

The following content was supplied by the authors as supporting material and has not been copy-edited or verified by JBJS.

Table 1. Studies Assessing Arthroscopic Performance

Author	Study Aim	Assessment Format	Resident Participants	Skills Assessed	Tools or Metrics	Results	Findings Relating to Assessment Tool(s)
STUDIES IN SIMULATED SETTING							
An et al ⁶ , 2018	To determine the effect of simulation training on gaze fixation strategies	SDM	16 PGY?	Diagnostic Knee Arthroscopy	Eye movements, PT	Correlation seen between PT and gaze fixation strategy with repeated instruction	Content, construct and criterion validity of gaze tracking is demonstrated
Angelo et al ¹⁴ , 2015	To assess construct validity of the tool on a cadaveric shoulder and establish a proficiency benchmark for arthroscopic bankhart repair	Expert assessor review of video footage	12 PGY 4-5	Arthroscopic Bankhart repair	TSC	IRR = 0.92, novice surgeons made 50% more errors and demonstrated increased performance variability and procedure time	Construct validity is demonstrated in the cadaveric model. A proficiency benchmark can be established as the mean performance of the expert group
Angelo et al ⁷ , 2015	To compare 3 training protocols for learning to perform and ABR	Expert assessor review of video footage	44 PGY 4-5	3-anchor ABR	Bankart procedure metric tool	A proficiency-based progression training curriculum and protocol coupled with the use of a shoulder model simulator produces a superior arthroscopic bankart skill set	Bankart procedure metric tool appears construct valid in this setting
Alvand et al ⁷⁴ , 2012	To assess the ability of novel visual parameters to objectively discriminate between various levels of arthroscopic experience	Live observation, SDM	15 PGY?	Diagnostic Knee Arthroscopy	HMA, hand position checking, instrument loss, triangulation time	Significant difference in performance between the three groups was seen with visual parameters, GRS and motion analysis	Visual parameters are construct and criterion valid in this setting

Alvand et al ⁷⁵ , 2013	Determine if a modified GRS can be used to assess the learning curve during simulated arthroscopic knee meniscal repair	Expert assessor review of video footage and SDM	19 PGY2-4	Arthroscopic knee meniscal repair	GRS, BAKSS, HMA	Moderate correlation observed between GRS and HMA parameters	Arthroscopic modified GRS can be used for modelling skill development in knee meniscal repair
Bayona et al ⁷⁶ , 2014	To assess the validity of the IGARS using a VR shoulder simulator	Expert assessor review of video footage	39 PGY?	Diagnostic Shoulder Arthroscopy	IGARS	IGARS demonstrated high internal consistency (Cronbach's alpha =0.92). IGARS can distinguish between different levels of experience on a VR simulator	IGARS demonstrates face, content and construct validity. It has high internal consistency and excellent IRR
Bhattacharyya et al ⁸ , 2017	Evaluate the effectiveness of cognitive task analysis for training in diagnostic knee arthroscopy	Expert assessor review of video footage	16 PGY?	Diagnostic Knee Arthroscopy	ASSET	The Cognitive Task Analysis showed improved technical skill compared to the control group as measured by the ASSET	ASSET is construct valid in this setting
Brusalis et al ³² , 2017	Evaluate a low-fidelity simulation model for ACL graft preparation	Live observation	10 PGY 1-4	ACL graft preparation	Error-focussed scoring	The simulator trained group performed significantly better than controls in overall performance, critical steps and errors	Assessment tool was construct valid in this model
Camp et al ⁹ , 2016	Compare the impact of cadaveric simulation training compared to VR training in diagnostic knee arthroscopy	Expert assessor review of video footage	45 PGY 1-5	Diagnostic Knee Arthroscopy	PT, ASSET	PT and ASSET scores improved in both the cadaveric and VR groups as compared to the controls	ASSET and PT are construct and concurrent valid in this setting
Cannon et al ³³ , 2014	To assess the construct validity of a VR arthroscopy simulator	SDM	PGY 1, 6 PGY 5	Diagnostic Knee Arthroscopy	PT	The mean PT correlated with experience, and was significantly different	Construct validity is demonstrated. Educational benefit is

						between PGY 1 and both PGY 5 and expert groups. The difference between PGY 1 and 5 was approaching significance (0.055). the level of completeness was NS suggesting all groups achieved a level of consistency in their performance of the procedure. Times improved across 3 trial repetitions in all groups	seen in repetitious practice.
Cetinkaya et al ¹⁰ , 2017	Evaluate the impact of a simulation course on psychomotor skill performance	Live observation	70 PGY?	Basic arthroscopy skills	Psychomotor test metrics	Scores improved significantly for time to completion and error rate for all psychomotor tests post-course	Construct validity of psychometric testing tools for measuring basic arthroscopic skills is demonstrated
Chong et al ¹¹ , 2016	Define the early learning of arthroscopic knot tying	Biomechanical testing	3 PGY 1, 3 PGY 3	Arthroscopic knot tying	FPA, PT	Scores improved in the inexperienced group across 3 stages of simulator training	Construct and concurrent validity of knot tensile strength testing demonstrated
Colaco et al ³⁴ , 2017	To assess the construct validity of three skill deconstructed models	Live observation	12 PGY 1-10+, 7	Basic arthroscopy skills	PT, Hand position checking	Average PT and hand position checking frequency correlated inversely with experience level	Construct validity of simulator and educational impact demonstrated
Coughlin et al ³⁵ , 2015	To assess construct validity of the simulator model using an aggregate assigned score for the six component tasks	Live observation, Expert assessor review of video footage	12 PGY 2-3, 16 PGY 4-5	Basic arthroscopy skills	Aggregate arthroscopic skills score	Construct validity demonstrated by significant improvement in scores by increasing levels of training between all groups. The model was	Construct validity and reliability demonstrated

						highly reliable with ICC =0.99 for both IRR and intra-rater reliability	
Dwyer et al ³⁶ , 2015	To validate dry knee simulator model for assessing performance of ACLR	Live observation	29 PGR 1-5	ACLR	ASSET	Internal reliability using the total ASSET score was very high (>0.9). Construct validity demonstrated in significant observed score differences by level (p<0.05)	ASSET is reliable and construct valid
Dwyer et al ⁷⁷ , 2016	To determine if an OSATS is valid for assessing residents performance of sports surgery procedures in a competency-based model	Expert assessor review of video footage	27 PGY 1-5	Partial meniscectomy, Whipstitch hamstring graft, ACL femoral tunnel drilling, labral suture passing, insertion of rotator cuff anchors, passing rotator cuff sutures, tie sliding arthroscopic knot	ASSET, OSATS TSC and GRS	A significant difference by PGY was seen for the overall GRS, total ASSET score and total checklist score, as well as for each procedure (p<0.001)	Construct and concurrent validity demonstrated
Dwyer et al ³⁷ , 2017	To evaluate the use of dry models to assess performance of arthroscopic rotator cuff repair (RCR) and labral repair (LR) on a dry model	Live observation, expert assessor review of video footage	LR: 39 PGY 1-5, RCR: 35 PGY 1-5	Arthroscopic RCR and LR	TSC, ASSET non-OSATS GRS	Internal consistency and IRR using total ASSET score was high (>0.9). Construct validity of the model was demonstrated	ASSET demonstrated internal consistency, IRR, construct validity and concurrent validity
Elliot et al ⁷⁸ , 2012	To develop a scoring system to evaluate individual proficiency at diagnostic knee arthroscopy	Live observation (remote for blinding)	20 PGY 1-5)	Diagnostic Knee Arthroscopy	PT, Arthroscopic Skills Assessment form	Statistically significant differences in total score, procedure time and number of missed structures between	The tool is face, content and construct valid for assessing diagnostic knee arthroscopy

						participants grouped by experience level	performance in the simulated setting
Escoto et al ³⁸ , 2012	To evaluate the construct validity of a force-sensing simulator for knee arthroscopy skill assessment	SDM	10PGY? 'novices'	Diagnostic and therapeutic knee arthroscopy	PT, Instrument collision force	Performance difference between experts and novices was seen in 9 of 14 tasks. The difference in force applied to the femur is statistically significant for the two oscillating shaving tasks with the novices applying more force than the experts on average	Force applied to joint structures has construct validity in assessing arthroscopic skill level
Fucentese et al ³⁹ , 2015	To determine the face and construct validity of a new VR simulator for therapeutic and diagnostic knee arthroscopy	SDM	33 PGY?	Diagnostic and therapeutic knee arthroscopy	PT, Instrument path length	Face validity and educational impact of simulator was good. Experts performed better than novices in all domains however performance difference between the three groups individually did not reach significance	SDM not construct valid in this model
Garfield Roberts et al ⁶³ , 2017	To assess the face and construct validity of a passive haptic VR simulator	SDM and expert assessor review of video footage	25 PGY?Ω	Diagnostic and therapeutic knee and shoulder arthroscopy	PT, path length, novel SDM: scratching score, % of normal meniscus removed and % of lesion removed. ASSET GRS	ASSET GRS score, PT and path length demonstrated construct validity. Good IRR with ASSET. Mixed construct validity evidence for the novel metrics	ASSET and established SDM metrics showed good construct validity
Gomoll et al ⁴⁰ , 2007	Test the construct validity of a shoulder arthroscopy simulator	SDM	25 PGY2-5	Diagnostic shoulder arthroscopy	PT, Instrument path length, collisions	A significant association was observed between all tested parameters and level of surgical experience	SDM are construct valid in this model

Gomoll et al ¹² , 2008	To assess the skill retention/improvement 3 years after arthroscopic simulator training	SDM	10 PGY?	Diagnostic shoulder arthroscopy	PT, Instrument path length, Collisions and Injuries	Subjects improved significantly across all 4 parameters at 3 year retest	PT, path length demonstrated construct validity. Collisions and injuries did not.
Howells et al ⁴¹ , 2008	Test the construct validity of a motion analysis system for assessing performance of simple arthroscopic tasks	SDM	20 PGY?	Diagnostic and therapeutic knee arthroscopy	HMA	Significant performance differences seen between surgeons and non-surgeons ($p < 0.0001$) and between senior and junior surgeons ($p < 0.05$) – trend towards decreased PT and improved economy of movement with increasing arthroscopic experience	HMA has construct validity in this setting
Insel et al ⁴² , 2009	To develop and validate an objective model for assessing basic arthroscopic proficiency	Live observation	59 PGY 1-5	Diagnostic Knee Arthroscopy	BAKSS	Strong correlation between GRS scores, YIT and number of previous arthroscopies performed ($r = 0.88$, $p < 0.01$). TSC scores were moderately correlated with YIT ($r = 0.73$, $p < 0.01$) and number of previous arthroscopies ($r = 0.64$, $p < 0.01$)	The BAKSS tool is construct valid for assessing proficiency in diagnostic knee arthroscopy
Jackson et al ¹³ , 2012	Demonstrate learning curve for meniscal repair and determine impact of task repetition on retention of this skill	Live observation SDM	19 PGY?	Arthroscopic meniscal repair	HMA	All subjects demonstrated a clear learning curve during the initial learning phase. There was no loss of skill seen after a 6 month break on task repetition	HMA can be used to model learning curves in arthroscopic meniscal repair
Khanduja et al ⁴³ , 2017	To test the construct validity of the hip diagnostics module of	SDM	10 PGY 1-6	Diagnostic hip arthroscopy	PT, error rating	Increased experience in hip arthroscopy was reflected by	Construct validity of SDM demonstrated

	a virtual reality hip arthroscopy simulator					significantly better performance on the simulator across 2 tasks	
Koehler et al ⁸⁰ , 2013	To evaluate the content and concurrent validity and reliability of the ASSET for performance of diagnostic knee arthroscopy	Expert assessor review of video footage	28 PGY 1-5	Diagnostic knee arthroscopy	ASSET	Significant score differences between novice, intermediate and advanced groups was seen. Scores were strongly correlated between raters ($r=0.91$, $p<0.01$) and for sequential procedures by each surgeon ($r=0.79$, $p<0.01$)	Construct and concurrent validity, inter-rater reliability and test-re-test reliability were demonstrated for diagnostic knee arthroscopy in the cadaveric model
Koehler et al ⁷⁹ , 2013	To test the validity and reliability of the ASSET as a pass-fail examination of arthroscopic skill	Expert assessor review of video footage	28 PGY?	Diagnostic knee arthroscopy	ASSET	Raters agreed on pass-fail rating for 56/60 videos (ICC=0.83). Logging >80 arthroscopic cases or performing more than 35 knee arthroscopies was predictive of passing	ASSET is valid and reliable as a pass-fail examination of diagnostic arthroscopy of the knee in the simulation setting
Martin et al ⁴⁴ , 2011	Evaluate the correlation between timed task performance in an arthroscopic shoulder simulator and timed task performance in a cadaveric shoulder arthroscopy model	Live observation, SDM	15 PGY?	Diagnostic shoulder arthroscopy	PT	PT on the simulator was strongly correlated with PT on the cadaveric model ($r=0.736$, $p<0.001$), and experience predicted performance in both models ($p=0.016$)	PT is a construct valid measure of arthroscopic skill in VR and cadaveric simulation settings
Martin et al ⁴⁵ , 2012	To evaluate the correlation between timed task performance on an arthroscopic shoulder simulator and resident experience	SDM	27 PGY 1	Diagnostic shoulder arthroscopy	PT	PT correlates with experience. For every PGY increment, there was a 16 second improvement in the time required to complete the simulator task ($p<0.005$)	PT as measured on the simulator is a construct valid measure of arthroscopic skill

