Topics in Training

A Validated Orthopaedic Surgical Simulation Model for Training and Evaluation of Basic Arthroscopic Skills

Ryan P. Coughlin, MD, Thierry Pauyo, MD, J. Carl Sutton III, MD, Larry P. Coughlin, MD, and Stephane G. Bergeron, MD, MPH

Investigation performed at the Division of Orthopaedic Surgery, McGill University, Montreal, Quebec, Canada

Background: To our knowledge, there is currently no validated educational model to evaluate and teach basic arthroscopic skills that is widely accessible to orthopaedic residency training programs. The primary objective was to design and to validate a surgical simulation model by demonstrating that subjects with increasing level of training perform better on basic arthroscopic simulation tasks. The secondary objective was to evaluate inter-rater and intra-rater reliability of the model.

Methods: Prospectively recruited participants were divided by level of training into four groups. Subjects performed six basic arthroscopic tasks using a box model: (1) probing, (2) grasping, (3) tissue resection, (4) shaving, (5) tissue liberation and suture-passing, and (6) knot-tying. A score was calculated according to time required to complete each task and deductions for technical errors. A priori total global score, of a possible 100 points, was calculated by averaging scores from all six tasks using equal weights.

Results: A total of forty-nine participants were recruited for this study. Participants were grouped by level of training: Group 1 (novice: fifteen medical students and interns), Group 2 (junior residents: twelve postgraduate year-2 or postgraduate year-3 residents), Group 3 (senior residents: sixteen postgraduate year-4 or postgraduate year-5 residents), and Group 4 (six arthroscopic surgeons). The mean total global score (and standard deviation) differed significantly between groups ($p < 0.001$): 29.0 ± 13.6 points for Group 1, 40.3 ± 12.1 points for Group 2, 57.6 ± 7.4 points for Group 3, and 72.4 ± 3.0 points for Group 4. Pairwise comparison with Tukey correction confirmed construct validity by showing significant improvement in overall performance by increasing level of training between all groups ($p < 0.05$). The model proved to be highly reliable with an intraclass correlation coefficient of 0.99 for both inter-rater and intra-rater reliability.

Conclusions: A simulation model was successfully designed to teach and evaluate basic arthroscopic skills showing good construct validity. This arthroscopic simulation model is inexpensive, valid, and reliable and has the potential to be implemented in other training programs.

Peer Review: This article was reviewed by the Editor-in-Chief and one Deputy Editor, and it underwent blinded review by two or more outside experts. It was also reviewed by an expert in methodology and statistics. The Deputy Editor reviewed each revision of the article, and it underwent a final review by the Editor-in-Chief prior to publication. Final corrections and clarifications occurred during one or more exchanges between the author(s) and copyeditors.

Disclosure: None of the authors received payments or services, either directly or indirectly (i.e., via his or her institution), from a third party in support of any aspect of this work. None of the authors, or their institution(s), have had any financial relationship, in the thirty-six months prior to submission of this work, with any entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. Also, no author has had any other relationships, or has engaged in any other activities, that could be perceived to influence or have the potential to influence what is written in this work. The complete **Disclosures of Potential Conflicts of Interest** submitted by authors are always provided with the online version of the article.

J Bone Joint Surg Am. 2015;97:1465-71 • http://dx.doi.org/10.2106/JBJS.N.01140

There has been a recent emphasis to incorporate medical simulation models into structured educational curricula in response to ongoing concerns over work-hour restrictions, patient safety, and the impact of fellowship training on residency education[1,2]. A survey of surgical residents found that 26% of trainees worried about not feeling confident to operate independently before the end of their training[3]. Although arthroscopic procedures are among the most commonly performed surgical procedures in orthopaedics, they are technically challenging for most learners and difficult for educators to teach[4]. Senior residents have reported feeling that they are less prepared in arthroscopic surgical procedures compared with open surgical procedures and thinking that there is insufficient time dedicated to arthroscopic training[5].

The American Board of Orthopaedic Surgery (ABOS) and the Orthopaedic Surgery Residency Review Committee (RRC) of the Accreditation Council for Graduate Medical Education recently approved mandates to implement surgical simulation training in all orthopaedic residency programs[6]. The changes in program requirements were primarily targeted to postgraduate year 1 (PGY-1) by introducing basic motor skills training commonly used in the initial management of orthopaedic patients in the emergency department and operating room. The Surgical Skills Task Force developed a structured educational curriculum consisting of seventeen simulation modules that is now offered on the ABOS web site[6]. These serve as a guide to individual residency programs to facilitate the development of their own skills program at their respective institutions. However, the content of these modules and the performance metrics have yet to be evaluated for validity and reliability. The ABOS acknowledges the need for dedicated studies to refine and to improve future models[6].

