# The Arthroscopic Surgical Skill Evaluation Tool (ASSET)

**Ryan J. Koehler, MSEd [Medical Student]**,
University of Rochester School of Medicine, Rochester, New York

**Simon Amsdell, MD [Resident]**,
University of Rochester, Department of Orthopaedic Surgery, Rochester, New York

**Elizabeth A Arendt, MD [Professor]**,
University of Minnesota, Department of Orthopaedic Surgery, Minneapolis, Minnesota

**Leslie J Bisson, MD [Clinical Associate Professor]**,
University of Buffalo, Department of Orthopaedic Surgery, Buffalo, NY

**Jonathan P Braman, MD [Assistant Professor]**,
University of Minnesota, Department of Orthopaedic Surgery, Minneapolis, Minnesota

**Aaron Butler, BS [Medical Student]**,
University of Rochester School of Medicine, Rochester, New York

**Andrew J Cosgarea, MD [Professor]**,
Johns Hopkins, Department of Orthopaedic Surgery, Baltimore, Maryland

**Christopher D Harner, MD [Professor]**,
University of Pittsburgh, Department of Orthopaedic Surgery, Pittsburgh, Pennsylvania

**William E Garrett, MD [Professor]**,
Duke University, Department of Orthopaedic Surgery, Durham, North Carolina

**Tyson Olson, BS [Medical Student]**,
University of Rochester School of Medicine, Rochester, New York

**Winston J. Warme, MD [Associate Professor]**, and
University of Washington, Department of Orthopaedic Surgery, Seattle, Washington

**Gregg T. Nicandri, MD [Assistant Professor]**
University of Rochester, Department of Orthopaedic Surgery, Rochester, New York

## Abstract

**Background—**Surgeries employing arthroscopic techniques are among the most commonly performed in orthopaedic clinical practice however, valid and reliable methods of assessing the arthroscopic skill of orthopaedic surgeons are lacking.

Corresponding Author: Gregg Nicandri, Department of Orthopaedic Surgery, University of Rochester, 601 Elmwood Ave., Box 665, Rochester, NY 14642 Gregg_Nicandri@urmc.rochester.edu.

**Hypothesis**—The Arthroscopic Surgery Skill Evaluation Tool (ASSET) will demonstrate content validity, concurrent criterion-oriented validity, and reliability, when used to assess the technical ability of surgeons performing diagnostic knee arthroscopy on cadaveric specimens.

**Study Design**—Cross-sectional study; Level of evidence, 3

**Methods**—Content validity was determined by a group of seven experts using a Delphi process. Intra-articular performance of a right and left diagnostic knee arthroscopy was recorded for twenty-eight residents and two sports medicine fellowship trained attending surgeons. Subject performance was assessed by two blinded raters using the ASSET. Concurrent criterion-oriented validity, inter-rater reliability, and test-retest reliability were evaluated.

**Results**—Content validity: The content development group identified 8 arthroscopic skill domains to evaluate using the ASSET. Concurrent criterion-oriented validity: Significant differences in total ASSET score ($p < 0.05$) between novice, intermediate, and advanced experience groups were identified. Inter-rater reliability: The ASSET scores assigned by each rater were strongly correlated ($r = 0.91$, $p < 0.01$) and the intra-class correlation coefficient between raters for the total ASSET score was 0.90. Test-retest reliability: there was a significant correlation between ASSET scores for both procedures attempted by each individual ($r = 0.79$, $p < 0.01$).

**Conclusion**—The ASSET appears to be a useful, valid, and reliable method for assessing surgeon performance of diagnostic knee arthroscopy in cadaveric specimens. Studies are ongoing to determine its generalizability to other procedures as well as to the live OR and other simulated environments.

## Introduction

The Accreditation Council for Graduate Medical Education (ACGME) recently announced the Next Accreditation System. This new system expands upon the six core competencies and will require programs to demonstrate that residents are progressing toward proficiency by attaining milestones in general and specialty-specific areas throughout training[26]. In orthopaedic surgery, three of the proposed 16 specialty-specific milestones areas are managed primarily using arthroscopic techniques (ACL injury, meniscal injury, and rotator cuff injury). Knowledge of a procedure does not always equate to the ability to successfully perform that procedure and we feel that determining a surgeon's overall proficiency should include an assessment of technical skill. This assessment must be reliable and valid, it should be unbiased, and it should be feasible for programs to administer at little additional cost[3, 25].

In several surgical specialties, global rating scales have been developed to assess surgical skill[3]. The Objective Structured Assessment of Technical Skill (OSATS) global rating scale has been used to evaluate open procedural skills in general surgery[23]. Due to the different set of skills required to perform minimally invasive procedures, similar global rating scales have been created and validated for laparoscopic and endoscopic procedures[1, 2, 15, 33, 35].

Insel et al. utilized a global rating scale and task-specific checklist in their Basic Arthroscopic Knee Skill Scoring System which was created to assess surgeon performance of arthroscopic partial meniscectomy in cadaveric knees[21]. This assessment tool demonstrated construct validity, but their use of a single, non-blind observer did not allow for reliability testing[21]. In a similar process, Hoyle et al. created the Global Rating Scale for Shoulder Arthroscopy (GRSSA) to assess video recordings of diagnostic shoulder arthroscopy performed on live patients[18]. The GRSSA demonstrated construct validity but poor inter-rater reliability[18]. Due to their design, both of these global rating scales lack generalizability in that they can only be used to assess partial meniscectomy and shoulder arthroscopy respectively.