Martin et al ¹⁴ , 2015	To determine if low-fidelity arthroscopic simulation training improves basic ankle arthroscopy performance	Expert assessor review of video footage	29 PGY 1-5	Diagnostic ankle arthroscopy	ASSET	Simulation group outperformed the control group in total ASSET and checklist scores ($p < 0.001$)	ASSET is construct valid
Martin et al ¹⁰ , 2016	To evaluate the correlation between timed task performance on a VR shoulder arthroscopy simulator and participation in an expert arthroscopy course	SDM	99 PGY (mean) 3	Diagnostic shoulder arthroscopy	PT, path length	Significant improvements in PT and path length (camera and probe tip) were seen after training	PT and path length are construct valid in this VR model
McCarthy et al ⁴⁷ , 1999	To evaluate the construct validity of the Sheffield Knee Arthroscopy Trainer	SDM	6 PGY?	Diagnostic knee arthroscopy	PT, probe collisions	Experienced surgeons performed best with fewer instrument collisions	PT and error counting is construct valid measure of arthroscopic skill
McCarthy et al ⁴⁸ , 2006	To evaluate the Sheffield Knee Arthroscopy Training System	SDM	13 PGY?	Diagnostic knee arthroscopy	PT, path length, collisions	Experienced surgeons performed better with shorter procedure time, probe path length and found more pathology. After SKATS training, novices showed improvements across all domains	SDM construct valid in this simulator model
Middleton et al ⁸¹ , 2016	To compare three GRS tools for the assessment of simulated arthroscopic skills	SDM, expert assessor review of video footage	21 PGY?	Knee: diagnostic arthroscopy, basic triangulation and medial meniscectomy. Shoulder: diagnostic arthroscopy, basic triangulation task, advanced	ASSET, BAKSSS, IGARS	All GRS demonstrated construct validity with significant differences between each skill level and arthroscopic task. IRR was high for all. Correlation with time taken and path length was significant for all	No single GRS tool demonstrated superiority

				triangulation task			
Nwachukwu et al ⁸² , 2016	To test the ability of a procedure specific checklist to detect performance improvement over time	Live assessment	21 PGY 3	Diagnostic knee arthroscopy, diagnostic shoulder arthroscopy	Diagnostic Shoulder Arthroscopy Checklist, Diagnostic Knee Arthroscopy Checklist	Mean time to checklist (procedure) completion improved pre- and post- intervention (6/52)	The checklists are construct valid to measure operative efficiency. They do not measure quality of performance.
Olson et al ⁸³ , 2013	To test the construct validity and reliability of a modified BAKSSS (GRS only) using mixed-level assessors	Live assessment	23 PGY1-5	Diagnostic knee arthroscopy	BAKSSS GRS	The modified BAKSSS demonstrated construct validity with junior residents achieving lower scores (mean 20 vs mean 30 for senior residents).	The BAKSSS GRS showed construct validity and fair inter-rater reliability
Pedowitz et al ⁴⁹ , 2002	To evaluate the construct validity of a novel shoulder VR simulator	SDM	22 PGY?	Diagnostic shoulder arthroscopy	Instrument path length/ratio, error scoring (collisions/injuries)	Test time and path ratio differed significantly as a function of surgical experience	Limited construct validity of SDM
Pedowitz et al ⁸⁴ , 2015	To a) assess a new biomechanical assessment of arthroscopic knots and to b) establish proficiency benchmarks using the fundamentals of arthroscopy trainer	Biomechanical testing	44 PGY4-5	Arthroscopic knot tying	FPA	Performance was inconsistent between experience levels, and failure rate could not predict experience level	The results suggest that FPA of arthroscopic knot quality is not construct valid, or that there might be performance heterogeneity amongst even the most experienced arthroscopic surgeon
Phillips et al ⁸⁵ , 2017	Evaluate the use of a TSC, ASSET, PT and GRS for assessing performance of arthroscopic hip labral repair in a dry model	Expert assessor review of video footage	37 PGY NS	Arthroscopic hip labral repair	TSC, ASSET, novel GRS, PT	Dry models to assess the performance of arthroscopic labral hip repair by residents is both reliable and valid	ASSET demonstrated IC, IRR, construct validity and concurrent validity

Pollard et al ⁵⁰ , 2012	To compare learning curves in diagnostic hip arthroscopy on a low fidelity simulator, in lateral and supine positions	SDM	20 PGY (1-4)	Diagnostic hip arthroscopy	PT, HMA	Both groups demonstrated learning in all parameters (p<0.001). Junior participants achieved performance parity with senior participants after 12 attempts	Construct validity of SDM demonstrated
Price et al ⁸⁶ , 2015	Evaluate the number of arthroscopies needed to achieve consultant level performance	Expert assessor review of video footage, SDM	28 PGY (1-8)	Diagnostic knee arthroscopy	BAKSSS, GRS, HMA, PT	There was significant improvement in performance with increasing experience (p<0.05)	BAKSSS, PT and HMA is construct and concurrent valid
Rahm et al ⁵¹ , 2016	To assess face and construct validity of a VR based model for shoulder arthroscopy	SDM	25 PGY?	Diagnostic hip arthroscopy	PT, Instrument path length	Simulator positively rated as educationally valuable. Experts were significantly faster and demonstrated better economy of movement	Construct validity of metrics demonstrated
Rahm et al ¹⁵ , 2018	Test a standardized, competency based training protocol on a VR arthroscopy simulator	Expert assessor review of video footage, SDM	20 PGY 1-5	Diagnostic knee and shoulder arthroscopy	ASSET, PT, Instrument path length	The residents performance significantly improved post-training in ASSET and SDM parameters. Expert ASSET score was significantly higher than the residents post-training score,	ASSET was construct valid. SDM showed a mixed validity picture; for shoulder it was construct valid, but not for knee
Rebolledo et al ¹⁶ , 2015	To compare performance of arthroscopic simulator trained and didactic-trained residents in diagnostic knee arthroscopy	Live observation	14PGY 1-2	Diagnostic shoulder arthroscopy	PT, IGI	Participants trained on the simulator outperformed the didactic trained group on the cadaveric model test by PT (-35%, p=0.02) and IGI (-35%, p=0.01)	Educational value of simulator demonstrated. IGI appears to be construct and concurrent valid but not formally evaluated here
Rose et al ⁵² , 2015	To assess the construct validity of three skill-	SDM	10 PGY NS	Basic arthroscopy skills	PT, HMA	2 assessments (Steady and Probe and Track a Moving Target) demonstrated construct	VR task deconstruction could be used for assessment and development of

	deconstructed VR models					validity. 1 assessment (Steady and Telescope) did not	early arthroscopic skills
SladeShantz et al ⁸⁷ , 2013	To determine whether a global assessment of arthroscopic skills was valid for blinded assessment of cadaveric diagnostic arthroscopy	Expert assessor review of video footage	13 PGY 1-3	Diagnostic knee arthroscopy	OOAS	The agreement between global assessment scores was strong (ICC=0.80). Internal consistency was excellent (Cronbach's alpha=0.97) and test re-test reliability was strong. The test is construct valid and able to discriminate between various levels of training (p<0.0001)	The OOAS is reliable and construct valid in this setting
Srivastava et al ⁵³ , 2004	To evaluate the construct validity of a novel shoulder simulator	SDM	21 PGY?	Diagnostic shoulder arthroscopy	PT, collisions/ errors	The expert group performed the task more quickly (p=0.013) and more accurately (p=0.002)	Construct validity of procedural accuracy demonstrated. PT not construct valid in this model
Tuijthof et al ⁵⁴ , 2010	To evaluate face and construct validity of a knee arthroscopy simulator	SDM	8 PGY?	Diagnostic knee arthroscopy, diagnostic shoulder arthroscopy	PT	The experienced (consultant) group were more time efficient than the residents in task completion for each repetition (p<0.05). The participants reported overall good face validity of the simulator	Face and construct validity of the simulator is demonstrated
Tashiro et al ⁵⁵ , 2008	To test the construct validity of a knee arthroscopy simulator with force evaluation	SDM	24 PGY?	Diagnostic and therapeutic knee arthroscopy	PT, instrument path length, instrument forces	The experienced group performed both tasks more efficiently and competently than the less experienced groups. Path length was shorter, velocity faster and forces applied lower	SDM construct valid in this model

Tofte et al ⁵⁶ , 2017	To validate a VR arthroscopy simulator	SDM	35 PGY 1-6	Diagnostic knee and shoulder arthroscopy	PT, instrument path length	Significant correlations between experience level and performance as measured by PT and instrument path length	PT and SDM construct valid in this model
Wong et al ¹²³ , 2015	To test the construct validity of the Arthroscopic Knot Trainer	Expert assessor review of video footage	21 PGY 1-3, 11 PGY 4-5	Arthroscopic knot tying	PT, number of knots tied in pre-specified time limit	The simulator could discriminate performance by level of experience. Positive qualitative feedback on educational value	PT construct valid in this model
STUDIES IN LIVE THEATRE SETTING							
Cannon et al ¹⁷ , 2014	To assess the transfer validity of skills learnt on a VR simulator to the operating room	Expert assessor review of video footage	54 PGY 3	Diagnostic knee arthroscopy	TSC and GRS	The simulator trained group performed significantly better as measured by the TSC (p=0.031), including probing skills (p=0.016) but not visualisation skills (p=0.34), compared to the control group (the TSC weighted probing skills double the weight of visualisation skills). The GRS failed to reach significance, probably because of an extreme outlier	The TSC has a mixed validity picture, and the weightings of probing skills vs visualisation skills may need adjusting. GRS not construct valid in this setting
Dunn et al ¹⁸ , 2015	To test the impact of a simulation training intervention on a) the initial skill improvement, and b) retention of skill longitudinally in performance of diagnostic shoulder arthroscopy	Live observation	17 PGY 1-5	Diagnostic shoulder arthroscopy	ASSET	No performance difference between simulator trained and untrained at baseline, the simulator trained group improved as compared to baseline in mean ASSET (p=0.023) and PT (P=0.01). The training effect was lost by 12 months follow-up.	ASSET is construct and concurrent valid in the live theatre setting

Gallagher et al ⁸⁸ , 2018	To determine the IRR of a TSC	Expert assessor review of video footage	44 PGY 4-5	Arthroscopic Bankhart Repair	TSC	Mean IRR = 0.93 (range 0.84-0.99)	High IRR demonstrated, tool suitable for high stakes assessment
Garfield Roberts et al ³¹ , 2019	Investigate transfer validity of simulation training using elbow-worn motion sensors	SDM	30 PGY 2-3	Diagnostic Knee Arthroscopy	Elbow motion analysis, PT	The intervention group outperformed the control group in all metrics	Intra-operative elbow motion analysis metrics are construct valid in the live theatre setting
Hodgins et al ⁸⁹ , 2014	To describe the learning curve for diagnostic knee arthroscopy	Live observation	20 PGY 1-5	Diagnostic Knee Arthroscopy	BAKSSS, GRS, LC-CUSUM	Competency as assessed by the TSC was achieved by 40% of trainees after a median of 16 procedures and by the GS for 1 trainee	LC-CUSUM is an effective method to evaluate procedural competence in arthroscopic training and can provide objective feedback and benchmarks in the learning phase
Howells et al ¹⁹ , 2008	To evaluate the transfer validity of arthroscopic skills from simulator training to the operating theatre	SDM, live observation	20 PGY 1-2	Diagnostic Knee Arthroscopy	HMA, OSATS GRS and PBA	Simulator trained group performed significantly better than untrained group using PBA and OSATS GRS demonstrating the transfer of skills from simulator to live theatre	HMA has construct and criterion validity in this setting
Hoyle et al ⁹⁰ , 2012	Develop and validate a new GRS for shoulder arthroscopy	Expert assessor review of video footage	10 PGY?	Diagnostic shoulder arthroscopy	GRSSA	Good construct validity, mixed reliability	GRSSA is construct valid for assessing performance in diagnostic shoulder arthroscopy, and can be used remotely from the procedure
Koehler et al ⁹¹ , 2015	To determine the validity and reliability of using ASSET to assess arthroscopic skill in the operating room	Expert assessor review of video footage	8 PGY 3-5	Diagnostic shoulder and knee arthroscopy	ASSET	The senior group achieved significantly higher mean ASSET scores compared to junior group for both procedures. ICC for total scores was good (knee = 0.81, shoulder	ASSET is feasible, reliable and construct valid for assessing diagnostic arthroscopy in live theatre

						= 0.84). Raters concurred on pass-failure evaluation in 89% of procedures	
Talbot et al ¹² , 2015	To assess the reliability and validity of the shoulder OPAT when performing diagnostic shoulder arthroscopy	Live observation and expert assessor review of video footage (for IRR assessment)	6 PGY 3-10	Diagnostic shoulder arthroscopy	Shoulder OPAT	Internal consistency = 0.77, IRR=0.6, IRR=0.82. Face, content, construct and concurrent validities were demonstrated.	Shoulder OPAT fulfils several aspects of reliability and validity and is perceived as superior to PBA
Waterman et al ²⁰ , 2016	To assess the training impact of a shoulder arthroscopy simulator model	SDM, Expert assessor review of video footage	22 PGY?	Diagnostic shoulder arthroscopy	ASSET, PT instrument path length	The simulator trained group showed improvement in ASSET score in the live theatre environment. There was no significant improvement in PT seen in the live environment	ASSET is construct valid in this setting.

