

COMPARISON OF THREE VIRTUAL REALITY ARTHROSCOPIC SIMULATORS AS PART OF AN ORTHOPEDIC RESIDENCY EDUCATIONAL CURRICULUM

Kevin D Martin, DO, MAJ, MC¹, Craig C Akoh, MD², Annunziato Amendola, MD², Phinit Phisitkul, MD²

ABSTRACT

Purpose: Orthopedic education continues to move towards evidence-based curriculum in order to comply with new residency accreditation mandates. There are currently three high fidelity arthroscopic virtual reality (VR) simulators available, each with multiple instructional modules and simulated arthroscopic procedures. The aim of the current study is to assess face validity, defined as the degree to which a procedure appears effective in terms of its stated aims, of three available VR simulators.

Methods: Thirty subjects were recruited from a single orthopedic residency training program. Each subject completed one training session on each of the three leading VR arthroscopic simulators (ARTHRO mentor-Symbionix, ArthroS-Virtamed, and ArthroSim-Toltech). Each arthroscopic session involved simulator-specific modules. After training sessions, subjects completed a previously validated simulator questionnaire for face validity.

Results: The median external appearances for the ARTHRO Mentor (9.3, range 6.7-10.0; p=0.0036) and ArthroS (9.3, range 7.3-10.0; p=0.0003) were statistically higher than for ArthroSim (6.7, range 3.3-9.7). There was no statistical difference in intraarticular appearance, instrument appearance, or user friendliness between the three groups. Most simulators reached an appropriate

level of proportion of sufficient scores for each category ($\geq 70\%$), except for ARTHRO Mentor (intraarticular appearance-50%; instrument appearance-61.1%) and ArthroSim (external appearance-50%; user friendliness-68.8%).

Conclusion: These results demonstrate that ArthroS has the highest overall face validity of the three current arthroscopic VR simulators. However, only external appearance for ArthroS reached statistical significance when compared to the other simulators. Additionally, each simulator had satisfactory intraarticular quality. This study helps further the understanding of VR simulation and necessary features for accurate arthroscopic representation. This data also provides objective data for educators when selecting equipment that will best facilitate residency training.

INTRODUCTION

Orthopedic education continues to move towards an evidence-based curriculum in order to comply with new Accreditation Council for Graduate Medical Education (ACGME) and Resident Review Committee (RRC) requirements.^{1,2} Orthopedic training programs throughout the country are working to incorporate new simulation mandates into their residency programs. The RRC defines simulation as: "the imitation of the operation of a real world process or system." Orthopedic educators have been using simulation models such as cadavers, wood models, Styrofoam anatomic models, and box trainers. Historically, these methods have been successfully used in residency skills training. Unfortunately, the cost of maintaining a functioning cadaveric lab and arthroscopic equipment can be inhibitory. Direct faculty instruction and resident education is often limited due to clinical and surgical obligations while remaining compliant with the mandated 80 hour work week.³⁻⁵ These factors have created the need for the development of arthroscopic knee and shoulder virtual reality (VR) simulators to allow trainees to practice basic skills and receive objective feedback on performance.

VR simulation technology has drastically improved over the past decade from unrealistic low fidelity models⁶ to current high fidelity models with nearly unlimited procedures and tasks to improve basic arthroscopic

¹Army Orthopaedic Surgery, Foot & Ankle
Evans Army Community Hospital
Fort Carson, Colorado, 80913

²University of Iowa Hospitals and Clinics
Department of Orthopedics and Rehabilitation
Iowa City, IA 52242

Corresponding Author:
Craig C Akoh, MD
Department of Orthopedics and Rehabilitation
University of Iowa
200 Hawkins Drive, 01008 JPP
Iowa City, IA 52242
Telephone: 319-356-2595
Craig-Akoh@uiowa.edu

Conflict of Interest Statement: Each author certifies that he or she has no commercial associations (eg, consultancies, stock ownership, equity interest, patent/licensing arrangements, etc) that might pose a conflict of interest in connection with the submitted article.

