ORIGINAL REPORTS

Effect of Feedback Modality on Simulated Surgical Skills Learning Using Automated Educational Systems– A Four-Arm Randomized Control Trial

Recai Yilmaz, MD,[*,†] *Ali M. Fazlollahi, MSc,*[*,†] *Alexander Winkler-Schwartz, MD, PhD,*[*,‡] *Anna Wang, BSc,*[*] *Hafila Hassan Makhani, BSc,*[*] *Ahmad Alsayegh, MBBS, MSc,*[*,‡,§] *Mohamad Bakhaidar, MBBS, MSc,*[*,‡,§] *Dan Huy Tran, MSc,*[*] *Carlo Santaguida, MD,*[‡] *and Rolando F. Del Maestro, MD, PhD*[*,†,‡]

[*]Neurosurgical Simulation and Artificial Intelligence Learning Centre, Department of Neurology & Neurosurgery, Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada; [†]Faculty of Medicine and Health Sciences, McGill University, Montreal, Canada; [‡]Department of Neurology and Neurosurgery, Montreal Neurological Institute and Hospital, McGill University, Montreal, Quebec, Canada; and [§]Division of Neurosurgery, Department of Surgery, Faculty of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia

OBJECTIVE: To explore optimal feedback methodologies to enhance trainee skill acquisition in simulated surgical bimanual skills learning during brain tumor resections.

HYPOTHESES: (1) Providing feedback results in better learning outcomes in teaching surgical technical skill when compared to practice alone with no tailored performance feedback. (2) Providing more visual and visuospatial feedback results in better learning outcomes when compared to providing numerical feedback.

DESIGN: A prospective 4-parallel-arm randomized controlled trial.

SETTING: Neurosurgical Simulation and Artificial Intelligence Learning Centre, McGill University, Canada.

PARTICIPANTS: Medical students ($n = 120$) from 4 Quebec medical schools.

RESULTS: Participants completed a virtually simulated tumor resection task 5 times while receiving 1 of 4 feedback based on their group allocation: (1) practice-alone without feedback, (2) numerical feedback, (3) visual feedback, and (4) visuospatial feedback. Outcome measures were participants' scores on 14-performance metrics and the number of expert benchmarks achieved during each task. There were no significant differences in the first task which determined baseline performance. A statistically significant interaction between feedback allocation and task repetition was found on the number of benchmarks achieved, $F(10.558, 408.257)=3.220$, $p < 0.001$. Participants in all feedback groups significantly improved their performance compared to baseline. The visual feedback group achieved significantly higher number of benchmarks than the practice-alone group by the third repetition of the task, $p = 0.005$, 95%CI [0.42 3.25]. Visual feedback and visuospatial feedback improved performance significantly by the second repetition of the task, $p = 0.016$, 95%CI [0.19 2.71] and $p = 0.003$, 95%CI [0.4 2.57], respectively.

CONCLUSION: Simulations with autonomous visual computer assistance may be effective pedagogical tools in teaching bimanual operative skills via visual and visuospatial feedback information delivery. (J Surg Ed

Funding: This work was supported by a Brain Tumor Foundation of Canada Brain Tumor Research Grant, a Medical Education Research Grant from the Royal College of Physicians and Surgeons of Canada, the Franco Di Giovanni Foundation, and the Montreal Neurological Institute and Hospital, along with grants from the Fonds de recherche du Québec−Santé − Formation de doctorat, and a Max Binz Fellowship from McGill University Internal Studentships. A prototype of the NeuroVR used in this study was provided by the National Research Council of Canada, Boucherville, Quebec, Canada.

Previous Presentation: Portions of this work were presented at the American Association of Neurological Surgeons (AANS) Annual Meeting in Los Angeles, California, USA, on April 21-24, 2023

Data Availability Statement: The dataset analyzed in this study is available from the corresponding author on a reasonable request. A sample raw simulation data file is available online: https://doi.org/10.6084/m9.figshare.15132507.v1.

Correspondence: Inquiries to Recai Yilmaz, MD, Neurosurgical Simulation and Artificial Intelligence Learning Centre, Montreal Neurological Institute and Hospital, 300 Rue Léo Pariseau, Suite 2210, Montreal, Quebec, Canada, H2X 4B3.; e-mail: recai.yilmaz@mail.com

  1931-7204/$30.00
https://doi.org/10.1016/j.jsurg.2023.11.001

000:1−13. © 2023 Association of Program Directors in Surgery. Published by Elsevier Inc. All rights reserved.)

KEY WORDS: neurosurgery, simulation, surgical education, technical skill, virtual reality

COMPETENCIES: Neurosurgery, Simulation, Surgical Education, Learning Feedback, Technical Skills, Virtual Reality

INTRODUCTION

In medical education, advancing educational technologies promise to support trainee learning.[1] Among these, computer-assisted tools, such as artificial intelligent tutors, emerged as appropriate candidates to guide independent learning, and some offered advantages over traditional learning.[2] In surgical education, simulation platforms equipped with automated feedback systems allow learners to practice their bimanual surgical skills in a risk-free environment without the need for supervision.[3,4] This liberates instructors' time to be invested in other aspects of patient care or surgical education such as mentorship. A key technical advantage of these computer-assisted systems is their ability to differentiate the expertise level of surgeons with granularity and precision.[3,5] This not only presents new perspectives to understand the composites of expertise, but increases efficiency in trainee learning by providing quantifiable learning objectives, for which specific feedback and actionable goals can be directed to improve performance.[2] In addition, these systems can provide trainees with detailed visuospatial information about their bimanual performance which may increase their three-dimensional appreciation of surgical performance on anatomical structures.[6]

In medical education, extensive research is conducted to design effective curricula.[7-9] Teaching methodologies focus on increasing trainee engagement in learning while the students efficiently master their skills. Although quantifying surgical bimanual skills serves the purpose of providing objective feedback, this data can be presented to learners in a variety of formats such as numerical, visual, spatial, video, haptic, and auditory.[3,10,11] However, because of the relative recency of these educational tools in surgical simulation training, more research is needed to evaluate the effectiveness of various feedback modalities to maximize efficiency in teaching technical skills. This randomized control trial investigated the effect of four feedback protocols including numerical, visual, and visuospatial feedback along with practice alone with no tailored performance feedback, as a control, to evaluate the rate of acquisition of technical skills of medical students. The objectives were: (1) To explore the effect of feedback to the learning rate in surgical simulation training in comparison with practice without feedback. (2) To determine how more visual and spatial feedback modalities compare with numerical feedback.