*TSC part-validated in live theatre environment, study took place in simulation lab. ∞ Validation involved multiple surgical specialities including (T&O), specialty specific data was not provided.

Abbreviations; ACLR = Anterior Cruciate Ligament Reconstruction, TSC = Task Specific Checklist, ASSET = Arthroscopic Surgical Skill Evaluation Tool, PT = Procedure Time, FPA = Final Product Analysis, SDM = Simulator Derived Metrics, IGARS = Imperial Global Arthroscopy Rating Scale, GRS = Global Rating Scale, BAKSSS = Basic Arthroscopic Knee Skills Scoring System, GRSSA = Global Rating Scale for Shoulder Arthroscopy, HMA = Hand Motion Analysis, VR = Virtual Reality, PGY = Postgraduate Year, NS = Not Specified, RCR = Rotator cuff repair, LR = Labral repair, OPAT = Objective practical assessment tool, OAAS = Objective Assessment of Arthroscopic Skills, LC-CUSUM = Cumulative Summation Test for Learning Curve, IGI = Injury Grading Index, NOF = Neck of Femur, TAD = Tip Apex Distance, COR = Cut Out Rate.

Table 2. Studies Assessing Open Surgical Performance

Author	Study Aim	Assessment Format	Resident Participants	Skills Assessed	Tools or Metrics	Results	Findings Relating to Assessment Tool(s)
STUDIES IN SIMULATED SETTING							
Akhtar et al ⁵⁸ , 2015	To assess the construct validity of a VR trauma simulator for performing DHS fixation of trochanteric fracture	SDM	10 PGY1-4, 10 PGY5-12	DHS fixation of trochanteric fracture	FPA, fluoroscopy time	Statistically significant differences seen across 5 of 6 performance indices (excluding total procedure time) which correlated with frequency of exposure of operating	Construct validity of FPA demonstrated in the simulated model
Anderson et al ⁹³ , 2016	To determine if OSATS predicts surgical quality of the procedure	Live observation, Biomechanical testing, FPA	21 PGY 1-2 (A) + 30 PGY 2-5 (B)	A) Tibial Plafond Fracture Reduction B) Distal radius ORIF	OSATS, FPA	OSATS did not correlate with the articular reduction quality (A) or the integrity of the mechanical fixation (B)	OSATS do not effectively assess the quality of the surgical result
Aoude et al ⁵⁹ , 2016	To assess the utility of computer-assisted surgery in pedicle screw placement	Expert assessor review of CT images	24 PGY 1-4	Pedicle screw placement	FPA	Experience level did not predict screw placement accuracy	FPA not construct valid in this setting
Backstein et al ⁹⁴ , 2004	To compare the effect of three types of feedback on resident post-test performance; control group (no feedback), video and self-review and expert feedback	Expert assessor review of video footage	29 PGY 1-5	Plating of long bone fracture, TBW olecranon, Z-plasty	Video assessment with feedback	GRS scores across the 3 groups and tasks were not significantly different	Video feedback assessment failed to demonstrate an improvement in technical skills
Bergeson et al ⁹⁵ , 2008	To evaluate the early learning curve of pedicle screw placement	Live observation	3 (PGY1 and 3)	Thoracic pedicle screw placement	FPA	Acceptable placement accuracy levels were achieved by the fourth attempt	Construct validity and educational impact of FPA is demonstrated
Bernard et al ⁹⁶ , 2016	Compare the reliability and validity of OSATS checklist, GRS score and subjective pass/fail to assess resident	Live observation	23 PGY 1-5	Open surgical approaches to the shoulder (deltopectoral, lateral deltoid-splitting and posterior)	OSATS checklist, GRS, pass-fail assessment	Concurrent validity shown between OSATS and GRS for the 3 shoulder approaches. OSATS has superior reliability compared with GRS and pass/fail	OSATS checklists are a valid and reliable assessment of technical skills across 3 open

	operative skill in shoulder surgery						surgical shoulder approaches
Blyth et al ⁶⁰ , 2008	To test the construct validity of a computer-based VR simulator	SDM	6 PGY1-2, 6PGY3-5	DHS	PT, Composite score derived from simulator metrics	Accuracy, number of x-rays and speed were significantly different between experience levels, critical error rate was inversely associated with experience level. PT was not predictive of experience	x-ray use, speed and critical error rate demonstrate construct validity. PT did not in this model
Burns et al ⁹⁷ , 2017	To evaluate the educational value of low fidelity ORIF simulation and objective measures of performance using biomechanical means	Biomechanical testing	16 PGY 1	Ulnar ORIF	FPA	Pre-to post-simulation operative success rates were significantly improved (p<0.001) and were maintained at 3/12 follow-up	Biomechanical construct failure testing is a valid method of measuring operative success in the simulated setting
Butler et al ^{*21} , 2017	Evaluate a training course for interns on closed reduction and pinning of paediatric supracondylar fractures	Live observation and written examination	19 PGY 1-5	Closed reduction and pinning of paediatric supracondylar fractures	TSC and written exam	Post-training there was no difference in MCQ scores and TSC score comparing interns and senior residents (significant difference prior)	Written examination and TSC construct valid in this model
Christian et al ⁶¹ , 2018	Assess whether VR simulation platform can distinguish between novice and experienced surgeons when performing percutaneous hip pinning	SDM	17 PGY2-5	Percutaneous fixation of valgus impacted fractured NOF	FPA, PT, fluoroscopy time	Significant association between performance and experience in 10/15 outcome measures.	FPA and PT construct valid in this model. Fluoroscopy time not construct valid.
Froelich et al ⁶² , 2011	Evaluate the construct validity of a haptic VR surgical simulator for assessing performance in centre-centre	SDM	15 PGY1-5	Guidewire insertion for DHS	FPA, PT, fluoroscopy time	No difference seen between groups for procedure time or TAD. Significant difference observed in final wire position on lateral view (p=0.01), wire passes	Mixed picture of construct validity although some measures show discriminatory ability. Potential

	guidewire insertion during intertrochanteric proximal femoral fracture fixation					(p=0.03) and fluoroscopy time (p=0.05)	for educational impact
Giurin et al ⁶⁴ , 2018	To evaluate a low fidelity simulator for nail bed repair	Live observation	12 PGY?	Nail-bed repair	PT	PT was significantly faster for the experienced group (p=0.01)	PT is construct valid in this model
Gottschalk et al ²² , 2015	To analyse the training effect of 3D simulation on lateral mass screw placement	Expert assessor review of images	15 PGY 1-6	Lateral mass screw placement	FPA	Subjects in the 3D training groups showed significantly improved drilling trajectories as compared to the controls (p<0.0001)	FPA as measured by aggregate mean difference from ideal has construct validity in this model
Hohn et al ⁶⁵ , 2014	To evaluate a novel set of low fidelity bone training models	Live observation	15 PGY 1-5	Basic orthopaedic surgical skills	FPA	The model was feasible. No significant difference in performance between experience levels was seen	FPA was not construct valid in this low fidelity model
LeBlanc et al ¹²⁴ , 2013	To compare task performance on a novel fracture simulator model compared to a bench-top model	Live observation	22 PGY 1-5	Surgical fixation of ulnar fracture	OSATS, PT	Both simulators distinguished between different experience levels, participants performed significantly better on the virtual simulator compared to bench model in all measures except PT (<0.05)	Construct validity for OSATS demonstrated, not for PT in this model
Leong et al ⁶⁶ , 2008	To evaluate the validity of three low cost models of fracture fixation in the assessment of technical skills	HMA, Live observation	15 PGY?	DCP application, IM nail insertion, wrist ex-fix application	HMA, OSATS GRS	OSATS scored differed significantly between the 3 groups and 3 procedures with high IRR (Cronbach's alpha = 0.88). Motion analysis distinguished between the three groups on the DCP model but a ceiling effect was observed in the IM nail and ex-fix procedures	The DCP cadaveric porcine model has face, content and construct validity. The IM nail and Ex fix models are less sensitive and further work is needed

Lopez et al ⁶⁷ , 2015	To assess the construct validity of a cost effective psychomotor assessment tool; the Fundamentals of Orthopaedic Surgery (FORS)	Live observation	58 PGY?	Basic psychomotor surgical tasks	FORS	Stratification by experience level was seen in all 6 tasks	Construct validity demonstrated. Educational benefit seen in repetitious practice
MacEwan et al ⁹⁸ , 2016	To compare the O-SCORE to the OSATS in assessing performance of a simulated radius ORIF	Expert assessor review of video footage	19 PGY 1-5	ORIF midshaft radius	O-SCORE, OSATS checklist, GRS	O-SCORE demonstrates accurate and reproducible results compared to current gold standard tools (OSATS/GRS) when used in a randomised blinded fashion	Construct, concurrent validity and reliability of O-SCORE demonstrated
Mayne et al ⁶⁸ , 2016	To evaluate a distal radius fracture simulator	Expert assessor review of video footage	20 PGY 1-5	Distal radius ORIF	OSATS, FPA	Significant performance differences between junior and senior residents was seen with OSATS ($p < 0.001$). No difference in FPA was seen. IRR was high for all measures	OSATS is construct valid. FPA is not in this model.
Nousiainen et al ²⁴ , 2012	To compare performance change in naïve participants undergoing training in cannulated hip pinning on a simulator, using either computer assisted or conventional fluoroscopic techniques	SDM	52 PGY1/Medical Students	Cannulated hip pinning for fractured neck of femur	FPA, Image intensifier use	All participants improved in hardware placement accuracy after the training ($p < 0.001$) and the skill level was retained at post-testing, retention testing at 4/52 and transfer testing onto the opposite study arm. No significant change in guidewire parallelism was seen. Radiation time and dose varied by intervention but not reported by experience level	FPA and image intensifier use were construct valid in this setting
Pedersen et al ⁹⁹ , 2014	To develop and validate a tool to assess hip fracture	SDM	10 PGY1	3 internal fixation procedures for	Hip fracture performance score	The combined score could discriminate between experience level.	Construct validity and reliability of

	fixation performance and set pass/fail benchmark			undisplaced femoral neck fracture		Inter-case reliability was 0.83. Non of the individual SDM demonstrated discriminatory abilities	the score demonstrated
Putnam et al ¹⁰⁰ , 2015	Assess whether standardised OITE, OSATS and simulation based computer-animated testing can predict biomechanical construct performance following volar plating of distal radial fractures in cadaver	Live observation, SDM and biomechanical testing	15 PGY 3-4	Volar plating of distal radius fracture	Fracture specific knowledge test, OITE, OSATS, GRS, TSC	No statistically significant correlation seen between performance on biomechanical testing and that of knowledge tests, OITE, GRS or TSC	Traditional written and computer-based testing methods failed to predict which participants fracture constructs would pass biomechanical testing
Qassemyar et al ⁶⁹ , 2014	Evaluate the face and construct validity of a low-fidelity hand trauma examination model	Live observation	19 (PGY 3-4)	Hand trauma skills; Z-plasty, metacarpal fracture fixation, tendon repair	OSATS	There was significant performance differences seen between microsurgery experience levels	OSATS was construct valid in this model
Rambani et al ⁷⁰ , 2014	To validate a desktop simulator for spinal surgery	Live observation and SDM	12 PGY?	Lumbar pedicle screw insertion	TSC	Improved performance in simulator trained group	TSC appears to show construct validity in this model
Ruder et al ²⁵ , 2017	Evaluate the effectiveness of a training session in orthopedic drilling technique	Live observation	5 PGY 1	Bone drilling	FPA	Plunge depth was significantly reduced in both groups post-training (p<0.05)	FPA was construct valid in this model
Sonnadara et al ²⁶ , 2011	To assess whether an intensive lab based skills course at the start of orthopedic residency is effective for teaching core skills	Live observation	22 PGY 1	Basic orthopedic surgical skills	OSATS	Residents in the simulation trained group performed better in in both checklist and GRS components of OSATS in post-training assessment	OSATS are a construct valid
Sonnadara et al ²⁷ , 2012	To examine retention rates for basic surgical skills	Live observation	18 PGY1-3	Basic orthopedic skills	OSATS	Mean GRS score for competency based curriculum (intervention)	OSATS can be used longitudinally to

