

TOPICS IN TRAINING

Using the Arthroscopic Surgery Skill Evaluation Tool as a Pass-Fail Examination

Ryan J. Koehler, MSED, and Gregg T. Nicandri, MD

Investigation performed at the Kenneth E. DeHaven Arthroscopic Surgical Skills Laboratory, Department of Orthopaedic Surgery, University of Rochester, Rochester, New York

Background: Examination of arthroscopic skill requires evaluation tools that are valid and reliable with clear criteria for passing. The Arthroscopic Surgery Skill Evaluation Tool was developed as a video-based assessment of technical skill with criteria for passing established by a panel of experts. The purpose of this study was to test the validity and reliability of the Arthroscopic Surgery Skill Evaluation Tool as a pass-fail examination of arthroscopic skill.

Methods: Twenty-eight residents and two sports medicine faculty members were recorded performing diagnostic knee arthroscopy on a left and right cadaveric specimen in our arthroscopic skills laboratory. Procedure videos were evaluated with use of the Arthroscopic Surgery Skill Evaluation Tool by two raters blind to subject identity. Subjects were considered to pass the Arthroscopic Surgery Skill Evaluation Tool when they attained scores of ≥ 3 on all eight assessment domains.

Results: The raters agreed on a pass-fail rating for fifty-five of sixty videos rated with an interclass correlation coefficient value of 0.83. Ten of thirty participants were assigned passing scores by both raters for both diagnostic arthroscopies performed in the laboratory. Receiver operating characteristic analysis demonstrated that logging more than eighty arthroscopic cases or performing more than thirty-five arthroscopic knee cases was predictive of attaining a passing Arthroscopic Surgery Skill Evaluation Tool score on both procedures performed in the laboratory.

Conclusions: The Arthroscopic Surgery Skill Evaluation Tool is valid and reliable as a pass-fail examination of diagnostic arthroscopy of the knee in the simulation laboratory.

Clinical Relevance: This study demonstrates that the Arthroscopic Surgery Skill Evaluation Tool may be a useful tool for pass-fail examination of diagnostic arthroscopy of the knee in the simulation laboratory. Further study is necessary to determine whether the Arthroscopic Surgery Skill Evaluation Tool can be used for the assessment of multiple arthroscopic procedures and whether it can be used to evaluate arthroscopic procedures performed in the operating room.

Disclosure: One or more of the authors received payments or services, either directly or indirectly (i.e., via his 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.

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the views of the National Center for Research Resources or the National Institutes of Health.

The use of competency-based graduate medical education requires new assessment tools to evaluate proficiency. In orthopaedic surgery, assessment of competency includes evaluation of technical skill. Any assessment of technical skill should at a minimum be validated and should demonstrate reliability¹. However, an assessment that demonstrates validity and reliability cannot be assumed to be adequate for use as a pass-fail examination. Published assessments of arthroscopic technical skill can demonstrate validity and reliability, but all fall short of being suitable for use as a pass-fail examination because of a lack of specific criteria for passing the assessment²⁻⁵. This shortcoming leaves any potential user with the task of subjectively defining what a passing score will be and limits the use of these tools.

The Arthroscopic Surgery Skill Evaluation Tool (ASSET) was developed as a video-based assessment of technical skill that would be suitable for evaluating multiple procedures in both the simulation laboratory and the operating room. "Passing" criteria were established during content validation with the intention of creating an examination tool that could be used to assess the arthroscopic technical competency of orthopaedic residents in training⁶. This type of examination would provide information that may be useful for residency programs, the Residency Review Committee for Orthopaedic Surgery, or the American Board of Orthopaedic Surgery as it could serve as an indication of whether test takers have some of the basic technical surgical skills necessary for safe practice. We have previously demonstrated that the ASSET was both valid and reliable for assessing diagnostic arthroscopy of the knee in the simulated environment⁶.

The purposes of this study were to describe how the ASSET could be used as a pass-fail evaluation of diagnostic arthroscopy of the knee and to test the feasibility of using this tool for examination. Our hypothesis is that the ASSET would allow for reliable and valid pass-fail assessment of this basic arthroscopic skill.