The Arthroscopic Surgery Skill Evaluation Tool (ASSET) was created to be used as an assessment of global arthroscopic technical skill. It was designed to be generalizable to multiple arthroscopic procedures as well as both the live OR and simulated environments. The ASSET includes eight domains that can be evaluated using intraoperative procedural video recorded through the arthroscopic camera. The purpose of this study was twofold; (1) to describe the development and content validation of the ASSET and (2) to evaluate its validity and reliability. Our hypothesis was that the ASSET will demonstrate content validity, concurrent criterion-oriented validity, inter-rater reliability, and test-retest reliability, when used to assess the technical ability of surgeons performing diagnostic knee arthroscopy on cadaveric specimens in the surgical skills laboratory.

## Materials and Methods

### Development of the ASSET Global Rating Scale

This study was approved by our institutional review board. A modified Delphi method was used to develop the content of the ASSET global rating scale[13, 36]. Content development group members were recruited by the principle investigator, using personal contacts known to have an interest in arthroscopic skills education. An e-mail explaining the study and the role of each subject was sent to 8 experts at 7 orthopaedic residency programs. Subjects were informed of the estimated time commitment for the project and offered a $100 gift card as compensation for participation. Of the 8 experts contacted, 7 responded that they wished to participate in this study.

A conference call was held in October 2011 to introduce the study and provide pertinent background information. The content development group was asked to reach a consensus regarding basic criteria to which an ideal arthroscopic evaluation tool would adhere. An initial version of the ASSET was developed based on that criteria following an extensive review of the surgical education literature and preliminary testing. Similar to global rating scales validated in other surgical fields, a 5-point Likert-type scale with descriptors at 1, 3, and 5 was found to be the most practical[17, 21, 23, 33, 35]. However, rather than assigning arbitrary numbers to the descriptors, the ASSET was designed to correspond to the levels of the Dreyfus model of skill acquisition with "1" representing the novice level, "3" representing the competent level, and "5" representing the expert level of arthroscopic skill performance [6, 8]. It was thought that this would help improve reliability and would allow for its use within the context of an already accepted model of professional skill acquisition.

Group members were provided with the initial version of the ASSET and given access to intraoperative video examples of diagnostic knee arthroscopy. Group members were instructed to use the ASSET to rate the videos and complete an online survey (SurveyMonkey.com, Palo Alto, CA) that asked them to identify any omitted domains which they considered important to include on an assessment of arthroscopic skill. Each member was also asked to make any necessary changes to the domains contained on the initial ASSET. Survey and email responses were collected for 2 weeks, summarized by the principle investigators, and presented anonymously back to the group. A subsequent conference call was held to ensure that member opinions were being correctly interpreted by all. Based on group suggestions, a new version of the ASSET was drafted and sent to all members for review and critique. After 30 days, suggestions were reviewed by the principal investigators and incorporated into a final version of the ASSET that was presented to the group for a final vote.

### Development of the ASSET Task Specific Checklist for Diagnostic Knee Arthroscopy

A checklist for diagnostic arthroscopy of the knee was also created using the Delphi method. Since diagnostic arthroscopy of the knee may be performed in a variety of ways it was felt that this checklist was necessary to standardize the procedure being evaluated by the ASSET. The checklist was not designed to contain all steps that constitute a complete diagnostic arthroscopy of the joint. Rather, the checklist was created to determine the minimum acceptable set of steps that must be included for a video-based competency assessment of the procedure using the ASSET. For this project, diagnostic arthroscopy of the knee was selected because it represents one of the first arthroscopic procedures performed by residents. It is expected that task specific checklists will be created so that ASSET may be used in the evaluation of more complex surgical procedures such as menisectomy, anterior cruciate ligament reconstruction, and rotator cuff repair.

In October 2011, group members were asked to deconstruct diagnostic arthroscopy of the knee listing all steps they considered essential using an online survey (SurveyMonkey.com, Palo Alto, CA). Responses were collected for 2 weeks and used to create a master list of all steps. A second online survey was then sent that instructed group members to select the steps from the master list that they considered essential to include for video-based assessment of diagnostic knee arthroscopy. Group members were given 30 days to review the list and vote. All steps that achieved 100% group consensus were removed from further analysis. Steps that did not achieve 100% agreement were presented back to the group along with the majority opinion. Group members were given 30 days to provide their opinion for or against the remaining steps. Following this, the procedure checklist was created from items that >80% of respondents considered essential and were modified based on the feedback provided. A final yes/no vote was taken on the checklist in its entirety.

### Establishing the Validity and Reliability of ASSET

A recruitment email was sent to all orthopaedic surgery residents at our institution describing the research study. Residents were asked to self-enroll and complete an online survey (Survey Monkey, Palo Alto, CA) which asked them to report the number of knee arthroscopies recorded in their ACGME case log and to estimate the number of prior

arthroscopic knee procedures in which they had actively performed a portion of the procedure. The postgraduate year of each resident was obtained from the roster kept by the department. Two orthopaedic faculty members were also recruited through email and participated in this study. Prior to entering the skills lab, all subjects viewed an intraoperative recording of an orthopaedic faculty member performing a diagnostic arthroscopy of the knee. After watching the video, subjects were assigned to begin on either a right or left cadaveric specimen using alternative assignment based on the date of participation. Each subject was asked to complete all of the tasks outlined by the content development group's previously agreed upon task specific checklist for diagnostic knee arthroscopy. A cheat sheet with brief verbal directions for each task was fixed to the arthroscopic monitor for subjects to reference during the procedure. A trained observer provided assistance to each subject by placing varus and valgus stress on the knee. Verbal instructions from the observer were limited to the instructions provided on the cheat sheet. Upon completing the lab and demographics survey, subjects were provided with a $10 gift card as compensation. Video of each subject performing the diagnostic knee arthroscopy was recorded directly through the arthroscopic camera. No views external to the joint or audio was recorded to prevent identification of any subject from the video recordings. Each recording began when the patellofemoral joint could be visualized and continued until all tasks had been performed or a reasonable amount of time had been given (25 minutes). All recorded videos were transferred from the hard drive connected to the arthroscopic tower and assigned a random identification number for subsequent review and rating. All videos (n=60) were independently reviewed by two raters blinded to subject identity. Both raters were involved in the content validation of the assessment tool, and both had prior experience using a global rating scale to assess arthroscopic skill.

