AMJ Australasian Medical Journal

[AMJ 2019;12(8):239-245]

Evaluating the skill curve of psychomotor skill acquisition in arthroscopy, using a virtual reality simulator in trainee doctors

Arkie Ariyana[1,2], Martin Richardson[1,2], and Dean McKenzie[2]

1. The University of Melbourne, Australia
2. Epworth Hospital, Australia

RESEARCH

Please cite this paper as: Ariyana A, Richardson M, McKenzie D. Evaluating the skill curve of psychomotor skill acquisition in arthroscopy, using a virtual reality simulator in trainee doctors. AMJ 2019;12(8):239–245.
https://doi.org/10.35841/1836-1935.12.8.239-245

Corresponding Author:
Dr. Arkie Ariyana
The University of Melbourne, Australia
E-mail: arkie.ariyana@gmail.com

ABSTRACT

Background
Virtual reality (VR) simulation with haptic feedback has emerged as one of the most promising methods of teaching basic techniques in minimally invasive surgery (MIS). Simulation training in MIS procedures allows trainees to receive feedback without putting patients at risk and can assist in decreasing error rates and iatrogenic injury.

Aims
The aim of this study is to evaluate the learning curve in acquiring psychomotor skills required for shoulder arthroscopy using a VR device with haptic feedback.

Methods
The TolTech ArthroSim (Touch of Life Technologies, Inc., Aurora, Colorado) was selected as the simulator. Six repetitions of the same arthroscopic procedure were performed in a single one-hour session. The metrics included in this study were time taken to complete procedures in seconds. 24 trainee doctors without prior arthroscopy exposure were recruited in this study

Results
There is a relevant reduction in time taken throughout the training sessions. The largest reduction in time taken occurred in the first two repetitions going from 422.61s to 242.19s a reduction of 42.69 per cent. This reduction in time taken levels out after the fourth repetition.

Conclusion
There is a significant but steep learning curve in trainee doctors learning arthroscopy using VR simulator with haptic feedback which correlated with the first 40 minutes of training. Despite a large variation in innate arthroscopic skill, on average participants were able to significantly improve by the end of the study. There is also a reduction in the variation of arthroscopic skills between participants in later repetitions.

Key Words
Arthroscopy, virtual reality, education, orthopaedics

What this study adds:

1. What is known about this subject?
Multiple studies have validated VR simulators as an effective training tool with transferability to real life operations

2. What new information is offered in this study?
There is a significant but steep learning curve in trainee doctors learning arthroscopy using VR which correlated with the first 40 minutes of training.

3. What are the implications for research, policy, or practice?
This study could change the practice of surgical education by employing VR as an effective tool for training junior doctors.

Background

VR simulation with haptic feedback has emerged as one of the most promising methods of teaching basic techniques in minimally invasive surgery (MIS). Simulation training in MIS procedures allows trainees to receive feedback without putting patients at risk and can assist in decreasing error rates and iatrogenic injury. With the continued improvement in realism and increased availability of modern simulators such as VR devices with haptic feedback, we may be able to compensate for the reduced time in theatre experience of surgeons in training. Multiple studies have also shown that most VR devices on the market provided a satisfactory level of realism.[1] VR devices were also able to distinguish between expert and novice users, indicating that arthroscopic skills gained in the operating theatre can be assessed in virtual reality.[2-6] There is also strong evidence for the transferability to proficiency in the operating theatre.[7-9]

There appears to be a learning curve associated with the acquisition of the skills required for arthroscopic surgery. Alvand et al. looked at arthroscopy naïve medical students and revealed a variation in the innate arthroscopic skill between individuals.[10] In this study, 20 students were asked to perform 30 repetitions of two arthroscopic tasks using a benchtop shoulder model in a single session. The results showed that most students were able to gain competence by the end of the study. A similar study by Rahm et al. showed a similar outcome whereby arthroscopic naïve students were able to significantly improve within four 30 minutes virtual arthroscopy knee simulator training but plateaued thereafter within the setting studied.[11]

The aim of this study is to evaluate the learning curve in acquiring psychomotor skills required for shoulder arthroscopy using a VR device with haptic feedback.

