



# High-fidelity hybrid simulation not only optimizes skills acquisition but improves nontechnical skills

Cristián Ruz<sup>1</sup> · Sebastián Irrarrázaval<sup>1</sup> · Pablo Besa<sup>1,2</sup>  · Catalina Vidal<sup>1</sup> · Claudio Nazar<sup>3</sup> · Julian Varas<sup>2</sup> · Ignacio Villagrán<sup>4</sup> · Eduardo Fuentes-López<sup>5</sup> · Arnoldo Riquelme<sup>6,7</sup>

Received: 14 February 2022 / Revised: 6 May 2022 / Accepted: 6 June 2022 / Published online: 15 July 2022  
© The Author(s), under exclusive licence to Association for Surgical Education 2022

## Abstract

**Objective** Knee arthrocentesis is frequently performed as a diagnostic or therapeutic procedure. Although considered a key competency for medical doctors, most students never execute the procedure during their training. We aimed to assess technical and nontechnical skills for knee arthrocentesis through high-fidelity hybrid simulation.

**Design** Trainees received educational documentation prior to training. Trainees took a medical history and obtained informed consent from a patient-actor, then encountered a simulated knee to execute the procedure. We adapted a direct observation scale to assess technical and nontechnical skill performance. Personalized feedback was received after each session. Performance among trainees (learning curves) and between trainees and experts was compared using a mixed-effects model.

**Setting** Pontifical Catholic University of Chile Simulation Center.

**Participants** Medical students and general physicians were recruited for training; orthopedic surgeons were recruited as experts.

**Results** Trainees significantly improved from the first session to the second and third. The third session was the learning curve plateau. Performance obtained in the third and fourth sessions were similar to expert performance. The assessment tool evaluated technical and nontechnical skills with high internal consistency and showed high interobserver reliability.

**Discussion** Learning curve analysis showed that high-fidelity simulation allowed trainees to become proficient in technical and nontechnical skills required to perform a safe knee arthrocentesis.

**Level of evidence** Level II (prospective cohort study).

**Keywords** Knee arthrocentesis · Medical education · Simulated training · Procedure skills · Nontechnical skills · Hybrid simulation · DOPS scale

## Introduction

Knee arthrocentesis is a procedure performed to aspirate fluids or inject medications into the knee joint cavity. It can be used as a diagnostic or therapeutic procedure [1, 2].

Diagnostically, knee arthrocentesis is a cornerstone in the differential diagnosis of inflammatory knee effusions [3]. As a treatment, joint injection of steroids or other medications plays a key role in the symptomatic relief of knee osteoarthritis (OA).

✉ Pablo Besa  
pablobesa@gmail.com

<sup>1</sup> Orthopedic Surgery Department, Faculty of Medicine, Pontificia Universidad Católica de Chile, Diagonal Paraguay N° 362, 3rd floor, Santiago, Chile

<sup>2</sup> Simulation and Experimental Surgery Centre, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>3</sup> Anesthesiology Department, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>4</sup> Physical Therapy, Department of Health Sciences, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>5</sup> Speech Therapy Career, Center for Medical Education and Health Sciences, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>6</sup> Health Sciences, Faculty of Medicine, Pontificia Universidad Católica de Chile, Santiago, Chile

<sup>7</sup> Department of Gastroenterology, Pontificia Universidad Católica de Chile, Santiago, Chile

While the American Board of Orthopedic Surgery (ABOS) included arthrocentesis as a core competency for orthopedic residents [4], other guidelines advise general physicians to learn the competency, regarding it as a significant milestone to obtain during undergraduate medical degree formation [5]. In our postgraduate orthopedic program, only 2 of the last 16 accepted residents (13%) had performed a knee arthrocentesis at the undergraduate level. Most last-year medical students feel ill-prepared to execute the procedure [5]. Furthermore, previous studies have shown that the success rate of palpation-guided therapeutic knee arthrocentesis among orthopedic surgery residents can be as low as 55% [6, 7].

Several medical education methodologies have been described to train knee arthrocentesis. These models have focused on the procedure's technical aspects, leaving nontechnical dimensions out of the training scenario [8, 9]. This leads to an incomplete and underperforming educational model [10, 11]. The inclusion of nontechnical skills promotes meaningful experiences that have shown to improve learning [12]. Moreover, exclusion can induce the belief that nontechnical competencies are less relevant when approaching a patient for a procedure [12]. In addition, previously published teaching models have lacked explicit instruments to assess trainees, limiting their reproducibility and external validity [4, 9].