METHODS

Setting

This four-parallel arm randomized controlled trial (trial registration: ISRCTN17590019) was conducted at the Neurosurgical Simulation and Artificial Intelligence Learning Centre, McGill University, Montreal, Canada. Medical students in first to fourth year from 4 universities in the Province of Quebec were invited to participate in the trial. Data was collected between July 2019 and October 2020, in 60-minute simulation sessions with no follow-up (Fig. 1). One hundred and twenty medical students participated in the trial, and no exclusion criteria were applied. No changes were made to the methods after trial commencement. An online random number generator was used to determine participant group allocation. Study procedures were in accordance with the ethical standards of the responsible committee on human experimentation (institutional and national) and with the Declaration of Helsinki.[12] COVID-19 public health measurements and the Montreal Neurological Institute and Hospital's protocols were followed to ensure participants' and researchers' safety during the conduct of the study. The time frame of the trial was predetermined with no restrictions on the number of simulation sessions that could take place. The trial participation was terminated with the restrictions imposed by changes to public health protocols due to COVID-19 pandemic in October 2020 while the number of participants sufficed a statistical power of 0.99 for between- within-group interaction. This study was approved by the McGill University Health Centre Research Ethics Board, Neurosciences-Psychiatry. An approved consent form was signed by all study participants before trial participation. All participants filled a pre-questionnaire related to demographics and previous simulation experience and surgical exposure (Table 1). A postquestionnaire was completed after the trial for the rating of the simulation learning (Supplementary Table 2). This report adheres to guidelines for the reporting of multi-arm parallel group randomized trials, extension of the CONSORT 2010 Statement.[13] Study interventions involved no harm to participants. Participants were informed that their information will be anonymized, and despite the careful measures taken to avoid the chance

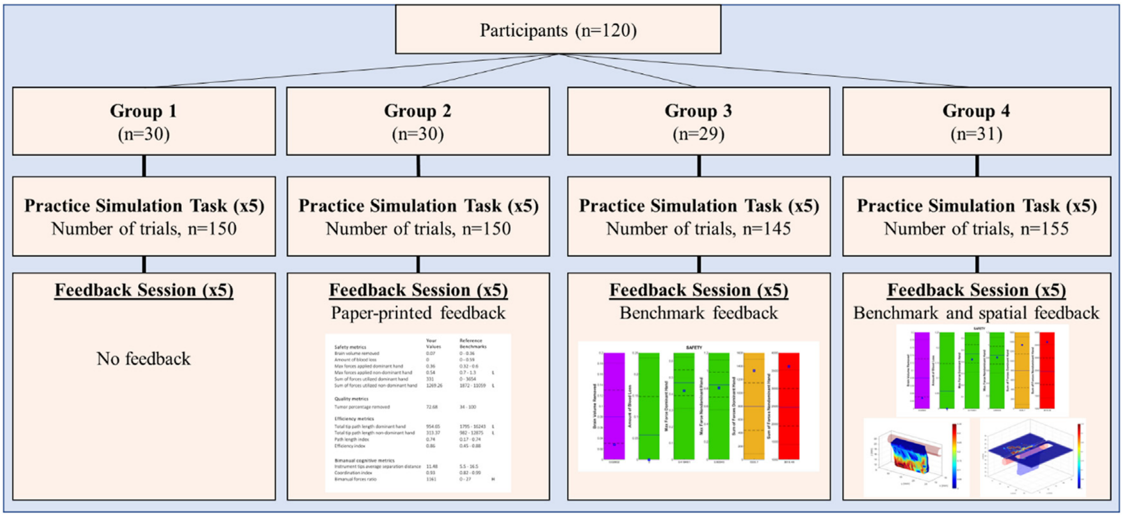

FIGURE 1. Flow diagram. One hundred twenty students were randomly allocated into 4 different feedback groups including practice-alone with no tailored performance feedback group. No participant/data was excluded from the analysis.

TABLE 1. Participant Characteristics

	Group 1 Practice Alone With No Feedback (*n* = 30)	**Group 2 No Visual Feedback (*n* = 30)**	**Group 3 Visual Feedback (*n* = 29)**	**Group 4 Visuospatial Feedback (*n* = 31)**	**All Participants (*n* = 120)**
Mean age ± SD (range)	23.6 ± 4.8 (19-44)	22.8 ± 3.3 (19-31)	22.4 ± 2.6 (19-28)	23.6 ± 3.5 (18-33)	23.1 ± 3.6 (18-44)
Male/female	18/12	18/12	18/11	17/14	71/49
Handedness (right/left/ambidextrous)	27/3/0	28/2/0	24/4/1	29/1/1	108/10/2
Medical school:					
McGill University	24	22	21	25	92
University of Montreal	5	6	4	5	20
University of Sherbrooke	1	2	3	1	7
University of Laval	0	0	1	0	1
Year in medical school:					
1st	16	21	18	20	75
2nd	10	6	7	8	31
3rd	3	2	2	1	8
4th	1	1	2	2	6
Level of interest in surgery, median (range)	4 (2-5)	4 (1-5)	4 (1-5)	4 (1-5)	4 (1-5)
Completed surgical rotation (Y/N)	2/28	1/29	2/27	2/29	7/113
Playing video games:					
Not at all	12	13	13	13	51
Occasionally (less than 2 hours per week)	9	9	7	9	34
Often (2-10 hours per week)	6	8	6	6	26
Very often (more than 10 hours per week)	3	0	3	3	9
Playing musical instruments:					
I don't play any musical instrument	11	14	9	17	51
Yes, I am at beginner level	6	4	6	3	19
Yes, I am at intermediate level	6	7	8	6	27
Yes, I am at advanced level	6	4	4	5	19
Yes, I am at master level	1	0	2	0	3
Previously used virtual reality simulation (Y/N)	1/29	2/28	0/29	2/29	5/115

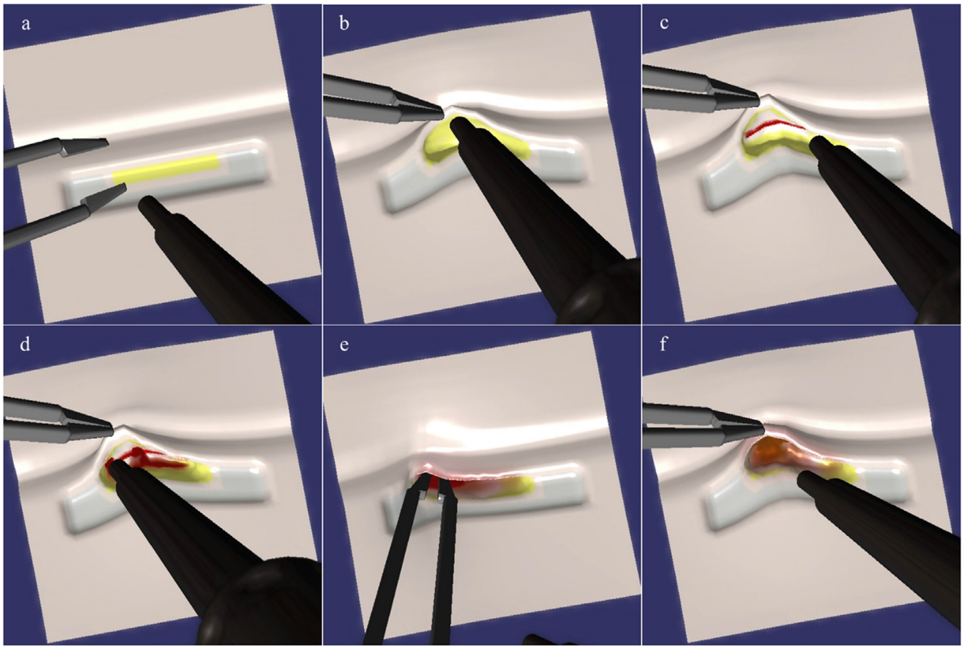

FIGURE 2. Simulated scenario. (a) The virtually simulated task involved the subpial resection of a rectangular yellow tumor using an ultrasonic aspirator in the dominant hand and bipolar forceps in the non-dominant hand. (b) The goal of the task was to remove the tumor completely while minimizing injury to surrounding tissues. (c) There was a blood vessel with ability to bleed, located posterior to the tumor. (d) Any damage to this blood vessel resulted in bleeding. (e) Ultrasonic aspirator was used to aspirate the blood and bipolar was used to cauterize the bleeding vessel. (f) The appearance of the tissue after successful cauterization.

that they may be identified, their trial performance would have no influence on their academic evaluation.

