Surgery in Practice and Science 14 (2023) 100198

Contents lists available at ScienceDirect

Surgery in Practice and Science

journal homepage: www.sciencedirect.com/journal/surgery-in-practice-and-science

Development and validation of an objective assessment of surgical skill in arthroscopic management of meniscal tear: A pilot study

Rachel S. Bronheim [a], Majd Marrache [a], Alexander E. Loeb [a], Johnathan A. Bernard [b], Dawn M. LaPorte [a,*]

[a] Department of Orthopaedic Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD, USA
[b] OrthoVirginia, Herndon, VA, USA

ARTICLE INFO

Keywords:
Meniscus tear
Resident competency
Surgical education
Surgical skill

ABSTRACT

Introduction: As resident evaluation moves to a competency-based system, validated tools for assessment of surgical skill are increasingly important. We created and validated a checklist to measure resident surgical skill for arthroscopic management of meniscal tear.

Materials and Methods: Using a Delphi survey method, we created an objective, structured assessment of surgical skill for treatment of meniscal tears. The Meniscus Treatment Task List (MTTL) comprises 5 domains: diagnostic arthroscopy, medial meniscectomy, lateral meniscectomy, medial meniscal repair, and lateral meniscal repair. Orthopaedic surgery residents were recruited to perform diagnostic arthroscopy, partial meniscectomies, and all-inside meniscal repairs with cadaveric models. Arthroscopic videos were graded by fellowship-trained surgeons using the MTTL and the validated Arthroscopic Surgical Skill Evaluation Tool (ASSET) global rating scale (GRS). Postgraduate year (PGY), operative time, and case logs were recorded for each resident. Data were analysed using bivariate correlation, analysis of variance, pairwise comparison, Pearson's correlation coefficient, and intraclass correlation coefficient. α=0.05.

Results: Twenty-two orthopaedic surgery residents (PGY1–PGY4) participated. MTTL scores were higher in the PGY4 class than in the PGY1 class (mean difference, 11 points, $p = 0.04$). Operative time was inversely correlated with number of cases logged ($r = -0.53, p = 0.01$), number of arthroscopic cases logged ($r = -0.50, p = 0.02$), and MTTL score ($r = -0.46, p = 0.03$). MTTL score was positively correlated with number of cases ($r = 0.44, p = 0.04$) and number of arthroscopic cases logged ($r = 0.50$, $p = 0.02$). MTTL scores were positively correlated with the ASSET GRS ($r = 0.71$, $p<0.001$). Intraclass correlation coefficient of 0.89 and Pearson's correlation coefficient of 0.89 demonstrated strong interrater reliability of MTTL scores ($p<0.01$).

Conclusions: This pilot study demonstrates the validity and reliability of the MTTL for assessing resident proficiency in arthroscopic management of meniscal tears in cadaveric specimens. Expansion of this model to other orthopaedic procedures for objective assessment of surgical skill may be useful.

Introduction

Orthopaedic surgery residency is designed to produce safe, competent, and confident surgeons. Faculty and peer evaluations play a large role in resident development, and objective, specific feedback is crucial for development and improvement of surgical skills. Historically, methods of evaluation have been largely subjective, including one-on-one feedback, self-assessment, non-validated scales, and faculty group discussion [1]. The field of orthopaedic education has begun to transition from this traditional model to a more objective and standardized model of resident training and assessment. The value of standardization of resident assessments lies in the ability to grade and compare residents across programs, to more easily identify technical deficiencies, and to ensure that graduates of residency programs have a sufficient level of surgical skill to safely practice orthopaedic surgery.

Notably, competency-based assessment of resident surgical skill has garnered interest from the Accreditation Council for Graduate Medical Education (ACGME), the American Board of Orthopaedic Surgery (ABOS), the American Orthopaedic Association (AOA), and the American Academy of Orthopaedic Surgeons (AAOS) [2,3]. As part of this

* Corresponding author at: Department of Orthopaedic Surgery, The Johns Hopkins University, 601 N. Caroline St., Baltimore, MD 21287, USA.
E-mail address: dlaport1@jhmi.edu (D.M. LaPorte).

https://doi.org/10.1016/j.sipas.2023.100198
Received 18 January 2023; Received in revised form 27 June 2023; Accepted 28 June 2023
Available online 29 June 2023

effort to enhance clinical and surgical competence, the ACGME established core competencies and milestones to provide a common framework for orthopaedic surgery resident education [4].