Materials and Methods

This study was approved by our institutional review board and was granted an educational study exemption. Twenty-eight orthopaedic surgery residents and two orthopaedic sports medicine surgeons participated in this study. Each resident subject was asked to complete an online demographics survey that asked for hand dominance, the number of knee arthroscopies logged in the Accreditation Council for Graduate Medical Education (ACGME) Resident Case Log System, and the estimated number of arthroscopic knee procedures in which they had actually performed a portion of the case (SurveyMonkey, Palo Alto, California). Each resident's year in training was obtained from the current roster maintained by the Department of Orthopaedic Surgery at our institution. Upon completion of the survey and participation in the laboratory, subjects were provided with a \$10 gift card as compensation for their time.

Subjects were recorded while they performed diagnostic arthroscopy of both a right cadaveric knee and a left cadaveric knee in our institution's arthroscopic skills laboratory. Prior to entering the laboratory, subjects were asked to watch a video of the procedure being performed and were provided with a list of tasks that they would be asked to complete for this study (Table I). These steps were identical to the task-specific checklist for diagnostic knee arthroscopy that was utilized in the initial validation study⁶.

Subjects were randomly assigned by their date of participation to perform their first procedure on either a right or left knee. They were provided

TABLE I Task-Specific Checklist for Diagnostic Arthroscopy of the Knee

Inspect the suprapatellar pouch
Evaluate the patellofemoral articulation
Evaluate the patella (medial and lateral and inferior and superior)
Inspect the lateral gutter
Inspect the popliteus tendon and recess
Inspect the medial gutter
Inspect and probe the medial femoral condyle
Inspect and probe the medial tibial plateau
Inspect and probe the anterior, middle, and posterior medial meniscus
Inspect and probe the anterior cruciate ligament, posterior cruciate ligament, and ligamentum
Inspect and probe the lateral femoral condyle
Inspect and probe the lateral tibial plateau
Inspect and probe the anterior, middle, and posterior lateral meniscus
Evaluate passive tracking of the patella in the trochlear groove

with the checklist of steps to perform and an assistant who could assist with providing varus and valgus forces to the knee, but verbal assistance was limited to the information provided on the checklist. The subjects were asked to record their performance with use of the arthroscopic camera starting once adequate visualization of the patella was achieved. No video was recorded outside of the joint. Recording continued until all tasks were completed or a substantial amount of time (twenty-five minutes) had passed. The video recordings were then randomly presented to two raters (R.J.K. and G.T.N.) blind to subject identity and were evaluated with use of the ASSET. Both raters were from the same institution as each other and the trainees. The raters were instructed to use only the eight domains of the ASSET that assess technical skill. The ninth domain, Added Complexity of Procedure, was not rated, as the same cadaveric knees were used for all subjects (i.e., the complexity was the same for each arthroscopy). Each rater blindly reviewed all videos independently. The raters were required to watch each video in its entirety once, but could rewind, fast-forward, and rewatch videos as needed. Each rater had prior experience using a video-based assessment of arthroscopic skill and both raters were involved with the development of the ASSET.

The ASSET includes eight domains for assessing technical skill using procedural video recorded through the arthroscopic camera and one additional domain of Added Complexity of Procedure (Table II). All domains were developed by a content committee that consisted of eight surgeons with experience in arthroscopic education; one (G.T.N.) of these surgeons was an author of the current study. Eight of the domains were considered to be the basic domains of technical skill that could be assessed during multiple procedures performed in both the simulation laboratory and the operating room. The Dreyfus model of skill acquisition was used as the framework for constructing the ASSET with descriptive weighting at the novice, competent, and expert levels⁷. The content committee developed the descriptive weightings with the intent that an individual would be considered to pass the assessment only when he or she achieved a score of ≥ 3 for each domain. Any score of < 3 (competent) in any one domain would be considered a fail. A detailed description of each domain was included in the supplement to the initial article discussing the validation of the assessment tool⁶. Because the intent was to design the ASSET for use in the simulation laboratory and the operating room, a ninth domain of assessment (Added Complexity of Procedure) was included to control for the different degrees of difficulty encountered because of patient or cadaveric specimen factors.