Method

Settings and subjects

This study was undertaken in a simulation laboratory of a university teaching hospital, Epworth HealthCare Richmond. Medical students from the University of Melbourne were invited to take part in the study through a cohort-wide announcement and a post on a student group on a social media platform. Criteria for inclusion in the study were that each participant was a current medical student in their second to fourth years (final year). A pragmatic sample size of 24 students was selected for two groups of twelve patients each. As this is a pilot study the sample size was not chosen on the basis of statistical power analysis, however twelve persons per group is regarded as a practical rule of thumb.[12] A recent empirical study has shown, moreover, that at least 10 persons per group is recommended for pilot studies expecting a medium effect size (i.e., difference in group means equal to half a standard deviation) with 80 per cent power and an alpha or probability level of 0.05.[13] The experience and gained in this study will be used to design a larger study in the future.

Participants were excluded if they had previously performed any arthroscopies or had any prior experience on arthroscopic simulators. Participants were randomly assigned (using a permuted block algorithm implemented in the Stata 15 statistical package, Stata Corporation, College Station, Texas, 2017) to either study group A (Locate & Palpate) or study group B (Arthroscopy) as described above. Each participant was provided with information about the study, and written consent obtained.

The simulation models

The TolTech ArthroSim (Touch of Life Technologies, Inc., Aurora, Colorado) was selected as the simulator. The TolTech ArthroSim is one of the leading high-fidelity VR arthroscopic on the market and has been validated to show satisfactory levels of face validity.[1] The ArthroSim is a VR arthroscopic simulator with a passive haptic feedback system. Only the shoulder model was used in this study. There was inbuilt software to measure performance metrics for user feedback (time taken, camera roughness, instrument roughness, camera path length, instrument path length).

Training regime

Before the start of the training program, participants were shown a Microsoft PowerPoint (Microsoft Corporation, Redmond, Washington) 64-Bit edition, 2017 presentation that detailed the task and fully demonstrated the anatomical landmarks on the shoulder model. This included labelled diagrams of each of the anatomical landmarks accompanied with explanatory text and an arthroscopic photograph. Participants could refer to the PowerPoint presentation as necessary during each session. Participants were informed to complete the tasks as efficiently as they could and were aware of the factors that will affect how their performance will be scored. The factors were time taken and movement economy (camera/instrument path length).

Participants were divided into two groups A and B. Both groups performed "Spheres inbox" as their first task to become familiar with handling the 30-degree angled arthroscope for five minutes. In this task, participants were required to locate and centre the spheres in boxes in different positions and angles. Afterward, Group A performed anatomical locate and palpate tasks. Group B performed arthroplasties. Six repetitions of each task were performed in a single one-hour session. Participants could follow the on-screen guidance provided by the VR simulator but received no guidance or feedback from the investigators.

The tasks

The diagnostic shoulder palpation task requires participants to locate and palpate anatomical landmarks for three seconds in a random order using a probe. In the acromioplasty task, participants were instructed to remove a set of spheres representing the abnormal acromion morphology using a bur. Green spheres represent the part of the bone that is to be removed. The orange sphere represents the risk zone and the red spheres represent bone that should never be removed.

Performance assessment

For each task, the performance of participants was measured using the inbuilt software. The metric included in this study was time taken to complete procedures in seconds. The goal was to compare the speed of completion throughout each repetition.

Statistical methods

The statistical analyses were performed at Epworth HealthCare Richmond by A/Prof Dean McKenzie using Stata version 15 (Stata Corporation, College Station, Texas, 2017). Boxplots (Nuzzo, 2016; Tukey, 1977) were included to summarize the raw data.[14,15] To compare the performances of the students, we compared time taken to complete the procedure using orthogonal polynomials[16] a system of weighting coefficients that allowed assessment of the linear, quadratic and cubic relationship between equally spaced ordered categories, in this case repetitions. We employed linear regression models, with a robust variance estimator, allowing for clustering or multiple repetitions for each participant.[17] Data from groups A and B were combined, as not statistically significant overall ('main') effect of group. We are, however, aware that the sample size of our pilot study is very small.

Results

Figure 1: Histogram showing frequency distributions for each repetition in time (seconds)

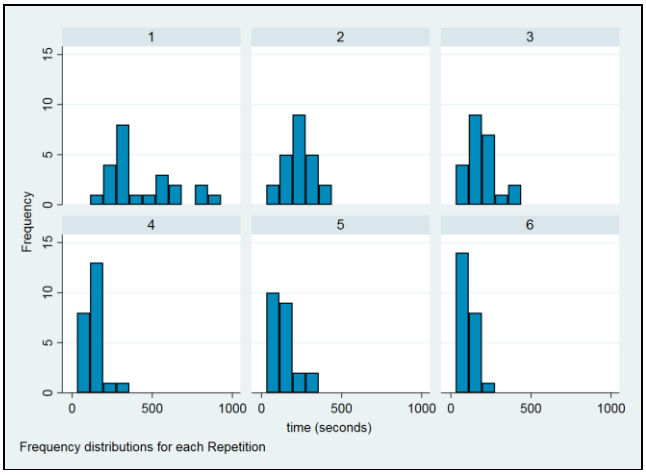


Figure 1 shows a wide variation in the distribution of time taken in the early repetitions, especially in the first repetition. In later repetitions this distribution is narrowed towards a shorter time taken to complete the procedure.