The initial milestones for orthopaedic surgery represented specific conditions and professional skills that the ACGME and ABOS deemed crucial for development of orthopaedic residents. Management of meniscal tear was one of the original milestones, and therefore represents a meaningful target for development of competency-based evaluation. Meniscal tear is now included in the Operative Management of Arthroscopically Treated Conditions Patient Care in Milestones 2.0 [5].

Several global rating scales (GRSs) have been developed to evaluate surgical skills ranging from open surgery to shoulder and knee arthroscopy [6–9]. However, these scales are not designed to assess the technical details of distinct procedures [10]. The Arthroscopic Surgical Skill Evaluation Tool (ASSET) is a GRS that has been validated for dry knee models, but not wet models [11,12]. To date, there is no objective, validated resident assessment tool specific to arthroscopic management of meniscal tears.

The purpose of this study was to create and validate a checklist to be used as an objective measure of resident surgical skill in arthroscopic management of meniscal tears.

Materials and methods

Development of the Meniscus Treatment Task List

The Objective Structured Assessment of Technical Skills (OSATS) checklist is a validated tool for grading technical proficiency. An OSATS Task-Specific Checklist was created using a Delphi survey method. The Delphi method is a survey process in which questionnaires are sent to a panel of experts, a summary of the answers is given as feedback by a survey facilitator, and the process is repeated until the panel is in agreement [13]. A questionnaire was circulated to 4 fellowship-trained sports medicine orthopaedic surgeons from 2 outside institutions. After 3 feedback rounds of editing and revision, one author summarized the opinion of the experts for accuracy, clarity, and comprehension.

Checklist domains developed using the Delphi process comprised a list of steps to be performed for each procedure. Through this iterative method, a comprehensive, itemized arthroscopic meniscal treatment checklist (Meniscus Treatment Task List [MTTL]) was created with 5 domains: diagnostic arthroscopy, medial meniscectomy, lateral meniscectomy, medial meniscus repair, and lateral meniscus repair. The checklist assessed 50 items, with each item valued at 1 point (Appendix A).

Participant recruitment and arthroscopy simulation

Of 30 residents who were emailed to request their participation in the study, 22 agreed. No compensation was provided for their participation. The following information was collected from each participant: year of residency, total case volume and arthroscopic case volume (as assessed by ACGME case log data), hand dominance, subspecialty interest, and number of sports medicine rotations completed. To reduce the risk of bias, each participant was assigned a number by a random number generator to de-identify participants and their study data.

The arthroscopy simulation consisted of 5 tasks to be performed on a cadaveric knee specimen: a diagnostic arthroscopy, medial and lateral meniscectomies, and medial and lateral meniscal repairs. In the meniscectomy tasks, participants were instructed to perform partial medial and lateral meniscectomies, simulating a white-white zone tear, and to resect approximately 30%–50% of the central meniscus. For the meniscal repair tasks, participants were instructed to place a single all-inside meniscus repair suture in simple fashion in both the medial and lateral menisci with the provided implant. All tasks were recorded using a standard arthroscopy tower for subsequent review by fellowship-trained sports medicine physicians.

Cadaveric, full-knee joint specimens were obtained from the Maryland Department of Health State Anatomy Board (Baltimore, Maryland) and the Anatomy Gifts Registry (Hanover, Maryland). One author inspected each knee specimen externally for anatomical deformities before simulation. All testing was performed at an on-campus cadaver and biomechanical laboratory. Arthroscopic equipment and video capturing capability were provided by Stryker Corporation (Kalamazoo, Michigan), and 50 all-inside meniscus repair suture devices (Fast-Fix 360 Meniscal Repair System, Smith & Nephew, Memphis, Tennessee) were provided by the manufacturer.