Ethical Review Statement: This article is IRB exempt and was HIPPA compliant

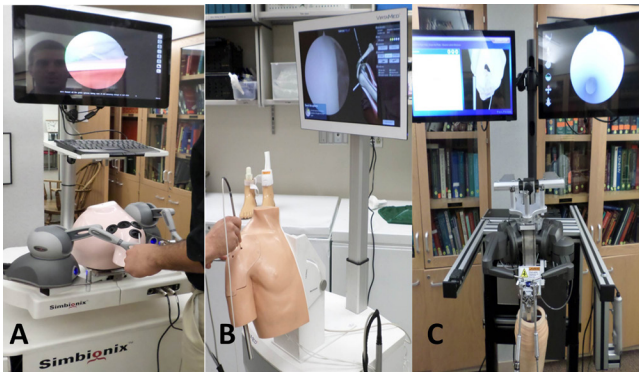


Figure 1. Arthroscopic simulators including A) ARTHRO mentor (Symbionix), B) ArthroS (Virtamed) and C) ArthroSim (Toltech).

skills.^{8,10,11} High fidelity models provide trainees with realistic anatomic features with tactile feedback while allowing manipulation of relevant structures using appropriately-sized arthroscopic instruments.⁷ VR simulators have achieved varying levels of high fidelity while incorporating task-specific modules that reinforce skills, such as simple triangulation, palpation, suture passage, and grasping techniques. Various curriculum and assessments give educators the ability to objectively track trainee performance, thus meeting ACGME requirements.

Based on the changing environment of orthopedic education and the development of several VR arthroscopic simulators, the aim of the current study is to assess face validity, defined as the degree to which a procedure appears effective in terms of its stated aims, of three available VR simulators. The current investigation is a prospective comparative study to evaluate the face validity of the three leading arthroscopic VR simulators as evaluated by orthopedic trainees and staff surgeons at a single training institution. The four variables that comprised face validity in this study were outer appearance, intraarticular appearance, instrument appearance, and user friendliness. We hypothesized that the three leading VR arthroscopic simulators would have no significant difference in face validity.

METHODS

Participants

We recruited thirty volunteer participants from a single orthopedic residency training program during June 2014. Participants included nineteen orthopedic residency trainees, two medical students, and nine arthroscopy-trained staff members. Academic standing was not affected by study involvement and individual performance was not evaluated for academic purposes. Per the study design, the subjects' arthroscopic experience varied by postgraduate year, prior training experiences, and years of practice. Hence, study participants provided

a wide variety of arthroscopic experience. Except for possible brief exposures at academic meetings, all subjects were naïve to each of the simulators. The simulation evaluations took place over a three-week period during June of 2014. Each simulator was randomly assigned to participating subjects during scheduled academic time.

Simulators

We prospectively collected data while evaluating three arthroscopic VR simulators. Three companies were sent an invitation to participate in the study which included utilization of their simulator over a two-day period with a representative present to provide instruction and technical support. The three companies that were invited all agreed to participate and included ARTHRO Mentor (Symbionix, Cleveland, OH), ArthroS (Virtamed, Zurich, Switzerland), and ArthroSim (Touch of Life Technologies, Aurora, CO) (Figure 1). Each simulator was then randomly assigned a test date and time and subjects randomly assigned to each simulator as previously described. For consistency, a training module that included examination and palpation of structures within the glenohumeral joint was selected to assess face validity and user-friendliness of each arthroscopic VR simulator.

Simulation Evaluation

Prior to the tutorial session, each participant reviewed the study objectives and scoring system. Each participant was given a two minute tutorial session, during which subjects were able to manipulate each simulator, similar to a previously established protocol.⁸ Participants were then given five minutes of undirected arthroscopy time to become further familiarized with each of the three simulators. Following the tutorial, participants completed 1) anatomic identification of specified structures within the glenohumeral joint and 2) palpation of specific anatomic landmarks using a specific glenohumeral module for each arthroscopic simulator. After completing these modules, each subject immediately and anonymously completed a previously established questionnaire for face validity variables, including external appearance of each simulator, intraarticular appearance, instruments, and user-friendliness (Table I). A sufficient representation of validity was defined as a median score of 7 per category.¹² All testing and questionnaires were proctored by a single individual to maintain consistency between groups.