Figure 2: Boxplot for time taken to complete procedure (in milliseconds) for groups A and B (combined) for each repetition

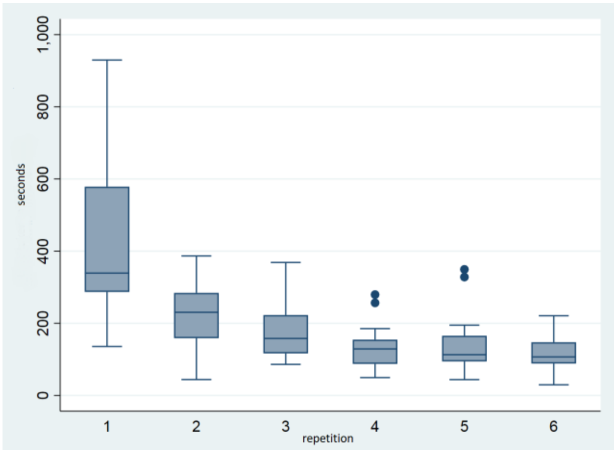
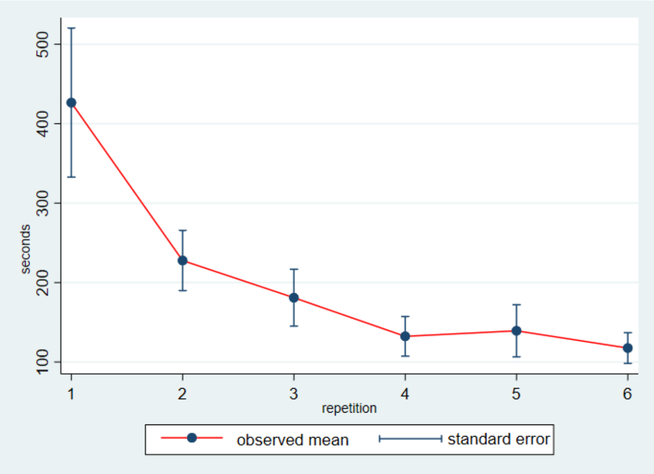


Figure 2 shows a reduction in the median time taken to complete the procedure throughout subsequent repetitions. The width of the interquartile range (the difference between the 25th and 75th percentiles of the distribution) also decreases over the repetitions. There was particularly high variation as measured by IQR in the early repetitions, especially in the first repetition of 291.48 seconds compared to the IQR for the last repetition of 58.79 seconds. All the participants improved, as shown by decreasing medians or 50th percentiles by the end of the study, decreasing from 339.14 seconds in the first repetition to 106.86 seconds in the last repetition. A similar pattern was noted in the raw means and standard deviations shown in Table 1 and the fitted means, standard errors and 95 per cent confidence intervals, obtained from the regression model, adjusted for clustering, shown In Table 1 and Figure 3.

Figure 3: Mean fitted value of time taken to complete procedure (in seconds) for groups A and B for each repetition



There was a relevant reduction in time taken throughout the training sessions. The largest reduction in time taken occurred in the first two repetitions going from a fitted (obtained from the regression model) of 422.61s to 242.19s, a reduction of 42.69 per cent. This reduction in time taken levels out after the fourth repetition. The overall regression model, incorporating linear, quadratic and cubic relationships exhibited an R^2 of 0.497, indication that almost half of the variation in time taken was accounted for by repetition, this result being statistically significant ($F(3,22)=28.18$, $p<0.001$). The AIC (Akaike Information Criterion) and the generally more conservative BIC (Bayesian Information Criterion) indices are measures of model performance, as weighed against model complexity.[15] Both indices suggested that the increase in the number of parameters needed to add a cubic term to the regression model was balanced by an increase in model performance.

Discussion

Analysing the learning curve of the participants using the mean fitted values (using predicted means), we found a statistically significant improvement in the first four sessions for the time taken with a plateau thereafter. We also discovered a decrease in the variability of performance indicated by the reduction in standard deviations and IQRs from session one to session six. Overall participants were able to improve despite the large variation in the early repetitions.