The aims of this study were to (1) design and implement a high-fidelity hybrid simulation scenario of knee arthrocentesis, (2) compare last-year medical students and general physicians through a four-session workshop with experts in the knee arthrocentesis procedure in a simulated environment, and (3) adapt a DOPS scale to assess technical and nontechnical skills related to knee arthrocentesis.

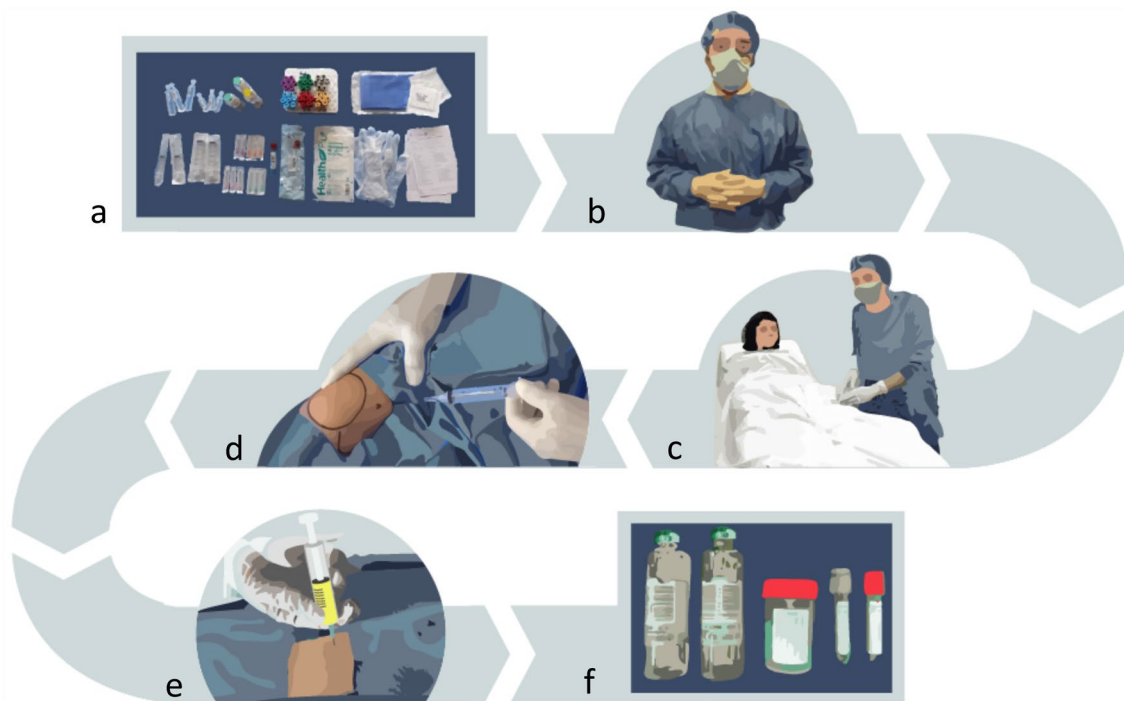
## Materials and methods

Institutional board review was obtained (no. 190107005). Last-year medical students and general physicians were recruited to complete four nonconsecutive training sessions in a 3-month period. We recruited six orthopedic surgeons with experience in knee arthrocentesis as experts against whom we could compare trainee improvement. Five days before encountering their initial training session, trainees received instructional documentation including written directions and rationale for performing a diagnostic knee arthrocentesis, a description of the training scenario including the assessment tool used to evaluate performance in technical and nontechnical competencies, and a video describing the procedure step by step.