Simple box trainers have been proven to be highly effective in teaching basic technical skills in areas of urology, gynecology, and general surgery[7-10]. However, no such model has been validated for arthroscopic skills training in orthopaedic surgery. The purpose of this study was to construct and to validate a box model consisting of training modules with reliable performance metrics that are directed at evaluating specific psychomotor elements that are fundamental to arthroscopy. Our primary objective was to assess construct validity of the model using a global score assigned for the overall model. The secondary objective was to evaluate inter-rater and intra-rater reliability using the same global score. To assess validity, we hypothesized that subjects with increasing level of training in arthroscopy would perform better on basic arthroscopic simulation tasks. Our second hypothesis was that the model would show moderate to high inter-rater and intra-rater reliability.

Materials and Methods

Full approval for this study was granted by the research ethics committee at our institution. A nominal group technique was used to generate ideas for designing a pilot arthroscopic simulation skills model and establishing evaluation criteria[11]. To establish face validity, the content development group consisted of a panel of four senior arthroscopic surgeons and senior orthopaedic residents. A preliminary list of basic arthroscopic skills and tasks was derived from expert opinion and review of the literature. After structured discussions, a refined list of arthroscopic tasks based on consensus opinion was generated. Each task was designed to incorporate different basic skills such as handling tissue with the dominant and the nondominant hand, camera spatial orientation, optimizing depth perception, and using multiple instruments while working in different planes. A modified version of the McGill Inanimate System for Training and Evaluation of Laparoscopic Skills (MISTELS) scoring metrics system was adapted to our model[12]. The maximum allotted time for each task was determined by ensuring that the members of the content development group could finish the exercises with sufficient time remaining.

Eligible subjects were voluntarily recruited in February and March 2014. All subjects signed an informed consent form and no stipend was given. The participants included fourth-year medical students applying to orthopaedic residency programs, orthopaedic residents, and subspecialty-trained arthroscopic surgeons, all from a single institution. Orthopaedic fellows and members of the content development group were excluded. To assess construct validity, participants were grouped by level of training: Group 1 (novice: medical students and PGY-1 residents), Group 2 (junior residents: PGY-2 and PGY-3 residents), Group 3 (senior residents: PGY-4 and PGY-5 residents), and Group 4 (arthroscopic surgeons). Each subject performed six basic arthroscopic tasks using a box model: (1) probing, (2) grasping, (3) tissue resection, (4) shaving, (5) tissue liberation and suture-passing, and (6) tissue approximation and arthroscopic knot-tying. The training box was opaque and measured 23 × 18 × 15 cm. The sides of the box were composed of a synthetic membrane in which standardized premade portals were placed. Portal placement varied according to the exercise being performed. The optical system consisted of a 30° arthroscope, camera, light source, and video monitor (Arthrex, Naples, Florida).

Before beginning each task, the participants watched a short, two-minute video showing proper performance of each exercise. Participants were videotaped using a camera focused exclusively on the video monitor. Audio was muted to prevent identification of any participant. Participants were assigned a random identification number selected from sealed envelopes for subsequent scoring. The grading was done by two independent and blinded reviewers (R.P.C. and J.C.S.) during two sessions held four weeks apart, a delay previously used for measuring intra-rater reliability[13,14]. A score was assigned for each task based on time to complete the task and deductions incurred for technical errors (Total Score = Timing Score – Penalty Score). The following is a complete description of how the scores were calculated. Given that some tasks had different maximum raw scores, the scores for each task were transformed to a scale from 0 to 100. A final global score of a possible 100 points was calculated by averaging the scores from all six tasks using equal weights.

Description of How Scores Were Calculated for Each Task

For each exercise, a timing score was calculated by subtracting the time to completion from a maximum allotted time in seconds. Precision was objectively scored by calculating a penalty score for each exercise as described below. A total score for each task was calculated by subtracting a penalty score from the timing score (Total Score = Timing Score – Penalty Score). For example, the maximum allotted time for task 1 is 180 seconds. If a participant were to complete task 1 in eighty seconds, he or she would obtain a timing score of 100 points. If he or she then received a 40-point deduction for a penalty score, the total score for task 1 would be 60 points. The lowest possible total score was 0, as no negative scores were assigned.