The results show that arthroscopically naïve trainee doctors have a steep but short learning curve in acquiring the psychomotor skills required for shoulder arthroscopy using VR with haptic feedback. Improvement in skill was seen throughout the study with most of the learning observed in the first four repetitions without noticeable improvement thereafter. This represents a total training time of approximately 40 minutes. This learning curve is consistent with some of the current literature which shows that arthroscopy skill curve plateaus after a surprisingly short amount of time.

Several reasons could explain this plateau. Firstly, it could be limited to the machine's ability to recreate more realistic environments with increasing difficulty. As a result, true psychomotor skill progression and rote learning via muscle memory is hard to distinguish. Secondly, it may be true that in this demographic psychomotor skill progression only takes four short attempts to solidify. A larger sample size with a more diverse population to include metrics such as age, sex, level of medication education and other predictive factors should be included in further studies. Finally, it is possible that early and dramatic improvement in learning curve could be the result of familiarity to the anatomical environment. Participants show a significant reduction in time to complete each procedure once they are familiar with the anatomical environment. Although a pre-test PowerPoint presentation was given to minimize this effect it may be relevant to exclude anatomical landmarks all together in future studies. This can be done in future studies by using tasks available in the software that does not involve any anatomy but require the same level psychomotor skills to complete.

A study by Alavan et al.[10] investigated the innate arthroscopic skills and learning curve patterns of arthroscopy naïve medical students using a dry benchtop model of the knee & shoulder. A similar demographic and sample size (N=20) were used. In this study, ten repetitions were done for three weeks for a total of 30 repetitions. The results show that seven out of 20 of the individuals performing the shoulder task were unable to achieve a predefined level of competence despite sustained practice at the tasks. He believes that there exists a considerable variability in innate arthroscopic ability amongst medical students. This is inconsistent with the results found in this study as most of our participants were able to improve significantly in a short amount of time. It may be important for further studies to include more types of tasks and more metrics such as instrument path length (a measure of movement economy) and tissue damage to distinguish between the "good" and "bad" performers.

Rahm et al.[11] studied the performance of medical students (N=20) using a haptic knee arthroscopy simulator. In his study, students were asked to perform eight sessions with six exercises lasting 30 minutes each for four weeks. He showed that a relevant improvement is found in the first four exercise sessions with no significant improvement afterwards. This equated to about two hours in training time to plateau. His study shows that although performance in the first 30 minutes of training couldn't predict future performance, the best performers in the last session were able to be predicted based on the performance of their fourth session. This data agrees with our study, a short and steep learning curve exists in learning arthroscopy. The present study could benefit from a prolonged period with more sessions divided per week as opposed to a one-day session. This way skill retention and progression could be examined with more data points.

We are mindful of the potential limitations of this study. Firstly, although this is a prospective pilot study, the small sample size of 24 was too small to detect subtle differences or to employ more sophisticated regression procedures.[18] The small sample size and the uneven group distribution was due to technical difficulty and time restriction during data collection, the original sample size was intended to be 60. High variation could also be due to small sample size. Other factors that may contribute to this variation could be investigated in larger future studies. Secondly, there has been no study to determine the best metrics to measure arthroscopic psychomotor skills using virtual reality devices. Therefore, it may be beneficial to create guidelines to test for arthroscopic proficiency using VR. Furthermore, there are no established guidelines for testing factors that influence arthroscopic skill acquisition or whether it is possible to predict performance using a questionnaire. Finally, current VR training devices are still limited in their ability to create increasingly difficult variations of the same procedures.

Despite the limitations, the study has revealed some interesting findings of training arthroscopic naïve medical trainee doctors and potential for the use of VR in the recruitment of doctors into the orthopaedic training program. Firstly, we have found that an arthroscopy naïve doctor can significantly improve in performing arthroscopy in VR after four sessions of the same procedure within a 40-minute training session. Secondly, despite the large variation in innate arthroscopy skill most participants were able to significantly improve after the end of the study. This shows that it may be unfair to predict an individual's arthroscopic skill based on the first attempt. Finally, no predictive factors have yet been found that may be able to predict a surgeon's ability to learn arthroscopy using VR.

Conclusion

There is a significant but steep learning curve in trainee doctors learning arthroscopy using VR simulator with haptic feedback which correlated with the first 40 minutes of training. Despite a large variation in innate arthroscopic skill, on average all participants were able to significantly improve by the end of the study. There is also a reduction in the variation of arthroscopic skills between participants in later repetitions.