## Training scenario

A high-fidelity hybrid simulation scenario was created. A patient-actor was trained with a script consisting of a 30-year-old patient arriving in the emergency department (ED) with two-day left knee pain associated with fever and joint inflammation. The patient-actor was stationed on a gurney; upon uncovering her left knee, trainees encountered a simulated knee (Sawbones®, Pacific Research Laboratories; Vashon, WA, USA). The model joint was a non-articulated knee with a partially mobile patella. On a side table, trainees had to select the required materials to perform the procedure, including hospital paperwork and informed consent. A health care assistant was posted to assist the trainee upon request but limited his participation to orders given by the trainee. During each session, trainees were required to take an abbreviated history and perform a physical examination of the patient. They had to explain the procedure and obtain written informed consent. After preparing the required instruments, they had to execute the procedure and load the laboratory test tubes. Finally, they completed hospital exam forms and gave the patients postprocedure recommendations (Fig. 1). A single orthopedic surgeon evaluated all trainees, and sessions were recorded for secondary evaluation to determine the inter-rater reliability of the evaluation tool. We used a specific direct observation of procedural skills (DOPS) scale designed for the scenario (supplemental material 1 and 2). After the procedure, all trainees were immediately conducted to a debriefing room to receive feedback from the surgeon who evaluated their performance. The surgeon had been previously trained to give effective feedback using the Pendleton model [13, 14]. Feedback was also registered. Each trainee's DOPS result constituted a point in their individual learning curve. Trainee learning curves were compared with expert performance to measure student proficiency in the training scenario. Proficiency was defined as the trainee's ability to safely conduct the procedure with careful consideration for the patient and following the best practices outlined in the educational material they received [15] measured through the de novo DOPS scale.

After feedback, each trainee completed a validated satisfaction scale [16]. This tool measured the trainees' perceptions regarding scenario realism, quality of the instructional material sent beforehand, feedback received, and perceived utility of the training session. One year after training, all participants were contacted to determine if they had performed any knee arthrocentesis. Those who had performed the procedure were given a questionnaire to measure how confident and prepared they felt to undertake the real-life procedure. Specifically, we asked if training



**Fig. 1** Key steps of the training scenario following the medical history and informed consent: **a** selecting required materials; **b** sterile environment preparation; **c** patient and skin preparation; **d** puncture site selection; **e** puncture and fluid extraction; and **f** tube filling and paperwork

had allowed them to perform patient consent and education, perform a safe knee arthrocentesis, fill laboratory tubes and paperwork, and explain postprocedure care to the patient. In addition, we asked the trainees to assess the perceived utility of participating in the training sessions.

### DOPS adaptation and validation

We adapted the new score from a DOPS previously validated in the same cultural setting [17]. The adaptation maintained the 11 items included in the original DOPS but adjusted their descriptors to assess knee arthrocentesis. We determined the content validity of the de novo DOPS by conducting a Delphi panel composed by experts in Orthopedic Surgery, Rheumatology and Emergency Medicine. Ratings and commentaries for each item were registered, and modifications were made for repeat expert assessment. We repeated expert consultation through the Delphi panel until we obtained at least 80% agreement on all items.

With the second evaluation performed by another orthopedic surgeon, inter-rater reliability was assessed [18]. Validity analysis was carried out with DOPS scale applications in consecutive sessions for each trainee. The construct validity was determined through an exploratory and confirmatory factor analysis. The exploratory factor analysis detected latent variables or constructs underlying the base of the observed variables [19, 20]. A confirmatory factor

analysis was further performed to validate the factor structure identified in the prior exploratory analysis [20]. The final scale graded each of the eleven items from 1 to 7 (with grades 1–3 as insufficient; 4 as standard setting; 5–6 partially accomplished; and 7 being completely accomplished).

### Statistical analysis

#### Adaptation and validation of the DOPS scale

Interrater reliability was measured with the weighted Kappa (wK) coefficient. Levels of agreement for wK were determined as proposed by Landis et al., considering wK values 0.00–0.20 as slight agreement; 0.21–0.40, fair agreement; 0.41–0.60, moderate agreement; 0.61–0.80, substantial agreement; and 0.81–1.00, almost perfect agreement [18].

In the exploratory factor analysis, the number of factors (or dimensions) was selected considering the Kaiser–Guttman [21] and Cattell [22] criteria. Thus, the factors with Eigenvalue above one, and those above the inflection point in the scree plot were retained. The determination coefficient ( $R^2$ ) was estimated to quantify the percentage of the scale's items' variance explained by the two factors identified in the exploratory analysis. Internal consistency for each dimension detected in factor analysis was performed using Cronbach's Alpha.