Simulation Setting

The NeuroVR (previously NeuroTouch) neurosurgical simulation platform (CAE Healthcare, Montreal, Canada) with haptic feedback was utilized.[14] The haptic feedback integrated in the instrument handles was to provide a more realistic experience for all participants of the study regardless of the feedback interventions they receive for learning. This haptic technology allowed the integration of learning feedback on instrument force utilization for the study groups as trainees interact with delicate tissues during the simulated performance. The simulated task was previously developed to replicate the subpial resection of brain tumors.[15] Participants performed this simulated subpial tumor resection task 5 times with 5 minutes given to complete each task. The simulated scenario included the subpial resection of a yellow rectangular tumor (Fig. 2) using a simulated ultrasonic aspirator and bipolar forceps to completely remove the tumor within the time limit while minimizing damage to the surrounding tissue which mimics the adjacent normal gyrus.[5,16] Both instruments were activated using pedals. Part of the tumor was placed under healthy brain tissue where lifting this simulated pial layer using the bipolar was necessary to gain access and remove the remaining underlying tumor. A blood vessel was incorporated into the simulation adjacent to the distal tumor wall and bleeding resulted from injury to this vessel. Bleeding was controlled utilizing the cauterizing function of the bipolar forceps (Fig. 2e). The NeuroVR platform recorded performance data in 20-millisecond increments (50 recording per second) involving time, the information of force applied by the 2 instruments, instrument tip location, amount of tissue and tumor removed, amount of bleeding, and pedal activation.

Expert Level Benchmarks

Expert level benchmarks were developed using previously validated 14-performance metrics,[3,5,17] described in the results section. The data used to develop these benchmarks was previously available in our center and was recorded during 14 neurosurgeons' performance on the same simulated tumor resection task. Using this dataset, expert mean, and standard deviation values were calculated for each performance metrics to define the limits of the expert level benchmark. A metric score between 1 standard deviation above and below the mean was considered within the benchmark for that task.

Feedback Setting

Four feedback protocols included (1) practice alone with no tailored performance feedback, (2) numerical feedback, (3) visual feedback, and (4) visuospatial feedback. All participants received standard verbal and written instructions before the start of the trial including how to use the simulator handles to carry out the simulated procedure and the feedback information they would be provided with. All participants were also informed concerning the 14-performance metrics that would be used to assess their performance. The data recorded by the simulator was used to calculate participants' metric scores and determine whether they are within the benchmarks. Participants were given 5 minutes between the tasks either to rest or receive the feedback information corresponding their group allocation. After each task, participants in Group-1 (n=30) received no tailored performance feedback. In Group-2 (n = 30), participants received a printed copy of their performance scores on the 14 metrics that was compared with expert level benchmarks (Supplementary Fig. 1). Any performance score falling above or below the expert benchmark was indicated with a letter "H" (higher) or "L" (lower), respectively. In Group-3 (n = 29), participants received a screen-based graphical representation of their performance scores on the 14 metrics. The graphics were green colored for each performance metrics if participant's score was within the benchmark, yellow if their score was between 1 and 2 standard deviations of the benchmark, or red if their score was outside 2 standard deviations of the benchmark (Supplementary Fig. 2). The graphics were also represented in purple for any performance score that was better than the benchmark. Participants in Group-4 (n = 31) received the same colored-graphical demonstration but additionally, they were shown two 3D spatial models that showed the anatomical structures of the tumor and pial surface. The amount of force applied on these tissues by the ultrasonic aspirator and the bipolar were shown according to the color scale ranging from red to blue, where red indicated a higher force applied (Supplementary Fig. 3). For all groups, the number of benchmarks achieved was calculated across 5 repetitions of the task. Automated feedback during the trial, data analysis and visualization were performed using MATLAB (The MathWorks Inc, Natick, Massachusetts, USA) release 2021a. All codes were written by the authors.

Hypotheses

(1) Participants in feedback groups will achieve significantly higher number of benchmarks than those who practice without feedback. (2) Participants who receive visual and visuospatial feedback will achieve significant improvement earlier across the 5 repetitions of the task than those who receive only numerical information.

Statistical Analysis

A priori sample size calculation, with a statistical power of 0.9, an effect size of 0.3, a correlation of 0.5 among repeated measures, and an alpha error probability of 0.05 for between groups comparison yielded a requirement of 25 participants in each group, and 100 participants in total. The participation of 120 students provided an achieved statistical power of 0.95. Two-way mixed Analysis of Variance (ANOVA) explored the interaction of feedback group assignment (between-groups) and task repetition (within-groups) on participants number of benchmarks achieved. There were no outliers, as assessed by visual examination of studentized residuals for values greater than ± 3. Data was normally distributed, as visually assessed by Normal Q-Q Plot. Levene's test showed homogeneity of variances, based on median ($p > 0.05$), and Box's test demonstrated homogeneity of covariances, $p = 0.948$. Mauchly's test of sphericity indicated that the assumption of sphericity was violated for the two-way interaction, $\chi 2(9) = 34.92$, $p < 0.001$. Therefore, the results with Greenhouse-Geisser correction are reported. Differences between-feedback groups were investigated using one-way ANOVA. Within-feedback group differences were analyzed using one-way repeated measures ANOVA. Between-feedback group post hoc analyses were done using Tukey HSD or Games Howell tests depending on the homogeneity or heterogeneity of variances, respectively. Within-group post hoc analyses were done using Bonferroni post hoc tests. Cohen's d effect sizes were reported for post hoc comparisons.[18] The variable "number of benchmarks achieved" was assumed as a ratio variable, having the meaningful 0 point (no success). As such, our analyses were done using parametric statistical tests described above. Non-parametric equivalent statistical analysis was also reported in the supplementary data (Supplementary Fig. 4). Statistical analysis was done using IBM SPSS Statistics, Version 27.

RESULTS

Participants

Participants' average age (mean [SD, min-max]) was 23.1 [3.6, 18-44] years and participant handedness was 108/10/2 (right-handed/left-handed/ambidextrous) (Table 1). Five participants previously used virtual reality simulation.

