TOPICS IN TRAINING

Assessing Performance in Shoulder Arthroscopy: The Imperial Global Arthroscopy Rating Scale (IGARS)

Sofia Bayona, PhD, MEd, MAS, BS, Kash Akhtar, FRCS(Tr&Orth), FRSA, MEd, MSc, MBBS,
Chinmay Gupta, PhD, FRCS(Tr&Orth), MA(Oxon), BMBCh, Roger J.H. Emery, FRCSE, MS,
Alexander L. Dodds, FRCS(Tr&Orth), MSc, BCh, MB, and Fernando Bello, PhD, BSc (Hons)

Investigation performed at the Musculoskeletal Laboratory and the Department of Surgery and Cancer, Imperial College London, London, United Kingdom

Background: Surgical training is undergoing major changes with reduced resident work hours and an increasing focus on patient safety and surgical aptitude. The aim of this study was to create a valid, reliable method for an assessment of arthroscopic skills that is independent of time and place and is designed for both real and simulated settings. The validity of the scale was tested using a virtual reality shoulder arthroscopy simulator.

Methods: The study consisted of two parts. In the first part, an Imperial Global Arthroscopy Rating Scale for assessing technical performance was developed using a Delphi method. Application of this scale required installing a dual-camera system to synchronously record the simulator screen and body movements of trainees to allow an assessment that is independent of time and place. The scale includes aspects such as efficient portal positioning, angles of instrument insertion, proficiency in handling the arthroscope and adequately manipulating the camera, and triangulation skills. In the second part of the study, a validation study was conducted. Two experienced arthroscopic surgeons, blinded to the identities and experience of the participants, each assessed forty-nine subjects performing three different tests using the Imperial Global Arthroscopy Rating Scale. Results were analyzed using two-way analysis of variance with measures of absolute agreement. The intraclass correlation coefficient was calculated for each test to assess inter-rater reliability.

Results: The scale demonstrated high internal consistency (Cronbach alpha, 0.918). The intraclass correlation coefficient demonstrated high agreement between the assessors: 0.91 ($p < 0.001$). Construct validity was evaluated using Kruskal-Wallis one-way analysis of variance (chi-square test, 29.826; $p < 0.001$), demonstrating that the Imperial Global Arthroscopy Rating Scale distinguishes significantly between subjects with different levels of experience utilizing a virtual reality simulator.

Conclusions: The Imperial Global Arthroscopy Rating Scale has a high internal consistency and excellent inter-rater reliability and offers an approach for assessing technical performance in basic arthroscopy on a virtual reality simulator.

Clinical Relevance: The Imperial Global Arthroscopy Rating Scale provides detailed information on surgical skills. Although it requires further validation in the operating room, this scale, which is independent of time and place, offers a robust and reliable method for assessing arthroscopic technical skills.

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: One or more 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 an aspect of this work. In addition, one or more of the authors, or his or her institution, has had a financial relationship, in the thirty-six months prior to submission of this work, with an entity in the biomedical arena that could be perceived to influence or have the potential to influence what is written in this work. 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.

Surgical training has undergone major changes recently with limits being placed on resident working hours¹ and an increasing focus on patient safety. This has led to a direct reduction in the amount of training opportunities available and may well result in less experienced trainees. Concomitantly, there is a need for appropriate methods of assessment and feedback in educating surgeons so that training can be closely guided and personalized to the individual. Assessments have a role in ensuring patient safety and may determine who is competent in performing a surgical procedure.

Surgeons are assessed using a number of methods to judge both technical and non-technical skills. High-quality assessment should be valid, reliable, repeatable, feasible, and cost-effective. It should include opportunities for feedback and have a continuous developmental impact on learning².

Using procedure-based assessments, checklists, and direct observation of procedural skills in surgery requires both the trainee and the assessor to be in the operating room. Therefore, this method is dependent on time and place. The operating room contributes to the validity of the assessment, but it presents an element of risk to the patient, as well as problems in terms of repeatability, making it difficult to compare performance or to set standards³. These factors impact the feasibility and reliability. Moreover, these settings usually make anonymous assessment impossible, risking conscious and unconscious bias on the part of the assessor.