## Learning curve analysis

A mixed-effects/multilevel model with a random intercept was constructed to study differences in consecutive DOPS results of each trainee. The use of multilevel models was based on the fact that each trainee's performance was assessed in repeated training sessions. Thus, as DOPS scores in consecutive sessions for the same subject are compared, a correlation among them is expected, producing biased estimates of the standard errors and confidence intervals. Mixed-effects/multilevel models can be used to obtain standard errors that take the clustering within subjects into account. Multilevel statistical modeling enables quantitative analysis of learning curves and has been proven to have higher statistical power than conventional repeated-measures analysis of variance (ANOVA) [19]. This statistical method has also been used in previous research to analyze how trainees acquire skills [28–30].

Given that residuals of the mixed-effects model did not have a normal distribution, the standard error was estimated using bootstrapping (10,000 replications). Thus, a 95% confidence interval (95% CI) was obtained using the bias-corrected and accelerated method. Mean scores and 95% CI were expressed for each training session.

All analyses were conducted on Stata version 16 (StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC).

## Results

### Adaptation and validation of the DOPS scale

A Delphi panel composed of 17 orthopedic surgeons, 4 rheumatologists and 2 emergency department physicians. They received the 11-item DOPS with knee arthrocentesis descriptors. Two rounds were necessary to obtain over 80% agreement on every item.

In the exploratory factor analysis, two factors (or dimensions) with an Eigenvalue greater than one were identified (bidimensional). These factors explained 91.7% of the variance observed in the de novo DOPS, and they could be classified into a technical domain (two items) and a nontechnical domain (nine items). Regarding the confirmatory factor analysis, all standardized factor loadings were above 0.3 (cutoff point) and statistically significant (Table 1). The  $R^2$  fluctuated between 0.28 (item one) and 0.93 (item ten). It means that at least 93% of the variability of item ten's scores was explained by the factor identified. Internal consistency for each dimension was  $\alpha = 0.86$  in the technical domain and  $\alpha = 0.70$  in the nontechnical domain. Interrater reliability was almost perfect (wK 0.87).

## Learning curve analysis

Twenty-eight trainees were recruited (10 last-year medical students and 18 general physicians). Three had clinical experience as general physicians, with 4 years of experience each, resulting in a mean experience of 0.43 years among the 28 trainees. Only two of them (7%; both general physicians) had performed a single knee arthrocentesis prior to the training sessions. We also recruited six orthopedic surgeons to serve as experts.

Performance significantly improved between the first session (mean score 5.89 [95% CI 5.70–6.07]) and second session (mean score 6.51 [95% CI 6.30–6.72];  $p < 0.01$ ) and between the second and third session (mean score 6.80 [95% CI 6.59–7.00];  $p < 0.05$ ). We found no difference between the third session and the fourth (mean score 6.78 [95% CI 6.48–7.00];  $p = 0.94$ ) (Fig. 2).

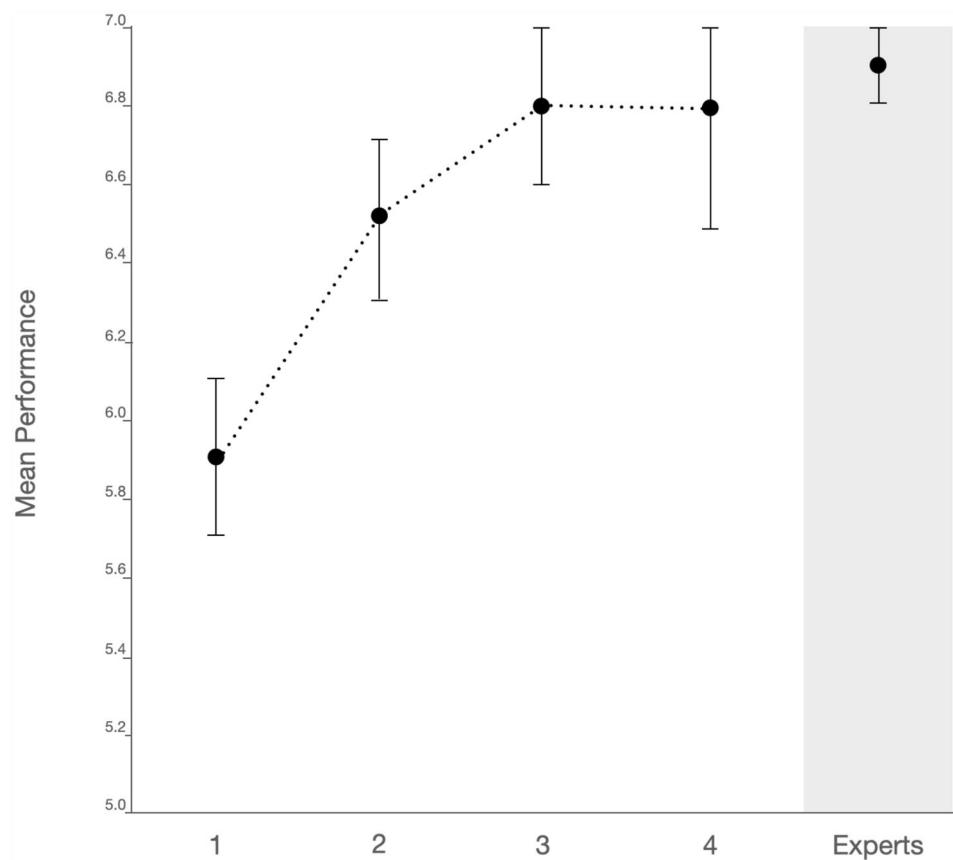
The second mixed-effects model was designed to compare each trainee session with the performance of the six expert orthopedic surgeons. We found that surgeons had a significantly higher mean score in their session (mean score 6.94 [95% CI 6.88–7.00]) than trainees did in their first ( $p < 0.01$ ), and second session ( $p < 0.01$ ). The third and fourth trainee sessions did not differ significantly with the experts' performance ( $p = 0.21$  and  $0.31$ , respectively).