	taught at the beginning of residency					group was 4.3 and maintained 6/12 later. Both intervention and senior control group performed better than the junior resident control group ($p < 0.001$)	assess skill retention
Sonnadara et al ²⁸ , 2013	Compare student-led and instructor-led techniques in improving performance in 4 skills during a surgical boot camp	Live observation	12 PGY1	Basic orthopedic skills	OSATS	Checklist and GRS scores were improved in the student led group	OSATS is construct valid
Shi et al ⁷¹ , 2018	Assess validity of VR simulator to teach lumbar pedicle screw placement by comparing VR trained and untrained resident performance on cadaveric specimens	Expert assessment of final product (scan images)	10 PGY? (procedure naïve)	Lumbar pedicle screw insertion	FPA	VR trained group had significantly lower screw penetration rates, improved penetration distances and placement acceptability compared to the untrained group	FPA in this setting is face, content and construct valid
Sugand et al ⁷² , 2018	To validate a digital fluoroscopic simulator for guidewire insertion	SDM	26 PGY 2-9	Dynamic Hip Screw guidewire insertion	PT, FPA	The expert group achieved significantly better TAD and COR. PT was not significantly different between experts and novices	TAD and COR had construct validity. PT did not in this model
Tonetti et al ²⁹ , 2009	To evaluate the educational value of a path simulator under fluoroscopic guidance in performing percutaneous iliosacral screw insertion	Live observation	23 PGY?	Percutaneous iliosacral screw insertion simulator	Number of x-rays used, Iatrogenic Index	Significant differences in the number of intra-operative x-rays was seen between experience, prior procedural and technical knowledge subgroups. Iatrogenic index scores were not significantly different between groups	Intra-operative x-ray use has construct validity in this setting. Iatrogenic index failed to show construct validity

Williams et al ¹⁰¹ , 2017	Test the utility of a battery of 5 motor skills tests (Grooved PegBoard test and 4 novel tests) to evaluate potential orthopedic residents	Live observation	30 PGY1-4 residents	Closed fracture reduction, drilling, dexterity and visuospatial skills	Psychometric test metrics	Residents performed better than non-residents ($p < 0.0001$) in every exercise but the drilling test	Concurrent validity demonstrated by comparison to Grooved PegBoard test (internal control). Construct validity demonstrated in 3 of 4 tests
Van Heest et al ¹⁰² , 2009	To evaluate the reliability and validity of 4 tests for assessing competence in performing CTR; knowledge test, detailed TSC, GRS, pass/fail assessment	Live observation	28 PGY 1-6	CTR	TSC, GRS, pass/fail test, knowledge test	Correlation between YIT and knowledge scores ($p < 0.001$) YIT and detailed checklist scores ($p = 0.002$), YIT and GRS scores ($p = 0.04$) and YIT and pass/fail ($p < 0.001$). No correlation was seen between YIT and PT.	Knowledge test and OSATS construct valid
Van Heest et al ¹⁰³ , 2012	To evaluate OSATS for 3 common upper extremity procedures	Live observation	27 PGY 2-5	TFR, CTR, DRFF	TSC, GRS, PT, pass/fail test	Construct validity between YIT and checklist scores demonstrated for TFR and CTR, between YIT and GRS scores for TFR and DRFF and between YIT and pass/fail for TFR. Criterion validity demonstrated between GRS, checklist scores and pass/fail for all procedures. Reliability poor. Participants rated OSATS as educationally valuable	OSATS construct valid and educationally valuable. Not reliable in this model
Xiang et al ³⁰ , 2014	Test the impact of a pre-operative planning simulator on junior surgeons' pedicle screw insertion accuracy	Live observation, SDM	2 PGY?	Pedicle screw insertion	PT, FPA	PT and positional accuracy improved with training (self-controls) and experts outperformed trainees	PT and FPA construct valid here

Yehyawit et al ⁷³ , 2013	1. To develop a physical model to improve articular fracture reduction skills. 2. To develop objective assessment methods and 3. To assess the construct validity of the simulation	HMA, Live observation	7 PGY1/2, 5PGY4/5	Complex articular fracture fixation	FPA, PT, HMA metrics, GRS	Significant difference seen in cumulative hand distance travelled between junior and senior residents. There was no difference seen in FPA, PT or GRS	Hand movement economy has construct validity in this setting
STUDIES IN LIVE THEATRE SETTING*							
Beard et al ¹⁰⁴ , 2011	Compare the acceptability, reliability and validity of three WBA methods of assessing surgical skill	Live observation	Not stated	Not stated	PBA	PBA showed high reliability for assessing same index procedure (G>0.8) Good correlation was seen with PBA scores and specialist training level, years of surgical training and recent index exposure (r=0.31-0.71). Within test correlations were strong between checklist and global ratings for PBA (r=0.73)	PBA showed construct and concurrent validity, reliability and feasibility
Davies et al ¹⁰⁵ , 2018	Develop and validate a new operative assessment tool that addresses current barriers	Live observation	49 PGY 3-10	CTD	GOSLE	Strong correlation between PBA and PBA ratings (r=0.87, p<0.001). Positive participant feedback on feasibility and educational impact	Psychometric performance of the GOSLE is comparable to the PBA but is preferred by trainees and trainers. High-quality, actionable feedback better supports formative assessment
Gofton et al ¹⁰⁶ , 2012	To pilot and evaluate the validity and reliability of the O-SCORE	Live observation	22 PGY1-5	ORIF Wrist, ORIF Hip, ORIF ankle, hemiarthroplasty,	O-SCORE	Total performance scores were discriminated by year, PGY 1-2 lower than PGY3 (p<0.001) and	O-SCORE successfully demonstrates construct validity,

				THA, knee arthroscopy		PGY 3 lower than PGY 4-5(p<0.02)	reliability, feasibility and educational impact across various PGY levels regardless of procedure type
Hawkes et al ¹⁰⁷ , 2017	To evaluate the Fracture Fixation Assessment Tool for assessing the quality of surgical fixation	Expert assessor review of final product (x-rays)	Unknown number of residents performed test cases	Various fracture fixation procedures; ORIF bimalleolar ankle/radius/ulnar/distal humerus/tibial plateau, TBW Olecranon/patella, IM nail tibia, DHS	The Fracture Fixation Assessment Tool	Internal consistency = 0.98, inter-rater reliability =0.71, CVR = 0.65	Internal consistency, reliability and content validity is demonstrated. May have applications in service evaluation as well as training
Hoffer et al ¹⁰⁸ , 1990	To see if psychomotor tests on entry and exit to specialist training correlate with faculty ranking of surgical skill	Live observation	8 PGY 1	Unspecified	The Crawford small-part dexterity test, steadiness hole test (unsupported), Bennett hand-tool test. Purdue peg board test and Minnesota rate manipulation test	The rankings from the Crawford and steadiness tests had significant correlation to subjective performance ranking by faculty	Psychomotor performance at entry and exit of specialist training appears to predict surgical skill as assessed by subjective measures
Marriott et al ¹⁰⁹ , 2010	Evaluate the validity, reliability and acceptability of the PBA	Live assessment	81 PGY? Across 6 specialities (7 in T&O)	Hip replacement and Knee replacement	PBA	Construct validity demonstrated by correlation of PBA scores with measures of surgical training and experience. Reliability of the adjusted total item score and GSS for any given procedure was achieved using 4 and 3 assessor judgements	PBA demonstrates construct validity, this evidence is limited to 2 T&O procedures performed by 7 trainees

Ω - Number includes medical students. DHS = Dynamic Hip Screw, ORIF = Open reduction internal fixation, TBW = Tension Band Wire, IM = Intramedullary, FORS = Fundamentals of Orthopaedic Surgery, THA = Total Hip Arthroplasty, TFR = Trigger finger release, CTR = Open carpal tunnel release, DRFF = Distal radius fracture fixation, Ex-fix = external fixator, DCP = dynamic compression plate, GSS = Global Summary Scale, GOSLE = Generic Operative Supervised Learning Event, PBA = Procedure Based Assessment, T&O = Trauma & Orthopaedics.

Table 3. Utility evidence of the assessment tools used to evaluate surgical competency in T&O

OBJECTIVE ASSESSMENT OF TECHNICAL SKILL										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
TASK SPECIFIC CHECKLISTS										
OSATS Task Specific Checklist ^{23, 26-28, 61, 68, 77, 96, 98, 102, 103, 127, 128}	Checklist used to evaluate performance of individual parts of a procedure – binary yes/no descriptors	Can be used by non-expert assessors. Useful for teaching trainees sequencing of procedural steps	Quality of performance and outcome are not captured. Rigid binary scoring does not allow for acceptable deviation from standard procedural steps. Early ceiling effect.	Y ⁹⁶	Y ⁹⁶	Y ^{26-28, 68, 77, 96, 98, 102, 103, 127} N ¹⁰⁰	Y ^{77, 96, 98, 102, 103}	Y ^{96, 98, 127, 128, N¹⁰³}	Y ^{26, 27, 96}	Y ^{28, 103}
Non-OSATS Task Specific Checklist ^{17, 33, 37, 73, 85, 88, 114}	Procedure deconstructed into steps, often by Delphi consensus. Largely binary descriptors.	As above	As above	Y ^{33, 37, 85, 114}	Y ^{33, 37, 85, 114}	Y ^{37, 85, 114} M ¹⁷ N ⁷³	Y ^{37, 85}	Y ^{37, 88, 114}	Y ^{37, 114}	-
GLOBAL RATING SCALES										
OSATS Global Rating Scale ^{26-28, 61, 66, 68, 69, 77, 93, 96, 98, 100, 102, 128}	Objective Assessment of 7 generic open surgical skill domains (respect for tissue, time and motion, instrument handling, knowledge of instruments, flow of operation, use of assistants). 5 point Likert scale with middle and extremes	Captures quality of performance, not procedure specific, can assess complex procedures where there is more than one acceptable method. Can discriminate between	Need expert surgeon evaluators. More time consuming. Does not assess skills specific to procedure	Y ⁹⁶	Y ⁹⁶	Y ^{19, 26-28, 66, 68, 69, 77, 96, 98, 102, 103} N ^{93, 100}	Y ^{77, 96, 98, 103} N ⁹³	Y ^{66, 96, 128, N¹⁰³}	Y ^{26, 27, 96}	Y ^{28, 103}

	anchored by explicit descriptors(1)	'competent' and 'expert' performance.								
Non-OSATS Global Rating Scale ^{37, 85}	5 point scale corresponding to Dreyful model of skill acquisition (novice, advanced beginner, competent, proficient, expert)	Captures quality of performance. Quick.	As above	NR	NR	Y ^{37, 85}	Y ⁸⁵	Y ³⁷	NR	Y ⁸⁵
Non-OSATS procedure specific GRS A ^{89, 127} and B ¹⁷	A) 2 item GRS on 5 point scale (1=very poor, 2 = clearly superior), items are overall performance and quality of final product. B) 7 global impression items evaluated on an adjectival scale from 0 to 4 (4 best)	Captures quality of performance	As above	NR	Y ¹⁷	Y ^{89, 127} N ¹⁷	Y ¹⁷	Y ^{17, 127}	NR	NR
ARTHROSCOPIC OBJECTIVE SKILLS ASSESSMENT TOOLS										
Basic Arthroscopic Knee Skill Scoring System (BAKSSS) ^{42, 74, 75, 81, 89}	Objective assessment of 10 arthroscopic skill domains. 5 point Likert scale with middle descriptive anchors	Captures quality of performance. Flexible in time, can be applied live or to recorded footage. GRS component appears to be reliable when used by inexperienced raters ⁸³	Time consuming, does not assess skills specific to the procedure. Consent/knowledge domain may not be appropriate for technical skills assessment	Y ^{42, 74, 75}	Y ^{42, 74, 75, 83*}	Y ^{42, 74, 81, 89}	Y ^{74, 75, 81, 89}	Y ⁸¹ M ^{83*}	Y ⁷⁴	-

