The Importance of Automated Real-Time Performance Feedback in Virtual Reality Temporal Bone Surgery Training

Myles Davaris[1], Sudanthi Wijewickrema[1], Yun Zhou[1], Patorn Piromchai[1,2],
James Bailey[1], Gregor Kennedy[1], and Stephen O'Leary[1]

[1] The University of Melbourne, Australia
[2] Khon Kaen University, Thailand

Abstract. Virtual reality (VR) is increasingly being used as a training platform in many fields including surgery. However, practice on VR simulators alone is not sufficient to impart skills. Provision of performance feedback is essential to enable skill acquisition by ensuring that mistakes are identified and corrected, strengths are reinforced, and insights into consequences of actions are provided. As such, for a simulation system to be an effective training platform and to enable self-directed learning, it is imperative that automated performance feedback is provided by the system. Although there has been increased interest in the development of feedback methodologies in VR-based surgical training in recent years, their effectiveness in practice has rarely been investigated. In this paper, we investigate the impact of performance feedback in a VR-based surgical training platform with respect to skill acquisition and retention through a randomized controlled trial. We show that feedback during training is essential for both acquisition and retention of surgical skills.

Keywords: Virtual Reality Surgical Training · Automated Performance Feedback · Temporal Bone Surgery

1 Introduction

Virtual reality (VR) has become the go-to technology when developing education systems in recent years. VR simulators are accepted as ideal for this task, as they offer risk-free, interactive, immersive, repeatable, and easily accessible platforms, using which standardized training programs can be developed. The effectiveness of VR-based systems in teaching skills and knowledge has been tested in different application domains, but the results have been mixed.

For example, Gamito et al. [13] showed that a VR-based serious games application can be used to significantly improve attention and memory functions in patients in cognitive rehabilitation. Mao et al. [27] discussed how VR can be used with a robot to improve the gait of subacute stroke patients. A case study on the use of VR in American football training [21] showed a 30% average improvement in scores after VR training. Sacks et al. [35] found that, in training construction safety, VR is better for stone cladding work and for cast-in-situ concrete work, but not for general site safety. Wijewickrema et al. [49] showed that there was no significant difference in knowledge after training on a VR ear anatomy simulator, when compared to training on the same content in the form of a presentation. In a comparison between virtual training and physical training for teaching a bimanual assembly task, Murcia-Lopez and Steed [30] found that there was no significant difference between groups.

This trend continues in the field of surgery as well, where some studies showed that VR-based training was better than traditional training, while others found no significant differences. In Hamilton et al. [16], Seymour et al. [40], and Grantcharov et al. [14] it was found that the skill level of surgical residents performing laparoscopic cholecystectomy improved with VR training. Ost at al. [32], Rowe et al. [34], and Blum et al. [5] showed that fellows and residents performed a bronchoscopy task faster and more skilfully after VR training than untrained controls. Sedlack et al. [37, 38] observed that residents and fellows performed better in colonoscopy after VR training. Ahlberg et al. [2] observed that medical student performance with VR training for laparscopic appendectomy was no better than that of non-trained controls. In Hogle et al. [20] it was seen that there was no significant difference in performance in surgical residents when performing cholecystectomy between intervention and control groups.

These mixed results show that it is impractical to form sweeping conclusions as to the effectiveness of VR in surgical education. As such, we also need to consider other factors such as the task being trained, the skill level of the student, the level of instruction, and even the design of the evaluation study that may affect evaluation results. In this paper, we explore one of these factors, namely, the effect of performance guidance/feedback in VR-based surgical simulation.

It has become evident that the sole availability of a surgical simulator is not sufficient for a meaningful educational experience and an appropriate curriculum should be available to utilize its full potential [44, 12]. One important aspect when designing an effective surgical curriculum is the provision of performance feedback [44, 45]. Feedback is essential for effective skill acquisition, and must be both timely and contextually relevant [10, 28]. Its purpose is to reinforce strengths, address weaknesses, and foster improvements in the learner by providing insights into the consequences of their actions and by highlighting the differences between intended and actual results [44]. It was shown in Hattie & Timperley [19] that the most effective forms of feedback provide cues or reinforcement to learners; are in the form of video-, audio-, or computer-assisted instructional feedback; and/or relate to goals.

