Netherlands Association for Medical Education Ethical Review Board (NERB dossier number: 1013) (Figure 1).

Randomization

Participants were randomly allocated to either the OCHRA feedback (n = 25) or OSATS feedback group (n = 25) using an Excel random group generator.

Surgical procedure

The Lichtenstein open inguinal hernia repair was chosen as a procedure containing multiple surgical steps and because of its spatial complexity, which requires a certain level of surgical anatomical knowledge and VSA of the learner. The first part of the surgery, until resecting the hernia sac, requires solely basic surgical skills such as incising,

dissecting, and ligating. The second part, the placement and fixation of the mesh, is more complex. Each participant performed the Lichtenstein open inguinal hernia repair 2 times on a validated simulation model.²⁵ Participants were given access to the online course 1 week before the experiment to prepare for the experiment. The course consisted of 3 components: an introductory description that included text and figures regarding the surgical anatomy, a stepwise textual description, and a video demonstration of the procedure on the identical model used during the experiment.²⁶ The video demonstration depicted all important steps that need to be undertaken during surgery. Video was accompanied by auditory explanation. Participants were able to retrieve the materials as many times as they wanted and were able to do it on their own pace.

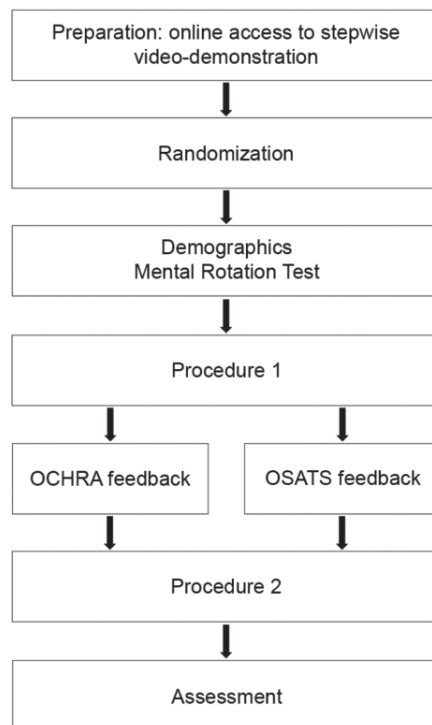


Figure 1. Flowchart of study design.

On the day of experiment, participants were given 30 minutes to complete each procedure.²⁷ The second procedure was performed directly after the provided feedback on the first procedure. Both procedures were recorded on video for blinded assessment. Participants were wearing a right-hand glove for the recording of motion by a motion tracking device (PST Base, PS-Tech B.V., Amsterdam, The Netherlands).

Demographic questionnaire

The questionnaire was administered before the experiment to account for factors that could possibly influence the performance. In a previous study, the time students studied for the open inguinal hernia repair with the use of a video demonstration had a significant modifying effect on surgical performance.²⁷ Therefore, study time was included in the questionnaire and was accounted for in the data analysis.

Assessment of VSA

VSA were measured by the mental rotation test (MRT) before the experiment. The MRT is a validated 24-item psychometric test and is the gold standard in assessing VSA in anatomical and surgical education.^{19,28, 29, 30} Participants were given 10 minutes to complete the test. The maximum possible score for the test was 24 points.

Interventions

In the OCHRA feedback group, postoperative feedback was provided using OCHRA. The OCHRA checklist is a reliable and valid instrument that has been successfully used in assessment of performance in various surgical procedures.^{8,31, 32, 33} It is a procedure-specific step-by-step skills assessment checklist that is characterized by a breakdown of a procedure into tasks.²⁶ Each step is assessed for being performed correctly and if errors are being made during the particular step. Provided feedback was based on the evaluation of each performed procedural step (*Supplementary material 11*). If a particular step was performed incorrectly, the error was discussed, and a proper execution of the step was explained. No points or final scores were awarded for the performance.