Throughout the simulation, the study coordinator was available to aid participants only as a simulated "industry representative,'' demonstrating the function of the meniscus repair suture device as needed. For each step of this objective structured clinical examination, participants were given no further instruction on technique, anatomy, or procedure.

Simulation evaluation

De-identified arthroscopic videos of each participant's simulation were distributed to 2 fellowship-trained sports medicine surgeons. Both evaluators were selected from outside institutions to ensure independent evaluation and minimize bias. Two grading instruments were used to evaluate the videos. Simulations were first scored with the ASSET, a validated GRS for arthroscopy [11] (Appendix B). Next, the simulations were scored with the MTTL. Residents earned 1 point per task completed on the MTTL, with a maximum possible score of 50 points. Arthroscopic inspection of the knee constituted approximately half of the points in the MTTL, with procedural skills constituting the other half.

Participants and the study coordinator were blinded to both the ASSET and MTTL scores so performances could be evaluated in an unbiased manner. Physicians who had participated in the Delphi process, graded participants' MTTL instruments, or performed the statistical analysis were not blinded to the ASSET and MTTL scores.

Statistical analysis

Descriptive statistics were used to report resident-level data, including operative experience and performance metrics by post-graduate year (PGY) of training (Table 1). Differences were calculated using 1-way analysis of variance (ANOVA). Construct validity was assessed using 1-way ANOVA with Welch correction to compare mean MTTL scores with PGY. When significant differences were observed, post-hoc testing using a custom contrasts post-hoc test was performed for all pairwise comparisons. Pearson correlation coefficient (r) was used to determine the relationship between mean MTTL scores and other variables (PGY, total case number, arthroscopic case number, and operative time). Concurrent criterion-related validity was assessed by comparing performance scored using the MTTL with the previously

Table 1
Operative experience and performance metrics by PGY of training for 22 orthopaedic surgery residents performing arthroscopic treatment of meniscal tears.

Parameter	Mean ± Standard Deviation				*p*-value
	PGY-1 ($N = 6$)	PGY-2 ($N = 5$)	PGY-3 ($N = 5$)	PGY-4 ($N = 6$)	
No. of cases logged	156 ± 78	660 ± 222	1106 ± 192	1251 ± 262	<0.001
No. of arthroscopic cases logged	3 ± 2	61 ± 18	127 ± 50	150 ± 48	<0.001
Operative time (minutes)	57 ± 27	46 ± 9.4	32 ± 10	33 ± 14	0.08
ASSET score	14 ± 5	17 ± 4	18 ± 3	26 ± 5	<0.001
MTTL score	24 ± 9	25 ± 8	27 ± 6	35 ± 2	0.03

ASSET, Arthroscopic Surgical Skill Evaluation Tool; MTTL, Meniscus Treatment Task List; PGY, post-graduate year.

validated ASSET GRS. Interrater reliability of the MTTL score was determined using the intraclass correlation coefficient for absolute agreement between a fixed, non-random set of raters.

An *a priori* confidence interval was set at 95%. A power analysis was conducted for a β of 80% and an α of 0.05, with a clinically meaningful difference of 10 points on the ASSET GRS or 2 points on the MTTL. All statistical analyses were performed using Stata/IC 15.1 (StataCorp LLC, College Station, TX).

Results

Participants were 6 PGY-1 residents, 5 PGY-2 residents, 5 PGY-3 residents, and 6 PGY-4 residents.

Construct validity

Mean (± standard deviation) MTTL scores increased from the PGY1 class (24 ± 9.3) to PGY2 (25 ± 8.2), PGY3 (27 ± 6.2), and PGY4 (35 ± 2.2), differing significantly when assessed with ANOVA ($p = 0.03$; Fig. 1). Post-hoc pairwise comparison testing of these scores showed a significant difference only between the PGY1 and PGY4 classes (mean difference 11, $p = 0.04$).