In recent years, research in developing automated feedback systems for VR simulation has grown. In temporal bone surgery, most simulators today provide some form of guidance/feedback to students during and/or after training. Bhutta [4] in a recent review of simulation platforms in temporal bone surgery identified several VR simulators that have in-built guidance systems. For example, the VR temporal bone surgery simulator developed by Stanford University [29, 41] provides interactive feedback on maintaining proper technique in the form of coloured dots [39]. The VOXEL-MAN TempoSurg simulator [24, 33] provides step-by-step procedural guidance for performing an operation by showing the desired end product of each step and a textual explanation in a separate panel.

The virtual temporal bone simulator developed by Ohio State University [48, 47] has an integrated intelligent tutor which provides functions such as structure identification and an expert demo mode (replaying of a pre-recorded expert procedure with customizable viewing parameters). The Visible Ear simulator [42, 43] supports an integrated tutor function which provides step-by-step procedural guidance through the green-lighting of steps on the temporal bone and a separate panel with information about the current step and the end-product view.

The University of Melbourne VR Temporal Bone Surgery Simulator [31] provides both technical and procedural guidance during training. For example, step-by-step guidance on how to perform a cortical mastoidectomy is presented as highlighted areas on a temporal bone [54]. Copson et al. [7] discuss a similar implementation of procedural guidance, based on the same simulator, where visual cues are presented one step at a time along with verbal explanations of each step. Methods of providing (verbal, auditory) feedback on surgical technique for this platform are discussed in Zhou et al. [58, 57] and Ma et al. [25, 26].

In most validation studies, the simulation system as a whole, inclusive of the automated guidance system, has been evaluated, usually with respect to a control group [56, 47] or in a pre-post comparison [11, 7]. A few studies exist that test the effect of feedback. For example, Wijewickrema et al. [53, 50] evaluated how feedback on technique and procedure respectively affected the performance of medical students performing a cortical mastoidectomy. Here, the control groups received no feedback while the intervention groups did. No post-tests were conducted and performance was measured during training. As such, whether surgical skills were properly acquired and retained were not tested. In another study, Wijewickrema et al [52] compared the effect of two different feedback generation methods. However, it has not been clearly shown what improvements can be expected with respect to surgical skill acquisition and retention, by providing automated guidance/feedback.

Here, we aim to bridge this gap through a user study comparing the performance of students trained on a VR simulator with and without automated guidance/feedback. To this end, we use the previously developed guidance/feedback system of the University of Melbourne temporal bone surgery simulator [52], which provides guidance/feedback on different forms of surgical skill.

2 Background

The VR platform used in this research is the University of Melbourne temporal bone surgery simulator (see Fig. 1). The virtual operating space consists of a model of a temporal bone and a surgical drill. The virtual model is generated using a segmented micro-CT scan of a human cadaveric temporal bone. The virtual drill reflects the movements of a haptic device which also provides tactile feedback to the user. The impression of depth is achieved through NVIDIA 3D vision technology. A MIDI controller is used as a convenient input device to change environment variables such as magnification level and burr size. Using the VR simulator, surgeons can perform ear operations to remove disease and improve hearing. This often involves removing parts of the temporal bone or operating on the middle or inner ears and requires safe navigation around anatomical structures such as the facial nerve, sigmoid sinus, and dura.

Fig. 1: The University of Melbourne VR temporal bone surgery simulator.

As surgical skills are multi-faceted, when expert surgeons teach trainees, they provide guidance/feedback on different aspects of the same. To emulate this, the simulation system considers four main aspects of skill that need to be acquired: procedural knowledge, knowledge of landmarks/surgical limits, manipulation of environmental variables, and drill handling/technical skills. The following sections discuss how guidance/feedback is provided in order to teach these skills.

Procedural guidance: Procedural guidance is provided using the step-by-step guidance method of Wijewickrema et al. [54]. Each step of the surgery is highlighted on the temporal bone and the next step is only provided once the current step is completed. Fig. 2 illustrates how the second step of a cortical mastoidectomy is highlighted (in green) after the first step has been drilled.

Advice on landmarks/surgical limits: There are inherent cues (for example, changes in colour and smoothness of the bone) that inform surgeons when they are nearing an anatomical structure. However, these cues may be too subtle for a novice to detect. Therefore, it is necessary to provide more obvious warnings to this effect. Following the work of Wijewickrema et al. [51], the system currently provides verbal warnings when a trainee is drilling within a specified distance of a structure. Further, to enable learning of the anatomical structures, functionality to make the temporal bone transparent, so that the underlying structures can be viewed, is also available.