TABLE II The ASSET Global Rating Scale

Domain		Global Rating			
Safety	1 (Novice)	2	3 (Competent)	4	5 (Expert)
	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	1 (Novice)	2	3 (Competent)	4	5 (Expert)
	Narrow field of view, inadequate arthroscope or light source positioning		Moderate field of view, adequate arthroscope or light source positioning		Expansive field of view, optimal arthroscope or light source positioning
Camera Dexterity	1 (Novice)	2	3 (Competent)	4	5 (Expert)
	Awkward or graceless movements, fails to keep camera centered and correctly oriented		Appropriate use of camera, occasionally needs to reposition		Graceful and dexterous throughout procedure with camera always centered and correctly oriented
Instrument Dexterity	1 (Novice)	2	3 (Competent)	4	5 (Expert)
	Overly tentative or awkward with instruments, unable to consistently direct instruments to targets		Careful, controlled use of instruments, occasionally misses targets		Confident and accurate use of all instruments
Bimanual Dexterity	1 (Novice)	2	3 (Competent)	4	5 (Expert)
	Unable to use both hands or no coordination between hands		Careful, controlled use of instruments, occasionally misses targets		Uses both hands to coordinate camera and instrument positioning for optimal performance
Flow of Procedure	1 (Novice)	2	3 (Competent)	4	5 (Expert)
	Frequently stops operating or persists without progress, multiple unsuccessful attempts prior to completing tasks		Steady progression of operative procedure with few unsuccessful attempts		Obviously planned course of procedure, fluid transition from one task to the next with no unsuccessful attempts
Quality of Procedure	1 (Novice)	2	3 (Competent)	4	5 (Expert)
	Inadequate or incomplete final product		Adequate final product with only minor flaws that do not require correction		Optimal final product with no flaws
Autonomy	1 (Novice)	2	3 (Competent)		
	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 (additional domain)	1	2	3		
	No difficulty	Moderate difficulty (mild inflammation or scarring)	Extreme difficulty (severe inflammation or scarring, abnormal anatomy)		

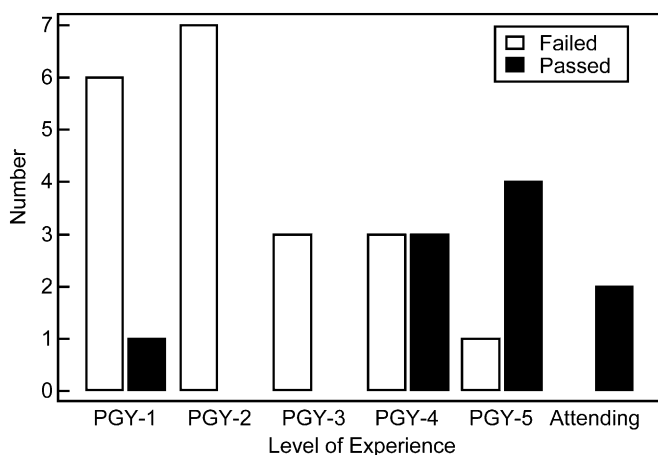


Fig. 1

Fig. 1 Bar graph showing the number of residents passing and failing on the basis of the level of experience. The results for two sports medicine faculty members are also shown. **Fig. 2** Line graph showing the number of knee arthroscopies logged that predict passing the ASSET on both trials.