## Statistical Methods

The validity of the ASSET was measured in several ways. Concurrent validity was assessed using Pearson's correlation coefficient (r) to determine the relationship between arthroscopic experience (PGY, ACGME weblog, estimated number of procedures performed) and total ASSET scores. Concurrent criterion-oriented validity was assessed by conducting a one-way analysis of variance for ASSET score with 3 levels of training; novice (PGY1-2), intermediate (PGY 3-4), advanced (PGY 5- Attending). If significant differences were observed, post-hoc testing was done using Student-Newman-Keuls test for all pairwise comparisons. The level of significance for all statistical tests was set at $p<0.05$.

The reliability of the ASSET was also measured in several ways. The internal consistency of the ASSET domains were assessed using Cronbach's alpha. The inter-rater reliability of the total ASSET score was determined using the intra-class correlation coefficient (ICC) for absolute agreement between a fixed, non-random set of raters. Inter-rater reliability on each ASSET domain was assessed using ICC. The test-retest reliability was assessed by comparing the correlation of the ASSET scores given for the first and second arthroscopy performed in the skills lab.

Descriptive statistics were used to report subject demographic variables.

# Results

## Content Validity

The content development group felt that an ideal arthroscopic evaluation tool would demonstrate validity and reliability, would be unbiased, would be practical and simple to administer, would assess both live and simulated surgery, and would be generalizable to multiple procedures and settings (OR and simulation lab). After completing the modified Delphi process the group unanimously approved the ASSET. This version included 8 domains for assessment with a ninth domain, "additional complexity of procedure", being included as a control measure (Figure 1). Each domain was designed to assess a unique facet of arthroscopic skill acquisition and every attempt was made to limit overlap. See appendix 1 for a detailed description of each domain. The total score of the ASSET is the sum of the 8 domains with a maximum of 38. It was the consensus of the group that for an individual to be considered competent for the technical portion of the procedure being assessed, they must achieve a minimum score of "3" in each of the 8 domains being assessed. The group also approved a task specific checklist for diagnostic knee arthroscopy (Figure 2).

## Validity of the ASSET

ASSET scores increased with level of training (Figure 3). The mean of the ASSET scores given for the first arthroscopy performed correlated with subject year in training (r=0.83, p<0.01) and the number ACGME knee arthroscopies reported (r=0.76, p<0.01). One-way analysis of variance revealed that there was a significant difference in the mean total ASSET score across the levels of training for both the first (p<0.01) and second (p<0.01) arthroscopy performed in the lab. Post-hoc, pairwise comparison of ASSET scores assigned for subjects' first arthroscopy performed in lab found that there was a significant difference between rater scores for each of the three levels of training (p<0.05) demonstrating concurrent criterion-oriented validity (Figure 4).

## Reliability of the ASSET

The inter-rater reliability of the total ASSET score was assessed for all videos rated (n=60) using ICC and found to be 0.90. There was no significant difference in the mean total ASSET scores assigned by each rater (p=0.93). Comparison of mean ASSET scores given for procedures performed on the left versus the right knee were not significantly different for rater 1 (p=0.50) or rater 2 (p=0.71). There was a positive correlation between rater scores that was significant (r=0.91, p<0.01). Cronbach's α demonstrated that the domains of the ASSET had good internal consistency for both raters (α=0.94). The inter-rater reliability was ICC = 0.75-0.87 for all ASSET domains except safety (κ= 0.52) (Table 1). There was a significant positive correlation between each subject's mean ASSET score for the first and second arthroscopy performed (r = 0.79, p<0.01) demonstrating test-retest reliability.

# Discussion

Over the last few decades there has been a significant shift towards competency based graduate medical education[5, 28]. In the future, valid and reliable assessments of technical skill are likely to play an increasingly important role in the evaluation of surgical

proficiency and the ASSET was designed specifically for this use. To be useful these assessment tools must be valid and reliable.

Our results indicate that the ASSET is a feasible, valid, and reliable assessment of diagnostic knee arthroscopy skill in the simulation lab. This tool does however, have several limitations. First, The ASSET was designed as a video-based assessment. Others have developed similar tools which employ live raters and argue that some important procedural skills may not be easily evaluated or well seen when using video. Our content development group however, felt that a video-based assessment offered significant advantages over using live raters. A video-based assessment tool facilitates unbiased assessment as scoring can take place out of the presence of and without identification of the examinee. This enables the evaluations to be done at any time or place and therefore decreases the burden on the rater. In addition the rater is able to pause or rewind the videotape during scoring which may lead to more accurate assessment and could improve reliability which has been poor in previously described arthroscopic assessments[1, 17, 19, 21]. Second, the ASSET is restricted in that only the intra-articular portion of the case can be evaluated. It was felt by the content development group that the tool should require only the use of inexpensive or readily available equipment and should not significantly interfere with standard operative protocol as the intent was for it to enable evaluation both in the simulated and live OR environments. Since the arthroscopic camera and video capture equipment is available in virtually all arthroscopic cases, we elected to record and evaluate only those portions of cases that are visualized by the arthroscopic camera. The extra-articular portions of a procedure such as portal placement or graft harvest must therefore be evaluated using other methods. An external camera may allow for adequate evaluation of these important skills, however its use may be more limited to assessment in the simulated environment where the position and type of camera can be standardized without affecting the surgical procedure. Third, we specifically developed the domains of the ASSET to allow for the assessment of multiple arthroscopic procedures. As a result the domains and weighted descriptors used were not specific to any procedure. We do recognize that this may affect the validity and utility of the assessment for some individual arthroscopic procedures. However, a similar assessment tool used for assessing laparoscopic skill, the Global Objective Assessment of Laparoscopic Skills (GOALS), has been shown to be valid for assessing technical skill during multiple procedures[15]. Because the process of rigorous validation is extensive, it was felt that one tool which could be used for multiple procedures would be preferable to tools developed and validated for each individual procedure. Fourth, though every effort was made to decrease ambiguity and increase reliability, the ASSET still remains a subjective evaluation. This is similar to other methods of assessment in surgery like the OSATS, GOALS, and ABOS step II. Research is ongoing regarding more objective measures of assessment and we feel that ultimately these types of objective assessment, combined with tools like the ASSET will provide the best evidence of a surgeon's competency[9, 12, 16, 24, 25, 30, 31]. Fifth, the domain safety demonstrated a reliability of ICC = 0.52 which is considered moderate by Landis and Koch criteria (Table 2)[22]. All other domains of ASSET demonstrated significantly higher levels or inter-rater agreement. Though this number is adequate, it does suggest some modification to this domain may improve the overall reliability of the ASSET. Safety was particularly hard to assess using this study design as the same cadaveric specimens were