Feedback on environmental settings: The ideal values for environmental settings such as magnification level and burr size differ according to which area of the temporal bone is currently being drilled. For example, at the start of the

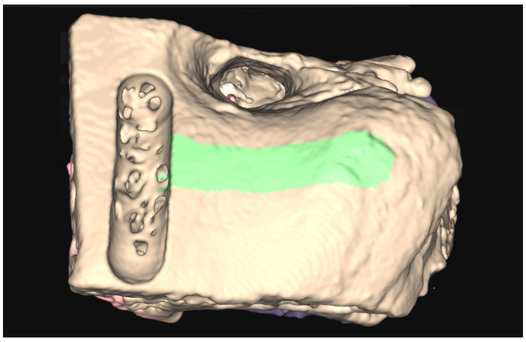

Fig. 2: Presentation of step-by-step procedural guidance.

procedure, in the central area, an overall view of the surgical space is required, and therefore, a lower magnification level is used. In contrast, when drilling deeper insider the mastoid, for example, near the facial nerve, more magnification will be required to get a better view. We use the method of generating regions in a temporal bone using morphological operations such as dilation and erosion as discussed in Wijewickrema et al. [51]. Once the regions are defined, to identify the valid ranges for each region, pre-collected expert data is used. Then, in real-time, if an environment setting is outside the pre-defined range of the relevant region, verbal auditory feedback is provided to inform the student of this. Fig. 3 illustrates how these regions are defined for the ourpose of providing feedback on environmental settings.

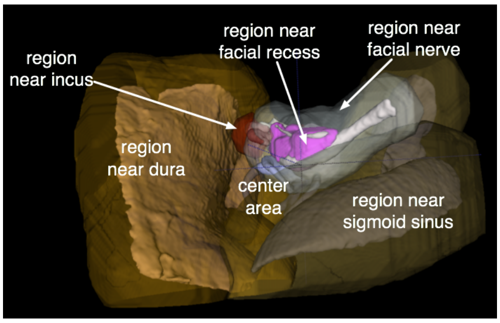

Fig. 3: Definition of regions where surgical technique is considered to be uniform.

Feedback on drill handling/technical/motor skills: Typically, surgical technique adopted at one stage of a surgery is different to that of another. For example, long strokes with a higher force can be used when drilling in an open area. However, more caution is warranted when drilling near a critical anatomical structure such as the facial nerve, and to avoid damage, less force should be used. Wijewickrema et al. [51] identified that the stages of a procedure is related to the regions being drilled. As such, in our system, different behaviour models are developed to identify poor technical skills per region in the following process. The same regions that were defined to provide feedback on environmental settings (as shown in Fig. 3) are used for this purpose.

The method based on random forests (RFs) discussed in Zhou et al. [57] is used here to generate feedback on drill handling. First, strokes are identified from the surgical trajectory using the method introduced in Hall et al. [15]. This enables the extraction of meaningful segments of the trajectory that can be used to define the quality of drill handling.

Once strokes are extracted, the values of metrics that define the quality of a stroke (such as speed and force) can be calculated for each stroke. We used such stroke metrics calculated for pre-collected expert and trainee data as features to train behaviour models that identified expert and trainee skill. Note that we made some modifications to the methods discussed in the original paper [57], when integrating them into our simulator. First, only the motion-based metrics (stroke length, duration, speed, acceleration, straightness, and force) were used as features when training the RF classifier. In contrast, in Zhou et al.'s work, they used motion-based metrics as well as environmental settings (simulator parameters) and proximity metrics (distance to structures). We separate the environmental settings from this and use a simple rule-based method of providing feedback on these, as discussed above. The distance measures are not required, once the regions around anatomical structures are defined. Second, in the original paper, the procedure was divided into stages, which were predicted using a pre-trained classifier, and behaviour models were trained for each stage. The prediction errors introduced by the stage detection was avoided here by using pre-defined regions instead, as discussed above.

If a stroke is classified as a trainee stroke, advice has to be provided on how to improve it so that expert-level behaviour can be learned. For this, we used the voting-based scheme used in Zhou et al. [57]. First, the expert stroke closest to the current trainee stroke was selected (from the pre-collected expert data) using a nearest neighbour strategy. In order to choose the specific feedback feature, the current trainee stroke and the closest expert stroke were classified by each tree in the RF. In a given tree, provided both strokes have been classified correctly, we computed the first feature (and direction: increase or decrease) on which the strokes were split into different branches and this feature received one vote. The feedback was then considered to be the feature and direction that received the most votes (for example, increase stroke length, decrease force etc.). This feedback is presented by the system to the user as verbal auditory instructions.