Significant positive correlations were demonstrated between mean MTTL score, total number of cases logged ($r = 0.44, p = 0.04$; Fig. 2), and total number of arthroscopic cases logged ($r = 0.50$, $p = 0.02$; Fig. 3). Additionally, mean MTTL was inversely correlated with operative time ($r = -0.46$, $p = 0.03$; Fig. 4). Subspecialty interest was not associated with MTTL score ($p = 0.15$). Evaluation of resident performance with the MTTL was positively correlated with the previously validated ASSET GRS, suggesting additional criterion-related validity ($r = 0.71, p < 0.001$; Fig. 5).

Interrater reliability of the MTTL

When using a Pearson correlation coefficient to analyse interrater reliability, the MTTL scores assigned by each rater were strongly correlated ($r = 0.89$, $p < 0.001$). They further demonstrated "good" to "excellent" interrater reliability with an intraclass correlation coefficient of 0.89 [14,15].

Discussion

In this study, we created a task-specific, objective, structured assessment of orthopaedic surgical skills using the Delphi method. The MTTL is a valid and reliable assessment of meniscal repair in a simulated environment. Specifically, we found excellent correlation of the MTTL with PGY of training, number of cases logged, number of arthroscopic cases logged, total operative time, and the ASSET GRS. Proof of validity of the MTTL instrument includes evidence of novice-expert differentiation, excellent inter-rater reliability, and strong correlation with performance on the previously validated ASSET GRS.

The significant correlation between MTTL performance and predictors of expertise, including PGY of training, number of cases logged, and number of arthroscopic cases logged, [16,17] establishes excellent construct validity in determining levels of competency. Further, higher MTTL score was associated with shorter operative time, which is similarly associated with surgical experience and studies using virtual reality simulators [18]. In addition, the MTTL demonstrated concurrent criterion-related validity as an objective assessment of resident performance when compared with the previously validated ASSET GRS.

Furthermore, the MTTL scoring system demonstrated "good" to "excellent" interrater reliability between 2 fellowship-trained orthopaedic sports medicine surgeons when evaluating the same participant. This result is largely related to the meticulous effort to create a precise, itemized checklist that would produce reproducible results. However, we also acknowledge the importance of having raters with a comprehensive understanding of the MTTL, as well as robust experience with the surgical technique being evaluated. It follows that, to ensure maximal reliability between raters of varied backgrounds in diverse training environments, establishment of a training module for evaluators may be necessary. Even with this caveat, the presence of significant agreement within our model between independent raters supports the consistency of our construct and its generalizability.

This study has several strengths. First, established orthopaedic surgery grading instruments have assessed general arthroscopic skills, but none has assessed procedures specifically highlighted as ACGME Milestones [11,18]. Second, use of the modified Delphi method in the development and design process ensured content validity with the inclusion of essential criteria and assessment of key competencies as agreed upon by a team of experts [10]. Third, the use of a cadaver model facilitated "gold-standard," high-fidelity simulation, allowing residents to demonstrate proficiency in a low-risk environment [12,19]. Finally, anonymity afforded by videos of arthroscopic procedures and the blinded administration of this study limited potential sources of bias.

The limitations of this study are also important to acknowledge. The simulation portion of the study included a small number of orthopaedic