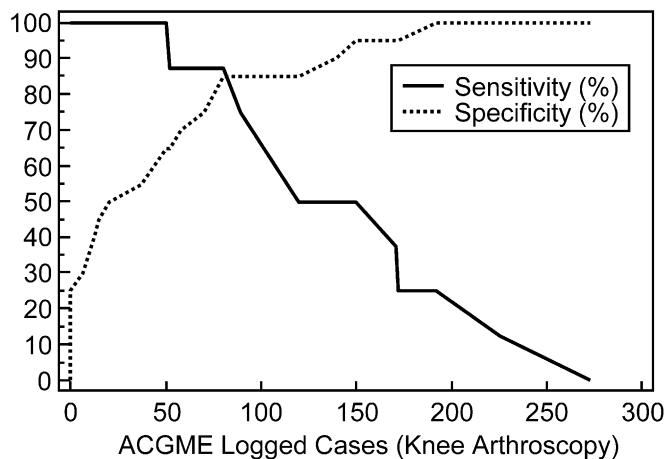


Fig. 2

Statistical Analysis

The agreement between raters was determined by using total percent agreement and by calculating the interclass correlation coefficient. For statistical analysis purposes, subjects were considered to pass an attempt when both raters assigned an ASSET score of ≥ 3 in all eight domains for the video being assessed. Receiver operating characteristic (ROC) analysis was completed to identify the number of ACGME cases logged and the actual number of arthroscopic knee cases performed that were most predictive of individuals who attained passing scores on both diagnostic arthroscopies performed. Statistical analysis was completed with use of MedCalc (MedCalc Software, Ostend, Belgium) with an a priori level of significance set at $\alpha = 0.05$.

Source of Funding

This research was supported in part by funding via an OMeGA Core Competency Innovations Grant. Additionally, this project was supported in part by Award Number TL1 RR 024135 from the National Center for Research Resources. Funds were used to pay for cadaveric specimens, additional supplies, statistical assistance, and part of the research coordinator's salary.

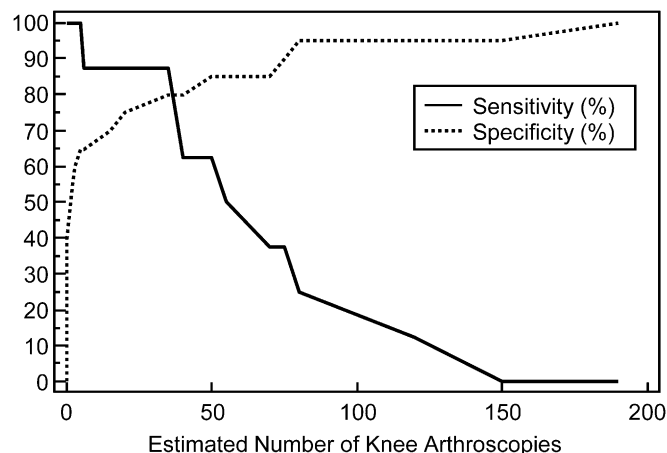


Fig. 3

Line graph showing the estimated number of knee arthroscopies performed that predict passing the ASSET on both trials.

Results

On the basis of the ASSET scores assigned, the raters agreed on a pass-fail rating for fifty-five (92%) of the sixty videos rated (Table III). This agreement was substantially greater than chance alone with an interclass correlation coefficient value of 0.83 (95% confidence interval [CI], 0.69 to 0.97). The likelihood of achieving a passing score on the ASSET increased as postgraduate year (PGY) in training increased (Fig. 1). Ten (33%) of thirty participants were assigned passing scores by both raters for both diagnostic arthroscopies performed in the laboratory (Fig. 1). ROC analysis demonstrated that logging more than eighty cases (sensitivity = 87.5, specificity = 85.0) or performing more than thirty-five arthroscopic knee cases (sensitivity = 87.5, specificity = 80.0) was predictive of attaining a passing ASSET score on both procedures performed in the laboratory (Figs. 2 and 3).

Discussion

To our knowledge, the ASSET is the first assessment of arthroscopic skill to have clearly defined criteria for passing. The criteria used for establishing a passing score on the ASSET were based on the framework of the Dreyfus model of skill acquisition. We chose this framework because it is a practical model for medical skill acquisition that can be used to guide competency-based assessment^{8,9}. With the ACGME's announcement of the Milestone Project, competency-based assessment will be an important part of how residency training programs demonstrate compliance¹⁰. In this investigation, we sought to determine whether the ASSET would be a feasible, valid, and reliable method of competency assessment when used as a pass-fail examination of a basic arthroscopic skill. We were able to demonstrate that the ASSET can be used as a valid and reliable pass-fail assessment of technical skill when surgeons perform diagnostic knee arthroscopy in the simulated environment.