3 Methodology

We conducted a randomized controlled trial of 40 medical students, with no prior surgical experience, to evaluate the importance of providing automated guidance/feedback in VR-based temporal bone surgery. This study was approved by the University of Melbourne Human Ethics Committee (#1135497.3). On the first day, they were first shown a video tutorial on how to perform a simple temporal bone surgery (cortical mastoidectomy) on our VR simulator. Then, they performed this surgery on the VR simulator with no automated guidance (pre-test). The pre-test was performed in order to gauge their initial skill level, to account for individual variations in skill acquisition and retention. Next, they were randomly allocated to one of two groups: control or feedback. Then, they underwent a training session, performing the same procedure on the simulator. The feedback group received automated guidance/feedback during this procedure, while the control group did not. On the second day, they underwent another training session similar to the previous day. After this, on the same day, the participants performed a post-test: the same procedure without guidance/feedback. They came back a week later (on the ninth day), and performed a cortical mastoidectomy without guidance/feedback as a retention test. We recorded all the procedures conducted by participants using screen capture software. The study design in illustrated in Fig. 4.

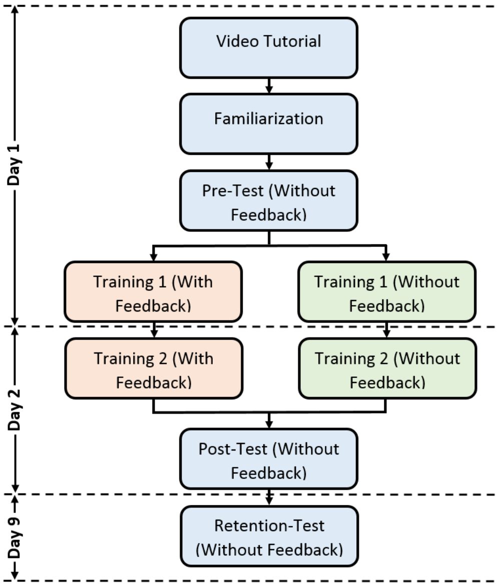

Fig. 4: Design of the evaluation study.

A blinded expert surgeon evaluated videos of the pre-, post- and retention tests based on a validated assessment scale [23]. This assessment scale was designed specifically for cortical mastoidectomy. It has been validated to be a feasible tool with high inter-rater agreement. This scale comprises two parts: checklist and global instruments, and assesses competency of the surgeon in performing the surgery as a whole. This takes into consideration all aspects of surgical skill, for example, knowledge of landmarks and procedure as well as technical skills.

The checklist instrument consists of 22 items (in 7 categories: initial bone cuts, defining anatomic limits, open antrum, digastric dissection, thin posterior EAC cortex (translucent), open facial recess, and posterior atticotomy). Each item is scored on a Likert scale ranging from 1 (unable to perform), through 3 (performs with minimal prompting), to 5 (performs easily with good flow). As such the minimum and maximum scores for this instrument are 22 and 110 respectively. However, as our study participants were medical students with no prior surgical experience, we did not teach them how to perform the 2 latter parts of the surgery (open facial recess and posterior atticotomy). As such only 13 items were relevant for this study, with minimum and maximum scores of 13 and 65 respectively. The global instrument comprises 10 items: understanding of objectives of surgery, interpretation of preoperative tests, use of otologic drill, knowledge of instruments, use of microscope, respect for surgical limits, time and motion, knowledge of specific procedure, Flow of operation, and overall surgical performance. The scoring is based on a 5-point Likert scale similar to that of the checklist instrument. As such, the minimum score is 10 and the maximum possible score is 50.

4 Results

First, we tested if the initial skill levels of the two groups were significantly different, using analysis of variance (ANOVA). We found that although the initial skill level of the feedback group was higher than that of the control group, these differences were not significant for either scores: checklist or global. Therefore, we can infer that the randomization procedure was successful.

To compare the level of surgical skill acquired after training by the two groups (post-test scores), taking into consideration the initial skill level of participants (pre-test scores) to account for individual aptitude, we performed an analysis of covariance (ANCOVA). A similar analysis was conducted to test for skill retention. Significant differences were observed between groups for both skill acquisition and retention, with the feedback group showing larger improvements in performance. The comparison results are shown in Table 1. In these ANCOVA analyses it was also tested what the effect of initial skill level was on skill acquisition and retention. It was seen that initial skill level was a significant factor in skill acquisition ($p = 0.026$ and $p = 0.007$ for checklist and global scores respectively), but not in skill retention.