used for all subjects and it was difficult to determine whether minor cartilage scuffing was pre-existing or was caused during that trial. This may therefore be easier to assess with less used specimens or in the live operating room setting.

A modified Delphi process similar to that employed in the development of other surgical assessments was utilized to determine the content of the ASSET[14, 27, 29]. Using this technique, the content development group was able to achieve a consensus opinion regarding the basic criteria to which an ideal arthroscopic skill evaluation would adhere as well as which specific domains of arthroscopic skill would be assessed by the ASSET. The convenience of using this method allowed for the entire process to be conducted electronically with experts from different geographic areas and at minimal cost. Multiple versions of the assessment were tested until the final version was created with a mix of domains and weighted descriptors that were felt to be clear and objective. The domains chosen are similar to those utilized by other global assessments of surgical skill but are specific to arthroscopy[23, 33, 35]. We felt that some of the previously proposed arthroscopic assessment tools had multiple domains that assessed similar skills which may place too much weight on a particular skill and could lead to confusion by the raters, effecting reliability. When developing the ASSET, every attempt was made to eliminate redundancy and maximize rater reliability. Other similar assessments of technical skill have included cognitive domains such as "knowledge of instruments" or "knowledge of procedure"[21, 23]. Though these domains are important, it was felt by the content development group that these could be better assessed by other methods and fell outside the realm of technical skill. The content of the ASSET is unique from other assessments in that it is meant to easily allow for unbiased (blinded) assessment, be practical and simple to administer, assess both live and simulated surgery, and be generalizable to multiple procedures and settings (OR and simulation lab). We are optimistic that this generalizability will enable collaboration between institutions, allow for the establishment of benchmarks for the attainment of procedural skills, and enable investigators to measure the effect that simulators and other training methodologies have on the acquisition of surgical skill.

The concurrent criterion-oriented validity of the ASSET was established utilizing methods similar to that of other assessments of surgical skill. Similar to the OSATS global rating scale which is widely used in general surgery to assess open and laparoscopic surgical skills as well as the GOALS which is used by the American College of Surgeons to assess their Fundamentals of Laparoscopic Surgery program, the ASSET was able to demonstrate that individuals with more surgical experience achieved higher scores on average than individuals with less surgical experience[23, 33].

An important measure of the reliability of an assessment is the agreement between different raters when evaluating the same subject. Without adequate inter-rater reliability, it becomes difficult to compare results when different raters are utilized. The ICC of the ASSET global rating scale matched or exceeded that which has been reported for other previously validated global assessments of surgical skill(Table 3)[23, 32-34]. We believe that this result was due to the fact that the domains and descriptors for the ASSET were chosen with the understanding that the reliability of the instrument was of significant importance. In addition, all raters had a thorough understanding of how to utilize the ASSET as well as practice with this and other

similar tools for the evaluation of surgical skill using intra-operative video. We do feel that an established rater training protocol will be necessary to maximize the reliability for ASSET between all raters and will therefore increase its generalizability. Another important measure of reliability is test-re-test reliability indicating that an individual will achieve a similar score when assessed a second time. The scores did correlate between an individual's first and second trial with an r=.79. We observed higher ASSET scores for the second knee arthroscopy when compared to the first. The increase in scores was greatest for subjects in PGY 1-3 (Figure 5) who have limited exposure to arthroscopy at our program. We believe this increase in scores most likely represents improved technical performance for this relatively basic procedure as this was not observed for the more experienced subjects.

It is currently unclear how best to use a skill assessment like the ASSET in demonstrating attainment of the ACGME core competencies. The Dreyfus model of skill acquisition describes developmental stages starting with novice and proceeding to advanced beginner, competent, proficient, and expert [10]. The trainee is assumed to progress along the continuum of the model with professional education and through the process of deliberate practice [11]. Some authors have attempted to describe this and similar models within the context of medical education and provide a picture of how this might be adapted to orthopaedic technical skill assessment[4, 7, 8, 20]. In an effort to allow some standardization within the overall assessment of surgical competency the ASSET was designed to follow this rubric (Table 4)[6, 8, 10]. It must be stated that controversy exists regarding how best to apply the Dreyfus model to medical education and further work is necessary in this area. It is also important to understand that the achievement of a competent score on the ASSET by an individual does not indicate that the individual is "competent" to perform the procedure being tested. It solely indicates that the examinee demonstrated a competent level of technical skill for that particular test and procedure. Assessment of overall surgical competence would require an integrative assessment of all facets of the procedure of which technical skill is only one component. Our hope is that the ASSET can be used in conjunction with other methods of competency evaluation to provide a clearer picture of the overall competence of the individual surgeon.