Arthroscopic Skills Assessment Form ⁷⁸	7 area TSC, points awarded for speed and deducted for cartilage injury. Scope score + time score = total score (max 100)	Values procedural care, timeliness and avoidance of injury. Can probably be used in-vivo	Needs expert evaluators	Y ⁷⁸	Y ⁷⁸	Y ⁷⁸	-	-	-	-
Objective Assessment of Arthroscopic Skills (OOAS) ⁸⁷	8 domain GRS with end and centre descriptive anchors	Rich descriptors, could detect subtle improvement, can show mastery	Anchor descriptors different for each domain and could be off putting for user	Y ⁸⁷	Y ⁸⁷	Y ⁸⁷	-	Y ⁸⁷	Y ⁸⁷	-
Arthroscopic Surgical Skill Evaluation Tool (ASSET) ^{8, 14, 15, 18, 20, 36, 37, 63, 77, 80, 81, 85, 91}	Objective assessment of 8 generic arthroscopic skill domains, using 5 point Likert scale with middle anchors	Discriminates between 'novice', 'competent' and 'expert'. Captures quality of performance. Generalisable to multiple arthroscopic procedures. Potential for use in high stakes pass-fail assessments.	Needs expert evaluators. More time consuming. Does not assess skills specific to procedure.	Y ⁸⁵	Y ^{80, 85, 91}	Y ^{8, 9, 14, 15, 18, 20, 36, 37, 63, 77, 80, 81, 85, 91}	Y ^{9, 18, 36, 37, 77, 80, 81, 85, 91}	Y ^{36, 37, 63, 80, 81, 91}	Y ^{36, 37, 80, 91}	Y ⁸⁵
Injury Grading Index Performance Scale ¹⁶	Psychometric scale graded from 1-10(1 best), similar to VAS, to subjectively evaluate for arthroscopic handling skills and intra-articular injury potential	Quick, easy to use, low cost	Requires expert assessor. Injury severity descriptors not defined, potential for subjectivity	-	-	P ¹⁶	P ¹⁶	-	-	-
Aggregate arthroscopic skills score ³⁵	Score based on PT and deductions for	May be educationally	Requires expert assessors, limited to simulated environment	Y ³⁵	Y ³⁵	Y ³⁵	N	Y ³⁵	Y ³⁵	Y ³⁵

	technical errors (total score = timing score – penalty score)	beneficial, score rewards efficiency and precision								
OPEN SURGERY OBJECTIVE SKILLS ASSESSMENT TOOLS										
O-SCORE: Ottawa Surgical Competency Operating Room Evaluation $\Psi^{98, 106}$	11 item tool. 8 items rated on a 5 point competency scale. 1 item assessing overall procedural competence, 2 feedback items	Can be used on wide range of open and arthroscopic procedures	Needs expert assessors	Y ¹⁰⁶	Y ¹⁰⁶	Y ^{98, 106}	Y ⁹⁸	Y ^{98, 106}	Y ¹⁰⁶	Y ¹⁰⁶
Generic Operative Supervised Learning Event (GOSLE) Ψ^{105}	8 point global rating scale of construct aligned descriptors, 3 boxes for free text feedback; 2 are compulsory (reinforcing areas of good practice and areas for improvement)	Quick to complete, can be used in live or simulated setting, not procedure specific, requires minimal training in its use, has educational impact	Expert assessors	Y ¹⁰⁵	Y ¹⁰⁵	Y ¹⁰⁵	Y ¹⁰⁵	Y ¹⁰⁵	Y ¹⁰⁵	Y ¹⁰⁵
Procedure Based Assessment ^{19, 104, 109}	Web or paper based assessment completed by trainer and trainee. Principle summative assessment tool in UK training	Low-cost, use in simulated or live environment, encourages trainee reflection and discussion	Cost to trainees (web subscription), need expert assessors. Blunt descriptors cannot distinguish mastery or higher order skills	-	-	Y $\infty^{19, 104}$	Y $\infty^{19, 104}$	Y $\infty^{104, 109}$	Y ∞^{104}	Y ∞^{104} M ¹⁰⁹

*Olson study(83); GRS component only

Ψ May also be used for arthroscopic procedures

∞ Beard et al(104) study involved multiple surgical specialities, by-speciality data was not presented

PROCEDURE SPECIFIC RATING SCALES										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
ARTHROSCOPIC PROCEDURES										
Shoulder OPAT (Objective Practical Assessment Tool) ⁹²	Objective assessment across 4 domains; EUA, procedural standards, glenohumeral joint arthroscopy and bursoscopy. Individual step descriptors within each domain awarded points on graded scale. Global summary level 1-5	Quick to complete, addresses perceived shortcomings of PBA ⁹² . Educational impact.	Weak inter-rater reliability ⁹² . Needs expert assessors	Y ⁹²	Y ⁹²	Y ⁹²	Y ⁹²	M ⁹²	Y ⁹²	Y ⁹²
Global Rating Scale for Shoulder Arthroscopy (GRSSA) ⁹⁰	Objective assessment across 6 domains on a scale of 1-5 with descriptive anchors	Quick to complete, can be used in a live or remote setting	Expert assessors. Limited scope for educational impact	Y ⁹⁰	Y ⁹⁰	Y ⁹⁰	-	N ⁹⁰	-	-
Diagnostic Shoulder Arthroscopy checklist and Diagnostic Knee Arthroscopy Checklist ⁸²	Checklist used to evaluate performance of individual component parts of a procedure – binary yes/no	Quick to complete, can be non-expert assessor	Does not evaluate quality of performance. May over-value speed	Y ⁸²	Y ⁸²	Y ⁸²	-	-	-	-
Imperial Global Arthroscopy Rating Scale (IGARS) ^{76, 81}	9 domains with 5 point Likert scale with middle and end anchors ,and final global assessment score	Can be used in-vivo or in simulated setting, and independent of time or place	Needs expert assessors	Y ⁷⁶	Y ⁷⁶	Y ^{76, 81}	Y ^{76, 81}	Y ^{76, 81}	Y ⁷⁶	-
Bankhart Procedure	45 steps in 13 phases, Y/N for performance. Errors scored (77 possible errors, 20 sentinel)	Could be used in vivo	Needs expert assessors	YΔ ¹²⁹	YΔ ¹²⁹	Y ⁷	-	Y ^{88, 114}	Y ¹¹⁴	Y ¹¹⁴

Metrics score ^{7, 114}										
OPEN SURGICAL PROCEDURES										
ACL error-focussed scale ³²	TSC with 9 critical steps, and GRS with error focus	Could be used in vivo	Needs expert assessors. Unclear how was adapted for use	-	-	Y ³²	-	-	-	-
DHS guidewire placement scoring system ¹¹⁸	7 domains, weighted according to the importance of the step. Maximum score 100	Could be used in vivo	Limited validity evidence, unclear how was developed and used.	-	-	Y ¹¹⁸ Φ	-	-	-	-
Lumbar pedicle screw placement scoring system ⁷⁰	8 domains, weighted according to the importance of the step. Maximum score 100	Could be used in vivo	Limited validity evidence. Unclear how was developed, and used. Evidence limited to use in computer based simulation	-	-	Y ⁷⁰ Φ	-	-	-	-
Hip fracture performance score ^{58, 99}	Weighted score based on SDM, expressed as 'percentage of maximum' to account for variability between procedures	Easy, quick, automated, no requirement for assessor	Restricted to simulator	-	-	Y ⁹⁹ , N ⁵⁸	-	Y ⁹⁹	-	-
Supracondylar fracture pinning checklist ²¹	15 item TSC of key procedural steps with descriptors, binary scoring	Easy, quick could use in live theatre or simulator	Needs expert assessor	-	-	Y ²¹	-	-	-	-

ΔValidity evidence from Angelo 2015 study, which was not included in the review as it includes consultants only

ΦLimited description, details of validity testing not reported fully

INDIVIDUAL PROCEDURAL METRICS										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Final Product Analysis	Objective assessment of final product quality eg. Guidewire position ^{61, 72} , tip-apex distance ^{58, 62} , wire-tip to subchondral distance ²⁴ , screw position ^{22, 30, 59, 71, 95} , articular reduction error ^{73, 93} , biomechanical properties ^{93, 97} articular surface congruency ⁷³ drill/probe accuracy ⁶⁵ , plunge depth ²⁵ , palmar tilt ⁶⁸	Easy to measure in simulated setting can be objective	Hard to measure in-vivo , case-specific, assessor rating can be subjective. Need expert rater	Y ⁷¹	Y ^{61, 71}	Y ^{22, 24, 30, 58, 61, 71, 72, 97} M ^{25, 62, 93} N ^{59, 65, 68, 73, 84}	Y ⁷¹ N ⁹³	-	-	Y ^{61, 95}
Fracture Fixation Assessment Tool ¹⁰⁷	Objective assessment of end product using radiographs across 4 domains; reduction, stability, implant position, impression (0-3 points max for each domain) ¹⁰⁷	Can be used in any time or place setting. Can be used for service evaluation as a well as training	Need expert assessor	-	Y ¹⁰⁷	-	-	Y ¹⁰⁷	Y ¹⁰⁷	-
Error Rating ^{43, 60, 67}	Reports frequency of errors i.e soft tissue/bony collisions	Can be used in any setting. Binary measure – easy to use	Need expert assessor. May be more suited to arthroscopic procedures in a simulated setting	-	Y ⁴³	Y ^{43, 60, 67}	Y ⁴³	-	-	-
Iatrogenic Index ²⁹	Score 1 to 5 (1 best) based on intra- and extra-ossesous aspect of trial and final screw trajectories	Can be used in simulated and live settings	Need expert assessor. Procedure specific	-	-	N ²⁹	-	-	-	-

Image Intensifier use ^{24, 29, 60, 72, 99} /fluoroscopy time ^{58, 61, 62, 99}	Number of x-rays used intra-operatively/time under fluoroscopy	Low cost, easy, can be used in simulated and live setting. Does not require expert assessor	Does not account for procedure complexity	-	-	Y ^{24, 29, 58, 60, 62, 72, 130} N ^{61, 99}	-	-	-	-
Procedure Time ^{6, 11, 12, 15, 18, 20, 30, 31, 33, 34, 38-40, 43-45, 47, 48, 50, 51, 53-57, 60-64, 67, 72, 73, 78, 86, 99, 102, 103, 110, 124}	Time for procedure or task completion	Easy to measure in simulated and in-vivo setting	Assumes speed = proficiency. In vivo patient and staff factors can affect procedure time	-	-	Y ^{6, 11, 12, 30, 31, 33, 34, 40, 43-45, 47, 48, 50, 51, 53-56, 61, 63, 64, 67, 78, 86, 110} M ^{15, 38, 39, 103} N ^{18, 20, 57, 60, 62, 72, 73, 99, 102, 124}	Y ^{6, 18, 47, 86} N ¹⁰²	-	-	-
Cumulative Summation Test for Learning Curve (LC-CUSUM) ⁸⁹	Derived from cumulative scores to determine when a trainee has reached a predefined level of competence	Allows for quantitative individualised assessment of learning and a graphical representation of the learning curve. Data can come from live or simulated setting	Requires sequential data points in series. Needs specialist statistican input to implement. Consensus required on pass threshold	-	-	-	-	-	Y ⁸⁹	-

MOVEMENT ANALYSIS										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Hand Motion Analysis Systems, hand and instrument motion analysis ^{13, 41, 50, 73-75, 86}	Sensors attached to the dorsum of surgeons hands/instrument tips record speed, distance and number of movements, can generate efficiency and economy scores	Sophisticated data profile enabling detection of subtle improvement in performance. May	Difficult to use in-vivo because of sterility considerations. Invasive for the surgeon being assessed	-	-	Y ^{13, 41, 50, 73-75, 86} M ⁶⁶	Y ^{41, 66, 74, 75, 86}	Y ⁷⁵	Y ⁷⁵	-

		be able to measure attainment of mastery								
Elbow Motion Analysis ³¹	Elbow worn sensors generate data on motion metrics; number of hand movements, number of minor movements, smoothness and time	As above. More feasible to use in-vivo as sensors sit at elbows (above sterile area)	Can be used in vivo	-	-	Y ³¹	-	-	Y ³¹	-
Gaze tracking using eye movements ⁶	Glasses worn by surgeon with sensors to track eye movements. Expressed as time and proportion fixed on screen vs. hands	Could be used in vivo. Sophisticated data profile enabling subtle performance improvement	Invasive for surgeon being assessed	-	-	Y ⁶	Y ⁶	-	Y ⁶	-
Hand position checking	Number of occurrences of hand position checking during procedure is recorded (fq is inversely proportional to experience)	Low/no-cost, easy, non-expert assessor, can be used in simulation and in-vivo	Does not assess quality of performance or outcome	-	-	Y ^{34, 74}	Y ^{34, 74}	-	Y ³⁴	-
Instrument loss ⁷⁴	Number of instances during which the tip of the arthroscopy probe was not visible on the arthroscopy display unit. Assesses visuospatial awareness and fine motor dexterity	Free of cost, easy to measure, can used in simulation and in-vivo	Does not assess quality of performance or outcome	-	-	Y ⁷⁴	Y ⁷⁴	-	-	-
Triangulation time ⁷⁴	Mean duration of instrument loss episodes during procedure (total duration of instrument loss in seconds divided by the number of instances of instrument loss)	Free of cost	Requires precise timekeeping, not easy for scrubbed trainer. Easier if video recorded procedure	-	-	Y ⁷⁴	Y ⁷⁴	-	-	-
Instrument path length/ratio ^{12, 15, 40, 48, 49, 51, 55, 56, 63, 110}	Composite measure of total distance travelled by camera and probe during an arthroscopic procedure, or ratio of measured path relative to the ideal path. Gives a measure of movement economy	Easy to obtain	Cannot be measured in vivo	-	-	Y ^{12, 40, 48, 49, 51, 55, 56, 63, 110} M ¹⁵	Y ^{40, 51, 55}	-	-	-
Instrument collision force ^{38, 55}	Precision, distribution of forces applied to the joint surface and efficiency of task completion can be characterised.	Sophisticated data profile, can see subtle improvements in performance	Cannot be measured in vivo	Y ³⁸	-	Y ⁵⁵ M ³⁸	-	-	-	P ³⁸
Collisions and injuries ^{12, 40, 47-49, 53, 63}	Number of times the probe/arthroscope contacts tissue,	Easy to obtain	Cannot be measured in vivo	-	-	Y ^{12, 40, 47, 48, 53} M ⁶³ N ⁴⁹	-	-	-	-

	and injuries defined as contacts beyond a threshold force									
--	---	--	--	--	--	--	--	--	--	--