In the OSATS feedback group, postoperative feedback was provided using the OSATS assessment tool (*Supplementary material 12*). OSATS is a 7-item global rating scale that focuses on the following overall competencies: (1) respect for tissue, (2) time and motion, (3) instrument handling, (4) knowledge of instruments, (5) use of assistance, (6) flow of operation, and (7) knowledge of procedure.⁶ The tool has been previously validated in a wide range of surgical procedures and disciplines with reasonable index of reliability.^{6,34,35} Provided feedback was based on the evaluation of each of the 7 competencies in the exact order of OSATS. Suboptimal performance and errors made within a competence were discussed based on an example followed by an explanation for the improvement. No points or final scores were awarded for the performance to avoid any bias that could be introduced by grading the performance during the feedback phase.

In both groups, feedback was provided immediately after performing the first procedure. The total feedback time was held constant in both conditions and was approximately 10 minutes. Feedback was provided by 1 of the 2 researchers who were trained in providing both types of feedback in the context of this experiment. Care was taken to ensure that the feedback was complete and that participants were able to ask questions and verify whether they understood the information properly.

Performance score

Video-recorded procedures were assessed blindly by 2 independent researchers using OSATS, as the most common assessment tool for surgical performance. A minimum of 1 and a maximum of 5 points could be awarded for each of the 7 competences. A maximum possible performance score for each procedure was 35 points. Both researchers were trained in assessment of recorded procedures. Training was facilitated by a surgeon who is an expert in this field. It included a comprehensive study of the procedure using the provided study material followed by execution of the procedure on the model themselves. After that, researchers were trained in assessment until they got sufficiently familiar with all aspects of OSATS. The actual assessment of recorded procedures was performed independently. In case of discrepancies, consensus was reached by re-evaluating the procedure. Additionally, 5% of procedures were randomly selected and assessed by the expert to detect any discrepancies in scoring. No differences in ratings were identified.

Motion tracking

Motion tracking analysis was performed using a combination of a commercially available optical tracker system (PST Base, PS-Tech B.V., Amsterdam, The Netherlands) and a customized glove for the dominant right hand. This could track 6 degrees of freedom position in Cartesian coordinates (X, Y, and Z axis) at a rate of 30 samples per second. Time to complete the task and path length were measured. These have shown to be excellent markers of surgical performance.^{11,36-38} Because not all participants were able to complete the procedure within 30 minutes, the completion of the step of hernia sac removal was chosen as the endpoint for the outcomes of motion tracking analysis.

Outcomes

The study outcomes were defined as the differences in mean improvement in performance score (as measured by the OSATS assessment tool; time (in seconds) and path length (in meters) between the first and the second procedure between 2 groups. Outcomes were stratified by MRT scores. Individuals who scored below the mean were assigned to the MRT-low group (n = 22). Students who scored above the mean were assigned to the MRT-high group (n = 28).

Statistical analysis

Because of the novelty of this study, no previous data were available to calculate the sample size. A sample size of 50 participants was assumed to be appropriate. Participants' baseline characteristics were summarized using descriptive statistics. Differences in baseline measurements were assessed with an independent *t* test for differences in means and χ^2 test for differences in proportions. The differences in mean performance scores of the first procedure between groups were assessed with an independent *t* test. The improvement between second and first procedure within a group was assessed with a paired *t* test. The difference in mean improvement (Δ) in performance scores between second and first procedure between groups were assessed with a 1-way ANCOVA. Δ Performance score was included as dependent variable, intervention group and study time as fixed factor (0–1 vs 1–2 vs 2–3 hours), and performance score on the first procedure and MRT score as covariates. Additionally, the outcomes were stratified by MRT score to evaluate the effect of intervention for different levels of VSA. The analyses were repeated for mean improvement in time (Δ time) and path length (Δ path length). Partial eta squared was calculated and used as an effect size (0.2 = small effect, 0.5 = moderate effect 0.8 = large effect). Analyses were performed using SPSS statistical software package version 25.0 for Windows (IBM Corp, Armonk, NY). Statistical significance was determined at the level of $p < .05$.

Results

A total of 50 medical students was included. There were no significant differences between groups on baseline characteristics, as shown in Table 1.