Task 1: Triangulation and Probing (Maximum Time, 180 Seconds; Maximum Score, 180 Points)

The participant had to identify three synthetic tissue tears of differing morphology hidden in separate locations. The probe was used to uncover a hidden symbol beneath each tear (Fig. 1-A). When the test proctor determined that the entire symbol was visualized, the participant was instructed to proceed. Five points were deducted each time the probe left the field of view. This error parameter reflects economy of movement and has been applied to other scoring

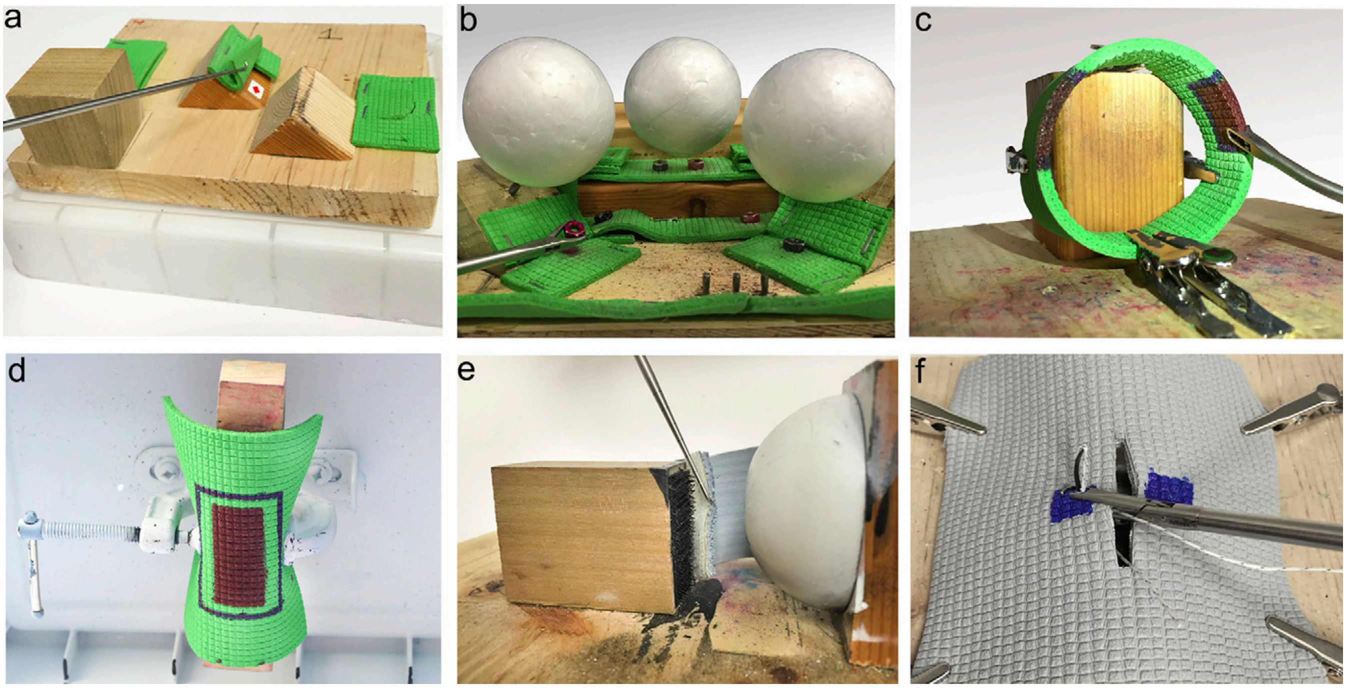

Fig. 1

Basic arthroscopic skills tasks. **Fig. 1-A** Triangulation and probing. **Fig. 1-B** Grasping and transferring objects. **Fig. 1-C** Tissue resection. **Fig. 1-D** Tissue-shaving. **Fig. 1-E** Tissue liberation and suture-passing. **Fig. 1-F** Tissue approximation and arthroscopic knot-tying.

systems[15]. Failure to identify all three symbols within the maximum allotted time resulted in a score of 0. This task simulated triangulation and probing used during diagnostic arthroscopy.