PSYCHOMOTOR TESTING										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Grooved PegBoard Test (Lafayette Instrument, Lafayette, IN) ^{10, 101, 108}	Time to complete the 25-hole board is recorded for dominant and non-dominant hand. Originally developed to test the dexterity of assembly line workers	Cheap, easy, portable, non-expert assessor	Results do relate to patient or operative outcome	-	-	Y ¹⁰¹ M ¹⁰	Y ¹⁰¹ N ¹⁰⁸	-	Y ¹⁰¹	-
Arthroscopic Knot Trainer ^{11, 57}	Non-anatomical, low-fidelity bench model with spring loaded eyelets to simulate tissue under tension during surgical repair	As above	Cannot use in-vivo	Y ⁵⁷	Y ⁵⁷	Y ^{11, 57}	Y ¹¹	-	-	Y ⁵⁷
The Crawford Small Parts Dexterity Test ¹⁰⁸	Tests hand eye co-ordination . Has application in the clinical setting for hand rehab by OT's. Part 1 measures dexterity in using forceps to insert metal pins in close-fitting holes and place collars over the pins. Part 2 measures dexterity in starting and setting screws in threaded holes in a metal plate	As above	As above	-	-	-	M ¹⁰⁸	-	-	-
Steadiness Hole Test ¹⁰⁸	Measures hand steadiness as the ability to hold a metal stylus in a fixed position within holes of different diameters without touching the sides	As above	As above	-	-	-	M ¹⁰⁸	-	-	-
Bennett hand-tool test ¹⁰⁸	Measures dexterity in handling ordinary mechanical tools. Developed for application in industrial apprentice training. The test consists of tools and two	As above	As above	-	-	-	N ¹⁰⁸	-	-	-

	uprights with bolts, the objective is to disassemble the bolts from one upright and re-assemble them on the corresponding rows on the other tes-board uprights										
Minnesota rate manipulation test ¹⁰⁸	Tests hand-eye co-ordination, originally developed as pre-employment screening tool, now mainly clinical application in rehab. Uses a hole-punched board and blocks, 5 sub-scores for complete score	As above	As above	-	-	-	N ¹⁰⁸	-	-	-	-
Autoscoring Mirror Tracer(Lafayette Instrument, Lafayette, IN) ¹⁰	Tests reversal ability. Participants trace a star pattern whilst watching only a mirror image	As above	As above	-	-	M ¹⁰	-	-	-	-	-
Purdue Pegboard Test (Lafayette Instrument, Lafayette, IN) ¹⁰	Measures gross hand and finger movements in assembly tasks	As above	As above	-	-	Y ¹⁰	-	-	-	-	-
O'Connor Tweezer Dexterity Test(Lafayette Instrument, Lafayette, IN) ¹⁰	Tests fine manual dexterity, participants have to use tweezers to place 1/16" pins in holes	As above	As above	-	-	Y ¹⁰	-	-	-	-	-
Etch-a-Sketch with overlay(Lafayette Instrument, Lafayette, IN) ¹⁰	Tests two hand and hand-eye coordination, participants have to trace a pattern on a standard Etch-a-Sketch	As above	As above	-	-	M ¹⁰	-	-	-	-	-
Two-arm coordination test (Lafayette Instrument, Lafayette, IN) ¹⁰	Tests coordination and balance. Participants trace an anodized star pattern with a two-handed triangular pointer.	As above	As above	-	-	M ¹⁰	-	-	-	-	-

SUBJECTIVE ASSESSMENTS										
Tool	Characteristics	Strengths	Limitations	Validity				Reliability	Feasibility	Educational impact
				Face	Content	Construct	Concurrent			
Pass/Fail judgement ^{96, 102, 103}	Subjective global judgement by expert assessor as to complete the tasks and	Any location and procedure	Expert assessors only. Subjective. Not reliable.	-	-	Y ^{96, 102, 103}	Y ^{96, 103}	N ⁹⁶	-	-

	perform each approach in a safe and controlled manner									
Video Feedback ⁹⁴	Operative performance is video recorded and then reviewed by the trainee, either with self –reflection or expert feedback	Encourages reflection, learning from own mistakes, used effectively in other industries (eg athletics)	Requires camera equipment set up, ethical issues with filming real operations	-	-	N ⁹⁴	N ⁹⁴	-	Y ⁹⁴	N ⁹⁴

References

1. Mitchell EL, Arora S, Moneta GL, Kret MR, Dargon PT, Landry GJ, Eidt JF, Sevdalis N. A systematic review of assessment of skill acquisition and operative competency in vascular surgical training. *J Vasc Surg*. 2014 May;59(5):1440-55. Epub 2014 Mar 19.
2. Okoro T, Sirianni C, Brigden D. The concept of surgical assessment: part 1 – introduction. *Bulletin of the Royal College of Surgeons of England*. 2010 Oct;92(9):322-3.
3. Okoro T, Sirianni C, Brigden D. The concept of surgical assessment: part 2 – available tools. *Bulletin of the Royal College of Surgeons of England*. 2010 Oct;92(9):324-6.
4. Shamseer L, Moher D, Clarke M, Ghersi D, Liberati A, Petticrew M, Shekelle P, Stewart LA; PRISMA-P Group. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015 Jan 2;350:g7647. Erratum in: *BMJ*. 2016 Jul 21;354:i4086.
5. National Institute for Health Research. PROSPERO international prospective register of systematic reviews. Accessed 15 April 2020. <https://www.crd.york.ac.uk/prospero/>
6. An VVG, Mirza Y, Mazomenos E, Vasconcelos F, Stoyanov D, Oussedik S. Arthroscopic simulation using a knee model can be used to train speed and gaze strategies in knee arthroscopy. *Knee*. 2018 Dec;25(6):1214-21. Epub 2018 Jun 20.
7. Angelo RL, Ryu RKN, Pedowitz RA, Beach W, Burns J, Dodds J, Field L, Getelman M, Hobgood R, McIntyre L, Gallagher AG. A proficiency-based progression training curriculum coupled with a model simulator results in the acquisition of a superior arthroscopic Bankart skill set. *Arthroscopy*. 2015 Oct;31(10):1854-71. Epub 2015 Sep 2.
8. Bhattacharyya R, Davidson DJ, Sugand K, Bartlett MJ, Bhattacharya R, Gupta CM. Knee arthroscopy simulation: a randomized controlled trial evaluating the effectiveness of the Imperial Knee Arthroscopy Cognitive Task Analysis (IKACTA) tool. *J Bone Joint Surg Am*. 2017 Oct 4;99(19):e103.
9. Camp CL, Krych AJ, Stuart MJ, Regnier TD, Mills KM, Turner NS. Improving resident performance in knee arthroscopy: a prospective value assessment of simulators and cadaveric skills laboratories. *J Bone Joint Surg Am*. 2016 Feb 3;98(3):220-5.
10. Çetinkaya E, Çift H, Aybar A, Erçin E, Güler GB, Poyanlı O. The timing and importance of motor skills course in knee arthroscopy training. *Acta Orthop Traumatol Turc*. 2017 Jul;51(4):273-7. Epub 2017 Jul 8.
11. Chong ACM, Pate RC, Prohaska DJ, Bron TR, Wooley PH. Validation of improvement of basic competency in arthroscopic knot tying using a bench top simulator in orthopaedic residency education. *Arthroscopy*. 2016 Jul;32(7):1389-99. Epub 2016 Apr 23.
12. Gomoll AH, Pappas G, Forsythe B, Warner JJP. Individual skill progression on a virtual reality simulator for shoulder arthroscopy: a 3-year follow-up study. *Am J Sports Med*. 2008 Jun;36(6):1139-42. Epub 2008 Mar 6.
13. Jackson WFM, Khan T, Alvand A, Al-Ali S, Gill HS, Price AJ, Rees JL. Learning and retaining simulated arthroscopic meniscal repair skills. *J Bone Joint Surg Am*. 2012 Sep 5;94(17):e132.
14. Martin KD, Patterson D, Phisitkul P, Cameron KL, Femino J, Amendola A. Ankle arthroscopy simulation improves basic skills, anatomic recognition, and proficiency during diagnostic examination of residents in training. *Foot Ankle Int*. 2015 Jul;36(7):827-35. Epub 2015 Mar 11.
15. Rahm S, Wieser K, Bauer DE, Waibel FW, Meyer DC, Gerber C, Fucentese SF. Efficacy of standardized training on a virtual reality simulator to advance knee and shoulder arthroscopic motor skills. *BMC Musculoskelet Disord*. 2018 May 16;19(1):150.
16. Rebolledo BJ, Hammann-Scala J, Leali A, Ranawat AS. Arthroscopy skills development with a surgical simulator: a comparative study in orthopaedic surgery residents. *Am J Sports Med*. 2015 Jun;43(6):1526-9. Epub 2015 Mar 13.
17. Cannon WD, Garrett WE Jr, Hunter RE, Sweeney HJ, Eckhoff DG, Nicandri GT, Hutchinson MR, Johnson DD, Bisson LJ, Bedi A, Hill JA, Koh JL, Reinig KD. Improving residency training in arthroscopic knee surgery with use of a virtual-reality simulator. A randomized blinded study. *J Bone Joint Surg Am*. 2014 Nov 5;96(21):1798-806.