Both groups improved significantly in terms of total OSATS score, time, and path length between the first and second time of performing the procedure (Table 2). Since not all participants were able to complete the procedure within 30 minutes, the completion of the step of hernia sac removal was chosen as the endpoint for the outcome measures time (s) and path length (m). This step was performed by 42 (84%) of participants. Path length data of 5 of the participants was lacking due to technical issues.

The mean improvement in performance scores was not significantly different between the 2 groups ($\beta = 2.1$; 95% IC [-0.41 to -4.5]; $\eta^2 = 0.06$; $p = .100$). However, the mean improvement in time ($\beta = -139.4$; 95% CI [-.262.5 to -16.5]; $\eta^2 = 0.13$; $p = .027$) and in path length ($\beta = -21.2$; 95% CI [-41.9 to -0.5]; $\eta^2 = 0.13$; $p = .046$) was significantly greater in the OCHRA feedback group.

Table 1. Baseline characteristics of the included participants.

	OCHRA feedback n = 25	OSATS feedback n = 25	p value
Sex, n (%)			
Male	10 (40)	13 (52)	.571
Female	15 (60)	12 (48)	
Age, mean \pm SD, in years	21.5 \pm 2.2	21.2 \pm 1.9	.537
Study phase, n (%)			
Bachelor students	15 (60)	14 (56)	.302
Master students	10 (40)	11 (44)	
Time spent studying online course, n (%)			
0 – 1 hours	8 (32)	5 (20)	.288
1 – 2 hours	16 (64)	16 (64)	
2 – 3 hours	1 (4)	4 (16)	
I liked the way the hernia repair was taught, median [IQR]	8.0 [7.0-9.0]	7.0 [6.5-8.7]	.104
I felt prepared after completing the online course, mean \pm SD	6.3 \pm 1.2	6.1 \pm 2.1	.679
Times seen open inguinal hernia repair surgery in real life, median [IQR]	0.0 [0.0-0.5]	0.0 [0.0-0.5]	.984
Other sources used to study, n (%)			
Not used	16 (64)	11 (44)	.256
Yes	9 (36)	14 (56)	
Time spent studying other sources, n (%)			
0 – 1 hour	8 (88.9)	13 (92.9)	1.00
1 – 2 hours	1 (11.1)	1 (7.1)	
Hours of sleep last night, median [IQR]	7.0 [6.0-8.0]	8.0 [7.0-8.0]	.471
Alcohol consumption last night, median [IQR]	0.0 [0.0-0.8]	0.0 [0.0- 0.0]	.402
Coffee consumption before surgical performance, median [IQR]	1.0 [0.0-1.0]	0.5 [0.0-1.0]	.879
Other circumstances that could have affected the surgical performance, n (%)			
Not used	18 (72)	22 (88)	.289
Yes	7 (28)	3 (12)	
Mental Rotation Test score, mean \pm SD	16.4 \pm 5.5	16.7 \pm 4.9	.872

OCHRA, Observational Clinical Human Reliability Analysis; OSATS, Objective Structured Assessment of Technical Skills; n, number of participants; SD, standard deviation; IQR, interquartile range.

Table 2. Differences in performance scores, time, and path length between 2 interventions.

	OCHRA feedback	OSATS feedback	p value
Performance score	n = 25	n = 25	
1 st procedure	17.4 ± 3.1	17.4 ± 3.8	.935
2 nd procedure	23.5 ± 5.4	21.8 ± 4.9	
Δ	6.2 ± 3.5*	4.4 ± 4.7*	.100 ^A
Time (sec)	n = 20	n = 22	
1 st procedure	1239.6 ± 274.8	1300.4 ± 382.3	.561
2 nd procedure	868.6 ± 151.6	1025.7 ± 286.4	
Δ	371.0 ± 223.4*	274.6 ± 341.6*	.027 ^A
Path length (m)	n = 19	n = 18	
1 st procedure	168.4 ± 61.5	168.9 ± 39.6	.977
2 nd procedure	112.4 ± 36.2	134.2 ± 36.3	
Δ	53.5 ± 42.4*	34.7 ± 39.0*	.046 ^A

Δ = delta, difference between 2nd and 1st procedure; sec, seconds; m, meters. *p < .001 paired t-test; ^Adifferences assessed with ANCOVA.