In conclusion, the ASSET appears to be a useful, valid, and reliable method for assessing surgeon performance of diagnostic knee arthroscopy in cadaveric specimens. Further study of the ASSET is currently underway to determine other measures of validity and reliability, as well as the feasibility of utilizing it to assess the technical skill of surgeons performing multiple procedures in both the operating room and simulation lab. Additional study is also needed to determine the role of technical skill evaluations within the overall context of the ACGME competency of patient care.

## References

1. Aggarwal R, Grantcharov T, Moorthy K, Milland T, Darzi A. Toward feasible, valid, and reliable video-based assessments of technical surgical skills in the operating room. Ann Surg. 2008; 247(2): 372–379. [PubMed: 18216547]
2. Aggarwal R, Grantcharov T, Moorthy K, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. Ann Surg. 2007; 245(6):992–999. [PubMed: 17522527]

3. Ahmed K, Miskovic D, Darzi A, Athanasiou T, Hanna GB. Observational tools for assessment of procedural skills: a systematic review. Am J Surg. 2011; 202(4):469–480. e466. [PubMed: 21798511]

4. Alderson D. Developing expertise in surgery. Med Teach. 2010; 32(10):830–836. [PubMed: 20854159]

5. Atesok K, Mabrey JD, Jazrawi LM, Egol KA. Surgical simulation in orthopaedic skills training. J Am Acad Orthop Surg. 2012; 20(7):410–422. [PubMed: 22751160]

6. Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. Health Aff (Millwood). 2002; 21(5):103–111. [PubMed: 12224871]

7. Breckwoldt J, Klemstein S, Brunne B, Schnitzer L, Arntz HR, Mochmann HC. Expertise in prehospital endotracheal intubation by emergency medicine physicians-Comparing 'proficient performers' and 'experts'. Resuscitation. 2011

8. 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):761–767. [PubMed: 18667892]

9. Danila R, Gerdes B, Ulrike H, Dominguez Fernandez E, Hassan I. Objective evaluation of minimally invasive surgical skills for transplantation. Surgeons using a virtual reality simulator. Chirurgia (Bucur). 2009; 104(2):181–185. [PubMed: 19499661]

10. Dreyfus, HL.; Dreyfus, SE.; Athanasiou, T. Mind over machine : the power of human intuition and expertise in the era of the computer. Oxford, UK: B. Blackwell; 1986.

11. Ericsson KA. Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med. 2004; 79(10 Suppl):S70–81. [PubMed: 15383395]

12. Escoto A, Trejos AL, Naish MD, Patel RV, Lebel ME. Force sensing-based simulator for arthroscopic skills assessment in orthopaedic knee surgery. Stud Health Technol Inform. 2012; 173:129–135. [PubMed: 22356973]

13. Fink A, Kosecoff J, Chassin M, Brook RH. Consensus methods: characteristics and guidelines for use. Am J Public Health. 1984; 74(9):979–983. [PubMed: 6380323]

14. Graham B, Regehr G, Wright JG. Delphi as a method to establish consensus for diagnostic criteria. J Clin Epidemiol. 2003; 56(12):1150–1156. [PubMed: 14680664]

15. Gumbs AA, Hogle NJ, Fowler DL. Evaluation of resident laparoscopic performance using global operative assessment of laparoscopic skills. J Am Coll Surg. 2007; 204(2):308–313. [PubMed: 17254935]

16. Howells NR, Brinsden MD, Gill RS, Carr AJ, Rees JL. Motion analysis: a validated method for showing skill levels in arthroscopy. Arthroscopy. 2008; 24(3):335–342. [PubMed: 18308187]

17. 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; 90(4):494–499. [PubMed: 18378926]

18. Hoyle AC, W C, Umaar R, Funk L. Validation of a global rating scale for shoulder arthroscopy: a pilot study. Shoulder and Elbow. 2012; 4:16–21.

19. Hoyle AC, Whelton C, Umaar R, Funk L. Validation of a global rating scale for shoulder arthroscopy: a pilot study. Shoulder & Elbow. 2012; 4(1):16–21.

20. Hsu DC, Macias CG. Rubric evaluation of pediatric emergency medicine fellows. J Grad Med Educ. 2010; 2(4):523–529. [PubMed: 22132272]

21. 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; 91(9):2287–2295. [PubMed: 19724008]

22. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1):159–174. [PubMed: 843571]

23. Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg. 1997; 84(2):273–278. [PubMed: 9052454]

24. Megali G, Sinigaglia S, Tonet O, Dario P. Modelling and evaluation of surgical performance using hidden Markov models. IEEE Trans Biomed Eng. 2006; 53(10):1911–1919. [PubMed: 17019854]

25. Memon MA, Brigden D, Subramanya MS, Memon B. Assessing the surgeon's technical skills: analysis of the available tools. Acad Med. 2010; 85(5):869–880. [PubMed: 20520044]

26. Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system--rationale and benefits. N Engl J Med. 2012; 366(11):1051–1056. [PubMed: 22356262]

27. 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; 201(2): 251–259. [PubMed: 20832048]

28. Pedowitz RA, Marsh JL. Motor skills training in orthopaedic surgery: a paradigm shift toward a simulation-based educational curriculum. J Am Acad Orthop Surg. 2012; 20(7):407–409. [PubMed: 22751159]

29. Peyre SE, Peyre CG, Hagen JA, et al. Laparoscopic Nissen fundoplication assessment: task analysis as a model for the development of a procedural checklist. Surg Endosc. 2009; 23(6): 1227–1232. [PubMed: 19057949]

30. Tanoue K, Uemura M, Kenmotsu H, et al. Skills assessment using a virtual reality simulator, LapSim, after training to develop fundamental skills for endoscopic surgery. Minim Invasive Ther Allied Technol. 2010; 19(1):24–29. [PubMed: 20095894]