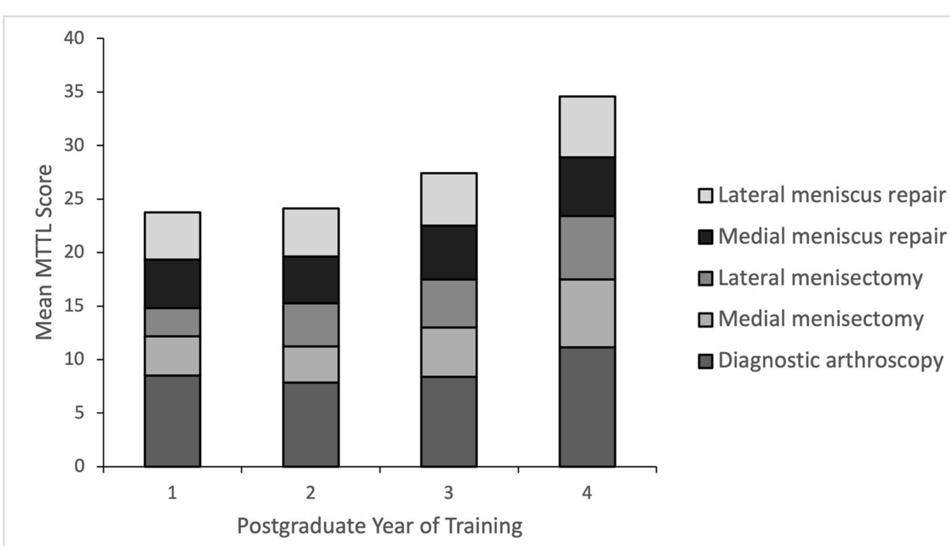

Fig. 1. Mean Meniscus Treatment Task List (MTTL) score by post-graduate year of training.

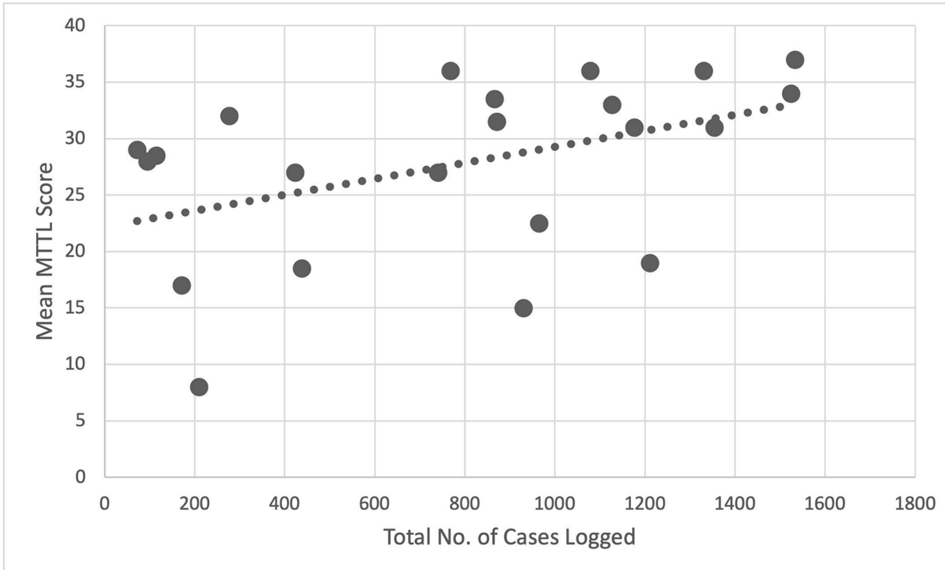

Fig. 2. Mean Meniscus Treatment Task List (MTTL) score by total number of cases logged. Coefficient of determination (r^2) = 0.19.