Effect of VSA

When outcomes were stratified by MRT scores, the greater improvement in time in the OCHRA feedback group was observed only among individuals with lower VSA ($\beta = -220.2$; 95% CI [-418.4 to -22.1]; $\eta^2 = 0.26$; $p = .032$) (Figure 2).

As shown in Figure 3, a similar trajectory was observed for the improvement in path length. However, this difference did not reach the significance level ($\beta = -28.2$; 95% CI [-56.8 to 0.42]; $\eta^2 = 0.24$; $p = .053$). Regardless of intervention, MRT scores were significantly associated with mean improvement in time ($\beta = -14.17$; 95% CI [-26.9 to -2.6]; $\eta^2 = 0.14$; $p = .019$), but not in path length ($\beta = -0.74$; 95% CI [-2.8 to 1.3]; $\eta^2 = 0.01$; $p = .469$) and OSATS scores ($\beta = 0.05$; 95% CI [-0.2 to 0.3]; $\eta^2 = 0.004$; $p = .670$).

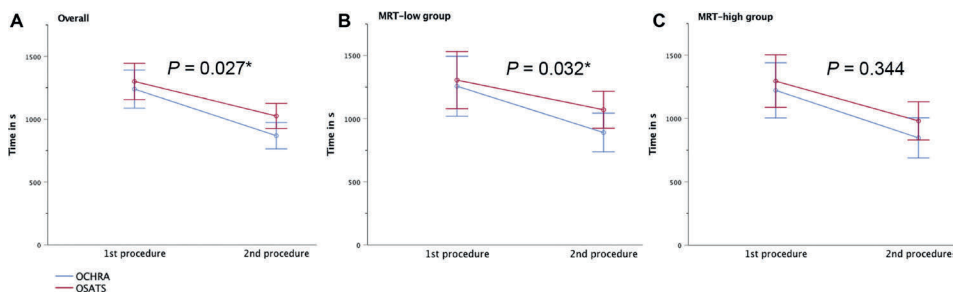


Figure 2. Differences in Δtime (s) between OCHRA feedback and OSATS feedback groups: (a) overall; (b) MRT-low group, and (c) MRT-high group; p < .05.

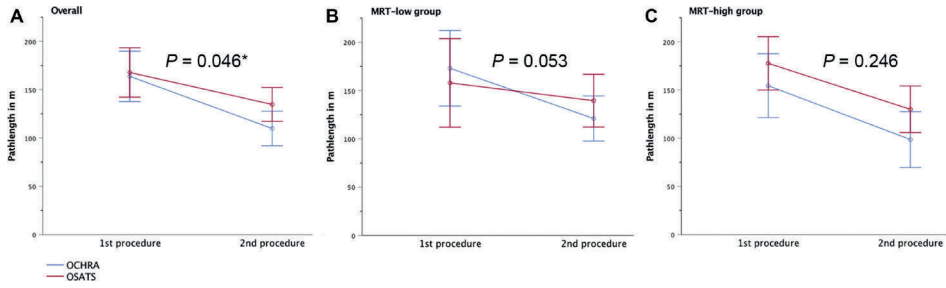


Figure 3. Differences in Δ path length (mm) between OCHRA feedback and OSATS feedback group: (a) overall; (b) MRT-low group, and (c) MRT-high group; $p < .05$.

DISCUSSION

The aim of this study was to investigate whether a task-specific, stepwise feedback checklist (OCHRA) leads to a greater improvement in surgical performance compared to a global rating scale feedback method (OSATS). The outcomes were evaluated in relation to VSA. The mean improvement in performance scores was not significantly different between the OCHRA and OSATS feedback groups. However, the OCHRA feedback showed a significant improvement on performance in terms of time and path length, as measured by the hand-motion analysis system. The effects of OCHRA feedback were present mainly among individuals with lower VSA.