Simulation and, in particular, virtual reality simulators allow trainees to gain experience in controlled, risk-free environments. They allow deliberate and repeated practice⁴ with no risk to the patient or doctor, and their role in orthopaedic training is increasing. However, they are not without limitations. Most available studies of virtual reality arthroscopy simulators do not investigate transferability of skills, and heterogeneity exists in the literature with respect to the comparison of simulator performance metrics other than completion time⁵. Still, the evidence base for the role of virtual reality simulators in surgical training is growing, particularly in laparoscopic surgery, which is becoming increasingly incorporated into the surgical curriculum^{6,7}.

Current assessment methodologies of intraoperative techniques are somewhat limited in scope⁸, and there is no specific assessment for virtual reality-simulated procedures such as shoulder and knee arthroscopy. Moreover, arthroscopic surgery has its own particular demands and requires a set of skills that are distinct from traditional open surgery. Hoyle et al.⁹ analyzed methods of assessing arthroscopy skills, and they inferred that the gold standard of technical skills evaluation is the Global Rating Scale (GRS). Their work was also based on a review of existing assessment tools: Global Operative Assessment of Laparoscopic Skills (GOALS)¹⁰, Objective Structured Assessment of Technical Skill (OSATS)^{11,12}, and Global Rating Index for Technical Skill (GRITS)¹³. However, the GRS proposed by Hoyle et al., although showing good internal consistency (Cronbach alpha, 0.89 to 0.96) and a good discriminatory ability ($p < 0.0001$), presented significant variability between raters ($p = 0.0002$ to 0.01), indicating a lack of inter-rater reliability.

Aims of the Study

Our study had two main aims. The first aim was to design a portable arthroscopic GRS independent of time and place to assess the technical competence of trainees when performing arthroscopy; this GRS was named the Imperial Global Arthroscopy Rating Scale (IGARS). The second aim was to validate the scale using tests of internal consistency and inter-rater reliability in the context of a virtual reality shoulder arthroscopy simulator with proven face, content, construct, and concurrent validity¹⁴⁻¹⁷.

Materials and Methods

IGARS Design

The IGARS for assessing technical performance was developed using a Delphi method. The Delphi method¹⁸ was applied with five surgeons (two of whom [C.G. and R.J.H.E.] were authors in this study) experienced in performing arthroscopic surgery and in teaching and assessing surgical trainees. This method consisted of a review of the literature to identify key elements in the assessment of arthroscopic technical skills, skills in other surgical specialties, and generic surgical technical skills. The resultant list was edited to remove any redundancy, and the experts generated new items to cover technical aspects specific to arthroscopy, such as appropriate portal position, depth perception, bimanual dexterity, and triangulation.

Following the Delphi methodology, an iterative process was undertaken to refine the rating scale and to ensure that all relevant aspects of arthroscopic skills were covered. Wording was reviewed to be clear and unambiguous. A 5-point scale was chosen for each item, with attention to explicitly describing the middle and extreme points to enhance reliability. Anchors were aimed to be specific, unequivocal, and as equally distributed as possible. These iterative rounds continued until consensus was achieved among all experts. In addition, a pilot study was undertaken using the IGARS with three experts and ten trainees. Item-total correlations were calculated, demonstrating that the IGARS was well received and further item reduction was not necessary.

Technical Requirements for Applying IGARS:

A Dual-Camera System

When reviewing surgical performance, we believed that it was important to observe the screen on which the interior of the joint is shown, as well as the global motion of the operator's hands and body movements. To achieve this, a system with two cameras was set up. The first camera recorded the intra-articular movements (from either the actual or the simulated arthroscope), and the second camera focused on the trainees' hands and body movements. The second camera avoided recording the face of the trainee to prevent individual identification and to therefore allow an unbiased analysis (Fig. 1). These technical requirements did not compromise the portability of our scale, as the dual-camera system could be set up in both real and simulated environments.

Initial Validation of the IGARS

The initial validation involved the evaluation of the face, content, and construct validities of the IGARS.