31. Tashiro Y, Miura H, Nakanishi Y, Okazaki K, Iwamoto Y. Evaluation of skills in arthroscopic training based on trajectory and force data. Clin Orthop Relat Res. 2009; 467(2):546–552. [PubMed: 18791774]

32. Van Heest A, Putnam M, Agel J, Shanedling J, McPherson S, Schmitz C. Assessment of technical skills of orthopaedic surgery residents performing open carpal tunnel release surgery. J Bone Joint Surg Am. 2009; 91(12):2811–2817. [PubMed: 19952242]

33. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg. 2005; 190(1):107–113. [PubMed: 15972181]

34. Vassiliou MC, Feldman LS, Fraser SA, et al. Evaluating intraoperative laparoscopic skill: direct observation versus blinded videotaped performances. Surg Innov. 2007; 14(3):211–216. [PubMed: 17928621]

35. Vassiliou MC, Kaneva PA, Poulose BK, et al. Global Assessment of Gastrointestinal Endoscopic Skills (GAGES): a valid measurement tool for technical skills in flexible endoscopy. Surg Endosc. 2010; 24(8):1834–1841. [PubMed: 20112113]

36. Williams PL, Webb C. The Delphi technique: a methodological discussion. J Adv Nurs. 1994; 19(1):180–186. [PubMed: 8138622]

| | 1 – Novice | 2 | 3 - Competent | 4 | 5- Expert |
|---|---|---|---|---|---|
| Safety | Significant damage to articular cartilage or soft tissue | | Insignificant damage to articular cartilage or soft tissue | | No damage to articular cartilage or soft tissue |
| Field of View | Narrow field of view, inadequate arthroscope or light source positioning | 2 | Moderate field of view, adequate arthroscope and light source positioning | 4 | Expansive field of view, optimal arthroscope and light source positioning |
| Camera Dexterity | Awkward or graceless movements, fails to keep camera centered and correctly oriented | 2 | Appropriate use of camera, occasionally needs to reposition | 4 | Graceful and dexterous throughout procedure with camera always centered and correctly oriented |
| Instrument Dexterity | Overly tentative or awkward with instruments, unable to consistently direct instruments to targets | 2 | Careful, controlled use of instruments, occasionally misses targets | 4 | Confident and accurate use of all instruments |
| Bi-Manual Dexterity | Unable to use both hands or no coordination between hands | 2 | Uses both hands but occasionally fails to coordinate movement of camera and instruments | 4 | Uses both hands to coordinate camera and instrument positioning for optimal performance |
| Flow of Procedure | Frequently stops operating or persists without progress, multiple unsuccessful attempts prior to completing tasks | 2 | Steady progression of operative procedure with few unsuccessful attempts prior to completing tasks | 4 | Obviously planned course of procedure, fluid transition from one task to the next with no unsuccessful attempts |
| Quality of Procedure | Inadequate or incomplete final product | 2 | Adequate final product with only minor flaws that do not require correction | 4 | Optimal final product with no flaws |
| Autonomy | 1 | | 2 | | 3 |
| | Unable to complete procedure even with intervention(s) | | Able to complete procedure but required intervention(s) | | Able to complete procedure without intervention |

**Added Complexity of Procedure**

| 1 | 2 | 3 |
|---|---|---|
| No difficulty | Moderate difficulty (mild inflammation or scarring) | Extreme difficulty (severe inflammation or scarring, abnormal anatomy) |

**Figure 1.**
ASSET Global Rating Scale

| Task Checklist - Diagnostic Arthroscopy (Knee) | Completed? | |
|---|---|---|
| Inspect suprapatellar pouch | ▢Yes | ▢No |
| Evaluate patellofemoral articulation | ▢Yes | ▢No |
| Evaluate patella (medial/lateral and inferior/superior) | ▢Yes | ▢No |
| Inspect the lateral gutter | ▢Yes | ▢No |
| Inspect the popliteus tendon and recess | ▢Yes | ▢No |
| Inspect medial gutter | ▢Yes | ▢No |
| Inspect and probe medial femoral condyle | ▢Yes | ▢No |
| Inspect and probe medial tibial plateau | ▢Yes | ▢No |
| Inspect and probe anterior, middle and posterior medial meniscus | ▢Yes | ▢No |
| Inspect and probe ACL/PCL/Ligamentum | ▢Yes | ▢No |
| Inspect and probe lateral femoral condyle | ▢Yes | ▢No |
| Inspect and probe lateral tibial plateau | ▢Yes | ▢No |
| Inspect and probe anterior, middle and posterior lateral meniscus | ▢Yes | ▢No |
| Evaluate passive tracking of patella in trochlear groove | ▢Yes | ▢No |