To test if there were skill improvements within groups before and after training, we used paired t-tests to compare the post- and retention test scores with pre-test scores. Cohen's d was used to calculate the effect sizes. Significant improvements with large effect sizes were observed in the feedback group with respect to both skill acquisition and retention, while the improvements in the control group were not significant with low effect sizes. Table 2 shows the results.

Table 1: Between-group analysis of skill acquisition and retention, taking initial skill level as a covariate. Statistically significant results are shown in bold.

Score	**Group**	**Adjusted Mean**	**F(1,37)**	**p > F**
Skill Acquisition				
Checklist	Feedback	37.08	27.73	< **0.001**
	Control	20.37		
Global	Feedback	33.13	24.84	< **0.001**
	Control	16.57		
Skill Retention				
Checklist	Feedback	28.46	13.73	< **0.001**
	Control	18.84		
Global	Feedback	25.17	13.24	< **0.001**
	Control	14.83		

5 Discussion

From the results of the between group analyses in Table 1, we observe that the group that received automated feedback during training performed significantly better than the control group with respect to both skill acquisition and retention. The within group analysis (Table 2) shows that the learning and retention rates of the control group was not significant with low effect sizes. In contrast, the feedback group had significant levels of skill acquisition and retention with high effect sizes. These results imply that, at the level of the participants (complete novices with no prior experience in surgery), task demonstration is not sufficient

Table 2: Within-group analysis of performance with respect to skill acquisition and retention. Statistically significant results are shown in bold.

Group	**Score**	**P-Value**	**Effect Size (d)**
Skill acquisition			
Feedback	Checklist	< **0.001**	1.74
	Global	< **0.001**	1.56
Control	Checklist	0.239	0.35
	Global	0.306	0.28
Skill retention			
Feedback	Checklist	**0.004**	0.90
	Global	**0.006**	0.85
Control	Checklist	0.386	0.27
	Global	0.552	0.19

to teach surgical skills, and performance feedback on different aspects of surgical skill is essential during training to ensure that skills are acquired and retained. These findings are in accordance with the principle of deliberate practice [10] which states that practice and relevant feedback are both important in gaining expertise. They also support other studies suggesting the benefit of real-time automated feedback in training [6, 8, 18, 55].

The significantly better performance after training with automated feedback further indicates that it was presented in such a way that it could be easily understood, thus retaining the cognitive load (burden placed on the human cognitive processing system [46]) at a manageable level to enable greater learning and performance [18]. That these results contradicts the guidance hypothesis which states that concurrent (real-time) feedback may lead to over-reliance and diminished performance [36], also indicates that the right amount of feedback was provided that supported learning, but discouraged over-reliance. However, as the students gain more experience in surgical procedures, it may be necessary to reduce the level of instruction, to avoid over-reliance [9, 22].

From the between group test results, we see that in addition to the presence or absence of feedback, the individual aptitude (as measured in the pre-test) also plays a significant role in the acquisition of surgical skills, but not in their retention. That individual differences affect complex skill acquisition seems obvious, but to what extent this is true is dependent on other factors as well. For example, Ackerman [1] discussed the ability-performance relationship as a function of task complexity, degree of task practice, and consistency of information-processing demands. This seems to indicate that with more practice, perhaps the significance of individual aptitude may lessen.

In this study, the training and evaluation of performance were both conducted on our VR simulation environment. As such, the results are not indicative of how participants would perform in a real-world scenario. However, there is evidence to suggests that skills learned in VR are transferable to real-world applications [17, 3]. This should be explored in future studies.

6 Conclusion

Here, we evaluated the effect of performance feedback on skill acquisition and retention in a VR-based surgical training platform. We observed that task demonstration and repeated practice to emulate the task on a VR simulator is not sufficient to acquire and retain surgical skills. Real-time feedback is not only helpful, but essential in the acquisition and retention of skills. Although in this paper, we established this for the test case of temporal bone surgery, the findings are in line with educational principles such as that of deliberate practice. This raises an important point which is often overlooked in VR education system design and development: the importance of appropriate feedback during training to ensure that the right skills are acquired and retained.