### Training satisfaction survey and follow-up

After debriefing, 24 of the 28 trainees (85%) answered a questionnaire to evaluate the training experience. They all agreed that (1) the instructional material they received prior to the training sessions was useful, (2) the training scenario allowed them adequate training, (3) they perceived the feedback received as useful, (4) the assessment tool allowed them to focus on improving specific tasks, (5) the addition of a clinical case at the beginning improved scenario fidelity, and (6) the use of a trained patient-actor allowed improvement of nontechnical (communication) skills. The only question that received less than 100% of agreement was the perception that the knee model used was realistic (only 19 of 24 agreed; 80%).

Seven of the 28 trainees (25%) had performed a real-patient knee arthrocentesis by the 1-year follow-up. All the students thought that the training was useful and allowed them to perform the procedure safely. They felt confident regarding patient consent and education, performing the procedure and explaining postprocedure cares to the patient. The only item where two trainees (29%) felt insecure was selecting the correct laboratory tubes and completing paperwork.

**Fig. 2** Mean performance and standard error (vertical axis) for each trainee session (one through four of the horizontal axis) and the experts' performance ("Experts" on the horizontal axis)



## Discussion

Our high-fidelity simulation scenario allowed student trainees to improve the technical and nontechnical skills required to perform a safe knee arthrocentesis.

A key to training and learning is repetition. Previous studies have focused on proving differences between students prior to training and after training [4, 9] or on measuring satisfaction after a single use [20]. Repeated training and evaluation increase student performance and avoid jumping to conclusions that could be due to chance [21, 22]. This is the first study to determine how many training sessions are required to achieve proficiency in knee arthrocentesis. Students' learning curves showed a learning plateau after three sessions. Measuring the fourth session allowed us to confirm that performance was sustained and not just a one-session peak [21]. The short learning curve could be explained by the detailed training documentation received prior to the sessions, the direct observation scale known to the trainees, and personalized feedback received after each session that allowed the trainees to focus on their mistakes and gave them advice on how to improve. This learning curve behavior is related to obtaining a proficiency in the scenario. Obtaining an expertise in

knee arthrocentesis requires experience and hence is not obtainable in a brief simulated training course [15].