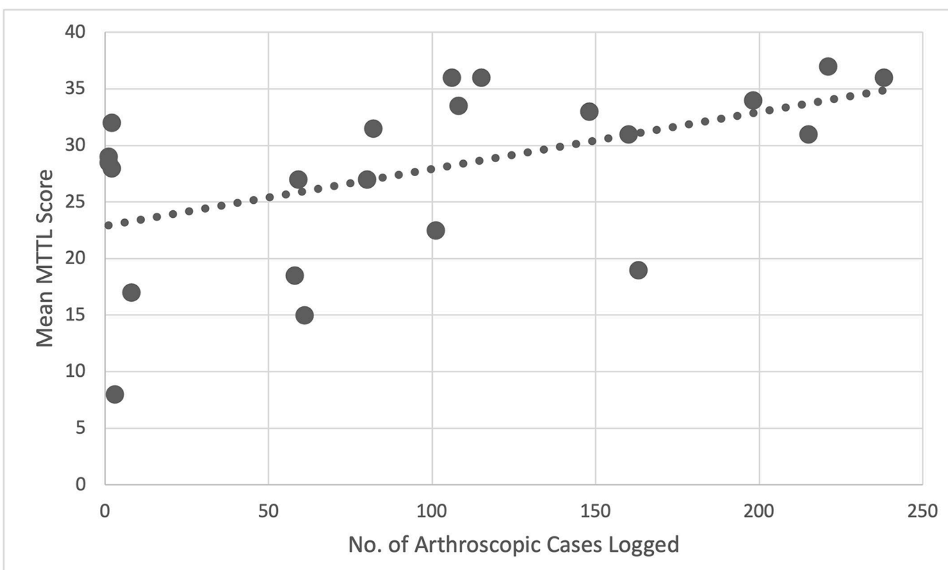

Fig. 3. Mean Meniscus Treatment Task List (MTTL) score by number of arthroscopic cases logged. Coefficient of determination (r^2) = 0.25.

surgery residents from a single institution, reducing the study's statistical power and generalizability. Expansion of this study across multiple institutions and a greater number of trainees would strengthen the external validity of the current work. The experiment lacked a true "expert" group of fellowship-trained surgeons, which may have further helped define our model's ability to delineate level of expertise by acting as a positive control. The experimental design did not enable us to perform retest-reliability testing given the limited number of cadaveric specimens. Evaluation was performed by 2 fellowship-trained surgeons; additional graders would provide greater confidence in the interrater reliability.

Similar studies using primarily video-based assessment have been criticized for failing to assess important procedural skills that may not be well visualized on video, such as management of the surgical field, incision, varus/valgus maneuvers to facilitate improved visualization, or closure [11]. However, considering that our purpose was to design an objective assessment model, the value of anonymous evaluation in our study outweighed such concerns. The study design enabled assessment of essential features of the procedure without introducing the bias associated with unblinding. This study was not designed to determine thresholds for a passing or failing score, which has been suggested as an important addition to GRSs and checklists [20]. Finally, the MTTL does not include a penalty for adverse events.

This study may have been complemented with a biomechanical assessment of strength of meniscal repair, because previous studies have shown lack of correlation between OSATS scores and biomechanical strength of a given construct; this is an avenue for further research [21, 22]. Although this study leveraged numerous measures to maximize objectivity in scoring, it is important to note that there is always some level of subjectivity present in any form of grading and scoring.

Conclusions

In this pilot study, the MTTL was shown to be a valid, objective, and

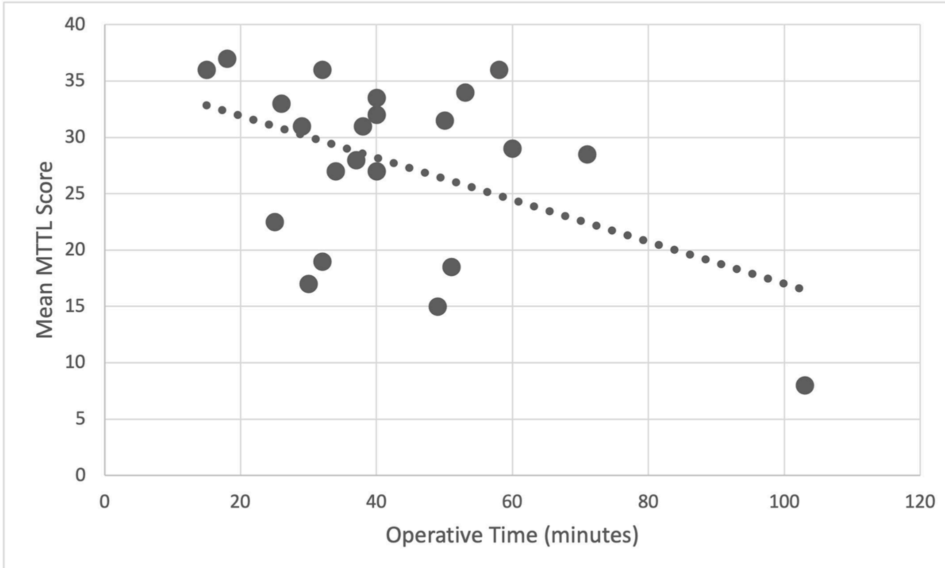

Fig. 4. Mean Meniscus Treatment Task List (MTTL) score by total operative time. Coefficient of determination (r^2) = 0.21.