Task 2: Grasping and Transferring Objects (Maximum Time, 360 Seconds; Maximum Score, 360 Points)

Three black objects and three red objects were positioned beneath Styrofoam balls at standardized locations (Fig. 1-B). The participant had to use a grasper to transfer the objects onto a series of pegs. The right hand was first used to transfer the black objects, followed by the left hand to transfer the red objects. The standardized locations for the objects were chosen so that it would be impossible to complete the task without changing hands. Twenty points were deducted for contacting the Styrofoam balls, as seen in motion analysis scoring with inadvertent tissue collision[16,17]. Ten points were deducted each time an object was dropped[12]. This task simulated grasping of intra-articular loose bodies inside a knee joint.

Task 3: Tissue Resection (Maximum Time, 420 Seconds; Maximum Score, 420 Points)

A piece of synthetic material was fixed in the shape of a loop. The material had two separate areas, each measuring 2 × 1.5 cm and containing three colored zones: an inner red zone, a middle blue zone, and an outer green zone (Fig. 1-C). The participant used an "up-biting" resector to completely remove the inner red zones without removing green material. The blue zone gave the participant a margin of error. Five points were deducted for each small square of green material that was removed and 5 points were deducted for each small red square that remained. This task simulated skills required in knee meniscal resection (Video 1).

Task 4: Tissue-Shaving (Maximum Time, 360 Seconds; Maximum Score, 360 Points)

A piece of synthetic material was suspended upside-down. The undersurface of the object had a marked area measuring 5 × 3.5 cm and contained an inner red zone, a middle blue zone, and an outer green zone (Fig. 1-D). The participant had to use a mechanically powered shaver to remove the inner red zone without damaging the outer green material. The blue zone gave the participant a margin of error. Five points were deducted for each small square of green material that was removed and 5 points were deducted for each small red square that remained. This task simulated skills required in shoulder acromioplasty.

Task 5: Tissue Liberation and Suture-Passing (Maximum Time, 300 Seconds; Maximum Score, 300 Points)

The participant used a tissue elevator to release synthetic material adherent to a wood block by Velcro (Fig. 1-E). The participant then used a suture-passing instrument to pass a number-2 suture through a standardized target on the material. Twenty-five points were deducted for passing suture outside the target. This task simulated skills required in shoulder labral repair.

Task 6: Tissue Approximation and Arthroscopic Knot-Tying (Maximum Time, 240 Seconds; Maximum Score, 240 Points)

Subjects first watched a three-minute video on how to perform an arthroscopic sliding knot[18]. A piece of synthetic material containing a premade tear measuring 3.5 cm was attached over a spherical surface. Both sides of the tear had marked targets through which the subject needed to pass two limbs of a suture (Fig. 1-F). The subject approximated the tear edges by performing a sliding-locking knot followed by three half hitches. Twenty points were deducted for failure to pass the suture through the marked targets. Failure to tie an arthroscopic sliding-locking knot resulted in a 50-point deduction. Twenty-five points were deducted for incomplete tissue approximation. Only the suture-passing portion of this exercise was timed. This task simulated skills required in rotator cuff repair.

Two trials were performed for each task to allow subjects to familiarize themselves with the model and handling of the instruments[19]. Only the scores from the second trial were used in the statistical analysis to allow for a learning curve. The data were initially assessed for normality, and

TABLE I Timing, Penalty, and Total Scores for Individual Tasks

	Task 1 (Max = 180)	Task 2 (Max = 420)	Task 3 (Max = 180)	Task 4 (Max = 360)	Task 5 (Max = 300)	Task 6 (Max = 240)	Total Global Score (Max = 100)
Group 1*							
Timing score	93 ± 60	101 ± 79	65 ± 76	96 ± 78	113 ± 93	165 ± 28	
Penalty score	21 ± 14	37 ± 26	38 ± 31	35 ± 28	3 ± 9	68 ± 9	
Total score	77 ± 53	74 ± 71	52 ± 65	74 ± 74	113 ± 93	97 ± 30	29 ± 13.6
Group 2*							
Timing score	135 ± 18	113 ± 70	82 ± 88	137 ± 80	180 ± 49	150 ± 47	
Penalty score	22 ± 11	35 ± 33	46 ± 36	28 ± 33	0 ± 0	50 ± 28	
Total score	114 ± 20	89 ± 70	73 ± 79	124 ± 76	180 ± 49	101 ± 53	40.3 ± 12.1
Group 3*							
Timing score	141 ± 13	206 ± 41	164 ± 64	216 ± 43	203 ± 75	175 ± 36	
Penalty score	11 ± 10	15 ± 15	13 ± 10	16 ± 14	3 ± 9	26 ± 27	
Total score	130 ± 19	191 ± 44	151 ± 64	200 ± 51	202 ± 78	149 ± 38	57.6 ± 7.4
Group 4*							
Timing score	160 ± 2	237 ± 37	229 ± 56	263 ± 33	242 ± 20	190 ± 14	
Penalty score	2 ± 3	6 ± 9	15 ± 4	6 ± 7	0 ± 0	0 ± 0	
Total score	157 ± 4	232 ± 39	215 ± 58	258 ± 29	242 ± 20	190 ± 14	72.4 ± 3
ANOVA†							
Timing score	0.0004	<0.0001	<0.0001	<0.0001	0.0012	0.13	
Penalty score	0.0013	0.0085	0.0038	0.047	0.49	<0.0001	
Total score	<0.0001	<0.0001	<0.0001	<0.0001	0.0015	<0.0001	<0.0001