The observed effectiveness of OCHRA feedback on surgical performance in a simplified hernia repair model, as a more precise and concise approach, is supported by the instructional alignment theory.³⁹ When training and assessment methods are aligned, the effects of instruction are up to 4 times greater than in nonaligned methods.³⁹ In the current study, participants prepared for the open inguinal hernia repair procedure using a stepwise video demonstration. As OCHRA feedback was based on the evaluation of the subsequent surgical steps instead of competencies as part of the OSATS feedback, a greater alignment between learning and feedback could be achieved. Although this did not result in a difference in outcome in terms of the surgical scores, differences were found for the time and path length. In this study, most participants could not finish the entire surgical procedure within the 30-minute timeframe. Possibly, differences in surgical scores would have been found if students did complete the entire surgical procedure. Additionally, the value of a checklist (OCHRA) and global rating scale (OSATS) assessments may depend on the level of learners' experience.⁴⁰ Global rating scales have been reported to be more useful for learners with higher levels of expertise, whereas checklists may be more useful for novice learners, such as the participants in this study.^{40,41}

The observed modifying effect of VSA on time and path length leads to important considerations. First, the findings are in line with previous research reporting positive association between VSA and hand motion.^{15,42-45} However, by treating VSA as a possible effect modifier, this study showed that this association was present only for individuals with lower levels of VSA. This effect, also referred to as the aptitude-treatment effect,^{46,47} has been repeatedly observed in the research field of anatomical education^{21-23,46} Therefore, it is instrumental to consider possible modifying effects of VSA on outcomes when designing new research. Second, the observed differences could be explained by the cognitive load theory.⁴⁸ Students with lower VSA are in general less effective in processing new spatial information in their working memory than students with higher VSA. However, in contrast to a global approach, the information from a task-specific stepwise feedback, building up on an already existing stepwise schema of a surgical procedure, could have decreased the cognitive load.⁴⁸ Subsequently, more working memory capacity could be created to process new procedural skills among low-performing individuals. This emphasizes the importance of an aptitude-based approach in learning and teaching surgical technical skills to novices. Lastly, the effect of VSA on OSATS scores was found to be not significant. This could be because of the inability of most participants to complete the entire procedure within the given timeframe.

OSATS was used both as an intervention and assessment scoring tool in this study. The rationale behind the choice to use the OSATS as an assessment scoring tool is that OSATS is considered to be the "gold standard" assessment tool for surgical performance and one of the few actually used in residency training and research.⁵⁻¹⁵ In The Netherlands, OSATS is incorporated within the surgical residency training.⁴⁹ Second, a systematic review comparing checklists with global rating scales as assessment tools reported that global rating scales might be better in capturing nuanced elements of expertise.⁴⁰ Other assessment tools for surgical performance, such as the recently reported Surgical Quality Assurance (SQA), could have been an option, and perhaps would have found differences in surgical performance.⁵⁰

The timing of feedback is still debated. Xeroulis et al distinguished feedback provided during the task (concurrent feedback) and feedback upon completing the task (summary feedback).³ The latter was found to be superior for learning basic surgical skills; however, Al Fayyad et al found the opposite. In their study, concurrent (immediate) feedback was perceived as superior in learning basic surgical skills compared to summary (delayed) feedback.⁵¹ In our study, summary feedback was chosen because the students operated on a simulation model without the risk of doing any harm. With an actual patient, a trainee needs guidance from a surgeon using concurrent feedback to avoid harmful errors.

Limitations

This study has several limitations. First, the sample size could not be calculated beforehand due to the novelty of the study aim and design. Although it was sufficient to reveal significant differences in terms of time and path length, the sample size could have been too small to detect significant differences in OSATS scores. Second, not all participants were able to complete the procedure within given 30 minutes. As the step of hernia sac removal was reached by most participants, it was used as the endpoint to ensure a justified comparison in terms of time and path length. Allowing participants to complete the entire procedure would have provided a better display of their performance. Third, the participants were medical students with low and slightly various levels of anatomical knowledge and technical skills, including suturing. Due to random allocation, these differences are expected to have little to no effect on outcomes. Additionally, the mean improvement in outcome measures was chosen instead of the absolute scores to account for those differences. Another limitation is the possible inability to generalize the conclusions to left-handed students because this study only included right-handed students. Furthermore, these findings cannot be generalized to other procedures outside of inguinal hernia repair. Last, the effect of OCHRA feedback was evaluated in a simulated environment. This study should be repeated among surgical residents with higher levels of anatomical knowledge and technical skills in a clinical setting on multiple procedures.