Face validity (the extent to which our IGARS appears to measure the issues relevant to arthroscopy¹⁹) and content validity (the degree to which elements of an assessment instrument are relevant, adequate, and representative of the target for a particular assessment purpose²⁰) were attained through the systematic methodology employed in developing the IGARS, which involved a group of experts at all stages of development.

Reliability and construct validity were tested by designing an experiment in which the IGARS was used as a means of assessing the technical performance of shoulder arthroscopy on a virtual reality simulator (ARTHRO Mentor; Simbionix, Cleveland, Ohio).

Fig. 1
Example of the output of the dual-camera system in a simulated environment. (Left) Virtual reality ARTHRO Mentor simulator intra-articular display. (Right) View of the participant's body movements using the simulator.

Experimental Design to Evaluate the Reliability and Construct Validity of the IGARS in a Simulated Environment

Three tests, composed of the most common tasks covering many of the key arthroscopic skills, were used to provide an observer with sufficient information to apply the IGARS: (1) diagnostic glenohumeral arthroscopy via a posterior portal, (2) probe examination of the subacromial space (arthroscope through a lateral portal, probe through an anterior portal), and (3) probe examination of a Bankart lesion (arthroscope through an anterior portal, probe through a posterior portal).

The time permitted for all three tests was limited to a maximum of twenty minutes. A separate pilot test verified that the majority of subjects could complete them within that time.

One participant and the corresponding data were removed from the study as the videos could not be edited to ensure anonymity. Consequently, the data of the study correspond to forty-nine subjects with a wide range of arthroscopic experience: twenty-eight novices, eleven intermediates (some experience being the main surgeon, but less than fifty procedures), and ten expert arthroscopists (more than fifty procedures as the main surgeon). Exclusion criteria were difficulties in understanding written English and any previous exposure to this virtual reality shoulder arthroscopy simulator. Ethical approval was obtained through the local research ethics committee.

Each subject was recorded performing the three tests and was assigned a unique identifier. The videos and data were reviewed to safeguard each subject's anonymity. This ensured a blinded study, so that neither the assessor nor the researcher performing the statistical data analysis could identify any participant.

A total of 147 videos (one for each test, forty-nine participants) were collected and were independently assessed by two blinded assessors using the IGARS. The assessment design allowed assessors to rate the videos at a time and location of their choice, and they were free to decide how many videos to assess in each sitting. The assessors completed the IGARS during and after watching each video. Completing each form took five minutes on average.

Statistical analysis was performed using SPSS 17.0 (SPSS, Chicago, Illinois). Validity and reliability of the scale were analyzed. The intraclass correlation coefficient was calculated to test inter-rater reliability. Results were analyzed using two-way analysis of variance with measures of absolute

agreement. Internal consistency was checked with the Cronbach alpha. Item-total correlations were calculated to confirm the final items. Skew was minimized by ensuring that each item measured different aspects of arthroscopic technical skills and by keeping the total number of items to the minimum necessary. All reliability measures were calculated for the whole IGARS. Construct validity was studied using the Kruskal-Wallis one-way analysis of variance.

Source of Funding

One author (S.B.) received grants from the Spanish Ministry of Education, Science and Innovation (TIN2010-21289) and the Cajal Blue Brain Project, Spanish partner of the Blue Brain Project initiative from École Polytechnique Fédérale de Lausanne in Switzerland. Two authors (S.B. and F.B.) received through their institutions a grant from the FP7 Marie Curie Actions for the Intra European Fellowship (grant PIEF-GA-2009-236642). Funds were used to pay for living expenses and salary for one author (S.B.) and for the administrative costs of the study.

Results

The first part of our study generated the IGARS as shown in Table I. The scale consists of nine dimensions (portal position and angles of instrument insertion, respect for tissue, proficiency in handling the arthroscope adequately using the camera and optic turns, proficiency using the instruments, depth perception, triangulation, efficiency and time, knowledge of the anatomy, knowledge of the procedure) and a final global assessment score. Each dimension is graded 1 to 5, with the middle and extreme points defined explicitly.