*The values are given as the mean and the standard deviation, in points. †The values are given as the p value.

continuous variables are reported as the mean and the standard deviation. Construct validity was measured using one-way analysis of variance (ANOVA) to assess differences in performance between groups for each task and total global score. A post hoc pairwise comparison with Tukey correction was used to compare the overall performance with level of training. Reliability was calculated using an intraclass correlation coefficient. Inter-rater reliability was measured between two blinded observers with use of the penalty scores and the total global score, and intra-rater reliability was measured for the same observer at four weeks apart. Significance was set at $p < 0.05$.

Source of Funding

There was no financial support for this study. The arthroscopic equipment was provided on loan from Arthrex (Naples, Florida).

Results

A total of forty-nine participants were voluntarily recruited for this study, which included thirty-six of forty residents from our training program. Participants were grouped by level of training: Group 1 (fifteen medical students and interns), Group 2 (twelve PGY-2 or PGY-3 residents), Group 3 (sixteen PGY-4 or PGY-5 residents), and Group 4 (six arthroscopic surgeons).

The mean timing scores, penalty scores, and total scores for each task are shown for each group in Table I. The ANOVA for each individual task showed a significant difference in scores according to the level of training ($p = 0.0015$ for task 5 and $p < 0.001$ for all other tasks). The mean total global score of a maximum 100 points differed

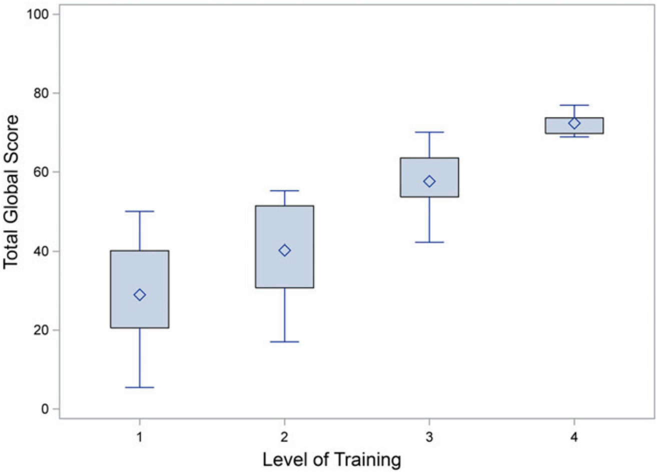

Fig. 2

Box-and-whisker plot comparing overall performance (total global score) by level of training. The box height represents the 25% to 75% interquartile range, and the means of each group are shown by the diamonds. The whiskers represent the range of maximum and minimum scores for each group.

TABLE II Pairwise Comparison of Overall Performance (Total Global Score) by Level of Training

Training Group	Total Global Score*
1 compared with 2	11.3 (0.3 to 22.3)
2 compared with 3	17.4 (6.6 to 28.2)
3 compared with 4	14.7 (1.1 to 28.3)

*The values are given as the group difference between the mean global score and the adjusted 95% confidence interval (determined with use of the Tukey correction), in points.