The findings of this study have implications for both practice and research. In this study, the open inguinal hernia repair was chosen as an exemplary procedure. It is unknown whether an inguinal hernia repair simulation is ideally suited to detect the effects of different types of feedback on study outcomes. The implementation of structured, stepwise feedback that is aligned with the learning activities should be considered especially in the early phases of surgical training. The aligned stepwise instruction using stepwise video demonstrations and procedure specific OCHRA checklist assessment can be transferred to other surgical procedures. The stepwise segmentation of a surgical procedure can be made using the step-by-step framework.²⁶ This stepwise description of a surgical procedure can then be used to create a procedure-specific OCHRA checklist. Moreover, an aptitude-based approach in teaching and learning of surgical procedural skills could be of benefit for individuals with lower VSA. As demonstrated, it is crucial to consider the modifying effect of VSA on surgical outcomes when setting up new research. In fact, when overall outcomes are not evaluated for different levels of VSA, the real differences may remain unrevealed.

CONCLUSIONS

In conclusion, a task-specific, stepwise feedback checklist (OCHRA) proves to be more effective in improving surgical skills, in terms of time and path length, among surgical novices compared to a global rating scale feedback (OSATS). The effects of a task-specific feedback are present mostly in individuals with lower levels of VSA.

REFERENCES

1. El Boghdady M, Alijani A. Feedback in surgical education. *The Surgeon* 2017;15(2):98-103.
2. Porte MC, Xeroulis G, Reznick RK, Dubrowski A. Verbal feedback from an expert is more effective than self-accessed feedback about motion efficiency in learning new surgical skills. *Am J Surg.* 2007;193(1):105-10.
3. Xeroulis GJ, Park J, Moulton C-A, Reznick RK, LeBlanc V, Dubrowski A. Teaching suturing and knot-tying skills to medical students: A randomized controlled study comparing computer-based video instruction and (concurrent and summary) expert feedback. *Surgery* 2007;141(4):442-9.
4. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. *Am J Surg.* 2011;202(4):469-80.
5. Van Hove P, Tuijthof G, Verdaasdonk E, Stassen L, Dankelman J. Objective assessment of technical surgical skills. *BJS* 2010;97(7):972-87.
6. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *BJS* 1997;84(2):273-8.
7. Niitsu H, Hirabayashi N, Yoshimitsu M, Mimura T, Taomoto J, Sugiyama Y, et al. Using the Objective Structured Assessment of Technical Skills (OSATS) global rating scale to evaluate the skills of surgical trainees in the operating room. *Surgery today.* 2013;43(3):271-5.
8. Tang B, Hanna GB, Joice P, Cuschieri A. Identification and categorization of technical errors by Observational Clinical Human Reliability Assessment (OCHRA) during laparoscopic cholecystectomy. *Arch Surg.* 2004;139(11):1215-20.
9. Biggs J. Enhancing teaching through constructive alignment. *Higher educ.* 1996;32(3):347-64.
10. Levin M, McKechnie T, Khalid S, Grantcharov TP, Goldenberg M. Automated Methods of Technical Skill Assessment in Surgery: A Systematic Review. *J Surg Educ* 2019;76(6):1629-39.
11. Datta V, Mackay S, Mandalia M, Darzi A. The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J Am Coll Surg.* 2001;193(5):479-85.
12. Shaharan S, Nugent E, Ryan DM, Traynor O, Neary P, Buckley D. Basic Surgical Skill Retention: Can Patriot Motion Tracking System Provide an Objective Measurement for it? *J Surg Educ* 2016;73(2):245-9.
13. Genovese B, Yin S, Sareh S, DeVirgilio M, Mukdad L, Davis J, et al. Surgical Hand Tracking in Open Surgery Using a Versatile Motion Sensing System: Are We There Yet? *The American surgeon.* 2016;82(10):872-5.
14. Corvetto MA, Fuentes C, Araneda A, Achurra P, Miranda P, Viviani P, et al. Validation of the imperial college surgical assessment device for spinal anesthesia. *BMC anesthesiology.* 2017;17(1):131.
15. Wanzel KR, Hamstra SJ, Caminiti MF, Anastakis DJ, Grober ED, Reznick RK. Visual-spatial ability correlates with efficiency of hand motion and successful surgical performance. *Surgery.* 2003;134(5):750-7.
16. Gordon HW. The cognitive laterality battery: tests of specialized cognitive function. *Int. J. Neurosci.* 1986;29(3-4):223-44.
17. Kozhevnikov M, Hegarty M. A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & cognition.* 2001;29(5):745-56.