This study had several limitations. First, our results were limited to the assessment of diagnostic arthroscopy of the cadaveric knee at a single institution. The ASSET was developed

TABLE III Overall Pass-Fail Agreement Between Raters

Rater 1	Rater 2	
	Pass	Fail
Pass	29	2
Fail	3	26

with an emphasis placed on creating a versatile assessment tool that could be utilized to assess multiple arthroscopic procedures and be employed in both the operating room and simulation laboratory. We are currently testing the validity and reliability of the ASSET as an assessment of cases in the operating room at our university and are testing the feasibility of using the tool at other institutions. This will give us a better picture of the overall generalizability of the tool. Second, the ASSET is restricted in that only the intra-articular portion of the case can be evaluated. The content development group thought that the tool should require only the use of inexpensive or readily available equipment and should not interfere with standard operative protocol as the intent was for it to enable evaluation both in the simulated and operating room environments. Because the arthroscopic camera and video-capture equipment are available in virtually all arthroscopic cases, we elected to record and to evaluate only those portions of cases that were 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. The Objective Assessment of Arthroscopic Skills is a video-based evaluation tool similar to the ASSET that was validated as a video-based assessment recording both the intra-articular portion of the case and the surgeon's hands with an external camera⁵. Although this use of an external camera may allow for adequate evaluation of important skills, its use may be limited to assessment in the simulated environment where the position and type of camera can be standardized without affecting the surgical procedure. Third, the raters in this study were involved in the development of the tool, which may have led to improved reliability. Raters with less familiarity with the tool may demonstrate less agreement. We are developing a standardized rater training for the ASSET to enhance the reliability of its use by evaluators outside of the content development group, and we believe that results similar to those of this study can be obtained by completing this training. We require that inexperienced raters read the original publication as well as rate six videos of individuals with varying levels of expertise (two novice, two intermediate, and two expert) in performing the procedure being assessed. Similar methods of rater training have been utilized for other assessments to improve reliability, and we are currently investigating our process^{6,11}. Fourth, it must be clear that the achievement of a passing score on the examination does not indicate that the surgeon is competent to perform the procedure being tested; it only indicates that the surgeon demonstrated a competent level of technical skill for that particular test on that occasion. It is unknown how many times a procedure should be evaluated to en-

sure accurate and reliable assessment of true competency. As there is no gold standard for assessing arthroscopic skill in the operating room, we are currently evaluating the use of the ASSET in the operating room to determine its utility in that training environment and to establish the predictive validity of scores obtained in the simulation laboratory. Fifth, although we were able to recruit twenty-eight of the thirty-five available residents, only three participants were from the PGY-3 class. We feel that, overall, the sample is representative; however, it would have been preferable to have more subjects from this group. In our opinion, PGY-3 residents are ideal subjects for this type of assessment, as this is typically when residents are increasingly performing basic arthroscopic procedures at our institution. Lastly, both raters in this study were from the same institution as the subjects. We attempted to control for this by blinding the raters to subject identity and by presenting the videos in a random order. Only the intra-articular portion of the case was rated and there was no way for the raters to identify the subjects. The ASSET was designed to enable subjects at one institution to be rated by an individual completely unknown to them and it is our hope that as others begin to use this scoring tool, this type of collaborative assessment will occur.

Using the ASSET as a pass-fail examination was reliable in this study. There was considerable agreement between raters with an interclass correlation coefficient of 0.83 (Table III). Assessment tools similar to the ASSET are reliable for assessing arthroscopic skill^{3,5}. However, neither assessment was developed or was tested as a pass-fail assessment^{3,5}. When initially developing the ASSET, we intended that it would be used as a pass-fail assessment because this was a major flaw in previously described arthroscopic assessments. For an assessment like the ASSET to be utilized by other institutions for examination in the future, demonstrating a level of reliability similar to that of this study will be essential.