TABLE III Inter-Rater and Intra-Rater Reliability Intraclass Correlation Coefficients for Penalty Scores and Total Global Scores

	Intraclass Correlation Coefficient for Penalty Score	Intraclass Correlation Coefficient for Total Global Score
Inter-rater reliability	0.96	0.99
Intra-rater reliability	0.99	0.99

significantly between groups ($p < 0.001$): 29.0 ± 13.6 points for Group 1, 40.3 ± 12.1 points for Group 2, 57.6 ± 7.4 points for Group 3, and 72.4 ± 3.0 points for Group 4 (Fig. 2). A pairwise comparison using a Tukey correction further showed significant improvement in overall performance by increasing level of training between all groups ($p < 0.05$) (Table II). Finally, the intraclass correlation coefficient for inter-rater reliability of the penalty scores was 0.96 and the mean intra-rater intraclass correlation coefficient for both reviewers was 0.99 (Table III). The intraclass correlation coefficient for the total global score was 0.99 for inter-rater reliability and 0.99 for intra-rater reliability.

Discussion

The purpose of this study was to develop and validate a model consisting of different novel box modules that simulate basic skills commonly used across a wide range of arthroscopic procedures. The intention was to create a model that could eventually be used as an educational tool to evaluate and to teach fundamental arthroscopic skills to orthopaedic trainees. The results of this study confirmed construct validity by showing significant improvement in overall performance between all groups according to the level of training. Internal consistency of the model was also demonstrated by showing that all six tasks discriminated well between groups when analyzed individually.

When creating an objective scoring system to assess surgical skills, time to completion has been a popular parameter[16,20,21]. Although speed is an important measure of efficiency, it cannot fully detail the psychomotor skills that the trainee lacks. With this in mind, we designed performance metrics that rewarded both efficiency and precision. To ensure that the scoring system was objective and reproducible, the reliability of the measuring system was tested by evaluating inter-rater and intra-rater reliability. The overall model proved to be highly reliable, with 0.99 intraclass correlation coefficients for both inter-rater and intra-rater reliability. The exceptionally high reliability of our model is partially attributed to the fact that timing scores accounted for one of two components in the calculation of the overall global score. The timing scores are objective and remained constant between raters. Therefore, only the penalty scores accounted for the variability between raters, which is used to measure the intraclass correlation coefficients. Some of the most inexperienced participants who incurred the highest number of penalties also scored 0 on their timing score because of their inability to complete the task in the allotted time. This resulted in a total score of 0 for that task, thus making obsolete any discrepancy between raters when recording penalty scores for these participants. This further minimized the subjectivity in the scoring metric and variability between raters, where one would expect to have the highest variability. However, this would only partially contribute to a high intraclass correlation coefficient when analyzing the raw data and the mean scores for Group 1, as few participants actually scored 0 during individual tasks (Table I). A more plausible explanation is that both timing and penalty scores in this model have a high intra-rater and inter-rater reliability. The fact that some of the participants scored 0 for some of the tasks could also result in a floor effect in our scoring system. However, the overall scores were confirmed to be normally distributed, ensuring no floor or ceiling effect in our scoring metrics.

The box trainer model designed in this study is inexpensive, is reusable, and requires little maintenance. Costs are given in U.S. dollars. The total cost of designing and building the model was $800. However, this value is an overestimate of the true cost to build the model as a substantial portion of the material was wasted on earlier prototypes. The modular synthetic tissue components used in the tissue resection and shaving tasks (tasks 3 and 4) cost $50 and provided enough material for approximately 100 trials. In contrast, human cadavers have limited availability, poor cadaveric tissue compliance, and high cost, all of which substantially limit their use[22]. The use of live animals is also problematic because of ethical concerns and the need for specialized facilities. Recent advances in virtual reality technology have demonstrated its potential for enhancing surgical skills training[16,21,23]. However, studies evaluating the preferences of surgical residents have found no difference between video box trainers and virtual reality simulators[24]. Financial considerations and high start-up costs are also obstacles to widespread adoption of virtual reality simulators[25]. In fact, 87% of orthopaedic program directors identified a lack of available funding as the most important barrier to implementing a formal surgical skills program[1].

Fidelity is the degree to which the device simulates reality. In the current study, the box model had inherent limitations typical of other low-fidelity models. This model was not designed to assess aspects such as fluid management, portal placement, and application of varus or valgus stresses to work in different intra-articular compartments of the knee. However, when learning basic skills, fidelity has been shown to be much less important than other factors, such as feedback, repetition, and individualized learning[26]. Studies in other subspecialties have shown no difference in the performance and learning of surgical skills between low-fidelity and high-fidelity models[27]. This model was intended to teach essential motor skills such as visual-spatial perception and hand-eye coordination. This objective is supported by the fact that low-fidelity box models have proven learning benefits in other surgical disciplines[7,9,10]. MISTELS[12] is a low-fidelity model that has been successfully validated and implemented into a variety of surgical training programs, including its incorporation into the Fundamentals of Laparoscopic Surgery (FLS) training course, a certification requirement by the American Board of Surgery[28]. Low-fidelity models, when applied correctly to appropriate learners, can confer the same learning benefit as high-fidelity models[27].