The second part of our study refers to the results in terms of validity and reliability. The IGARS was applied to 147 cases, and a reliability analysis was performed. Item-total correlations were calculated to confirm the final items. The item-total correlations were ≥ 0.88 , and the results did not show the need for further item reduction, supporting the findings of the initial pilot study.

TABLE I The Imperial Global Arthroscopy Rating Scale (IGARS)* for Assessing Arthroscopic Performance

Observer ID: _____

Video ID: _____

1. Makes the portals (or inserts the arthroscope and instruments) with the correct angles of insertion to access the targeted space
 - 1- Many unsystematic attempts and very inefficient access to the targeted point with randomly chosen angles of insertion.
 - 2-
 - 3- Several attempts before getting into the targeted space. Once found, maintains the correct arthroscope and instrument angles most of the time.
 - 4-
 - 5- Efficient angles of insertion, perfect access to the targeted space within the joint.
2. Respect for tissue (looking at the smoothness of the movements)
 - 1- Often used unnecessary force on tissue. Damage caused by inappropriate use of the instruments.
 - 2-
 - 3- Careful handling of tissue but occasionally rough movements or caused inadvertent damage.
 - 4-
 - 5- Consistently handled tissues appropriately, with minimal damage.
3. Proficiency in handling the arthroscope
 - 1- The view of the anatomy is not coherent (for example, randomly oriented upside-down). Incoherent turns of the optics and the camera. Repeatedly makes tentative or awkward movements with the arthroscope.
 - 2-
 - 3- Competent use of the arthroscope, coherent view of the anatomy although occasionally appeared stiff or awkward. Inefficient use of the optics turn.
 - 4-
 - 5- Fluid moves with the arthroscope, adequate use of camera and optics turns, and no awkwardness. If adequate, takes advantage of using different portals.
4. Proficiency in handling the instruments (excluding the arthroscope)
 - 1- Repeatedly makes tentative or awkward movements with instruments.
 - 2-
 - 3- Competent use of instruments, although occasionally appeared stiff or awkward.
 - 4-
 - 5- Fluid moves with instruments and no awkwardness.
5. Depth perception
 - 1- Constantly overshoots target. Slow to correct.
 - 2-
 - 3- Some overshooting or missing of target.
 - 4-
 - 5- Accurately directs instruments in the correct plane to target.
6. Bimanual dexterity and triangulation
 - 1- Noticeably awkward with non-dominant hand, poor coordination between hands. Unable to locate or great difficulty in locating the instrument.
 - 2-
 - 3- Uses both hands but does not maximize interaction between hands. Locates the instrument after a moderate time, but in a systematic way.
 - 4-
 - 5- Expertly uses both hands in complementary manner to provide optimum performance. Locates the instrument quickly and efficiently.
7. Efficiency and time
 - 1- Many unnecessary, inefficient movements. Constantly changing focus or persisting without progress.
 - 2-
 - 3- Slow, but planned movements are reasonably organized with few unnecessary or repetitive movements.
 - 4-
 - 5- Confident, clear economy of movement and maximum efficiency.

continued

TABLE I (continued)

8. Knowledge of anatomy and orientation

- 1- Deficient knowledge of the anatomy. Feeling of being lost within the joint. Clear mistakes when recognizing or looking for anatomical structures.
- 2-
- 3- No mistakes when recognizing the anatomy. Good orientation and knowledge of where the arthroscope and instruments are. Moderate knowledge of the anatomical structures and pathologies.
- 4-
- 5- Demonstrated familiarity with all the anatomical structures and pathologies. Excellent orientation.

9. Knowledge of specific procedure and/or tasks to be performed (related to each particular exercise)

- 1- Deficient knowledge of the operative steps, even with help.
- 2-
- 3- Able to follow the operative steps (or tasks) with moderate help.
- 4-
- 5- Demonstrated familiarity with all aspects of the operation and the operative steps to be performed.

10. Overall individual performance and competence (related to each particular exercise)

- 1- Very poor
- 2-
- 3- Competent
- 4-
- 5- Clearly superior

Observations:

Strengths:

Suggestions for improvements:

*The portable scale is designed to be used in both the operating room and in simulated environments. It allows an anonymous assessment that is independent of time and place.