Data and Performance Metrics

Data from 120 participants, from a total of 600 trials, was available for analysis (, Flow diagram) and there was no missing data. Participants' performance progress was tracked across 5 repetitions of the task on 14-performance metrics from 4 categories (1) safety, (2) quality, (3) efficiency, and (4) bimanual cognitive. Safety category included 6 metrics: (1) brain volume removed (cc), (2) amount of blood loss (cc), (3) maximum force applied with dominant hand (N), (4) maximum force applied with non-dominant hand (N), (5) sum of forces applied with dominant hand (N), and (6) sum of forces applied with non-dominant hand (N). Quality category included only tumor percentage removed. Efficiency category included 4 metrics: (1) total tip path length dominant hand (mm), (2) total tip path length non-dominant hand (mm), (3) path length index, (4) efficiency index. Bimanual cognitive category included (1) average instrument tips separation distance (mm), (2) coordination index, and (3) bimanual forces ratio. Descriptions of the performance metrics can be found on Supplementary Table 1.

Learning Curves

No statistical difference was found between groups at baseline performance ($p = 0.121$). There was a statistically significant interaction between the feedback group allocation and the number of repetitions of the task on the number of benchmarks achieved, $F(10.558, 408.257) = 3.220$, $p < 0.001$, effect size (partial $\eta2) = 0.077$, $\varepsilon = 0.88$ (Fig. 3). Group-3 made the quickest improvement where the number of benchmarks achieved was significantly higher than Group-1 by the third repetition of the task ($p = 0.005$, 95%CI [0.42 3.25], effect size (Cohen's d)=0.878). Group-4 outperformed Group-1 by the fourth repetition of the task ($p = 0.002$, 95%CI [0.54 3.00], effect size=1.035) while Group-2 did not outperform Group-1 within the 5 repetitions. In the final repetition of the task, Group-4 achieved 9.19 ± 1.66 (mean $\pm$ standard deviation) of the 14 benchmarks, Group-3 achieved 9.10 ± 1.82, Group-2 achieved 8.40 ± 2.06 while Group-1 achieved 7.30 ± 1.69 of the 14 benchmarks. Group-3 and Group-4 improved significantly from their baseline performance by the second repetition of the task ($p = 0.016$, 95%CI [0.19 2.71], effect size=0.746; and $p = 0.003$, 95%CI [0.4 2.57], effect size=0.885, respectively). Group-2 improved significantly from their baseline performance by the third repetition of the task ($p = 0.004$, 95%CI [0.42 3.04], effect size = 0.886) while Group-1 had no statistically significant improvement during the 5 repetitions.

Learning curves were also assessed for the 14-performance metrics. In the fifth repetition of the task, around 90% of participants in all groups, including no-tailored-

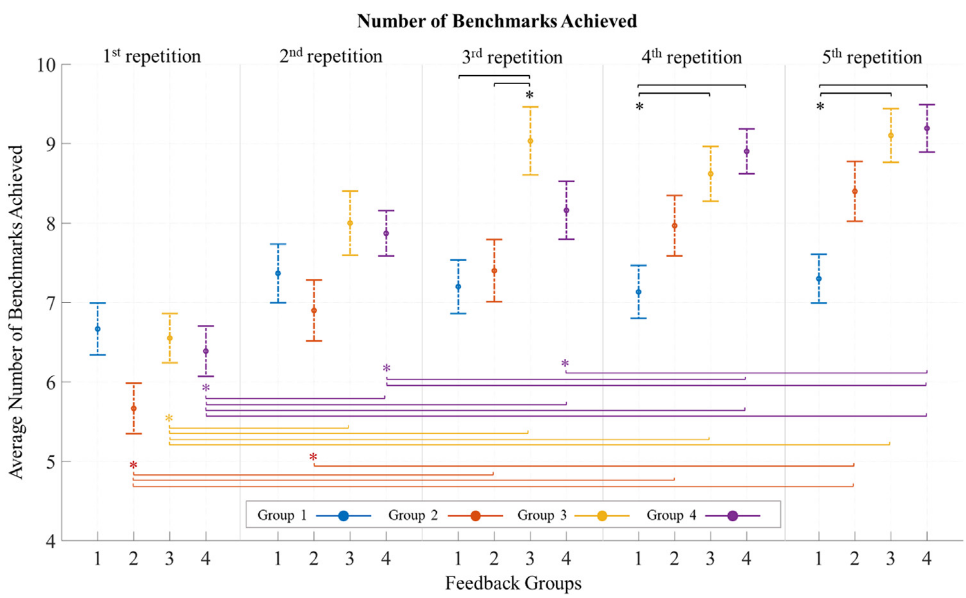

FIGURE 3. Number of benchmarks achieved. X-axis represents the 4 feedback groups. Each feedback group is color-coded (see the legend). Y-axis represents the average number of benchmarks achieved by each feedback group. *Horizontal lines represent statistically significant difference (p < .05). For within group differences, horizontal lines are represented with the respected color of the group. Vertical lines represent standard error bars. Group 3 and Group 4 improved significantly compared to the baseline performance by the second repetition. Group 2 improved significantly compared to baseline performance by the third repetition. Group 3 outperformed practice-alone Group 1 by the third repetition. Group 4 outperformed practice-alone Group 1 by the fourth repetition.

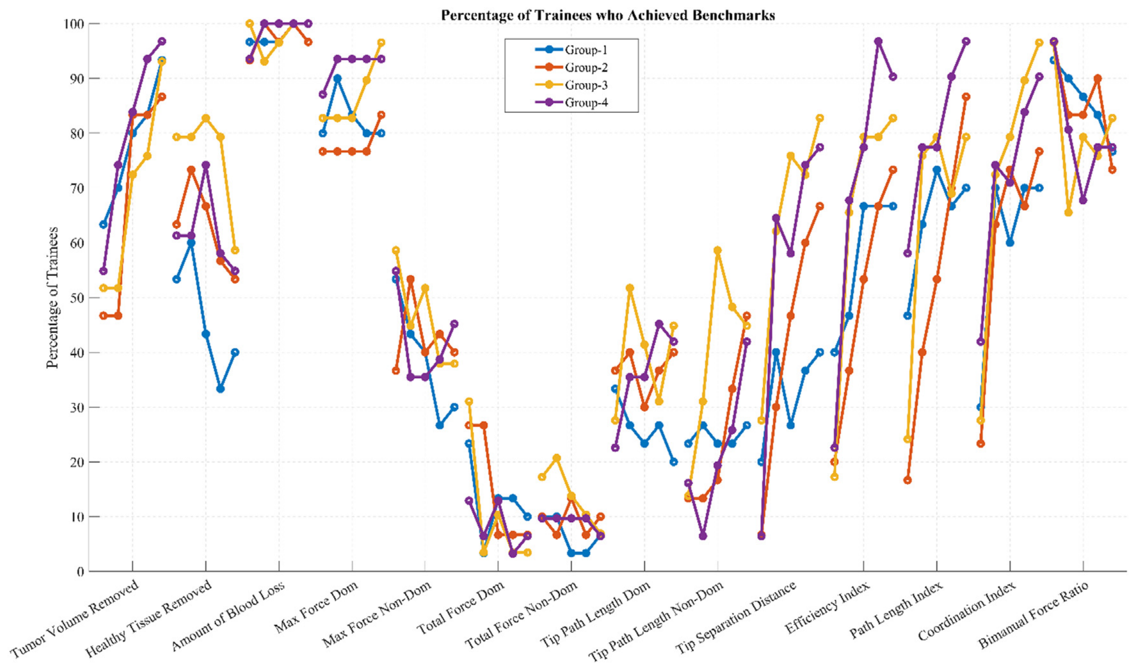

FIGURE 4. Percentage of trainees who achieved benchmarks. X-axis shows each of the 14-performance metrics on which the trainees were assessed. Each feedback group is color-coded (see the legend). Y-axis represents the percentage of trainees who achieved the benchmarks. There are 5 percentages shown for each performance metric across 5 trials, from the first repetition of the task/baseline performance to the fifth repetition.