References

1. Martin KD, Akoh CC, Amendola A, et al. Comparison of three virtual reality arthroscopic simulators as part of an orthopedic residency educational curriculum. Iowa Orthop J. 2016;36:20–5.
2. Bayona S, Fernandez-Arroyo JM, Martin I, et al. Assessment study of insight ARTHRO VR ((R)) arthroscopy virtual training simulator: Face, content, and construct validities. J Robot Surg. 2008;2(3):151–8.
3. Rahm S, Germann M, Hingsammer A, et al. Validation of a virtual reality-based simulator for shoulder arthroscopy. Knee Surg Sports Traumatol Arthrosc. 2016;24(5):1730-7.
4. Garfjeld Roberts P, Guyver P, Baldwin M, et al. Validation of the updated ArthroS simulator: face and construct validity of a passive haptic virtual reality simulator with novel performance metrics. Knee Surg Sports Traumatol Arthrosc. 2017;25(2):616–25.
5. Gomoll AH, Pappas G, Forsythe B, et al. Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. Am J Sports Med. 2008;36.
6. Cannon WD, Nicandri GT, Reinig K, et al. Evaluation of skill level between trainees and community orthopaedic surgeons using a virtual reality arthroscopic knee simulator. J Bone Joint Surg Am. 2014;96(7):e57.
7. Howells NR, Gill HS, Carr AJ, et al. Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. J Bone Joint Surg Br. 2008;90.
8. Rebolledo BJ, Hammann-Scala J, Leali A, et al. Arthroscopy skills development with a surgical simulator: A comparative study in orthopaedic surgery residents. Am J Sports Med. 2015;43(6):1526–9.

9. Henn RF, III, Shah N, Warner JJP, et al. Shoulder arthroscopy simulator training improves shoulder arthroscopy performance in a cadaveric model. Arthroscopy. 2013;29(6):982–5.
10. Alvand A, Auplish S, Gill H, et al. Innate arthroscopic skills in medical students and variation in learning curves. J Bone Joint Surg Am. 2011;93(19):e115(1-9).
11. Rahm S, Wieser K, Wicki I, et al. Performance of medical students on a virtual reality simulator for knee arthroscopy: an analysis of learning curves and predictors of performance. BMC Surg. 2016;16(1):14.
12. Julious SA. Sample size of 12 per group rule of thumb for a pilot study. Pharmaceut Stat. 2005;4:287–291.
13. Whitehead AL, Julious SA, Cooper CL, et al. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. Stat Methods Med Res. 2016;25:1057–1073.
14. Nuzzo RL. The box plots alternative for visualizing quantitative data. PM R. 2016;8(3):268–72.
15. Tukey J. Exploratory data analysis. Reading, Mass: Addison-Wesley Pub. Co.; 1977.
16. Cohen J, Cohen P, West S, et al. Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. 3rd ed. Hoboken, New Jersey: Erlbaum; 2003.
17. Williams RL. A note on robust variance estimation for cluster-correlated data. Biometrics. 2000;56(2):645–6.
18. Takezawa K. Introduction to nonparametric regression. Hoboken, New Jersey: Wiley-Interscience; 2006.

PEER REVIEW

Not commissioned. Externally peer reviewed.

CONFLICTS OF INTEREST

The authors declare that they have no competing interests.

FUNDING

None

ETHICS COMMITTEE APPROVAL

This study was approved by the University of Melbourne Education Human Ethics Advisory Groups (Ethics ID: 1750272.1) and the Epworth Research Development and Governance Unit prior to study commencement.

Table 1: Raw values (in seconds) of time to complete procedure for groups A and B combined

Repetition	Mean time (s)	Median	IQR[1]	SD[2]	Fitted mean[3]	Std. Error[3]	95% Conf. Interval	
1	426.5	339.14	291.49	212.82	422.61	44.37	334.47	518.53
2	227.71	230.36	125.44	85.97	242.19	17.92	190.53	264.89
3	180.98	157.9	106	81.21	161.99	16.93	145.86	216.09
4	132.36	128.89	67	56.59	141.39	11.79	107.89	156.83
5	139.31	113.04	70.91	74.38	139.78	15.05	107.15	171.47
6	117.6	106.86	58.8	43.89	116.51	9.15	98.62	136.58

[1]Interquartile range [2]Standard deviation [3]obtained from regression model [*]The fitted mean, standard error and 95% CI were obtained from the regression model, taking clustering (multiple repetitions) into account. Mean, median, IQR and SD were calculated from the raw data.