Fig. 2

A box plot showing the IGARS score. The portable IGARS was able to distinguish between subjects according to their level of experience. The first group had twenty-eight novices, the second group had eleven intermediates, and the third group had ten experts in arthroscopy. The dark lines in the middle of the boxes show the median, the top lines of the boxes denote the 75th percentile, and the bottom lines of the boxes denote the 25th percentile. The whiskers denote 1.5 times the interquartile range or, if no case has a value in that range, the minimum or maximum values. The asterisks denote extreme values or outliers.

The global intraclass correlation coefficient between assessors for the IGARS was high (0.91), and the scale showed high internal consistency (Cronbach alpha, 0.918), suggesting that all of the items in the scale are consistent and measure aspects of the same construct: arthroscopic technical skills. There was no ceiling or floor effect, with only a small proportion (<15%) of subjects scoring the minimum (1 point) or maximum (5 points) score as their total score.

The results of the Kruskal-Wallis one-way analysis of variance to test construct validity showed significance (chi-square test, 29.826; $p < 0.001$), proving that the IGARS was able to distinguish subjects according to their level of experience (Fig. 2). Additionally, a moderate Spearman correlation ($r = 0.675$, $p < 0.001$) was found between the number of arthroscopies done as the main surgeon and the score of the scale.

Discussion

With reduction in the resident training hours and patient safety being of paramount importance, there is a need for valid and reliable methods of assessing technical competence in surgical training. The IGARS, developed in our study, is usable in both real and simulated environments and is designed to overcome the limitation of assessment methods that are dependent on time and place. The incorporation of two synchronized cameras as part of the assessment, one focused on the video screen

and one focused on the trainee's hands, allowed for a comprehensive assessment of the trainee's performance that may also be used for self-assessment, reflection, and evaluation purposes (see Video 1). Additionally, these settings are reproducible in the operating room, permitting assessors, even from different institutions, to remotely and blindly rate trainees at their convenience and to stop or review relevant sections. Blinding of those undertaking rating and data analysis removed any inherent bias and supports the use of this method as a tool for assessment and feedback.

A Delphi method was used to obtain the final version of the scale, similarly to other studies for different surgical assessments^{21,22}. Reliability results improved considerably compared with previous arthroscopic assessments^{9,23}. Additionally, the recording by two synchronized cameras allows for anonymous assessment and increases the information obtained with respect to other similar studies in which a single videotape of the arthroscopic view was used^{9,24}. The external camera complements the view of the interior of the joint and allows the observation of relevant information, such as the angle of insertion of the instruments, the trainee's position, and the way that the trainee grasps and handles the instruments and the arthroscope.

The IGARS was designed to be feasible. It takes, on average, five minutes to complete after observing the performance, which is similar to the time taken to complete the other GRSs (three minutes for the Global Rating Scale for Shoulder Arthroscopy (GRSSA)⁹ and between two and five minutes for GRITS¹³).

The number of cases to be assessed, namely, 147 procedures (forty-nine subjects, three procedures each), is higher than in previous validation studies of other GRSs: ten videos for GRSSA⁹, sixty-eight subjects (two procedures each) for the GRS of the Basic Arthroscopic Knee Skill Scoring System²³, and thirty-two procedures for GRITS¹³.

The results of a validation study carried out on a virtual reality shoulder arthroscopy simulator show high values for the Cronbach alpha, indicating that the IGARS is reliable as an assessment tool to measure arthroscopic technical skills. The intra-class correlation coefficient between assessors was 0.91 ($p < 0.001$), which is considerably higher than for the other GRSs (ranging from 0.64 to 0.72 for OSATS¹¹ and 0.89 for GOALS¹⁰), demonstrating a high degree of consistency between assessors. Thus, IGARS has great potential as a means of assessing virtual reality simulator performance, with high internal consistency and inter-rater reliability.

The Orthopaedic Competence Assessment Project (OCAP)²⁵ is a combined initiative of the Education Committee of the British Orthopaedic Association (BOA) and the Specialist Advisory Committee (SAC) in Trauma & Orthopaedics in the United Kingdom that aims to improve training through a competency-based portfolio of coaching and assessment. This is done through the use of performance-based assessments that are formal, structured assessments of clinical activity and focus predominantly on common surgical procedures. However, there are limitations to this approach as it is difficult to stan-

dardize operations, and the initial steep part of the learning curve may potentially put patients at risk.