Teaching combined technical and nontechnical abilities has also proven to be effective through simulation [12, 23]. Simulated procedures and situations offer a structured teaching method that is replicable, objective, and safe (for the patient and student) [24, 25]. The inclusion of nontechnical aspects to a procedure usually considered technical has many proven benefits [12]. First, it allows for the teaching and learning of nontechnical aspects that are usually left out of technical training. Doctor–patient communication plays a key role in their relationship and in ideal clinical practice [26]. Furthermore, communication and nontechnical skills have been proven trainable [27]. Second, including these skills allowed us to incorporate the written consent as part of the procedure. The task of explaining the risks, benefits, and steps of the procedure to the patients helped our trainees incorporate significant theoretical concepts of knee arthrocentesis. Finally, as mentioned previously, hybrid training based on clinical scenarios adds meaning to the interaction. The presence of a real patient-actor increases the stakes for the student while remaining teaching friendly. This method has been proven to improve the learning experience [28]. Nontechnical aspects had not been trained in previous studies regarding knee arthrocentesis [8, 9, 20]. In our study,

the nontechnical dimension also improved between training sessions.

The scenario required creating an adaptation of a direct observation of procedure skills (DOPS) scale that proved to be consistent, reliable, and useful for trainee learning (assessment for learning) [29, 30]. Designing a simulated training scenario requires an objective measurement tool. Other studies have used generic observation tools or unvalidated checklists [31]. The main limitation for the use of a non-procedure-specific tool is that it lacks the nuance of the procedure and certainly does not incorporate nontechnical skills, limiting specific item-by-item feedback and therefore improvement. We decided to adapt a similar tool. Delfino et al. had already created a direct observation tool for a specific technical procedure (tracheal intubation) including nontechnical skills in their tool [17]. We decided to adapt the scale, maintaining the 11 items but changing the descriptors used in each of them. First, we determined that the tool had internal consistency and external reliability (among observers). This determination is important because it assures the reproducibility of our results [32, 33]. The tool we created also proved to be bidimensional. Obtaining two dimensions through exploratory and confirmatory factor analysis reaffirmed the notion that we were measuring technical and nontechnical competencies in the same procedure. Finally, we decided to make the DOPS scale available to trainees prior to the sessions. Traditionally, training and teaching scenarios prefer to maintain the items and descriptors of the scale away from the students. The use of the assessment tool for session preparation improves understanding of the key elements of each step, allowing students to learn from the assessment tool, thereby streamlining feedback and improvement [34]. This improvement might help explain the high initial rating for trainees' first session (mean score 5.89) and probably contributed to the steep learning curve they followed.

A 1-year follow-up found that trainees had performed the real-life procedures confidently given the abilities trained. Student perception has been proven to impact learning [35]. Positive perception leads to improved performance and increases the completion of training programs [36]. Our scenario had a positive trainee perception just after training and 1-year posttraining. Trainees positively valued the feedback given to them in the debriefing room. The feedback was structured, and the surgeon giving the feedback received prior training regarding feedback structure and technique. We believe our experience and the previous literature [37–40] make the inclusion of personalized feedback a key element in simulated training.

One of our main limitations was that the training sessions were not evenly spaced between students nor for single students. This is a frequent limitation when training has no curricular integration. After its initial success, the program will be included in student training, allowing for sessions at

regular intervals. Secondarily, although subjective trainee evaluations rated the model favorably, the knee model used is relatively simple and did not allow the students to scope the impact knee variability has on increasing arthrocentesis difficulty.

In this research, we have opted to apply multilevel statistical models to consider repeated-measures obtained from each trainee and compare groups. Therefore, using multilevel models, differences in DOPS performances between sessions, and among trainees and experts could be obtained. Identifying the point in the learning curve where the slope flattens out (inflection point) is critical [26]. This point represents a progressively higher effort to achieve trainees' learning gains before reaching a performance level similar to an expert (27). Thus, using this statistical approach, we could graphically represent the learning curves and identified their critical points. Future research could be conducted incorporating an interaction between individual-level (e.g., last-year medical student versus a specialist resident) and group-level variables (e.g., different simulations scenarios) through multilevel statistical models.

## Conclusion

A high-fidelity simulation scenario allowed student trainees to improve the technical and nontechnical skills required to perform a safe knee arthrocentesis. After three training sessions, the trainee performance was similar to the experts we assessed. The scenario required creating an adaptation of a direct observation of procedure skills (DOPS) scale that proved to be consistent and reliable.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s44186-022-00027-y>.

**Acknowledgements** We want to recognise the work of Valentina Riquelme, Faculty of Arts, Pontificia Universidad Católica de Chile, for her contribution with the illustration.

**Data availability** The datasets generated during and analyzed during the current study are available from the corresponding author upon request.