feedback group, were within the tumor percentage removed benchmark (Fig. 4). All groups removed significantly more tumor in the fifth repetition of the task compared to baseline performance ($p < 0.05$) (Fig. 5a). With only feedback groups, participants achieved the benchmarks >50% of the time with the metrics healthy tissue removed and instrument tip separation distance. Group-1 caused significantly more healthy tissue damage than Group-3 in the third to fifth repetitions of the task ($p = 0.002$ 95%CI [0.03 0.16], effect size=0.998) (Fig. 5b). Participants in Group-4 had a statistically significant lower instrument tip separation distance (using the 2 instruments together) than Group-1 at the fourth and fifth repetitions of the task ($p < 0.001$ 95%CI [-4.97 -1.21], effect size=1.133), and this was also observed in participants in Group-3 from the second to fifth repetitions of the task ($p = 0.029$ 95%CI [-5.81 -0.23], effect size = 0.862) (Fig. 5c). Group-3 and Group-4 improved significantly in efficiency index by the second repetition of the task ($p < 0.001$ 95%CI [0.12 0.25], effect size = 1.780) and ($p < 0.001$ 95%CI [0.08 0.21], effect size=1.432, respectively) while the remaining groups improved significantly by the third repetition (Fig. 5d). The learning curves and statistical comparison of the metric scores of the remaining 10 performance metrics can be found in Supplementary Figure 5.

In the postquestionnaire 5-point Likert scale, participants rated their simulation learning experience (Supplementary Table 2). Students' rating in Group-3 and Group-4 for the question "How beneficial do you think the simulator and training system is for learning about surgery?" was 5.0 [3-5] (median [range]) while in Group-2 and Group-1, it was 4.0 [3-5]. Participants in feedback groups rated "How beneficial was it to your performance to know which metrics you were being assessed on?" 5.0 [3-5] while no-tailored-feedback group rated 4.0 [2-5].

DISCUSSION

In surgery, advanced computer technologies allow for the collection of vast amounts of data concerning technical skill, accurate skill assessment, and provide error detection and tailored feedback.[3-5,19,20] These systems used in virtual reality simulation training have been shown to enhance learner skills, and provide a more efficient training than remote post hoc human instruction. [2]

To put this work in context, providing trainees with efficient training feedback while challenging them in realistically replicated operative tasks required a series of components. First, virtual reality platforms with realistic surgical procedures and extensive data recording capacity were developed.[14,21-24] Second, performance metrics encompassing critical features concerning the surgical procedure such as safety, efficiency, and

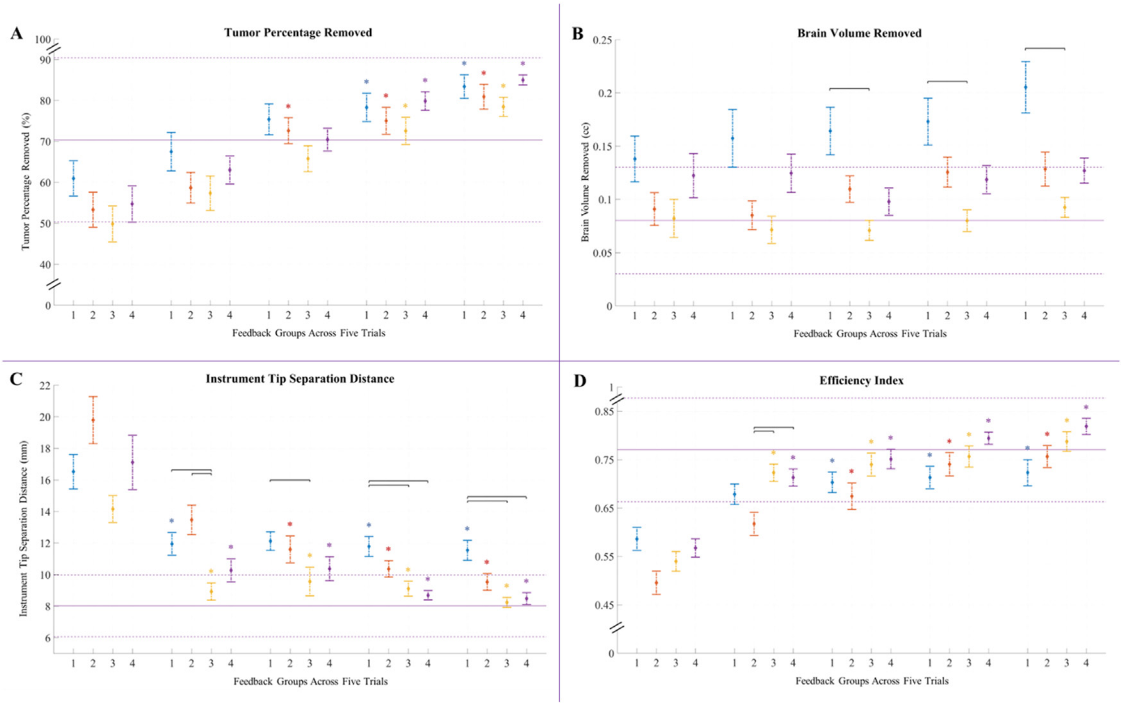

FIGURE 5. Performance metrics learning curves. The learning curves of 4 performance metrics. X-axes represent the task repetition from the first repetition/baseline performance to the fifth repetition for the 4-feedback groups. The purple straight horizontal line indicates the mean expert value for each performance metric while the 2 dotted purple lines one standard deviation above and below the mean indicate the boundaries of the expert benchmark. *Asterisks indicate significantly different values from the first repetition/baseline performance of that group. Horizontal square brackets show significant differences between feedback groups at the same repetition of the task. Axis brakes were indicated along y-axis. The learning curves of remaining 10 of the 14-performance metrics are shown in supplementary data.

performance quality along with bimanual dexterity and movement were developed to differentiate expertise groups and outline expert level performance benchmarks.[17,25,26] Spatial analysis of surgical performance using 3D tumor and tissue models has demonstrated differences between expert and novice level performances.[6,27] Third, artificial intelligence methodologies were employed to provide a comprehensive performance assessment and outline performance metrics critical to achieve expert level performance.[5,28,29] Fourth, feedback systems provided trainees with expert level performance benchmarks to improve bimanual skills, based on virtual reality artificial intelligence platforms.[3,4] After completing these steps, the current work explored the educational utility of these systems in improving trainee skills. We explored the efficacy of various instruction modalities by comparing numerical, visual, and visuospatial feedback.