References

1. Ackerman, P.L.: Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. Journal of experimental psychology: General **117**(3), 288 (1988)
2. Ahlberg, G., Heikkinen, T., Iselius, L., Leijonmarck, C.E., Rutqvist, J., Arvidsson, D.: Does training in a virtual reality simulator improve surgical performance? Surgical endoscopy and other interventional techniques **16**(1), 126–129 (2002)
3. Beyer-Berjot, L., Aggarwal, R.: Toward technology-supported surgical training: the potential of virtual simulators in laparoscopic surgery. Scandinavian Journal of Surgery **102**(4), 221–226 (2013)
4. Bhutta, M.: A review of simulation platforms in surgery of the temporal bone. Clinical Otolaryngology **41**(5), 539–545 (2016)
5. Blum, M.G., Powers, T.W., Sundaresan, S.: Bronchoscopy simulator effectively prepares junior residents to competently perform basic clinical bronchoscopy. The Annals of thoracic surgery **78**(1), 287–291 (2004)
6. Chang, J.Y., Chang, G.L., Chien, C.J.C., Chung, K.C., Hsu, A.T.: Effectiveness of two forms of feedback on training of a joint mobilization skill by using a joint translation simulator. Physical therapy **87**(4), 418–430 (2007)
7. Copson, B., Wijewickrema, S., Zhou, Y., Piromchai, P., Briggs, R., Bailey, J., Kennedy, G., O'Leary, S.: Supporting skill acquisition in cochlear implant surgery through virtual reality simulation. Cochlear Implants International **18**(2), 89–96 (2017)
8. Day, T., Iles, N., Griffiths, P.: Effect of performance feedback on tracheal suctioning knowledge and skills: randomized controlled trial. Journal of Advanced Nursing **65**(7), 1423–1431 (2009)
9. Driscoll, M.P., Driscoll, M.P.: Psychology of learning for instruction (2005)
10. Ericsson, K.A.: Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Academic medicine **79**(10), S70–S81 (2004)
11. Francis, H.W., Malik, M.U., Varela, D.A.D.V., Barffour, M.A., Chien, W.W., Carey, J.P., Niparko, J.K., Bhatti, N.I.: Technical skills improve after practice on virtual-reality temporal bone simulator. The Laryngoscope **122**(6), 1385–1391 (2012)
12. Fried, G.M.: Lessons from the surgical experience with simulators: incorporation into training and utilization in determining competency. Gastrointestinal Endoscopy Clinics **16**(3), 425–434 (2006)
13. Gamito, P., Oliveira, J., Coelho, C., Morais, D., Lopes, P., Pacheco, J., Brito, R., Soares, F., Santos, N., Barata, A.F.: Cognitive training on stroke patients via virtual reality-based serious games. Disability and rehabilitation **39**(4), 385–388 (2017)
14. Grantcharov, T.P., Kristiansen, V., Bendix, J., Bardram, L., Rosenberg, J., Funch-Jensen, P.: Randomized clinical trial of virtual reality simulation for laparoscopic skills training. British Journal of Surgery **91**(2), 146–150 (2004)
15. Hall, R., Rathod, H., Maiorca, M., Ioannou, I., Kazmierczak, E., OLeary, S., Harris, P.: Towards haptic performance analysis using k-metrics. Haptic and Audio Interaction Design pp. 50–59 (2008)
16. Hamilton, E., Scott, D., Fleming, J., Rege, R., Laycock, R., Bergen, P., Tesfay, S., Jones, D.: Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills. Surgical Endoscopy and Other Interventional Techniques **16**(3), 406–411 (2002)