Statistical Analysis

Data processing and statistical analyses were performed based on previous studies.¹¹ Variables were assessed for normality using the Kolmogorov-Smirnov test, histograms, and normal probability plots. Results revealed that questionnaire variables were not normally distributed. Therefore, nonparametric analyses were

Table I. Participant Post Assessment Questionnaire

<u>Outer Appearance</u>										
What is your opinion of the outer appearance of this simulator? (circle one)	1 (unreasonable)	2	3	4	5	6	7	8	9	10 (reasonable)
Is it clear in which joint you will be operating?	1 (unclear)	2	3	4	5	6	7	8	9	10 (very clear)
Is it clear in which joint you will be operating?	1 (unclear)	2	3	4	5	6	7	8	9	10 (very clear)
<u>Intraarticular Appearance</u>										
How realistic is the intra-articular anatomy?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
How realistic is the texture of the structures?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
How realistic is the color of the structures?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
How realistic is the size of the structures?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
How realistic is the size of the intra-articular joint space?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
How realistic is the arthroscopic image?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
<u>Instruments</u>										
How realistic do the instruments look?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
How realistic is the motion of your instruments?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
How realistic does the tissue feel when you are probing?	1 (unrealistic)	2	3	4	5	6	7	8	9	10 (realistic)
<u>User-Friendliness</u>										
How clear are the instructions to start an exercise on the simulator?	1 (unclear)	2	3	4	5	6	7	8	9	10 (very clear)
How clear is the presentation of your performance by the simulator?	1 (unclear)	2	3	4	5	6	7	8	9	10 (very clear)
Is it clear how you can improve your performance?	1 (unclear)	2	3	4	5	6	7	8	9	10 (very clear)
How motivating is the way the results are presented to improve your performance?	1 (not motivating)	2	3	4	5	6	7	8	9	10 (very motivating)

completed to assess face validity. Specifically, scores for external appearance, intraarticular appearance, instrument appearance, and user-friendliness were compared among the three simulators using the Kruskal-Wallis and Wilcoxon Rank Sum tests. These scores were presented as medians with an associated range. Differences in the proportion of simulators that met sufficient were compared using Fisher’s exact test. Proportion of training level within each group were compared using Fisher’s exact test. Statistical significance was reached with a p-value <0.05 while employing Bonferroni adjustment. Post-hoc analysis between groups reached statistical significance with a p-value of < 0.017 (0.05/3). All analyses

were completed using SAS software version 9.3 (SAS Institute, Cary, NC).

RESULTS

In total, 28/30 (93.3%) subjects enrolled in this study completed the designated modules. Each of the individual simulator evaluation groups was similar in size and demographics with no significant differences when comparing proportion of training level between groups (p = 0.097) (Table II). Eleven out of 28 subjects (39.3%) completed questionnaires for all three simulators. Seventeen out of 28 subjects (60.7%) completed one or two evaluations due to scheduling conflicts. All medical students (100%)

Table II. Distribution of Participants Completing Simulation Tasks

	Faculty (n=11)	Resident (n=15)	Medical Student (n=2)	Total
Simulator				
Arthro Mentor	9	7	2	18
Arthro S	9	7	2	18
Arthro Sim	6	8	2	16
Simulators completed				
One	4	11	0	15
Two	1	1	0	2
Three	6	3	2	11
Total	11	15	2	

and 6 out of 11 arthroscopy-trained staff subjects (54.5%) completed all three simulator evaluations.

The median external appearance score for the ARTHRO Mentor (9.3, range 6.7-10.0; p=0.0036) and ArthroS (9.3, range 7.3-10.0; p=0.0003) were statistically higher than for ArthroSim (6.7, range 3.3-9.7) (Table III). However, a difference in external appearance failed

to reach statistical significance when comparing ARTHRO Mentor and ArthroS (p = 0.2587). The median intraarticular appearance score for each of the three simulators did not reach statistical significance. Similarly, there were no significant between-test differences in instrument appearance or user friendliness. All face validity median parameters met a sufficient realistic representation score of at least 7, except for intraarticular appearance for ARTHRO mentor and outer appearance for ArthroSim.

Overall, most simulators reached sufficient frequency scores for each face validity category. Both ARTHRO Mentor (94.4%, p= 0.056) and ArthroS (100%, p= 0.007) had statistically higher percentages for proportion of sufficient scores for external appearance when compared to ArthroSim (50%) (Table IV). User friendliness for ArthroSim was rated as sufficient by 11 out of 16 participants (68.8%). ARTHRO Mentor did not reach an appropriate level of proportion of sufficient scores for intraarticular appearance (50%) and instrument appearance (61.1%). Overall, there were no statistical differences between simulators when comparing the frequency of sufficient scores achieved for intraarticular appearance, instrument appearance, and user friendliness.