We were also able to demonstrate the validity of the ASSET as a pass-fail examination by showing that as the participants' level of experience (postgraduate year) increased, a greater proportion achieved a passing scores for both knee arthroscopies performed in the laboratory (Fig. 1). The numbers of procedures that were most predictive of attaining a "passing" score on the ASSET from both raters for both attempts are outlined in Figures 2 and 3. We identified those residents who had actively participated in more than thirty-five arthroscopic knee cases as those who were most likely to pass and to be considered to have competent skills performing this procedure. Although the exact number of procedures required to become competent at diagnostic arthroscopy has not been established, others have suggested that competency in diagnostic arthroscopy of the knee is established after a number of cases similar to what we identified in this study¹². Video 1 (online) is an example of a diagnostic arthroscopy performed by a subject who received a failing ASSET score from both raters and Video 2 (online) is an example of diagnostic arthroscopy performed by a subject who received a passing ASSET score from both raters.

The use of the ASSET as a pass-fail examination has many potential applications. A test like the ASSET could provide formative feedback to trainees on the technical performance

of arthroscopic procedures, it could be used to test residents in the simulation laboratory to determine when they demonstrate sufficient technical skill to safely perform arthroscopic procedures, and hospitals could use an assessment like this in issuing surgeon privileges. Furthermore, assessment of technical skill may become part of the requirements for board certification in orthopaedic surgery in the future. The American Board of Surgery requires successful completion of the Fundamentals of Laparoscopic Surgery course, which includes a formal assessment of technical skill on a low-fidelity surgical simulation training box, to take the General Surgery Qualifying Examination¹³. The ASSET may prove to be useful for several of these applications, but further investigation is required before it can be used for such a high-stakes examination.

In conclusion, this study demonstrates that using the ASSET as a pass-fail examination of diagnostic knee arthroscopic

skill appears to be feasible, reliable, and valid. Further evaluation of the ASSET is currently under way for other arthroscopic procedures in both the simulation laboratory and the operating room, and its validity and reliability as a pass-fail test when used at multiple institutions on a large number of subjects have yet to be determined. ■

Ryan J. Koehler, MEd
Gregg T. Nicandri, MD
Department of Orthopaedic Surgery,
University of Rochester,
601 Elmwood Avenue,
Box 665, Rochester, NY 14642.
E-mail address for G.T. Nicandri: Gregg_Nicandri@urmc.rochester.edu

References

- Aggarwal R, Darzi A. Technical-skills training in the 21st century. *N Engl J Med*. 2006 Dec 21;355(25):2695-6.
- 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.
- 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.
- 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.
- Slade Shantz JA, Leiter JR, Collins JB, MacDonald PB. Validation of a global assessment of arthroscopic skills in a cadaveric knee model. *Arthroscopy*. 2013 Jan;29(1):106-12. Epub 2012 Nov 20.
- Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, Cosgarea AJ, Hamer CD, Garrett WE, Olson T, Warne WJ, Nicandri GT. The Arthroscopic Surgical Skill Evaluation Tool (ASSET). *Am J Sports Med*. 2013 Jun;41(6):1229-37. Epub 2013 Apr 2.
- Dreyfus HL, Dreyfus SE, Athanasiou T. Mind over machine: the power of human intuition and expertise in the era of the computer. Oxford: B. Blackwell; 1986.
- Batalden P, Leach D, Swing S, Dreyfus H, Dreyfus S. General competencies and accreditation in graduate medical education. *Health Aff (Millwood)*. 2002 Sep-Oct;21(5):103-11.
- 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 Aug;83(8):761-7.
- Nasca TJ, Philibert I, Brigham T, Flynn TC. The next GME accreditation system—rationale and benefits. *N Engl J Med*. 2012 Mar 15;366(11):1051-6. Epub 2012 Feb 22.
- 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.
- O'Neill PJ, Cosgarea AJ, Freedman JA, Queale WS, McFarland EG. Arthroscopic proficiency: a survey of orthopaedic sports medicine fellowship directors and orthopaedic surgery department chairs. *Arthroscopy*. 2002 Sep;18(7):795-800.
- American Board of Surgery. Training requirements for general surgery certification. 2013. http://www.absurgery.org/default.jsp?certgsqe_training. Accessed 2013 Sep 25.