In this study, the training sessions were organized based on time (number of repetitions) rather than defining a specific target proficiency level that trainees to achieve. This decision was influenced by the diverse training outcomes assessed and the time required for trainees to achieve proficiency in all 14-expert level benchmarks was unknown. Based on the results seen in Figure 4, achieving all 14 benchmarks would have been very challenging in a single training session, even for groups who received more efficient learning feedback.

Although Group-3 and Group-2 received the same metric information except for the application of color, Group-3 performed significantly better than Group-2 during the third repetition of the task. Additionally, Group-3 outperformed baseline performance in the second repetition of the task while Group-2 did not achieve the same success. The link between human color perception and psychological functioning is well studied.[30] In achievement contexts, such as education or athletic contests, psychologists have suggested that different colors cue learners' emotions and cognition which yields behavioral changes that can either optimize or impair performance.[31,32] Our results indicated that the colored visualization of the feedback information is critical in achieving more efficient training. In the future, computer assisted teaching systems including artificial intelligence applications may benefit from incorporating visually enriched feedback methodologies, which provides a more engaging learning feedback to maximize trainee surgical skill acquisition.[2,4,33] Similar training applications can provide benefits across different procedural medical disciplines.

In this study, 14-performance metric benchmarks were utilized to assess the simulated surgical performance and track improvement across the 5 repetitions of the tumor resection task. Some of the 14-performance metrics showed improvement for all groups regardless of feedback (Fig. 5 and Supplementary Fig. 5) because they may have epitomized some of the obvious goals of this surgical task. As such, all participants removed significantly more tumor (tumor percentage removed), achieved greater efficiency (efficiency index) and used their non-dominant hand more efficiently (coordination index, instrument tip separation distance) in the fifth repetition of the task (Fig. 5). However, feedback provided faster learning for the intervention groups and better performance improvement.

Although some of the performance metrics were expected to improve, the goal with some of the other metrics such as brain volume removed, was to stay within the benchmark (Fig. 5b) and to remove more tumor while not damaging the healthy tissue. Accordingly, both Group-1 and Group-3 removed the same amount of tumor, around 80%, while Group-3 harmed significantly less healthy tissue, used their dominant hand more precisely (lower total tip path length), and had significantly lower scores in instrument tip separation during the fifth repetition of the task. These results may indicate that feedback is necessary to achieve an appreciation of the complex interplay between multiple factors during tumor surgery to meet the goals of the task more safely and efficiently.

Real-time intelligent systems are being developed and tested in surgical bimanual skills training using virtual reality simulation.[3,34] Although this study has shown visual systems to be efficient for post hoc feedback, in future directions of this work, auditory instructions may be an alternative for real-time feedback applications to prevent visual distractions. Systems with audio, visual, and video feedback are combined in our current trials (ClinicalTrials.gov, NCT05168150) to provide engaging feedback information to trainees which may improve the amount of information received by trainees and their skill acquisition.[35]

The tailored information provided by the intelligent systems is important; however, the major advantage of computer systems in skills acquisition may be achieved by optimal combinations of visual and auditory feedback components (e.g., video). In a randomized controlled trial involving the resection of a simulated brain tumor resection task, participants were instructed by the Virtual Operative Assistance on 4 performance metrics selected by a support vector machine algorithm along with feedback demonstration videos.[2] Participants improved on a composite score based on 16 performance metrics, and on 8 of these 16 metrics changed significantly without receiving specific metric-based instructions.[36] Although the mechanism behind this extended effect is currently under investigation, a possible explanation is the breadth of extrinsic information contained in the feedback video demonstrations.[10] The ability to use both visual and auditory information may be the main advantage of these feedback systems in skills acquisition. To optimize the effectiveness in new feedback applications, it may be imperative to prioritize the pedagogical aspect of technical skill training and integrate informative, engaging, and easy to understand feedback information with the intelligent training systems.

This study has several limitations. (1) The training outcome in our simulation setting was limited to bimanual skills improvement. However, surgical operative room involves many other factors which can affect surgeon's performance and patient outcomes. Developing surgical simulation systems may provide a more immersive surgical training experience in the future. (2) Surgical trainees may be the most relevant trainee cohort for the testing of surgical training simulators. However, this study recruited medical students, a study cohort that may provide some advantages while also imposing limitations. Learning experience may differ as expertise develops.[37] Medical students' different interest level and procedural knowledge compared to surgical trainees may affect their surgical training interaction and skill acquisition; however, their limited experience provides a greater room for improvement in skill acquisition, a scenario closer to that of a fresh surgical trainee who has just started training. Additionally, a medical student cohort provides a large number of participants to obtain statistical power, which is difficult to obtain with the limited number of surgical trainees available. For these reasons, medical students may be a better cohort than surgical trainees especially for the development and testing phases of simulation and training systems. Once, these systems are well established, their efficacy in teaching and assisting surgical trainee cohorts should be confirmed in multi-institution trials. (3) Cognitive overload may limit the amount of information understood by the trainee. Cognitive load theory in education suggests that an optimal learning environment finds a balance between learners' intrinsic cognitive capacity, their motivation, and the extrinsic load of the instructional milieu.[38] Novice medical learners are also demonstrated to be at greater risk of overload in surgical simulation training.[39] In this application, training involved 1 session, in which learners sequentially removed 5 tumors, and were expected to improve on 14 performance metrics. The amount of information needed to master these 14 performance features in 1 session may overwhelm trainee cognitive capacity and limit skill acquisition. Cognitive overload may have limited the amount of improvement especially with the participants in Group-4 since providing extra visuospatial information to this group did not achieve better results. One can speculate that the ability of trainees in Group-4 to adequately review the complex additional visual and spatial information available to them in only the limited five-minute feedback session may have been difficult. This could have resulted in increased trainee stress, leaving less time for critical learning methods such as self-reflection and improvement planning.[40] Results of Group-3 may support this conclusion as this group made a faster improvement without the 3D spatial information, having a significantly greater number of benchmarks achieved than the baseline by the third repetition of the task. To prevent cognitive overload, longitudinal training settings with structured training goals in multiple sessions and/or different instruction methodologies may provide a better performance improvement.[41,42] These longitudinal settings may integrate visual and visuospatial feedback to achieve efficient learning settings as outlined in this study and help to assess and compare retention of skills. (4) Our focus in this study was to maximize efficiency in learning with visual assistance. This study did not incorporate tailored auditory, video, tactile feedback, or other possible feedback modalities. Computer systems may incorporate different feedback mechanisms, not being limited to visual feedback, while the feedback can be adjusted to user preference. Future studies may compare different feedback modalities and explore multimodal learning.[43] Using the haptic technology of the simulator, a tailored tactile feedback, such as vibration, can be implemented to inform the trainee when they apply too much force on delicate tissues.