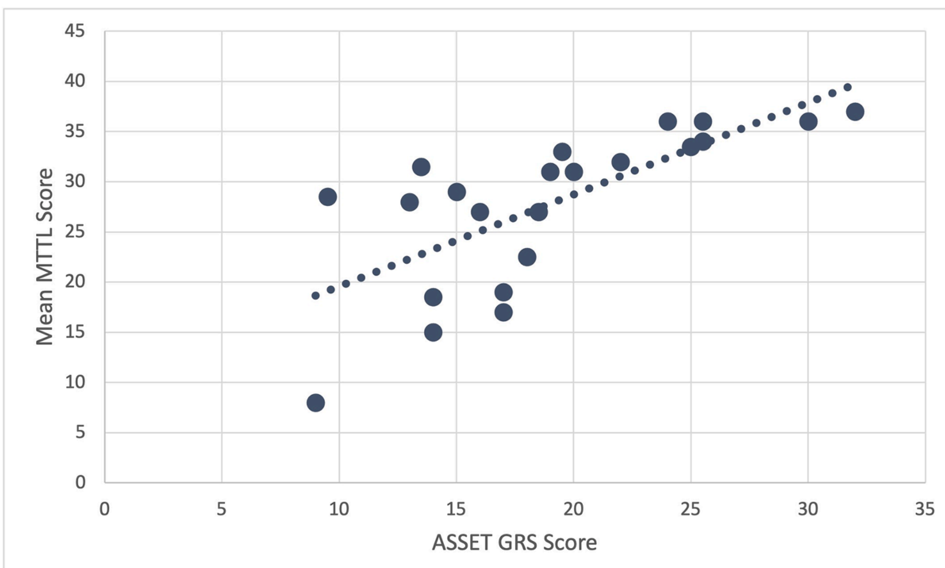

Fig. 5. Mean Meniscus Treatment Task List (MTTL) score by mean Arthroscopic Surgical Skill Evaluation Tool (ASSET) global rating scale (GRS) score. Coefficient of determination (r^2) = 0.51.

reliable method for assessing resident surgical competence in arthroscopic management of meniscal tears in a cadaveric model. Once these findings are externally validated, the MTTL may be used, in combination with other grading tools, to facilitate an objective assessment of resident technical competency. The MTTL may also be a valuable tool to assess resident progress over time, as the evaluation is standardized and simple to repeat. Furthermore, the Delphi model for creation of resident assessment tools may be applied to other common orthopaedic procedures for objective evaluation of surgical skill.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This study was supported in part by a grant from the American Board of Orthopaedic Surgery.

Acknowledgments

We acknowledge the American Board of Orthopaedic Surgery for grant support and Nathan Coleman, MD (OrthoVirginia, Herndon, VA) for his assistance as a grader and participant in the Delphi process. For their editorial assistance, we thank Rachel Box, MS, in the Editorial Services group of The Johns Hopkins Department of Orthopaedic Surgery.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sipas.2023.100198.