This model is currently one of the largest studies evaluating a surgical simulation model in orthopaedic surgery[29-32]. Our validated and reliable model is appropriate for teaching basic arthroscopic skills. As programs begin to adopt competency-based curricula, valid and reliable simulation tools that are practical and affordable will be essential to help improve resident learning and structured assessment. ■

Ryan P. Coughlin, MD
Thierry Pauyo, MD
J. Carl Sutton III, MD
Division of Orthopaedic Surgery,
Montreal General Hospital,
McGill University Health Centre,
1650 Cedar Avenue Room B5 159.6, Montreal,
Quebec H3G 1A4, Canada.
E-mail address for R.P. Coughlin: ryan.coughlin@mail.mcgill.ca.
E-mail address for T. Pauyo: thierry.pauyo@mail.mcgill.ca.
E-mail address for J.C. Sutton: joseph.sutton@mail.mcgill.ca

Larry P. Coughlin, MD
Division of Orthopaedic Surgery,
St. Mary's Hospital Center,
3830 Avenue Lacombe, Montréal,
Quebec H3T 1M7, Canada.
E-mail address: lpcoughlin@hotmail.ca

Stephane G. Bergeron, MD, MPH
Department of Orthopaedic Surgery,
Jewish General Hospital,
3755 Côte-Sainte-Catherine Road, Pavilion E,
Room E-003, Montreal,
Quebec H3T 1E2, Canada.
E-mail address: stephane.bergeron@mail.harvard.edu

References

1. Karam MD, Pedowitz RA, Natividad H, Murray J, Marsh JL. Current and future use of surgical skills training laboratories in orthopaedic resident education: a national survey. J Bone Joint Surg Am. 2013 Jan 2;95(1):e4.

2. Emery SE, Guss D, Kuremsky MA, Hamlin BR, Herndon JH, Rubash HE. Resident education versus fellowship training-conflict or synergy? AOA critical issues. J Bone Joint Surg Am. 2012 Nov 7;94(21):e159.

3. Bucholz EM, Sue GR, Yeo H, Roman SA, Bell RH Jr, Sosa JA. Our trainees' confidence: results from a national survey of 4136 US general surgery residents. Arch Surg. 2011 Aug;146(8):907-14.

4. Garrett WE Jr, Swiontkowski MF, Weinstein JN, Callaghan J, Rosier RN, Berry DJ, Harrast J, Derosa GP. American Board of Orthopaedic Surgery Practice of the Orthopaedic Surgeon: Part-II, certification examination case mix. J Bone Joint Surg Am. 2006 Mar;88(3):660-7.

5. Hall MP, Kaplan KM, Gorczynski CT, Zuckerman JD, Rosen JE. Assessment of arthroscopic training in U.S. orthopedic surgery residency programs—a resident self-assessment. Bull NYU Hosp Jt Dis. 2010;68(1):5-10.

6. American Board of Orthopaedic Surgery. ABOS Surgical Skills Modules for PGY-1 Residents. https://www.abos.org/abos-surgical-skills-modules-for-pgy-1-residents.aspx. Accessed 2014 Sep 6.

7. Fried GM, Feldman LS, Vassiliou MC, Fraser SA, Stanbridge D, Ghitulescu G, Andrew CG. Proving the value of simulation in laparoscopic surgery. Ann Surg. 2004 Sep;240(3):518-25; discussion 525-8.

8. Stefanidis D, Acker C, Heniford BT. Proficiency-based laparoscopic simulator training leads to improved operating room skill that is resistant to decay. Surg Innov. 2008 Mar;15(1):69-73. Epub 2008 Apr 2.

9. Dauster B, Steinberg AP, Vassiliou MC, Bergman S, Stanbridge DD, Feldman LS, Fried GM. Validity of the MISTELS simulator for laparoscopy training in urology. J Endourol. 2005 Jun;19(5):541-5.

10. Lentz GM, Mandel LS, Lee D, Gardella C, Melville J, Goff BA. Testing surgical skills of obstetric and gynecologic residents in a bench laboratory setting: validity and reliability. Am J Obstet Gynecol. 2001 Jun;184(7):1462-8; discussion 1468-70.