18. Vajsbaher T, Schultheis H, Francis NK. Spatial cognition in minimally invasive surgery: a systematic review. *BMC surgery*. 2018;18(1):94.
19. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and technical skills performance in health care: a systematic review. *Med Educ* 2015;49(11):1065-85.
20. Roach VA, Brandt MG, Moore CC, Wilson TD. Is three-dimensional videography the cutting edge of surgical skill acquisition? *Anat Sci Educ* 2012;5(3):138-45
21. Bogomolova K, van der Ham IJM, Dankbaar MEW, van den Broek WW, Hovius SER, van der Hage JA, et al. The Effect of Stereoscopic Augmented Reality Visualization on Learning Anatomy and the Modifying Effect of Visual-Spatial Abilities: A Double-Center Randomized Controlled Trial. *Anat Sci Educ* 2020;13:558-567
22. Cui D, Wilson TD, Rockhold RW, Lehman MN, Lynch JC. Evaluation of the effectiveness of 3D vascular stereoscopic models in anatomy instruction for first year medical students. *Anat Sci Educ* 2017;10(1):34-45.
23. Roach VA, Mistry MR, Wilson TD. Spatial visualization ability and laparoscopic skills in novice learners: evaluating stereoscopic versus monoscopic visualizations. *Anat Sci Educ* 2014;7(4):295-301.
24. Tchanchaleishvili V, Myers PO. Left-handedness—a handicap for training in surgery? *J Surg Educ* 2010;67(4):233-6.
25. Nazari T, Simons M, Zeb M, van Merriënboer J, Lange J, Wiggers T, et al. Validity of a low-cost Lichtenstein open inguinal hernia repair simulation model for surgical training. *Hernia* 2020;24:895-901.
26. Nazari T, Vlieger E, Dankbaar M, van Merriënboer J, Lange J, Wiggers T. Creation of a universal language for surgical procedures using the step-by-step framework. *BJS Open*. 2018;2(3):151-7.
27. Nazari T, van de Graaf FW, Dankbaar MEW, Lange JF, van Merriënboer J, Wiggers T. One Step at a Time: Step by Step Versus Continuous Video-Based Learning to Prepare Medical Students for Performing Surgical Procedures. *J Surg Educ* 2020;77(4):779-87.
28. Langlois J, Bellemare C, Toulouse J, Wells GA. Spatial abilities and anatomy knowledge assessment: A systematic review. *Anat Sci Educ* 2017;10(3):235-41.
29. Vandenberg SG, Kuse AR. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and motor skills* 1978;47(2):599-604.
30. Wanzel KR, Hamstra SJ, Anastakis DJ, Matsumoto ED, Cusimano MD. Effect of visual-spatial ability on learning of spatially-complex surgical skills. *Lancet* 2002;359(9302):230-1.
31. Foster J, Miskovic D, Allison A, Conti J, Ockrim J, Cooper E, et al. Application of objective clinical human reliability analysis (OCHRA) in assessment of technical performance in laparoscopic rectal cancer surgery. *Techniques in coloproctology*. 2016;20(6):361-7.
32. Miskovic D, Ni M, Wyles SM, Parvaiz A, Hanna GB. Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. *Surgical endoscopy*. 2012;26(3):796-803.
33. Talebpour M, Alijani A, Hanna G, Moosa Z, Tang B, Cuschieri A. Proficiency-gain curve for an advanced laparoscopic procedure defined by observation clinical human reliability assessment (OCHRA). *Surg Endosc* 2009;23(4):869-75.
34. Faulkner H, Regehr G, Martin J, Reznick R. Validation of an objective structured assessment of technical skill for surgical residents. *Academic medicine: journal of the Association of American Medical Colleges*. 1996;71(12):1363-5.
35. Vaidya A, Aydin A, Ridgley J, Raison N, Dasgupta P, Ahmed K. Current Status of Technical Skills Assessment Tools in Surgery: A Systematic Review. *J Surg Research*. 2020;246:342-78.