18. Dunn JC, Belmont PJ, Lanzi J, Martin K, Bader J, Owens B, Waterman BR. Arthroscopic shoulder surgical simulation training curriculum: transfer reliability and maintenance of skill over time. *J Surg Educ*. 2015 Nov-Dec;72(6):1118-23. Epub 2015 Aug 19.
19. 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.
20. Waterman BR, Martin KD, Cameron KL, Owens BD, Belmont PJ Jr. Simulation training improves surgical proficiency and safety during diagnostic shoulder arthroscopy performed by residents. *Orthopedics*. 2016 May 1;39(3):e479-85. Epub 2016 May 2.
21. Butler BA, Lawton CD, Burgess J, Balderama ES, Barsness KA, Sarwark JF. Simulation-based educational module improves intern and medical student performance of closed reduction and percutaneous pinning of pediatric supracondylar humeral fractures. *J Bone Joint Surg Am*. 2017 Dec 6;99(23):e128.
22. Gottschalk MB, Yoon ST, Park DK, Rhee JM, Mitchell PM. Surgical training using three-dimensional simulation in placement of cervical lateral mass screws: a blinded randomized control trial. *Spine J*. 2015 Jan 1;15(1):168-75. Epub 2014 Sep 4.
23. LeBlanc J, Hutchison C, Hu Y, Donnon T. A comparison of orthopaedic resident performance on surgical fixation of an ulnar fracture using virtual reality and synthetic models. *J Bone Joint Surg Am*. 2013 May 1;95(9):e60-5: S1-5.
24. Nousiainen MT, Omoto DM, Zingg PO, Weil YA, Mardam-Bey SW, Eward WC. Training femoral neck screw insertion skills to surgical trainees: computer-assisted surgery versus conventional fluoroscopic technique. *J Orthop Trauma*. 2013 Feb;27(2):87-92.
25. Ruder JA, Turvey B, Hsu JR, Scannell BP. Effectiveness of a low-cost drilling module in orthopaedic surgical simulation. *J Surg Educ*. 2017 May - Jun;74(3):471-6. Epub 2016 Nov 7.
26. Sonnadara RR, Van Vliet A, Safir O, Alman B, Ferguson P, Kraemer W, Reznick R. Orthopedic boot camp: examining the effectiveness of an intensive surgical skills course. *Surgery*. 2011 Jun;149(6):745-9. Epub 2011 Jan 14.
27. Sonnadara RR, Garbedian S, Safir O, Nousiainen M, Alman B, Ferguson P, Kraemer W, Reznick R. Orthopaedic boot camp II: examining the retention rates of an intensive surgical skills course. *Surgery*. 2012 Jun;151(6):803-7.
28. Sonnadara RR, Garbedian S, Safir O, Mironova P, Nousiainen M, Ferguson P, Alman B, Kraemer W, Reznick R. Toronto orthopaedic boot camp III: examining the efficacy of student-regulated learning during an intensive, laboratory-based surgical skills course. *Surgery*. 2013 Jul;154(1):29-33.
29. Tonetti J, Vadcard L, Girard P, Dubois M, Merloz P, Troccaz J. Assessment of a percutaneous iliosacral screw insertion simulator. *Orthop Traumatol Surg Res*. 2009 Nov;95(7):471-7. Epub 2009 Oct 3.
30. Xiang L, Zhou Y, Wang H, Zhang H, Song G, Zhao Y, Han J, Liu J. Significance of preoperative planning simulator for junior surgeons' training of pedicle screw insertion. *J Spinal Disord Tech*. 2015 Feb;28(1):E25-9.
31. Garfield Roberts P, Alvand A, Gallieri M, Hargrove C, Rees J. Objectively assessing intraoperative arthroscopic skills performance and the transfer of simulation training in knee arthroscopy: a randomized controlled trial. *Arthroscopy*. 2019 Apr;35(4):1197-1209.e1. Epub 2019 Mar 14.
32. Brusalis CM, Lawrence JTR, Ranade SC, Kerr JC, Pulos N, Wells L, Ganley TJ. Can a novel, low-cost simulation model be used to teach anterior cruciate ligament graft preparation? *J Pediatr Orthop*. 2017 Jun;37(4):e277-81.
33. Cannon WD, Nicandri GT, Reinig K, Mevis H, Wittstein J. Evaluation of skill level between trainees and community orthopaedic surgeons using a virtual reality arthroscopic knee simulator. *J Bone Joint Surg Am*. 2014 Apr 2;96(7):e57.
34. Colaco HB, Hughes K, Pearse E, Arnander M, Tennent D. Construct validity, assessment of the learning curve, and experience of using a low-cost arthroscopic surgical simulator. *J Surg Educ*. 2017 Jan - Feb;74(1):47-54. Epub 2016 Oct 5.
35. Coughlin RP, Pauyo T, Sutton JC 3rd, Coughlin LP, Bergeron SG. A validated orthopaedic surgical simulation model for training and evaluation of basic arthroscopic skills. *J Bone Joint Surg Am*. 2015 Sep 2;97(17):1465-71.
36. Dwyer T, Slade Shantz J, Chahal J, Wasserstein D, Schachar R, Kulasegaram KM, Theodoropoulos J, Greben R, Ogilvie-Harris D. Simulation of anterior cruciate ligament reconstruction in a dry model. *Am J Sports Med*. 2015 Dec;43(12):2997-3004. Epub 2015 Oct 12.
37. Dwyer T, Schachar R, Leroux T, Petrera M, Cheung J, Greben R, Henry P, Ogilvie-Harris D, Theodoropoulos J, Chahal J. Performance assessment of arthroscopic rotator cuff repair and labral repair in a dry shoulder simulator. *Arthroscopy*. 2017 Jul;33(7):1310-8. Epub 2017 Mar 25.

38. 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;173:129-35.
39. Fucentese SF, Rahm S, Wieser K, Spillmann J, Harders M, Koch PP. Evaluation of a virtual-reality-based simulator using passive haptic feedback for knee arthroscopy. *Knee Surg Sports Traumatol Arthrosc.* 2015 Apr;23(4):1077-85. Epub 2014 Feb 12.
40. Gomoll AH, O'Toole RV, Czarnecki J, Warner JJP. Surgical experience correlates with performance on a virtual reality simulator for shoulder arthroscopy. *Am J Sports Med.* 2007 Jun;35(6):883-8. Epub 2007 Jan 29.
41. Howells NR, Brinsden MD, Gill RS, Carr AJ, Rees JL. Motion analysis: a validated method for showing skill levels in arthroscopy. *Arthroscopy.* 2008 Mar;24(3):335-42.
42. 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.
43. Khanduja V, Lawrence JE, Audenaert E. Testing the construct validity of a virtual reality hip arthroscopy simulator. *Arthroscopy.* 2017 Mar;33(3):566-71. Epub 2016 Dec 16.
44. Martin KD, Belmont PJ, Schoenfeld AJ, Todd M, Cameron KL, Owens BD. Arthroscopic basic task performance in shoulder simulator model correlates with similar task performance in cadavers. *J Bone Joint Surg Am.* 2011 Nov 2;93(21):e1271-5.
45. 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 Nov 7;94(21):e160.
46. Martin KD, Akoh CC, Amendola A, Phisitkul P. Comparison of three virtual reality arthroscopic simulators as part of an orthopedic residency educational curriculum. *Iowa Orthop J.* 2016;36:20-5.
47. McCarthy A, Harley P, Smallwood R. Virtual arthroscopy training: do the "virtual skills" developed match the real skills required? *Stud Health Technol Inform.* 1999;62:221-7.
48. McCarthy AD, Moody L, Waterworth AR, Bickerstaff DR. Passive haptics in a knee arthroscopy simulator: is it valid for core skills training? *Clin Orthop Relat Res.* 2006 Jan;442:13-20.
49. Pedowitz RA, Esch J, Snyder S. Evaluation of a virtual reality simulator for arthroscopy skills development. *Arthroscopy.* 2002 Jul-Aug;18(6):E29.
50. Pollard TCB, Khan T, Price AJ, Gill HS, Glyn-Jones S, Rees JL. Simulated hip arthroscopy skills: learning curves with the lateral and supine patient positions: a randomized trial. *J Bone Joint Surg Am.* 2012 May 16;94(10):e68.
51. Rahm S, Germann M, Hingsammer A, Wieser K, Gerber C. Validation of a virtual reality-based simulator for shoulder arthroscopy. *Knee Surg Sports Traumatol Arthrosc.* 2016 May;24(5):1730-7. Epub 2016 Feb 9.
52. Rose K, Pedowitz R. Fundamental arthroscopic skill differentiation with virtual reality simulation. *Arthroscopy.* 2015 Feb;31(2):299-305. Epub 2014 Oct 11.
53. Srivastava S, Youngblood PL, Rawn C, Hariri S, Heinrichs WL, Ladd AL. Initial evaluation of a shoulder arthroscopy simulator: establishing construct validity. *J Shoulder Elbow Surg.* 2004 Mar-Apr;13(2):196-205.
54. Tuijthof GJM, van Sterkenburg MN, Siersevelt IN, van Oldenrijk J, Van Dijk CN, Kerkhoffs GMMJ. First validation of the PASSPORT training environment for arthroscopic skills. *Knee Surg Sports Traumatol Arthrosc.* 2010 Feb;18(2):218-24. Epub 2009 Jul 24.
55. 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 Feb;467(2):546-52. Epub 2008 Sep 13.
56. Tofte JN, Westerlind BO, Martin KD, Guetschow BL, Uribe-Echevarria B, Rungprai C, Phisitkul P. Knee, shoulder, and fundamentals of arthroscopic surgery training: validation of a virtual arthroscopy simulator. *Arthroscopy.* 2017 Mar;33(3):641-646.e3. Epub 2016 Dec 16.
57. Wong IH, Denkers M, Urquhart N, Farrokhyar F. Construct validity testing of the Arthroscopic Knot Trainer (ArK). *Knee Surg Sports Traumatol Arthrosc.* 2015 Mar;23(3):906-11. Epub 2013 May 18.
58. Akhtar K, Sugand K, Sperrin M, Cobb J, Standfield N, Gupte C. Training safer orthopedic surgeons. Construct validation of a virtual-reality simulator for hip fracture surgery. *Acta Orthop.* 2015;86(5):616-21.

59. Aoude A, Alhamzah H, Fortin M, Jarzem P, Ouellet J, Weber MH. The use of computer-assisted surgery as an educational tool for the training of orthopedic surgery residents in pedicle screw placement: a pilot study and survey among orthopedic residents. *Can J Surg.* 2016 Dec;59(6):391-8.
60. Blyth P, Stott NS, Anderson IA. Virtual reality assessment of technical skill using the Bonedoc DHS simulator. *Injury.* 2008 Oct;39(10):1127-33. Epub 2008 Jun 13.
61. Christian MW, Griffith C, Schoonover C, Zerhusen T Jr, Coale M, O'Hara N, Henn RF 3rd, O'Toole RV, Sciadini M. Construct validation of a novel hip fracture fixation surgical simulator. *J Am Acad Orthop Surg.* 2018 Oct 1;26(19):689-97.
62. Froelich JM, Milbrandt JC, Novicoff WM, Saleh KJ, Allan DG. Surgical simulators and hip fractures: a role in residency training? *J Surg Educ.* 2011 Jul-Aug;68(4):298-302. Epub 2011 Apr 16.
63. Garfield Roberts P, Guyver P, Baldwin M, Akhtar K, Alvand A, Price AJ, Rees JL. Validation of the updated ArthroS simulator: face and construct validity of a passive haptic virtual reality simulator with novel performance metrics. *Knee Surg Sports Traumatol Arthrosc.* 2017 Feb;25(2):616-25. Epub 2016 Apr 16.
64. Giurin I, Bréaud J, Rampal V, Rosello O, Solla F. A simulation model of nail bed suture and nail fixation: description and preliminary evaluation. *J Surg Res.* 2018 Aug;228:142-6. Epub 2018 Apr 11.
65. Hohn EA, Brooks AG, Leasure J, Camisa W, van Warmerdam J, Kondrashov D, Montgomery W, McGann W. Development of a surgical skills curriculum for the training and assessment of manual skills in orthopedic surgical residents. *J Surg Educ.* 2015 Jan-Feb;72(1):47-52. Epub 2014 Aug 6.
66. Leong JJH, Leff DR, Das A, Aggarwal R, Reilly P, Atkinson HDE, Emery RJ, Darzi AW. Validation of orthopaedic bench models for trauma surgery. *J Bone Joint Surg Br.* 2008 Jul;90(7):958-65.
67. Lopez G, Wright R, Martin D, Jung J, Bracey D, Gupta R. A cost-effective junior resident training and assessment simulator for orthopaedic surgical skills via fundamentals of orthopaedic surgery: AAOS exhibit selection. *J Bone Joint Surg Am.* 2015 Apr 15;97(8):659-66.
68. Mayne IP, Brydges R, Mokhtar J, Murnaghan ML. Development and assessment of a distal radial fracture model as a clinical teaching tool. *J Bone Joint Surg Am.* 2016 Mar 2;98(5):410-6.
69. Qassem Q, Boulart L. A 4-task skills examination for residents for the assessment of technical ability in hand trauma surgery. *J Surg Educ.* 2015 Mar-Apr;72(2):179-83. Epub 2014 Dec 10.
70. Rambani R, Ward J, Viant W. Desktop-based computer-assisted orthopedic training system for spinal surgery. *J Surg Educ.* 2014 Nov-Dec;71(6):805-9. Epub 2014 Jun 23.
71. Shi J, Hou Y, Lin Y, Chen H, Yuan W. Role of Visuo-haptic surgical training simulator in resident education of orthopedic surgery. *World Neurosurg.* 2018 Mar;111:e98-104. Epub 2017 Dec 15.
72. Sugand K, Wescott RA, Carrington R, Hart A, Van Duren BH. Teaching basic trauma: validating FluoroSim, a digital fluoroscopic simulator for guide-wire insertion in hip surgery. *Acta Orthop.* 2018 Aug;89(4):380-5. Epub 2018 May 10.
73. Yehyawi TM, Thomas TP, Ohrt GT, Marsh JL, Karam MD, Brown TD, Anderson DD. A simulation trainer for complex articular fracture surgery. *J Bone Joint Surg Am.* 2013 Jul 3;95(13):e92.
74. Alvand A, Khan T, Al-Ali S, Jackson WF, Price AJ, Rees JL. Simple visual parameters for objective assessment of arthroscopic skill. *J Bone Joint Surg Am.* 2012 Jul 3;94(13):e97.
75. Alvand A, Logishetty K, Middleton R, Khan T, Jackson WFM, Price AJ, Rees JL. Validating a global rating scale to monitor individual resident learning curves during arthroscopic knee meniscal repair. *Arthroscopy.* 2013 May;29(5):906-12.
76. Bayona S, Akhtar K, Gupte C, Emery RJH, Dodds AL, Bello F. Assessing performance in shoulder arthroscopy: the Imperial Global Arthroscopy Rating Scale (IGARS). *J Bone Joint Surg Am.* 2014 Jul 2;96(13):e112. Epub 2014 Jul 2.
77. Dwyer T, Slade Shantz J, Kulasegaram KM, Chahal J, Wasserstein D, Schachar R, Devitt B, Theodoropoulos J, Hodges B, Ogilvie-Harris D. Use of an objective structured assessment of technical skill after a sports medicine rotation. *Arthroscopy.* 2016 Dec;32(12):2572-2581.e3. Epub 2016 Jul 27.
78. Elliott MJ, Caprise PA, Henning AE, Kurtz CA, Sekiya JK. Diagnostic knee arthroscopy: a pilot study to evaluate surgical skills. *Arthroscopy.* 2012 Feb;28(2):218-24. Epub 2011 Oct 28.