**Figure 2.**
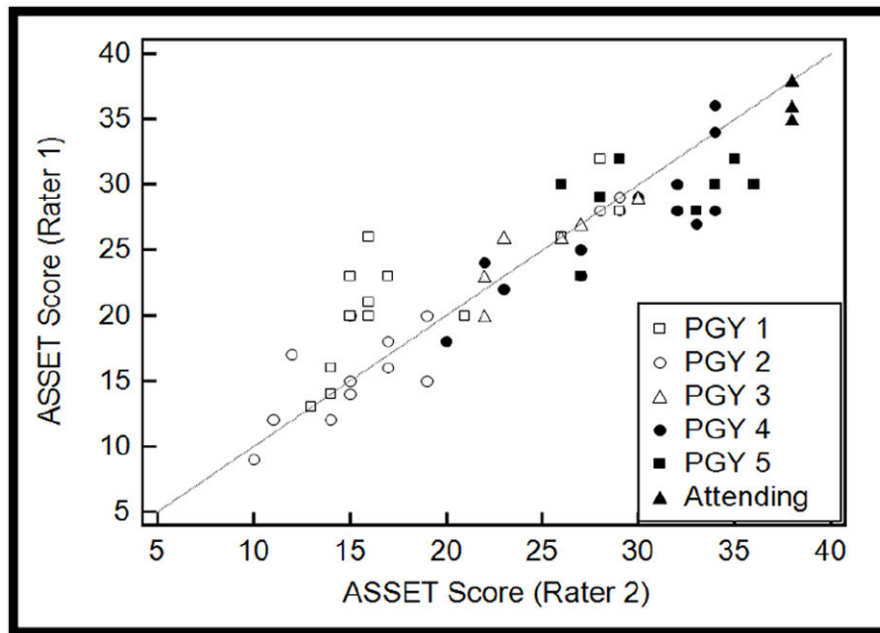Task Specific Checklist for Diagnostic Arthroscopy of the Knee

**Figure 3.**
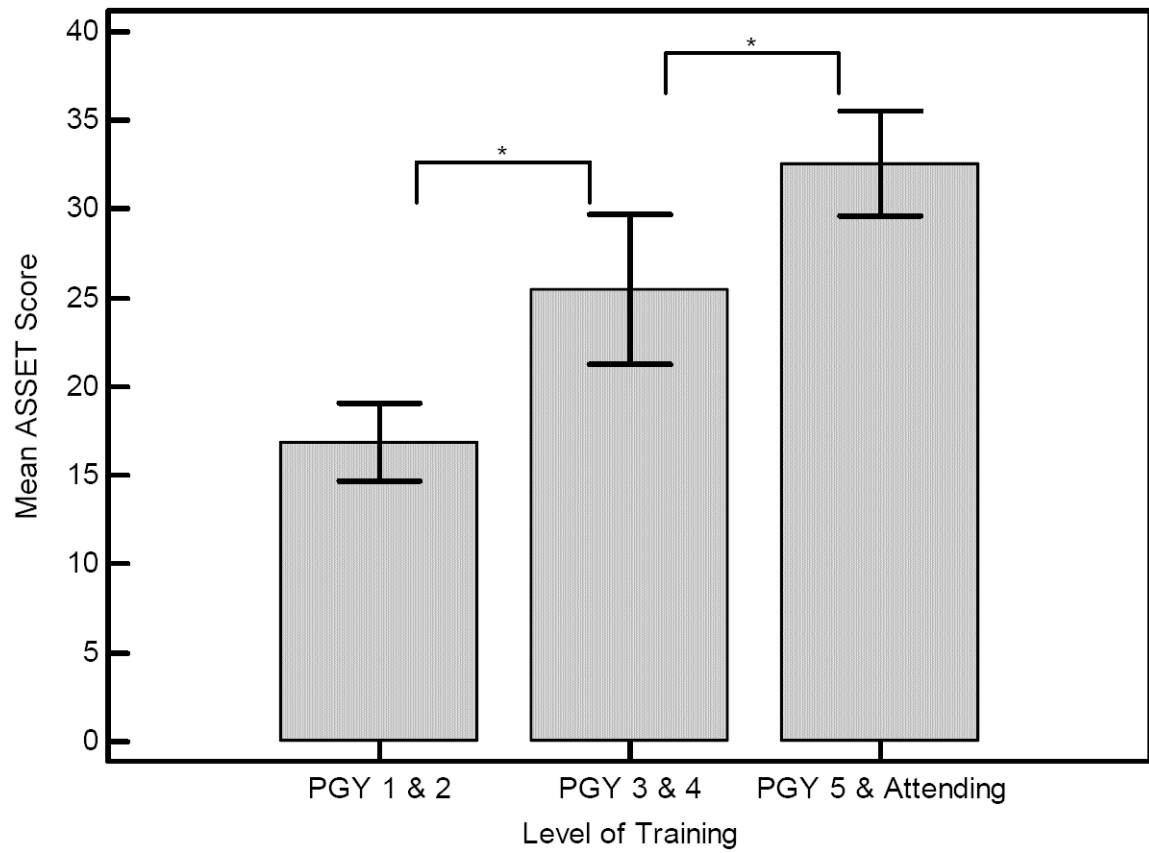ASSET Scores Increase with Level of Training

**Figure 4. Mean Total ASSET Scores Assigned for Each Level of Experience on First Arthroscopy Performed**

I-bars represent the 95% confidence interval of the mean.