Table III. Median Scores for Face Validity by Simulator*

Variables	ARTHRO mentor (S) (n=18)	ArthroS (V) (n=18)	ArthroSim (T) (n=16)	p-values
Outer Appearance	9.3 (6.7-10.0)	9.3 (7.3-10.0)	6.7 (3.3-9.7)	S vs V:0.2587 S vs T:0.0012† V vs T:<0.001†
Intraarticular Appearance	6.9 (4.3-10.0)	8.0 (6.2-10.0)	7.8 (3.7-9.0)	S vs V:0.0476 S vs T:0.1515 V vs T:0.8762
Instrument Appearance	7.3 (4.3-9.3)	8.3 (6.0-10.0)	8.2 (4.0-9.3)	S vs V:0.1924 S vs T:0.2320 V vs T:0.9861
User Friendliness	8.0 (5.8-9.8)	8.6 (5.8-10.0)	7.3 (2.5-9.8)	S vs V:0.1124 S vs T:0.2383 V vs T:0.0330

*Values expressed as mean (range). †Statistically significant when adjusted for multiple comparisons.

Table IV. Proportion of Sufficient Scores (≥7) by Simulator*

Variables	ARTHRO Mentor (S) (n=18)	ArthroS (V) (n=18)	ArthroSim (T) (n=16)	p-values
Outer Appearance	17 (94.4)	18 (100.0)	8 (50.0)	S vs V: 1.0000 S vs T: 0.0056 V vs T: 0.0007
Intraarticular Appearance	9 (50.0)	15 (83.3)	13 (81.3)	S vs V: 0.0750 S vs T: 0.0796 V vs T: 1.0000
Instrument Appearance	11 (61.1)	16 (88.9)	14 (87.5)	S vs V: 0.1212 S vs T: 0.1251 V vs T: 1.0000
User Friendliness	15 (83.3)	16 (88.9)	11 (68.8)	S vs V: 1.0000 S vs T: 0.4920 V vs T: 0.2143

*Values presented as no.(%).

DISCUSSION

The results of the current study indicate that there is a significant difference in external appearance for the three current arthroscopic simulators. However, ARTHRO Mentor failed to reach a face validity score of 7.0 for intraarticular appearance and the ArthroSim group failed to achieve a face validity score of 7.0 for external appearance. There continued to be a significant difference for external appearance even after adjusting for multiple comparisons. On the other hand, intraarticular appearance did not reach statistical significance after adjusting for multiple comparisons. ARTHRO Mentor tended to have worse face validity for intraarticular appearance when compared to the ArthroS group. To the best of our knowledge, this is the first three way head-to-head arthroscopic VR simulator study using previously validated metrics.

As simulation continues to be refined and incorporated into orthopedic education, residency curriculum across the country will be increasingly based on evidence and proven outcome measures of simulator education. The current study is intended to assist orthopedic educators in making evidence-based decisions in regards to VR simulation. The current study findings differ from a previous comparative VR simulator study that found the InsightArthroVR1Arthroscopy Simulator to score significantly better when compared to ArthroSim in user friendliness. The previous study also showed no statistical difference in external appearance and intraarticular appearance.¹² The addition of a third simulator in our study established a statistical difference in external appearance between the lower-performing ArthroSim group and the higher-performing ARTHRO Mentor and ArthroS groups. The most notable strength of the study is the number of simulators evaluated, as no previous study has completed a three-way head-to-head VR simulator trial using previously established metrics. A second strength is that nearly a quarter of subjects evaluating each simulator were made up of expert level staff physicians with extensive arthroscopic experience. Our study expounded upon the prior work by adding a third simulator that allowed subjects to see a more diverse spectrum of features and quality along with product upgrades from the previous study.

The ability to assess face validity will allow for residency programs to select the arthroscopic simulator that will fit their specific needs. As our institution moves towards meeting the ACGME requirements for resident experience, purchasing the simulator with the highest overall face validity will give residents the optimal arthroscopic training experience. However, other factors

such as price, service potential, and future curriculum development will affect the feasibility and ultimately our selection of an arthroscopic simulator.