## References

1. Shimizu M, Higuchi H, Takagishi K, Shinozaki T, Kobayashi T. Clinical and biochemical characteristics after intra-articular injection for the treatment of osteoarthritis of the knee: prospective randomized study of sodium hyaluronate and corticosteroid. *J Orthop Sci.* 2010;15(1):51–6.
2. Sánchez M, Anitua E, Azofra J, Aguirre JJ, Andia I. Intra-articular injection of an autologous preparation rich in growth factors for the treatment of knee OA: a retrospective cohort study. *Clin Exp Rheumatol.* 2008;26:910–3.

3. Margaretten ME, Kohlwes J, Moore D, Bent S. Does this adult patient have septic arthritis? *JAMA*. 2007;297(13):1478–88.
4. Seeley MA, Kazarian E, King B, Biermann JS, Carpenter JE, Caird MS, et al. Core concepts: orthopedic intern curriculum boot camp. *Orthopedics*. 2016;39(1):e62–7.
5. Telias NA, Innocenti DP, Orrego GF, Oliver CN, Huidobro DB, Orrego LM. Evaluación de percepción de conocimientos y destrezas de alumnos de medicina en traumatología y ortopedia. *Rev chil ortop traumatol*. 2011;52(2):63–70.
6. Curtiss HM, Finnoff JT, Peck E, Hollman J, Muir J, Smith J. Accuracy of ultrasound-guided and palpation-guided knee injections by an experienced and less-experienced injector using a superolateral approach: a cadaveric study. *PM R*. 2011;3(6):507–15.
7. Jackson DW, Evans NA, Thomas BM. Accuracy of needle placement into the intra-articular space of the knee. *J Bone Jt Surg Ser A*. 2002;84-A:1522–7.
8. Waikukul S, Chumtup W, Chotivichit A, Vanadurongwan B, Assawamongkolgul A, Rojanawanich V. A knee model for arthrocentesis simulation. *J Med Assoc Thai*. 2003;86(3):282–7.
9. Chiowchanwisawakit P, Ratanarat R, Srinonprasert V. Improving sixth year medical students' performance in knee arthrocentesis using a synthetic knee model. *Int J Rheum Dis*. 2015;18(7):742–50.
10. Greenberg CC, Regenbogen SE, Studdert DM, Lipsitz SR, Rogers SO, Zinner MJ, et al. Patterns of communication breakdowns resulting in injury to surgical patients. *J Am Coll Surg*. 2007;204(4):533–40.
11. Yule S, Flin R, Maran N, Rowley D, Youngson G, Paterson-Brown S. Surgeons' non-technical skills in the operating room: reliability testing of the NOTSS behavior rating system. *World J Surg*. 2008;32(4):548–56.
12. Moulton CA, Tabak D, Kneebone R, Nestel D, MacRae H, LeBlanc VR. Teaching communication skills using the integrated procedural performance instrument (IPPI): a randomized controlled trial. *Am J Surg*. 2009;197:113–8.
13. Pendleton D, Schofield T, Tate PHP. *The consultation: an press, approach to learning and teaching*. Oxford: Oxford University; 2003.
14. Ramani S, Krackov SK. Twelve tips for giving feedback effectively in the clinical environment. *Med Teach*. 2012;34(10):787–91.
15. Carraccio CL, Benson BJ, Nixon LJ, Derstine PL. From the educational bench to the clinical bedside: translating the Dreyfus developmental model to the learning of clinical skills. *Acad Med*. 2008;83(8):761–7.
16. Villagrán I, Tejos R, Chahuán J, Usilar T, Pizarro M, Varas J, et al. Percepción de estudiantes de pregrado de Medicina de talleres de simulación de procedimientos médico-quirúrgicos. 2018;146:649–58.
17. Delfino AE, Chandratilake M, Altermatt FR, Delfino AE, Chandratilake M, Altermatt FR. Validation and piloting of direct observation of practical skills tool to assess intubation in the Chilean context validation and piloting of direct observation of practical skills tool to assess intubation in the Chilean context. *Med Teach*. 2013;35:231–6.
18. Landis JR, Koch GG. Landis and Koch 1977 agreement of categorical data. *Biometrics*. 1977;33(1):159–74.
19. Memsoudis SG. Beyond repeated-measures analysis of variance data in anesthesia research. *Reg Anesth Pain Med*. 2012;37(1):99–105.