In conclusion, this randomized controlled trial allowed the comparison of different post hoc feedback modalities in surgical technical skills learning in the simulated environment. Simulations with autonomous visual and visuospatial feedback assistance provided trainees with a more effective way to master their bimanual operative skills.

AUTHOR CONTRIBUTIONS

R.Y. contributed to conceptualization of the study, data acquisition, methodology, development and implementation of the feedback systems, data analysis, writing the original draft, critical revision of the manuscript for important intellectual content, statistical analysis, visualization, and the codes used in this study.

A.M.F. contributed to conceptualization of the study, data acquisition, critical revision of the manuscript for important intellectual content.

A.W.S. contributed to conceptualization of the study, development of the feedback systems, statistical analysis, and critical revision of the manuscript for important intellectual content.

A.W. contributed to conceptualization of the study, piloting the study, implementation of the feedback systems, and critical revision of the manuscript for important intellectual content.

H.H.M. contributed to conceptualization of the study, piloting the study, data analysis, and critical revision of the manuscript for important intellectual content.

A.A, M.B., and D.H.T. contributed to conceptualization of the study, and critical revision of the manuscript for important intellectual content.

C.S contributed to conceptualization of the study, critical revision of the manuscript for important intellectual content, obtained funding, and supervision.

R.F.D. contributed to conceptualization of the study, methodology, development, and implementation of the feedback systems, writing the original draft, critical revision of the manuscript for important intellectual content, obtained funding, administration, and supervision.

ACKNOWLEDGMENTS

The authors would like to thank the medical students who participated in this study, and the staff at the Montreal Neurological Institute and Hospital who helped to ensure public safety during the time of the COVID-19 pandemic which made conducting this trial possible. The authors would like to thank Dr. Jose Andres Correa, Department of Mathematics and Statistics, McGill University for his assistance in conducting statistical analyses; and Denis Laroche, National Research Council of Canada, Boucherville, Quebec, Canada, for his technical assistance with the NeuroVR simulator. The authors thank Mina Khan, Zhen Yan Cao, Ian Langleben, and Mathilde Cloutier-Lachance for their help in reaching out medical students for trial participation. The authors thank the National Research Council of Canada, Boucherville, Quebec, Canada and Dr. Abdulrahman Sabbagh, Division of Neurosurgery, Department of Surgery, College of Medicine, King Abdulaziz University, Jeddah, Saudi Arabia for assistance in developing the scenarios used in this study. The authors acknowledge the National Neurosciences Institute, King Fahad Medical City, Riyadh, Saudi Arabia, for assistance in funding the creation of the simulated subpial tumor resection.

REFERENCES

1. Mirchi N, Ledwos N, Del Maestro RF. Intelligent tutoring systems: re-envisioning surgical education in response to COVID-19. *Can J Neurol Sci*.

2021;48:198–200. https://doi.org/10.1017/cjn.2020.202.

2. Fazlollahi AM, Bakhaidar M, Alsayegh A, et al. Effect of artificial intelligence tutoring vs expert instruction on learning simulated surgical skills among medical students: a randomized clinical trial. *JAMA Netw Open*. 2022;5(2):e2149008-e. https://doi.org/10.1001/jamanetworkopen.2021.49008.

3. Yilmaz R, Winkler-Schwartz A, Mirchi N, et al. Continuous monitoring of surgical bimanual expertise using deep neural networks in virtual reality simulation. *NPJ Digital Med*. 2022;5(1):54. https://doi.org/10.1038/s41746-022-00596-8.

4. Mirchi N, Bissonnette V, Yilmaz R, Ledwos N, Winkler-Schwartz A, Del Maestro RF. The virtual operative assistant: an explainable artificial intelligence tool for simulation-based training in surgery and medicine. *PLoS One*. 2020;15(2):e0229596. https://doi.org/10.1371/journal.pone.0229596.

5. Winkler-Schwartz A, Yilmaz R, Mirchi N, et al. Machine learning identification of surgical and operative factors associated with surgical expertise in virtual reality simulation. *JAMA Netw Open*. 2019;2(8):e198363. https://doi.org/10.1001/jamanetworkopen.2019.8363.

6. Yilmaz R, Ledwos N, Sawaya R, et al. Nondominant hand skills spatial and psychomotor analysis during a complex virtual reality neurosurgical task: a case series study. *Oper Neurosurg*. 2022;23(1):22–30. https://doi.org/10.1227/ons.0000000000000232.

7. Stojan J, Haas M, Thammasitboon S, et al. Online learning developments in undergraduate medical education in response to the COVID-19 pandemic: a BEME systematic review: BEME Guide No. 69. *Med Teach*. 2022;44(2):109–129. https://doi.org/10.1080/0142159x.2021.1992373.

8. Schwab B, Hungness E, Barsness KA, McGaghie WC. The role of simulation in surgical education. *J Laparoendosc Adv Surg Tech A*. 2017;27(5):450–454. https://doi.org/10.1089/lap.2016.0644.

9. Stefanidis D, Sevdalis N, Paige J, et al. Simulation in surgery: what's needed next? *Ann Surg*. 2015;261(5):846–853. https://doi.org/10.1097/sla.0000000000000826.

10. Fazlollahi AM. Artificial Intelligence Tutoring Compared to Expert Instruction in Surgical Simulation Training: A Randomized Controlled Trial. McGill University; 2021.

11. Rangarajan K, Davis H, Pucher PH. Systematic review of virtual haptics in surgical simulation: a valid educational tool? *J Surg Educ*. 2020;77(2):337–347. https://doi.org/10.1016/j.jsurg.2019.09.006.

12. Association WM. World medical association declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191–2194. https://doi.org/10.1001/jama.2013.281053.

13. Juszczak E, Altman DG, Hopewell S, Schulz K. Reporting of multi-arm parallel-group randomized trials: extension of the CONSORT 2010 statement. *JAMA*. 2019;321(16):1610–1620. https://doi.org/10.1001/jama.2019.3087.

14. Delorme S, Laroche D, DiRaddo R, Del Maestro RF. NeuroTouch: a physics-based virtual simulator for cranial microneurosurgery training. *Neurosurgery*. 2012;71(1 suppl Operative):32–42. https://doi.org/10.1227/NEU.0b013e318249c744.

15. Sabbagh AJ, Bajunaid KM, Alarifi N, et al. Roadmap for developing complex virtual reality simulation scenarios: subpial neurosurgical tumor resection model. *World Neurosurg*. 2020;139:e220–e2e9. https://doi.org/10.1016/j.wneu.2020.03.187.

16. Ledwos N, Mirchi N, Yilmaz R, et al. Assessment of learning curves on a simulated neurosurgical task using metrics selected by artificial intelligence. *J Neurosurg*. 2022;137:1160–1171. https://doi.org/10.3171/2021.12.JNS211563.