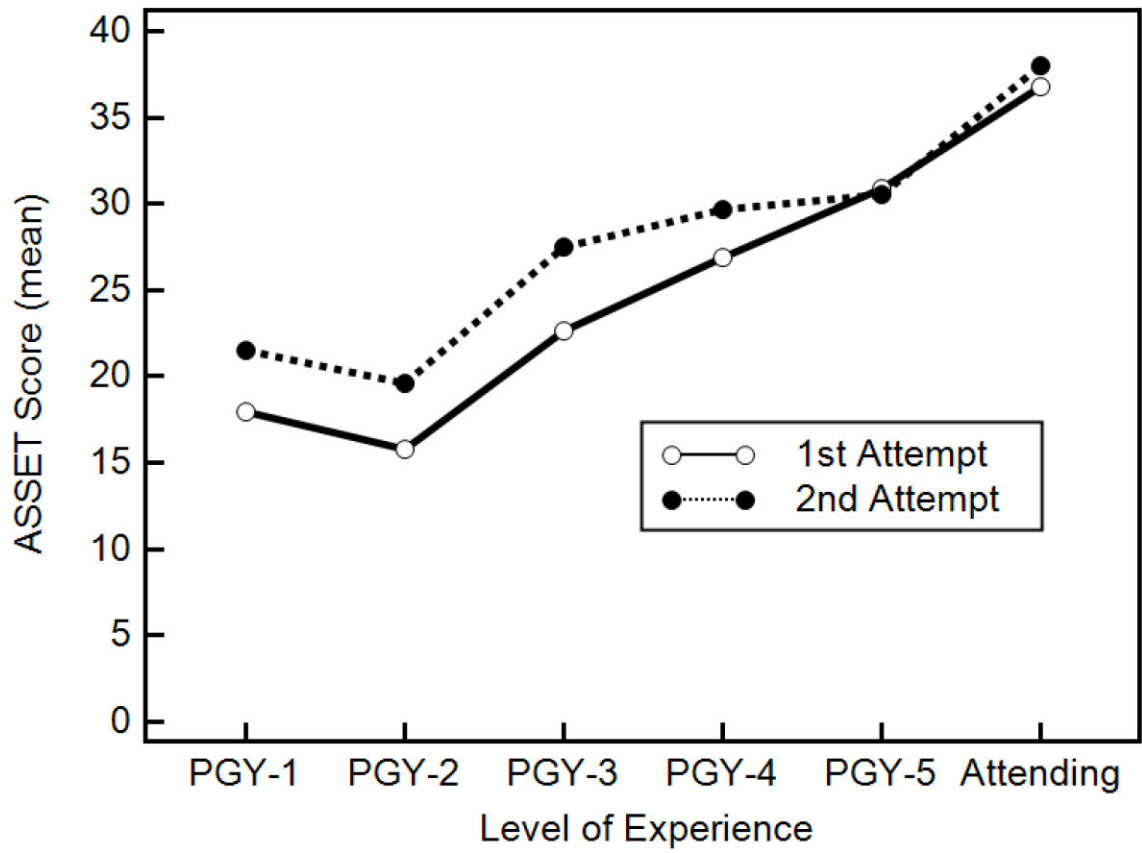* A significant difference was observed between experience levels (p<0.05).

**Figure 5.**
Comparison of Mean ASSET Score for First and Second Trial and Level of Training.

**Table 1**

Inter-rater Reliability for Each ASSET Domain

| ASSET Domain | ICC* |
|---|---|
| Safety | 0.52 (0.23 – 0.81) |
| Field of View | 0.78 (0.70 – 0.86) |
| Camera Dexterity | 0.75 (0.65 – 0.86) |
| Instrument Dexterity | 0.87 (0.82 – 0.92) |
| Bimanual Dexterity | 0.84 (0.76 – 0.91) |
| Flow of Procedure | 0.87 (0.80 – 0.94) |
| Quality of Procedure | 0.82 (0.74 – 0.90) |
| Autonomy | 1.00 |

*
Inter Class Correlation Coefficient (95%CI)

**Table 2**

Landis and Koch Levels of Reliability

| ICC Value | Level of Reliability |
|-----------|----------------------|
| <0.00     | Poor                 |
| 0.00-0.20 | Slight               |
| 0.21-0.40 | Fair                 |
| 0.41-0.60 | Moderate             |
| 0.61-0.80 | Substantial          |
| 0.81-1.00 | Almost Perfect       |

**Table 3**

Comparison of ASSET Reliability to Other Similar Global Rating Scales

| Assessment Tool | n | Raters | ICC# |
|---|---|---|---|
| ASSET (Diagnostic Knee Arthroscopy) | 60 videos | 2 | 0.90 |
| OSATS* Global Rating Scale (Basic Skill) | 12 stations | 2 | 0.70 to 0.72 |
| OSATS* Global Rating Scale (Carpal Tunnel Release) | 28 subjects | 2 | 0.69 |
| GOALS** Direct Observation (Cholecystectomy) | 19 subjects | 2 | 0.89 |
| GOALS** Video Observation (Cholecystectomy) | 10 videos | 4 | 0.68 |

#ICC=intra-class correlation coefficient.

*OSATS=Objective Structured Assessment of Technical Skills.

**GOALS=Global Operative Assessment of Laparoscopic Skill

**Table 4**
**Applying the Dreyfus Model to Surgical Competency**

| | |
|---|---|
| **Novice**<br>• Is rule driven<br>• Uses analytic reasoning and rules to link cause and effect<br>• Has little ability to filter or prioritize information, so synthesis is difficult at best and the big picture is elusive | Medical Student/Intern<br>• Not able to perform majority of procedure even with supervision |
| **Advanced Beginner**<br>• Is able to sort through rules and information to decide what is relevant on the basis of past experience<br>• Uses both analytic reasoning and pattern recognition to solve problems<br>• Is able to abstract from concrete and specific information to more general aspects of a problem | Junior Resident<br>• Performs majority of the procedure with supervision |
| **Competent**<br>• Emotional buy-in allows the learner to feel an appropriate level of responsibility<br>• More expansive experience tips the balance in clinical reasoning from methodical and analytic to more readily identifiable pattern recognition of common clinical problem presentations<br>• Sees the big picture<br>• Complex or uncommon problems still require reliance on analytic reasoning | Senior Resident/Fellow<br>• Performs complete procedure with supervision<br>• Likely able to perform safely without supervision but has not yet done so |
| **Proficient**<br>• Breadth of past experience allows one to rely on pattern recognition of illness presentation such that clinical problem solving seems intuitive<br>• Still needs to fall back to methodical and analytic reasoning for managing problems because exhaustive number of permutations and responses to management have provided less experience in this regard than in illness recognition<br>• Is comfortable with evolving situations; able to extrapolate from a known situation to an unknown situation (capable)<br>• Can live with ambiguity | Fellow/Practicing Orthopaedist<br>• Repeatedly performs task successfully and independently |
| **Expert**<br>• Thought, feeling, and action align into intuitive problem recognition and intuitive situational responses and management<br>• Is open to notice the unexpected<br>• Is clever<br>• Is perceptive in discrimination features that do not fit a recognizable pattern | Practicing Orthopaedist/Professor<br>• Demonstrates advanced skill and is adaptable to handling unexpected surgical situations |