17. Hammoud, M.M., Nuthalapaty, F.S., Goepfert, A.R., Casey, P.M., Emmons, S., Espey, E.L., Kaczmarczyk, J.M., Katz, N.T., Neutens, J.J., Peskin, E.G., et al.: To the point: medical education review of the role of simulators in surgical training. American journal of obstetrics and gynecology **199**(4), 338–343 (2008)
18. Hatala, R., Cook, D.A., Zendejas, B., Hamstra, S.J., Brydges, R.: Feedback for simulation-based procedural skills training: a meta-analysis and critical narrative synthesis. Advances in Health Sciences Education **19**(2), 251–272 (2014)
19. Hattie, J., Timperley, H.: The power of feedback. Review of educational research **77**(1), 81–112 (2007)
20. Hogle, N., Chang, L., Strong, V., Welcome, A., Sinaan, M., Bailey, R., Fowler, D.: Validation of laparoscopic surgical skills training outside the operating room: a long road. Surgical endoscopy **23**(7), 1476–1482 (2009)
21. Huang, Y., Churches, L., Reilly, B.: A case study on virtual reality american football training. In: Proceedings of the 2015 Virtual Reality International Conference. p. 6. ACM (2015)
22. Koedinger, K.R., Corbett, A.T., Perfetti, C.: The knowledge-learning-instruction framework: Bridging the science-practice chasm to enhance robust student learning. Cognitive science **36**(5), 757–798 (2012)
23. Laeeq, K., Bhatti, N.I., Carey, J.P., Della Santina, C.C., Limb, C.J., Niparko, J.K., Minor, L.B., Francis, H.W.: Pilot testing of an assessment tool for competency in mastoidectomy. Laryngoscope **119**(12), 2402–2410 (2009)
24. Leuwer, R., Petersik, A., Pflesser, B., Pommert, A., Tolsdorff, B., Höhne, K.H., Tiede, U.: Voxel-man temposurg a virtual reality temporal bone surgery simulator. Journal of Japan Society for Head and Neck Surgery **17**(3), 203–207 (2008)
25. Ma, X., Wijewickrema, S., Zhou, S., Zhou, Y., Mhammedi, Z., O'Leary, S., Bailey, J.: Adversarial generation of real-time feedback with neural networks for simulation-based training. arXiv preprint arXiv:1703.01460 (2017)
26. Ma, X., Wijewickrema, S., Zhou, Y., Zhou, S., OLeary, S., Bailey, J.: Providing effective real-time feedback in simulation-based surgical training. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 566–574. Springer (2017)
27. Mao, Y., Wei, X., Chen, P., Li, L., Chen, L., Xu, Z., Huang, D.: Effects of virtual reality with robot training on the gait of subacute stroke patients. Annals of Physical and Rehabilitation Medicine **61**, e180 (2018)
28. McGaghie, W.C., Issenberg, S.B., Petrusa, E.R., Scalese, R.J.: A critical review of simulation-based medical education research: 2003–2009. Medical education **44**(1), 50–63 (2010)
29. Morris, D., Sewell, C., Blevins, N., Barbagli, F., Salisbury, K.: A collaborative virtual environment for the simulation of temporal bone surgery. In: International conference on medical image computing and computer-assisted intervention. pp. 319–327. Springer (2004)
30. Murcia-Lopez, M., Steed, A.: A comparison of virtual and physical training transfer of bimanual assembly tasks. IEEE transactions on visualization and computer graphics **24**(4), 1574–1583 (2018)
31. O'leary, S.J., Hutchins, M.A., Stevenson, D.R., Gunn, C., Krumpholz, A., Kennedy, G., Tykocinski, M., Dahm, M., Pyman, B.: Validation of a networked virtual reality simulation of temporal bone surgery. The Laryngoscope **118**(6), 1040–1046 (2008)
32. Ost, D., De Rosiers, A., Britt, E.J., Fein, A.M., Lesser, M.L., Mehta, A.C.: Assessment of a bronchoscopy simulator. American journal of respiratory and critical care medicine **164**(12), 2248–2255 (2001)