79. Koehler RJ, Nicandri GT. Using the arthroscopic surgery skill evaluation tool as a pass-fail examination. *J Bone Joint Surg Am.* 2013 Dec 4;95(23):e1871-6.
80. Koehler RJ, Amsdell S, Arendt EA, Bisson LJ, Braman JP, Butler A, Cosgarea AJ, Harner 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.
81. Middleton RM, Baldwin MJ, Akhtar K, Alvand A, Rees JL. Which global rating scale? A Comparison of the ASSET, BAKSSS, and IGARS for the assessment of simulated arthroscopic skills. *J Bone Joint Surg Am.* 2016 Jan 6;98(1):75-81.
82. Nwachukwu B, Gaudiani M, Hammann-Scala J, Ranawat A. A checklist intervention to assess resident diagnostic knee and shoulder arthroscopic efficiency. *J Surg Educ.* 2017 Jan-Feb;74(1):9-15. Epub 2016 Aug 23.
83. Olson T, Koehler R, Butler A, Amsdell S, Nicandri G. Is there a valid and reliable assessment of diagnostic knee arthroscopy skill? *Clin Orthop Relat Res.* 2013 May;471(5):1670-6. Epub 2012 Dec 20.
84. Pedowitz RA, Nicandri GT, Angelo RL, Ryu RKN, Gallagher AG. Objective assessment of knot-tying proficiency with the Fundamentals of Arthroscopic Surgery Training Program Workstation and Knot Tester. *Arthroscopy.* 2015 Oct;31(10):1872-9. Epub 2015 Aug 19.
85. Phillips L, Cheung JJH, Whelan DB, Murnaghan ML, Chahal J, Theodoropoulos J, Ogilvie-Harris D, Macniven I, Dwyer T. Validation of a dry model for assessing the performance of arthroscopic hip labral repair. *Am J Sports Med.* 2017 Jul;45(9):2125-30. Epub 2017 Mar 29.
86. Price AJ, Erturan G, Akhtar K, Judge A, Alvand A, Rees JL. Evidence-based surgical training in orthopaedics: how many arthroscopies of the knee are needed to achieve consultant level performance? *Bone Joint J.* 2015 Oct;97-B(10):1309-15.
87. 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.
88. Gallagher AG, Ryu RKN, Pedowitz RA, Henn P, Angelo RL. Inter-rater reliability for metrics scored in a binary fashion—performance assessment for an arthroscopic Bankart repair. *Arthroscopy.* 2018 Jul;34(7):2191-8. Epub 2018 May 2.
89. Hodgins JL, Veillette C, Biau D, Sonnadara R. The knee arthroscopy learning curve: quantitative assessment of surgical skills. *Arthroscopy.* 2014 May;30(5):613-21.
90. Hoyle AC, Whelton C, Umaar R, Funk L. Validation of a global rating scale for shoulder arthroscopy: a pilot study. *Shoulder Elbow.* 2012 Jan;4(1):16-21.
91. Koehler RJ, Goldblatt JP, Maloney MD, Voloshin I, Nicandri GT. Assessing diagnostic arthroscopy performance in the operating room using the Arthroscopic Surgery Skill Evaluation Tool (ASSET). *Arthroscopy.* 2015 Dec;31(12):2314-9.e2. Epub 2015 Aug 28.
92. Talbot CL, Holt EM, Gooding BWT, Tennent TD, Foden P. The Shoulder Objective Practical Assessment Tool: evaluation of a new tool assessing residents learning in diagnostic shoulder arthroscopy. *Arthroscopy.* 2015 Aug;31(8):1441-9. Epub 2015 Apr 22.
93. Anderson DD, Long S, Thomas GW, Putnam MD, Bechtold JE, Karam MD. Objective Structured Assessments of Technical Skills (OSATS) does not assess the quality of the surgical result effectively. *Clin Orthop Relat Res.* 2016 Apr;474(4):874-81.
94. Backstein D, Agnidis Z, Regehr G, Reznick R. The effectiveness of video feedback in the acquisition of orthopedic technical skills. *Am J Surg.* 2004 Mar;187(3):427-32.
95. Bergeson RK, Schwend RM, DeLucia T, Silva SR, Smith JE, Avilucea FR. How accurately do novice surgeons place thoracic pedicle screws with the free hand technique? *Spine (Phila Pa 1976).* 2008 Jul 1;33(15):E501-7.
96. Bernard JA, Dattilo JR, Srikumaran U, Zikria BA, Jain A, LaPorte DM. Reliability and validity of 3 methods of assessing orthopedic resident skill in shoulder surgery. *J Surg Educ.* 2016 Nov-Dec;73(6):1020-5. Epub 2016 Jun 3.
97. Burns GT, King BW, Holmes JR, Irwin TA. Evaluating internal fixation skills using surgical simulation. *J Bone Joint Surg Am.* 2017 Mar 1;99(5):e21.
98. MacEwan MJ, Dudek NL, Wood TJ, Gofton WT. Continued validation of the O-SCORE (Ottawa Surgical Competency Operating Room Evaluation): use in the simulated environment. *Teach Learn Med.* 2016;28(1):72-9.
99. Pedersen P, Palm H, Ringsted C, Konge L. Virtual-reality simulation to assess performance in hip fracture surgery. *Acta Orthop.* 2014 Aug;85(4):403-7. Epub 2014 Apr 30.
100. 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 May-Jun;72(3):458-70. Epub 2014 Dec 24.

101. Williams JF, Watson SL, Baker DK, Ponce BA, McGwin G, Gilbert SR, Khoury JG. Psychomotor testing for orthopedic residency applicants: a pilot study. *J Surg Educ.* 2017 Sep - Oct;74(5):820-7. Epub 2017 Mar 7.
102. 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 Dec;91(12):2811-7.
103. 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 Feb;37(2):332-7: 337.e1-4.
104. Beard JD, Marriott J, Purdie H, Crossley J. Assessing the surgical skills of trainees in the operating theatre: a prospective observational study of the methodology. *Health Technol Assess.* 2011 Jan;15(1):i-xxi, 1-162.
105. Davies RM, Hadfield-Law L, Turner PG. Development and evaluation of a new formative assessment of surgical performance. *J Surg Educ.* 2018 Sep - Oct;75(5):1309-16. Epub 2018 Mar 24.
106. Gofton WT, Dudek NL, Wood TJ, Balaa F, Hamstra SJ. The Ottawa Surgical Competency Operating Room Evaluation (O-SCORE): a tool to assess surgical competence. *Acad Med.* 2012 Oct;87(10):1401-7.
107. Hawkes DH, Harrison WJ. Critiquing operative fracture fixation: the development of an assessment tool. *Eur J Orthop Surg Traumatol.* 2017 Dec;27(8):1083-8. Epub 2017 Mar 23.
108. Hoffer MM, Hsu SC. Hand function in selection of orthopedics residents. *Acad Med.* 1990 Oct;65(10):661.
109. Marriott J, Purdie H, Crossley J, Beard JD. Evaluation of procedure-based assessment for assessing trainees' skills in the operating theatre. *Br J Surg.* 2011 Mar;98(3):450-7. Epub 2010 Nov 24.
110. Martin KD, Patterson DP, Cameron KL. Arthroscopic training courses improve trainee arthroscopy skills: a simulation-based prospective trial. *Arthroscopy.* 2016 Nov;32(11):2228-32. Epub 2016 May 25.
111. Van Der Vleuten CPM. The assessment of professional competence: developments, research and practical implications. *Adv Health Sci Educ Theory Pract.* 1996 Jan;1(1):41-67.
112. Norcini J, Anderson B, Bollela V, Burch V, Costa MJ, Duvivier R, Galbraith R, Hays R, Kent A, Perrott V, Roberts T. Criteria for good assessment: consensus statement and recommendations from the Ottawa 2010 Conference. *Med Teach.* 2011;33(3):206-14.
113. van Hove PD, Tuijthof GJM, Verdaasdonk EGG, Stassen LPS, Dankelman J. Objective assessment of technical surgical skills. *Br J Surg.* 2010 Jul;97(7):972-87.
114. Angelo RL, Ryu RKN, Padowitz 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 Sep;31(9):1655-70. Epub 2015 Jul 31.
115. The Royal College of Obstetricians and Gynaecologists. OSATS. Accessed 15 April 2020. <https://www.rcog.org.uk/en/careers-training/about-specialty-training-in-og/assessment-and-progression-through-training/workplace-based-assessments/osats/>
116. The Royal College of Ophthalmologists. Objective assessment of surgical and technical skills (OSATS). Accessed 2019 Jun. <https://www.rcophth.ac.uk/curriculum/ost/assessments/workplace-based-assessments/objective-assessment-of-surgical-and-technical-skills-osats/>
117. Tsagakatakaki M, Choudhary A. Mersey Deanery ophthalmology trainees' views of the objective assessment of surgical and technical skills (OSATS) workplace-based assessment tool. *Perspect Med Educ.* 2013 Feb;2(1):21-7.
118. Rambani R, Viant W, Ward J, Mohsen A. Computer-assisted orthopedic training system for fracture fixation. *J Surg Educ.* 2013 May-Jun;70(3):304-8. Epub 2013 Feb 22.
119. Agha RA, Fowler AJ, Sevdalis N. The role of non-technical skills in surgery. *Ann Med Surg (Lond).* 2015 Oct 9;4(4):422-7.
120. Hunter AR, Baird EJ, Reed MR. Procedure-based assessments in trauma and orthopaedic training—the trainees' perspective. *Med Teach.* 2015 May;37(5):444-9. Epub 2014 Sep 4.
121. Pitts D, Rowley DI, Sher JL. Assessment of performance in orthopaedic training. *J Bone Joint Surg Br.* 2005 Sep;87(9):1187-91.
122. Bartlett JD, Lawrence JE, Stewart ME, Nakano N, Khanduja V. Does virtual reality simulation have a role in training trauma and orthopaedic surgeons? *Bone Joint J.* 2018 May 1;100-B(5):559-65.

123. Wong K, Stewart F. Competency-based training of basic trainees using human cadavers. *ANZ J Surg.* 2004 Aug;74(8):639-42.
124. LeBlanc F, Champagne BJ, Augestad KM, Neary PC, Senagore AJ, Ellis CN, Delaney CP; Colorectal Surgery Training Group. A comparison of human cadaver and augmented reality simulator models for straight laparoscopic colorectal skills acquisition training. *J Am Coll Surg.* 2010 Aug;211(2):250-5.
125. Karlen AI, Solberg EJ, Quanbeck DS, Van Heest AE. Orthopaedic surgery residency rotations and correlation with Orthopaedic In-Training Examination performance. *J Surg Educ.* 2018 Sep-Oct;75(5):1325-8. Epub 2018 Feb 12.
126. Swanson D, Marsh JL, Hurwitz S, DeRosa GP, Holtzman K, Bucak SD, Baker A, Morrison C. Utility of AAOS OITE scores in predicting ABOS Part I outcomes: AAOS exhibit selection. *J Bone Joint Surg Am.* 2013 Jun 19;95(12):e84.
127. Bradley CS, Moktar J, Maxwell A, Wedge JH, Murnaghan ML, Kelley SP. A reliable and valid objective structured assessment of technical skill for the application of a Pavlik harness based on international expert consensus. *J Pediatr Orthop.* 2016 Oct-Nov;36(7):768-72.
128. Moktar J, Popkin CA, Howard A, Murnaghan ML. Development of a cast application simulator and evaluation of objective measures of performance. *J Bone Joint Surg Am.* 2014 May 7;96(9):e76.
129. Angelo RL, Ryu RKN, Padowitz RA, Gallagher AG. Metric development for an arthroscopic Bankart procedure: assessment of face and content validity. *Arthroscopy.* 2015 Aug;31(8):1430-40.
130. Blyth P, Stott NS, Anderson IA. A simulation-based training system for hip fracture fixation for use within the hospital environment. *Injury.* 2007 Oct;38(10):1197-203. Epub 2007 Jul 19.