11. Centers for Disease Control and Prevention. Evaluation Briefs: Gaining Consensus Among Stakeholders Through the Nominal Group Technique. 2006 Nov. http://www.cdc.gov/healthyyouth/evaluation/pdf/brief7.pdf. Accessed 2014 Sep 6.

12. Derossis AM, Fried GM, Abrahamowicz M, Sigman HH, Barkun JS, Meakins JL. Development of a model for training and evaluation of laparoscopic skills. Am J Surg. 1998 Jun;175(6):482-7.

13. Grant JA, Miller BS, Jacobson JA, Morag Y, Bedi A, Carpenter JE; MOON Shoulder Group. Intra- and inter-rater reliability of the detection of tears of the supraspinatus central tendon on MRI by shoulder surgeons. J Shoulder Elbow Surg. 2013 Jun;22(6):725-31. Epub 2012 Nov 14.

14. Hermansson LM, Bodin L, Eliasson AC. Intra- and inter-rater reliability of the assessment of capacity for myoelectric control. J Rehabil Med. 2006 Mar;38(2):118-23.

15. Alvand A, Khan T, Al-Ali S, Jackson WF, Price AJ, Rees JL. Simple visual parameters for objective assessment of arthroscopic skill. J Bone Joint Surg Am. 2012 Jul 3;94(13):e97.

16. Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL. Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. J Shoulder Elbow Surg. 2004 Mar-Apr;13(2):196-205.

17. Mabrey JD, Reinig KD, Cannon WD. Virtual reality in orthopaedics: is it a reality? Clin Orthop Relat Res. 2010 Oct;468(10):2586-91.

18. Kim SH, Ha KI. The SMC knot—a new slip knot with locking mechanism. Arthroscopy. 2000 Jul-Aug;16(5):563-5.

19. McCarthy AD, Moody L, Waterworth AR, Bickerstaff DR. Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? Clin Orthop Relat Res. 2006 Jan;442(442):13-20.

20. Gomoll AH, O'Toole RV, Czarnecki J, Warner JJ. Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. Am J Sports Med. 2007 Jun;35(6):883-8. Epub 2007 Jan 29.

21. Pedowitz RA, Esch J, Snyder S. Evaluation of a virtual reality simulator for arthroscopy skills development. Arthroscopy. 2002 Jul-Aug;18(6):E29.

22. Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. N Engl J Med. 2006 Dec 21;355(25):2664-9.

23. Cannon WD, Eckhoff DG, Garrett WE Jr, Hunter RE, Sweeney HJ. Report of a group developing a virtual reality simulator for arthroscopic surgery of the knee joint. Clin Orthop Relat Res. 2006 Jan;442(442):21-9.

24. Shetty S, Zevin B, Grantcharov TP, Roberts KE, Duffy AJ. Perceptions, training experiences, and preferences of surgical residents toward laparoscopic simulation training: a resident survey. J Surg Educ. 2014 Sep-Oct;71(5):727-33. Epub 2014 May 1.

25. Sutherland LM, Middleton PF, Anthony A, Hamdorf J, Cregan P, Scott D, Maddern GJ. Surgical simulation: a systematic review. Ann Surg. 2006 Mar; 243(3):291-300.

26. McGaghie WC, Issenberg SB, Petrusa ER, Scalese RJ. A critical review of simulation-based medical education research: 2003-2009. Med Educ. 2010 Jan;44 (1):50-63.

27. Matsumoto ED, Hamstra SJ, Radomski SB, Cusimano MD. The effect of bench model fidelity on endourological skills: a randomized controlled study. J Urol. 2002 Mar;167(3):1243-7.

28. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, Hoffman K; SAGES FLS Committee. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery. 2004 Jan;135(1):21-7.

29. Martin KD, Belmont 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 Nov 2;93(21): e1271-5.

30. Martin KD, Cameron K, Belmont PJ, Schoenfeld A, Owens BD. Shoulder arthroscopy simulator performance correlates with resident and shoulder arthroscopy experience. J Bone Joint Surg Am. 2012 Nov 7;94(21):e160.

31. Moktar J, Popkin CA, Howard A, Murnaghan ML. Development of a cast application simulator and evaluation of objective measures of performance. J Bone Joint Surg Am. 2014 May 7;96(9):e76.

32. 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.