There is also a substantial cost of training in the operating room and it may be more appropriate to teach generic arthroscopic skills to trainees elsewhere. Simulation allows the training of surgeons in a controlled, repeatable, risk-free environment. The role of virtual reality simulation in surgical training continues to grow, and more appropriate tools for quantifying performance in this environment are required.

The IGARS was able to distinguish subjects as a function of their experience level with an approach that included observation by assessors, while guaranteeing anonymity of trainees and facilitating unbiased assessment independent of time and place.

The limitations to the current study included the non-homogeneous nature of the groups, as, for practical reasons, each group had a different number of participants. In addition, it is important to emphasize that the scores of a subject for a particular test only refer to his or her arthroscopic technical skills for that test. Moreover, the IGARS is constrained to assessing general arthroscopic technical skills. For procedure-specific skills, a complementary assessment tool should be used in conjunction with the IGARS. Lastly, the present study focused on the application of the IGARS to tasks performed on a virtual reality shoulder arthroscopy simulator; therefore, the scale still needs to be tested and validated on the real setting of the operating room.

Future Applications and Studies

The contribution of designing a portable GRS can be used as a model to develop GRSs in other surgical specialties. The versatility of designing a portable GRS that can be used both in the operating room and in simulated environments may be relevant for future research studies involving transferability of skills, as the same assessment instrument could be used in both settings.

In conclusion, the IGARS has the ability to safeguard the anonymity of the surgeon being assessed, thus removing familiarity bias from the assessment. The scale also has a high rate of internal consistency and inter-rater reliability and demonstrated construct validity, being able to distinguish subjects according to their level of experience. The IGARS is thus a promising assessment tool for arthroscopic surgical skills. If further validated in the operating room environment, the IGARS could be used in assessing the progress and competence of the orthopaedic trainees. ■

Note: We thank Jose Manuel Fernandez-Arroyo and Peter Reilly for their valuable insights into arthroscopic education.

Sofia Bayona, PhD, MEd, MAS, BS
Department of Computer Architecture and Technology,
Computing Sciences, and Artificial Intelligence,
Universidad Rey Juan Carlos,
Calle Tulipán s/n,
28933 Móstoles,

Madrid, Spain.

E-mail address: sofia.bayona@urjc.es

Kash Akhtar, FRCS(Tr&Orth), FRSA, MEd, MSc, MBBS
Chinmay Gupte, PhD, FRCS(Tr&Orth), MA(Oxon), BMBCh
Roger J.H. Emery, FRCSE, MS
Alexander L. Dodds, FRCS(Tr&Orth), MSc, BCh, MB

Fernando Bello, PhD, BSc (Hons)
Department of Surgery and Cancer,
Imperial College London,
Fulham Palace Road,
London,
W68RF United Kingdom.
E-mail address for C. Gupte: c.gupte@imperial.ac.uk