The most notable limitation of the present study is the small sample size. Sample size has been a constant problem with simulation studies due to program size and trainee availability, leaving nearly all simulation studies underpowered.^{6,8,9,11} However, our sample size of 30 is similar to previous head-to-head studies while comparing all three currently available arthroscopic VR simulators.¹² In order to limit selection bias, all available trainees participated. Unfortunately, training level affected participation with all three simulators. Only 20% of the resident subjects completed all three simulator modules. Resident availability for this study was likely limited due to clinical and educational responsibilities. A second limitation was that the order of exposure to the three simulators was not randomized. Individual randomization was not logistically possible given the limited availability of each simulator and participant. We were also not able to perform comparative timed-based task completion due to the vast differences in simulator designs. We also did not compare whether the skills obtained from each of the three arthroscopic simulators translated to realistic cadaveric models. Thirdly, the only conclusion that could be drawn from the results is that external appearance statistically differed between the three simulators. The transfer of arthroscopic simulator skills likely does not solely depend on external appearance. This also brings the question of which face validity variable is most important when selecting arthroscopic simulators for training. Although this study found there was a statistical difference in external appearance between the three simulators, other variables such as intraarticular appearance and user friendliness are likely important for face validity. Future studies including more rigorous head-to-head validation that evaluates the transfer of surgical skills to successful surgeries and improved patient outcomes are recommended.

In conclusion, these results demonstrate that ArthroS has the highest overall face validity of the three current arthroscopic VR simulators. However, only external appearance for ArthroS reached statistical significance when compared to the other simulators. Additionally, each simulator had satisfactory intraarticular quality. This study helps to further the understanding of VR simulation and necessary features for accurate arthroscopic representation. This data will help educators determine which simulator will best facilitate training for their individual program needs.

REFERENCES

1. Accreditation Council for Graduate Medical Education: ACGME program requirements for residency education in general surgery. Available at: <http://www.acgme.org/acwebsite/tabid/150/Program-and-Institutional-Accreditation/Surgical-Specialties/Surgery.aspx>. Accessed 4 Oct 2013.
2. Accreditation Council for Graduate Medical Education for Orthopaedic Surgery. <http://www.acgme.org/acgmeweb/tabid/140/Program-and-Institutional-Accreditation/Surgical-Specialties/Orthopaedic-Surgery.aspx>. Accessed 4 Oct 2013.
3. **Immerman I, Kubiak EN, Zuckerman JD.** Resident work-hour rules: a survey of residents' and program directors' opinions and attitudes. *Am J Orthop (Belle Mead NJ)* 2007; 36(12): E172-9.
4. **Damadi A, Davis AT, Saxe A, Apelgren K.** ACGME duty-hour restrictions decrease resident operative volume: a 5-year comparison at an ACGME-accredited university general surgery residency. *J Surg Educ* 2007; 64(5): 256-9.
5. **Schwartz SI, Galante J, Kaji A, et al.** Effect of the 16-hour work limit on general surgery intern operative case volume: a multi-institutional study. *JAMA Surg* 2013; 148(9): 829-833.
6. **Pedowitz RA, Esch J, Snyder S.** Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy* 2002;18:E29.
7. **Satava RM.** Virtual reality surgical simulator. The first steps. *Surg Endosc.* 1993;7:203-5.
8. **Martin KD, Belmont Jr. PJ, Schoenfeld AJ, Todd M, Cameron KL, Owens BD.** Arthroscopic basic task performance in shoulder simulator model correlates with similar task performance in cadavers. *J Bone Joint Surg Am.* 2011; 93(21): e1271-5.
9. **Cannon WD, Garrett WE Jr, Hunter RE, et al.** Improving Residency Training in Arthroscopic Knee Surgery with Use of a Virtual-Reality Simulator: A Randomized Blinded Study. *J Bone Joint Surg Am* 2014; 96(21):1798-1806.
10. **Cannon WD, Nicandri GT, Reinig K, Mevis H, Wittstein J.** Evaluation of skill level between trainees and community orthopaedic surgeons using a virtual reality arthroscopic knee simulator. *J Bone Joint Surg Am* 2014; Apr 2; 96(7): e57.
11. **Stunt JJ, Kerkhoffs GM, van Dijk CN, Tuijthof GJ.** Validation of the ArthroS virtual reality simulator for arthroscopic skills. *Knee Surg Sports Traumatol Arthrosc.* 2014 Jun 11.
12. **Tuijthof GJ, Visser P, Sierevelt IN, Van Dijk CN, Kerkhoffs GM.** Does perception of usefulness of arthroscopic simulators differ with levels of experience? *Clin Orthop Relat Res* 2011; 469(6):1701-8.
13. **Martin KD, Cameron K, Belmont Jr. PJ, Schoenfeld AJ, Owens BD.** Shoulder arthroscopy simulator performance correlates with resident and shoulder arthroscopy experience. *J Bone Joint Surg Am* 2012; 94: e160(1-5).