17. Alotaibi FE, AlZhrani GA, Sabbagh AJ, Azarnoush H, Winkler-Schwartz A, Del Maestro RF. Neurosurgical assessment of metrics including judgment and dexterity using the virtual reality simulator NeuroTouch (NAJD Metrics). *Surg Innov*. 2015;22(6):636–642. https://doi.org/10.1177/1553350615579729.

18. Lee DK. Alternatives to P value: confidence interval and effect size. *Korean J Anesthesiol*. 2016;69(6):555–562. https://doi.org/10.4097/kjae.2016.69.6.555.

19. Alkadri S, Ledwos N, Mirchi N, et al. Utilizing a multilayer perceptron artificial neural network to assess a virtual reality surgical procedure. *Comput Biol Med*. 2021;136:104770. https://doi.org/10.1016/j.compbiomed.2021.104770.

20. Reich A, Mirchi N, Yilmaz R, et al. Artificial neural network approach to competency-based training using a virtual reality neurosurgical simulation. *Oper*

Neurosurg. 2022;23(1):31-39. https://doi.org/10.1227/ons.0000000000000173.

21. Ribeiro de Oliveira MM, Nicolato A, Santos M, et al. Face, content, and construct validity of human placenta as a haptic training tool in neurointerventional surgery. *J Neurosurg*. 2016;124(5):1238-1244. https://doi.org/10.3171/2015.1.JNS141583.

22. Winkler-Schwartz A, Yilmaz R, Tran DH, et al. Creating a comprehensive research platform for surgical technique and operative outcome in primary brain tumor neurosurgery. *World Neurosurg*. 2020;144: e62-e71. https://doi.org/10.1016/j.wneu.2020.07.209.

23. Bowyer MW, Fransman RB. Simulation in General Surgery. Stefanidis D, Korndorffer JR Jr, Sweet R, editors. Comprehensive Healthcare Simulation: Surgery and Surgical Subspecialties, Cham: Springer International Publishing; 2019:171-183.

24. Ledwos N, Mirchi N, Bissonnette V, Winkler-Schwartz A, Yilmaz R, Del Maestro RF. Virtual reality anterior cervical discectomy and fusion simulation on the novel sim-ortho platform: validation studies. *Oper Neurosurg*. 2020;20(1):74-82. https://doi.org/10.1093/ons/opaa269.

25. AlZhrani G, Alotaibi F, Azarnoush H, et al. Proficiency performance benchmarks for removal of simulated brain tumors using a virtual reality simulator NeuroTouch. *J Surg Educ*. 2015;72(4):685-696. https://doi.org/10.1016/j.jsurg.2014.12.014.

26. Alotaibi FE, AlZhrani GA, Mullah MAS, et al. Assessing bimanual performance in brain tumor resection with NeuroTouch, a virtual reality simulator. *Oper Neurosurg*. 2015;11(1):89-98. https://doi.org/10.1227/NEU.0000000000000631.

27. Sawaya R, Bugdadi A, Azarnoush H, et al. Virtual reality tumor resection: the force pyramid approach. *Oper Neurosurg*. 2017;14(6):686-696. https://doi.org/10.1093/ons/opx189.

28. Bissonnette V, Mirchi N, Ledwos N, Alsidieri G, Winkler-Schwartz A, Del Maestro RF. Artificial intelligence distinguishes surgical training levels in a virtual reality spinal task. *J Bone Joint Surg Am*. 2019;101(23):e127. https://doi.org/10.2106/jbjs.18.01197.

29. Mirchi N, Bissonnette V, Ledwos N, et al. Artificial neural networks to assess virtual reality anterior cervical discectomy performance. *Oper Neurosurg*. 2020;19(1):65-75. https://doi.org/10.1093/ons/opz359.

30. Elliot AJ. Color and psychological functioning: a review of theoretical and empirical work. *Front Psychol*. 2015;6:368. https://doi.org/10.3389/fpsyg.2015.00368.

31. Elliot AJ, Maier MA. Color psychology: effects of perceiving color on psychological functioning in humans. *Annu Rev Psychol*. 2014;65(1):95-120. https://doi.org/10.1146/annurev-psych-010213-115035.

32. Hill RA, Barton RA. Red enhances human performance in contests. *Nature*. 2005;435(7040):293. https://doi.org/10.1038/435293a.

33. Roh TH, Oh JW, Jang CK, et al. Virtual dissection of the real brain: integration of photographic 3D models into virtual reality and its effect on neurosurgical resident education. *Neurosurg Focus*. 2021;51(2): E16. https://doi.org/10.3171/2021.5.FOCUS21193.

34. Yilmaz, R., Winkler-Schwartz, A., Mirchi, N., Del Maestro, R. Methods and systems for continuous monitoring of task performance. (2022). https://patents.google.com/patent/WO2022077109A1/

35. Howie EE, Dharanikota H, Gunn E, et al. Cognitive load management: an invaluable tool for safe and effective surgical training. *J Surg Educ*. 2023;80(3):311-322. https://doi.org/10.1016/j.jsurg.2022.12.010.

36. Ahmed Z, Lau CHH, Poole M, et al. Canadian conference for the advancement of surgical education (C-CASE) 2021: post-pandemic and beyond virtual conference abstracts. *Can J Surg*. 2021;64(6 suppl 1): S65-S79. https://doi.org/10.1503/cjs.018821.

37. Daley BJ. Novice to expert: an exploration of how professionals learn. *Adult Educ Q*. 1999;49(4):133-147. https://doi.org/10.1177/074171369904900401.

38. Sweller J. Cognitive load during problem solving: effects on learning. *Cogn Sci*. 1988;12(2):257-285. https://doi.org/10.1207/s15516709cog1202_4.

39. Andersen SA, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. Cognitive load in mastoidectomy skills training: virtual reality simulation and traditional dissection compared. *J Surg Educ*. 2016;73(1):45-50. https://doi.org/10.1016/j.jsurg.2015.09.010.

40. Marcovitch S, Jacques S, Boseovski JJ, Zelazo PD. Self-reflection and the cognitive control of behavior: implications for learning. *Mind Brain Educ*. 2008;2(3):136-141. https://doi.org/10.1111/j.1751-228X.2008.00044.x.

41. Andersen SAW, Mikkelsen PT, Konge L, Cayé-Thomasen P, Sørensen MS. The effect of implementing

cognitive load theory-based design principles in virtual reality simulation training of surgical skills: a randomized controlled trial. *Adv Simul*. 2016;1(1):20. https://doi.org/10.1186/s41077-016-0022-1.

42. Chan A, Singh S, Dubrowski A, et al. Part versus whole: a randomized trial of central venous catheterization education. *Adv Health Sci Educ*. 2015;20(4):1061–1071. https://doi.org/10.1007/s10459-015-9586-0.

43. Baykan Z, Naçar M. Learning styles of first-year medical students attending Erciyes University in Kayseri, Turkey. *Adv Physiol Educ*. 2007;31(2):158–160. https://doi.org/10.1152/advan.00043.2006.

SUPPLEMENTARY INFORMATION

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.jsurg.2023.11.001.