References

1. Bates T, Slade D. ASiT abstracts. The impact of the European working time directive on operative exposure. *Ann Royal Coll Surg Engl.* 2007;89(4):7-11.
2. Wilkinson JR, Crossley JG, Wragg A, Mills P, Cowan G, Wade W. Implementing workplace-based assessment across the medical specialties in the United Kingdom. *Med Educ.* 2008 Apr;42(4):364-73.
3. Norcini JJ. Work based assessment. *BMJ.* 2003 Apr 5;326(7392):753-5.
4. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Acad Med.* 2004 Oct;79(10 Suppl):S70-81.
5. Slade Shantz JA, Leiter JR, Gottschalk T, MacDonald PB. The internal validity of arthroscopic simulators and their effectiveness in arthroscopic education. *Knee Surg Sports Traumatol Arthrosc.* 2014 Jan;22(1):33-40. Epub 2012 Oct 2.
6. Aggarwal R, Grantcharov T, Moorthy K, Hance J, Darzi A. A competency-based virtual reality training curriculum for the acquisition of laparoscopic psychomotor skill. *Am J Surg.* 2006 Jan;191(1):128-33.
7. Aggarwal R, Crochet P, Dias A, Misra A, Ziprin P, Darzi A. Development of a virtual reality training curriculum for laparoscopic cholecystectomy. *Br J Surg.* 2009 Sep;96(9):1086-93.
8. Howells NR, Gill HS, Carr AJ, Price AJ, Rees JL. Transferring simulated arthroscopic skills to the operating theatre: a randomised blinded study. *J Bone Joint Surg Br.* 2008 Apr;90(4):494-9.
9. Hoyle AC, Whelton C, Umaar R, Funk L. Validation of a global rating scale for shoulder arthroscopy: A pilot study. *Shoulder & Elbow.* 2012 Jan;4(1):16-21.
10. Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg.* 2005 Jul;190(1):107-13.
11. Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, Brown M. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg.* 1997 Feb;84(2):273-8.
12. Kopta JA. An approach to the evaluation of operative skills. *Surgery.* 1971 Aug;70(2):297-303.
13. Doyle JD, Webber EM, Sidhu RS. A universal global rating scale for the evaluation of technical skills in the operating room. *Am J Surg.* 2007 May;193(5):551-5; discussion 555.
14. Bayona S, Fernandez-Arroyo JM, Martin I, Bayona P. Assessment study of insight ARTHRO VR® arthroscopy virtual training simulator: Face, content, and construct validities. *J Robot Surg.* 2008;2(3):151-8.
15. Angeline ME, Ho SSW. Increased training correlates with performance on a virtual reality simulator for shoulder arthroscopy. Mid-America Orthopaedic Association, Twenty-Seventh Annual Meeting; 2009 Apr 22-26; Florida.
16. Funk L. A technological chip on my shoulder – surgical training with simulators. *J Biomech.* 2010;43:7-8.
17. 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.
18. Linstone HA, Turoff M. The Delphi method: techniques and applications. Reading, MA: Addison-Wesley Pub. Co., Advanced Book Program; 1975.
19. Mohtadi NG, Griffin DR, Pedersen ME, Chan D, Safran MR, Parsons N, Sekiya JK, Kelly BT, Werle JR, Leunig M, McCarthy JC, Martin HD, Byrd JW, Philippon MJ, Martin RL, Guanche CA, Clohisy JC, Sampson TG, Kocher MS, Larson CM; Multi-center Arthroscopy of the Hip Outcomes Research Network. The Development and validation of a self-administered quality-of-life outcome measure for young, active patients with symptomatic hip disease: the International Hip Outcome Tool (iHOT-33). *Arthroscopy.* 2012 May;28(5):595-605, quiz 606-10.e1.
20. Bayona S, Espadero JM, Fernandez JM, Pastor L, Rodriguez A. Implementing virtual reality in the healthcare sector. In: Rao R, editor. *Virtual technologies for business and industrial applications: innovative and synergistic approaches.* 1st ed. Hershey, PA: Business Science Reference; 2010. p. 138-162, 350.
21. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. *J Clin Epidemiol.* 2003 Dec;56(12):1150-6.
22. Palter VN, MacRae HM, Grantcharov TP. Development of an objective evaluation tool to assess technical skill in laparoscopic colorectal surgery: a Delphi methodology. *Am J Surg.* 2011 Feb;201(2):251-9. Epub 2010 Sep 15.
23. Insel A, Carofino B, Leger R, Arciero R, Mazzocca AD. The development of an objective model to assess arthroscopic performance. *J Bone Joint Surg Am.* 2009 Sep;91(9):2287-95.
24. Mohtadi NG, Jager FL, Sasyniuk TM, Hollinshead RM, Fick GH. The reliability between surgeons comparing arthroscopic and video evaluation of patients with shoulder impingement syndrome. *Arthroscopy.* 2004 Dec;20(10):1055-62.
25. Guthrie H, Edwards D, Fetherston C, McMaster J. Orthopaedic competence assessment project deployed on operations. *J Bone Joint Surg Br.* 2011 Jan 01;93(SUPP 1):53.