36. Atesok K, Satava RM, Marsh JL, Hurwitz SR. Measuring surgical skills in simulation-based training. *JAAOS-Journal of the American Academy of Orthopaedic Surgeons*. 2017;25(10):665-72.
37. Binkley J, Bukoski AD, Doty J, Crane M, Barnes SL, Quick JA. Surgical simulation: markers of proficiency. *J Surg Educ* 2019;76(1):234-41.
38. Datta V, Chang A, Mackay S, Darzi A. The relationship between motion analysis and surgical technical assessments. *J Am Surg* 2002;184(1):70-3.
39. Cohen SA. Instructional alignment: Searching for a magic bullet. *Educational Researcher*. 1987;16(8):16-20.
40. Ilgen JS, Ma IW, Hatala R, Cook DA. A systematic review of validity evidence for checklists versus global rating scales in simulation-based assessment. *Med Educ* 2015;49(2):161-73.
41. Dankbaar ME, Stegers-Jager KM, Baarveld F, van Merriënboer JJ, Norman GR, Rutten FL, et al. Assessing the assessment in emergency care training. *PLoS One*. 2014;9(12):e114663.
42. Groenier M, Schraagen JM, Miedema HA, Broeders IA. The role of cognitive abilities in laparoscopic simulator training. *Adv Health Sci Educ* 2014;19(2):203-17.
43. Hedman L, Ström P, Andersson P, Kjellin A, Wredmark T, Felländer-Tsai L. High-level visual-spatial ability for novices correlates with performance in a visual-spatial complex surgical simulator task. *Surg Endosc* 2006;20(8):1275-80.
44. Schlickum M, Hedman L, Enochsson L, Henningsohn L, Kjellin A, Felländer-Tsai L. Surgical simulation tasks challenge visual working memory and visual-spatial ability differently. *World J Surg* 2011;35(4):710-5.
45. Hedman L, Klingberg T, Enochsson L, Kjellin A, Felländer-Tsai L. Visual working memory influences the performance in virtual image-guided surgical intervention. *Surg Endosc* 2007;21(11):2044-50.
46. Cook DA. The research we still are not doing: an agenda for the study of computer-based learning. *Acad Med* 2005;80(6):541-8.
47. Corraini P, Olsen M, Pedersen L, Dekkers OM, Vandenbroucke JP. Effect modification, interaction and mediation: an overview of theoretical insights for clinical investigators. *Clin Epidemiol* 2017;9:331-8.
48. van Merriënboer JJ, Sweller J. Cognitive load theory in health professional education: design principles and strategies. *Med Educ* 2010;44(1):85-93.
49. Hopmans CJ, den Hoed PT, van der Laan L, van der Harst E, van der Elst M, Mannaerts GH, et al. Assessment of surgery residents' operative skills in the operating theater using a modified Objective Structured Assessment of Technical Skills (OSATS): a prospective multicenter study. *Surgery* 2014;156(5):1078-88.
50. Tsai AY, Mavroveli S, Miskovic D, van Oostendorp S, Adamina M, Hompes R, et al. Surgical Quality Assurance in COLOR III: Standardization and Competency Assessment in a Randomized Controlled Trial. *Ann Surg* 2019;270(5):768-74.
51. Al Fayyadh MJ, Hassan RA, Tran ZK, Kempenich JW, Bunegin L, Dent DL, et al. Immediate auditory feedback is superior to other types of feedback for basic surgical skills acquisition. *J Surg Educ* 2017;74(6):e55-e61.