33. Reddy-Kolanu, G., Alderson, D.: Evaluating the effectiveness of the voxel-man temposurg virtual reality simulator in facilitating learning mastoid surgery. The Annals of The Royal College of Surgeons of England **93**(3), 205–208 (2011)
34. Rowe, R., Cohen, R.A.: An evaluation of a virtual reality airway simulator. Anesthesia & Analgesia **95**(1), 62–66 (2002)
35. Sacks, R., Perlman, A., Barak, R.: Construction safety training using immersive virtual reality. Construction Management and Economics **31**(9), 1005–1017 (2013)
36. Schmidt, R., Lee, T.: Motor Control and Learning: A Behavioral Emphasis. Human Kinetics (2011)
37. Sedlack, R.E., Kolars, J.C.: Computer simulator training enhances the competency of gastroenterology fellows at colonoscopy: results of a pilot study. The American journal of gastroenterology **99**(1), 33 (2004)
38. Sedlack, R.E., Kolars, J.C., Alexander, J.A.: Computer simulation training enhances patient comfort during endoscopy. Clinical Gastroenterology and Hepatology **2**(4), 348–352 (2004)
39. Sewell, C., Morris, D., Blevins, N.H., Dutta, S., Agrawal, S., Barbagli, F., Salisbury, K.: Providing metrics and performance feedback in a surgical simulator. Computer Aided Surgery **13**(2), 63–81 (2008)
40. Seymour, N.E., Gallagher, A.G., Roman, S.A., Obrien, M.K., Bansal, V.K., Andersen, D.K., Satava, R.M.: Virtual reality training improves operating room performance: results of a randomized, double-blinded study. Annals of surgery **236**(4), 458 (2002)
41. Sonny, C., Peter, L., Dong Hoon, L., et al.: A virtual surgical environment for rehearsal of tympanomastoidectomy. Studies in health technology and informatics **163**, 112 (2011)
42. Sørensen, M.S., Dobrzeniecki, A.B., Larsen, P., Frisch, T., Sporring, J., Darvann, T.A.: The visible ear: a digital image library of the temporal bone. ORL **64**(6), 378–381 (2002)
43. Sorensen, M.S., Mosegaard, J., Trier, P.: The visible ear simulator: a public pc application for gpu-accelerated haptic 3d simulation of ear surgery based on the visible ear data. Otology & Neurotology **30**(4), 484–487 (2009)
44. Stefanidis, D.: Optimal acquisition and assessment of proficiency on simulators in surgery. Surgical Clinics of North America **90**(3), 475–489 (2010)
45. Stefanidis, D., Heniford, B.T.: The formula for a successful laparoscopic skills curriculum. Archives of Surgery **144**(1), 77–82 (2009)
46. Van Merrienboer, J.J., Sweller, J.: Cognitive load theory and complex learning: Recent developments and future directions. Educational psychology review **17**(2), 147–177 (2005)
47. Wiet, G.J., Stredney, D., Kerwin, T., Hittle, B., Fernandez, S.A., Abdel-Rasoul, M., Welling, D.B.: Virtual temporal bone dissection system: Osu virtual temporal bone system: development and testing. The Laryngoscope **122**(S1), S1–S12 (2012)
48. Wiet, G.J., Stredney, D., Sessanna, D., Bryan, J.A., Welling, D.B., Schmalbrock, P.: Virtual temporal bone dissection: an interactive surgical simulator. OtolaryngologyHead and Neck Surgery **127**(1), 79–83 (2002)
49. Wijewickrema, S., Copson, B., Ma, X., Briggs, R., Bailey, J., Kennedy, G., OLeary, S.: Development and validation of a virtual reality tutor to teach clinically oriented surgical anatomy of the ear. In: 2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS). pp. 12–17 (June 2018). https://doi.org/10.1109/CBMS.2018.00010

50. Wijewickrema, S., Zhou, Y., Ioannou, I., Copson, B., Piromchai, P., Yu, C., Briggs, R., Bailey, J., Kennedy, G., O'Leary, S.: Presentation of automated procedural guidance in surgical simulation: results of two randomised controlled trials. The Journal of Laryngology & Otology **132**(3), 257–263 (2018)
51. Wijewickrema, S., Ioannou, I., Zhou, Y., Piromchai, P., Bailey, J., Kennedy, G., OLeary, S.: Region-specific automated feedback in temporal bone surgery simulation. In: Computer-Based Medical Systems (CBMS), 2015 IEEE 28th International Symposium on. pp. 310–315. IEEE (2015)
52. Wijewickrema, S., Ma, X., Piromchai, P., Briggs, R., Bailey, J., Kennedy, G., OLeary, S.: Providing automated real-time technical feedback for virtual reality based surgical training: Is the simpler the better? In: International Conference on Artificial Intelligence in Education. pp. 584–598. Springer (2018)
53. Wijewickrema, S., Piromchai, P., Zhou, Y., Ioannou, I., Bailey, J., Kennedy, G., Oleary, S.: Developing effective automated feedback in temporal bone surgery simulation. Otolaryngology–Head and Neck Surgery **152**(6), 1082–1088 (2015)
54. Wijewickrema, S., Zhou, Y., Bailey, J., Kennedy, G., O'Leary, S.: Provision of automated step-by-step procedural guidance in virtual reality surgery simulation. In: Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology. pp. 69–72. ACM (2016)
55. Xeroulis, G.J., Park, J., Moulton, C.A., Reznick, R.K., 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 **141**(4), 442–449 (2007)
56. Zhao, Y.C., Kennedy, G., Yukawa, K., Pyman, B., OLeary, S.: Improving temporal bone dissection using self-directed virtual reality simulation: results of a randomized blinded control trial. Otolaryngology–Head and Neck Surgery **144**(3), 357–364 (2011)
57. Zhou, Y., Bailey, J., Ioannou, I., Wijewickrema, S., Kennedy, G., OLeary, S.: Constructive real time feedback for a temporal bone simulator. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 315–322. Springer (2013)
58. Zhou, Y., Bailey, J., Ioannou, I., Wijewickrema, S., O'Leary, S., Kennedy, G.: Pattern-based real-time feedback for a temporal bone simulator. In: Proceedings of the 19th ACM Symposium on Virtual Reality Software and Technology. pp. 7–16. ACM (2013)