References

[1] Edgar L., McLean S., Hogan S., Hamstra S., Holmboe E. The milestones guidebook 2020.
[2] Nousiainen M, Incoll I, Peabody T, Marsh JL. Can we agree on expectations and assessments of graduating residents?: 2016 AOA critical issues symposium. J Bone Joint Surg Am 2017;99:e56. https://doi.org/10.2106/JBJS.16.01048.
[3] Nousiainen MT, McQueen SA, Hall J, Kraemer W, Ferguson P, Marsh JL, et al. Resident education in orthopaedic trauma: the future role of competency-based medical education. Bone Joint J 2016;98-B:1320–5. https://doi.org/10.1302/0301-620X.98B10.37031.
[4] Ames SE, Ponce BA, Marsh JL, Hamstra SJ. Orthopaedic surgery residency milestones: initial formulation and future directions. J Am Acad Orthop Surg 2020;28:e1–8. https://doi.org/10.5435/JAAOS-D-18-00786.
[5] Orthopaedic Surgery Milestones 2021.
[6] Atesok K, Mabrey JD, Jazrawi LM, Egol KA. Surgical simulation in orthopaedic skills training. J Am Acad Orthop Surg 2012;20:410–22. https://doi.org/10.5435/JAAOS-20-07-410.
[7] Lubowitz JH, Provencher MT, Brand JC, Rossi MJ. Expert surgeons can be distinguished from trainees, and surgical proficiency can be defined, using validated metrics and shoulder models. Arthroscopy 2015;31:1635–6. https://doi.org/10.1016/j.arthro.2015.06.040.
[8] Martin JA, Regehr G, Reznick R, MacRae H, Murnaghan J, Hutchison C, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. Br J Surg 1997;84:273–8. https://doi.org/10.1046/j.1365-2168.1997.02502.x.
[9] Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. Am J Surg 2005;190:107–13. https://doi.org/10.1016/j.amjsurg.2005.04.004.
[10] Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG. The Bankart performance metrics combined with a cadaveric shoulder create a precise and accurate assessment tool for measuring surgeon skill. Arthroscopy 2015;31:1655–70. https://doi.org/10.1016/j.arthro.2015.05.006.
[11] Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Bramen JP, et al. The arthroscopic surgical skill evaluation tool (ASSET). Am J Sports Med 2013;41:1229–37. https://doi.org/10.1177/0363546513483535.
[12] Dwyer T, Slade Shantz J, Kulasegaram KM, Chahal J, Wasserstein D, Schachar R, et al. Use of an objective structured assessment of technical skill after a sports medicine rotation. Arthroscopy 2016;32:2572–81. https://doi.org/10.1016/j.arthro.2016.05.037. e3.
[13] Rowe G, Wright G. Expert opinions in forecasting: the role of the Delphi technique. Principles of forecasting: a handbook for researchers and practitioners. Boston, MA: Springer US; 2001. p. 125–44.
[14] Cicchetti D. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instrument in psychology. Psychol Assess 1994;6:284–90. https://doi.org/10.1037/1040-3590.6.4.284.
[15] Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. J Chiropr Med 2016;15:155–63. https://doi.org/10.1016/j.jcm.2016.02.012.
[16] Hodgins JL, Veillette C, Biau D, Sonnadara R. The knee arthroscopy learning curve: quantitative assessment of surgical skills. Arthroscopy 2014;30:613–21. https://doi.org/10.1016/j.arthro.2014.02.021.
[17] Martin KD, Cameron K, Belmont PJ, Schoenfeld A, Owens BD. Shoulder arthroscopy simulator performance correlates with resident and shoulder arthroscopy experience. J Bone Joint Surg Am 2012;94:e160. https://doi.org/10.2106/JBJS.L.00072.
[18] 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;91:2287–95. https://doi.org/10.2106/JBJS.H.01762.
[19] Akhtar KSN, Chen A, Standfield NJ, Gupte CM. The role of simulation in developing surgical skills. Curr Rev Musculoskelet Med 2014;7:155–60. https://doi.org/10.1007/s12178-014-9209-z.
[20] VanHeest A, Kuzel B, Agel J, Putnam M, Kalliainen L, Fletcher J. Objective structured assessment of technical skill in upper extremity surgery. J Hand Surg Am 2012;37:332–7. https://doi.org/10.1016/j.jhsa.2011.10.050. 337.e1-4.
[21] Atesok K, Satava RM, Marsh JL, Hurwitz SR. Measuring surgical skills in simulation-based training. J Am Acad Orthop Surg 2017;25:665–72. https://doi.org/10.5435/JAAOS-D-16-00253.
[22] Putnam MD, Kinnucan E, Adams JE, Van Heest AE, Nuckley DJ, Shanedling J. On orthopedic surgical skill prediction–the limited value of traditional testing. J Surg Educ 2015;72:458–70. https://doi.org/10.1016/j.jsurg.2014.11.001.