20. Berman JR, Ben-Artzi A, Fisher MC, Bass AR, Pillinger MH. A comparison of arthrocentesis teaching tools: cadavers, synthetic joint models, and the relative utility of different educational modalities in improving trainees' comfort with procedures. *J Clin Rheumatol*. 2012;18(4):175–9.
21. Hopper AN, Jamison MH, Lewis WG. Learning curves in surgical practice. *Postgrad Med J*. 2007;83:777–9.
22. Varas J, Mejía R, Riquelme A, Maluenda F, Buckel E, Salinas J, et al. Significant transfer of surgical skills obtained with an advanced laparoscopic training program to a laparoscopic jejunojejunostomy in a live porcine model: feasibility of learning advanced laparoscopy in a general surgery residency. *Surg Endosc Other Interv Tech*. 2012;26(12):3486–94.
23. LeBlanc VR, Tabak D, Kneebone R, Nestel D, MacRae H, Moulton CA. Psychometric properties of an integrated assessment of technical and communication skills. *Am J Surg [Internet]*. 2009;197(1):96–101.
24. Atesok K, Mabrey JD, Jazrawi LM, Egol KA. Surgical simulation in orthopaedic skills training. *J Am Acad Orthop Surg*. 2012;20(7):410–22.
25. Corvetto M, Bravo MP, Montaña R, Utili F, Escudero E, Boza C, et al. Simulación en educación médica: una sinopsis. *Rev Med Chil [Internet]*. 2013;141(1):70–9.
26. Berner JE, Ewertz E. The importance of non-technical skills in modern surgical practice. *Cir Esp*. 2019;97(4):190–5.
27. Ranjan P, Kumari A, Chakrawarty A. How can doctors improve their communication skills? *J Clin Diagnostic Res*. 2015;9(3):1–4.
28. Stroud L, Cavalcanti RB. Hybrid simulation for knee arthrocentesis: Improving fidelity in procedures training. *J Gen Intern Med*. 2013;28(5):723–7.
29. Epstein RM. Medical education—assessment in medical education. *N Engl J Med*. 2007;356(4):387–96.
30. Naeem N. Validity, reliability, feasibility, acceptability and educational impact of direct observation of procedural skills (DOPS). *J Coll Physicians Surg*. 2013;23(1):77–82.
31. Downing SM. Validity: on the meaningful interpretation of assessment data. *Med Educ*. 2003;37:830–7.
32. Downing SM. The metric of medical education reliability: on the reproducibility of assessment data. *Med Educ*. 2004;38:1006–12.
33. Khan KZ, Ramachandran S, Gaunt K. The Objective Structured Clinical Examination (OSCE): AMEE guide no. 81. Part I: an historical and theoretical perspective. *Med Teach*. 2016;35(81):1437–46.
34. Lockyer J, Carraccio C, Chan M, Hart D, Smee S, Touchie C, et al. Core principles of assessment in competency-based medical education. *Med Teach*. 2017;39(6):609–16.
35. Schuwirth LWT, Van der Vleuten CPM. Programmatic assessment: from assessment of learning to assessment for learning. *Med Teach [Internet]*. 2011;33(6):478–85. <https://doi.org/10.3109/0142159X.2011.565828>.
36. Johnston S, Coyer FM, Nash R. Kirkpatrick's evaluation of simulation and debriefing in health care education: a systematic review. *J Nurs Educ*. 2018;57(7):393–8.
37. Kirkpatrick, Donald L. *Evaluating training programs: the four levels*. 3rd edition. San Francisco: Emeryville, CA: Berrett-Koehler; Publishers Group West, 1994. p. 52–62.
38. Norcini J, Burch V. 31 Workplace-based assessment as an educational tool: AMEE Guide. *Med Teach*. 2007;29(9–10):855–71.
39. Herrmann-Werner A, Nikendei C, Keifenheim K, Bosse HM, Lund F, Wagner R, et al. “Best practice” skills lab training vs a “see one, do one” approach in undergraduate medical education: an RCT on students’ long-term ability to perform procedural clinical skills. *PLoS ONE*. 2013;8(9):e76354.
40. Farquharson AL, Cresswell AC, Beard JD, Chan P. Randomized trial of the effect of video feedback on the acquisition of surgical skills. *Br J Surg*. 2013;100:1448–53.