AMERICAN
ORTHOPAEDIC
ASSOCIATION

AOA
1887

AOA Critical Issues in Education

Do Resident Surgical Volumes and Level of Training Correlate with Improved Performance on Psychomotor Skills Tasks: Construct Validity Testing of an ASSH Training Platform (STEP)?

Jeffrey J. Olson, MD, Bo Zhang, MD, Diana Zhu, MD, Evan T. Zheng, BA, George S.M. Dyer, MD, FAOA, Tamara D. Rozental, MD, FAOA, and Dawn M. LaPorte, MD, FAOA

Investigation performed at Johns Hopkins Department of Orthopaedic Surgery, Beth Israel Deaconess Medical Center Department of Orthopaedic Surgery, Brigham and Women's Hospital Department of Orthopaedic Surgery

Background: The Surgical Training and Educational Platform (STEP) was developed by the American Society for Surgery of the Hand (ASSH) as a cost-effective set of surgical simulation modules designed to assess critical skills in hand surgery. Previous study demonstrated that STEP can differentiate between novice trainees and board-certified, certificate of added qualification hand surgeons. The purpose of this study was to assess construct validity of STEP by testing its ability to differentiate psychomotor skill level among intermediate trainees.

Methods: We evaluated 30 residents from 2 orthopaedic residency programs on 8 modules: (1) lag screw fixation, (2) depth of plunge during bicortical drilling, (3) flexor tendon repair, (4) phalangeal fracture pinning, (5) central axis scaphoid fixation, (6) full-thickness skin graft harvest, (7) microsurgery, and (8) wrist arthroscopy. Spearman correlation was used to correlate total and task-specific scores to case log numbers, months in training, and number of hand surgery rotations.

Results: Senior residents had significantly higher mean number of total cases in their total case log (mean difference 96.2, 95% confidence interval [CI] 67.5-124.8, p < 0.01) and number of task-specific cases. Moderate correlation was observed between case log numbers and scaphoid fixation score (rs = 0.423, 95% CI 0.07-0.69) and total score (rs = 0.584, 95% CI 0.25-0.79). Moderate correlation was observed between months in training with: scaphoid fixation (rs = 0.377, 95% CI 0.01-0.66) and microsurgery (rs = 0.483, 95% CI 0.13-0.73); strong correlation was seen with total score (rs = 0.656, 95% CI 0.35-0.83). Moderate correlation with number of hand surgery rotations was observed with tendon

repair (rs = 0.362, 95% CI −0.01 to 0.65), skin graft (rs = 0.385, 95% CI 0.01-0.66), wrist arthroscopy (rs = 0.391, 95% CI 0.02-0.67), microsurgery (rs = 0.461, 95% CI 0.10-0.71), and scaphoid fixation (rs = 0.578, 95% CI 0.25-0.79); and strong correlation was seen with total score (rs = 0.670, 95% CI 0.37-0.84).

Discussion/Conclusion: The STEP is a validated ASSH education tool that provides a cost-effective simulation for the assessment of fundamental psychomotor skills in hand surgery. Total STEP score correlated with total task-related case volumes as well as months in training and number of hand rotations. Scoring could be modified to improve the fidelity of assessing surgical performance. And, while both time and human resource consuming to perform, administer, and score, this study demonstrates construct validity of STEP in assessing the progression of surgical skill through residency.

In 2003, the Accreditation Council for Graduate Medical Education (ACGME) implemented national work-hour restrictions in response to concern over an increase in sleep deprivation–related resident medical errors[1]. Since implementation, studies have demonstrated decreased trainee operative case volume and decreased satisfaction with surgical training among attendings and residents alike[2,3]. Subsequently, educators in orthopaedic surgery have sought to expand resident and fellow education outside of the operating room to focus on improving skill and increasing trainees repetitions without sacrificing patient safety. Arthroscopy and other task-based simulators, cadaveric dissections, and virtual reality simulators have been developed and implemented to address this need[4-8]. To date, adjunctive training in hand surgery has been limited to microsurgical simulation[9-12]. Furthermore, no platform has been formally recognized by the American Board of Orthopaedic Surgery or the American Academy of Orthopaedic Surgery as a reliable training or assessment tool. High costs, challenges with consensus, and lack of rigorous validation and objective measure of skill are cited as reasons for difficulty in gaining broad acceptance[13-17].

From 2017 to 2019, the American Society for Surgery of the Hand (ASSH) created a taskforce that developed Surgical Training and Educational Platform (STEP) in an effort to create a comprehensive psychomotor training and assessment tool for fundamental skills of hand surgery. Recognizing the limitations of previous platforms, STEP involved low-cost, high-fidelity simulation tasks, an objective metric-based scoring system inspired by the Foundations of Laparoscopic Surgery (FLS), and rigorous validation predicated on Messick validity framework[18,19]. Face (does the simulator reproduce the real-world skills) and content (does the simulator represent all facets of what it aims to measure) validity of the platform have since been demonstrated[20]. The initial study results demonstrated that board-certified (CAQ) hand surgeons significantly outperformed novice (postgrade year [PGY]-1) trainees in all simulation tasks[20]. Additional efforts have been made to use STEP for assessment with the hopes of establishing threshold levels for each task. To do this, carefully designed experiments are required to be performed to establish STEP in this new role. Toward that end, construct validity must be firmly established before next establishing inter-rater reliability and transfer validity, that is, evidence of responsiveness to training.

The purpose of this study was to assess construct validity of STEP by testing its ability to differentiate psychomotor skill level between intermediate trainees: junior and senior residents. We hypothesized that senior residents would outperform junior residents on simulation tasks, and further, that the residents' surgical case volume and length of training would positively correlate with their performance on the tasks.

Materials and Methods

Study Design

Residents from the Johns Hopkins and Harvard combined orthopaedic surgery residency programs participated in the STEP simulator testing. Both junior (PGY 2-3) and senior residents
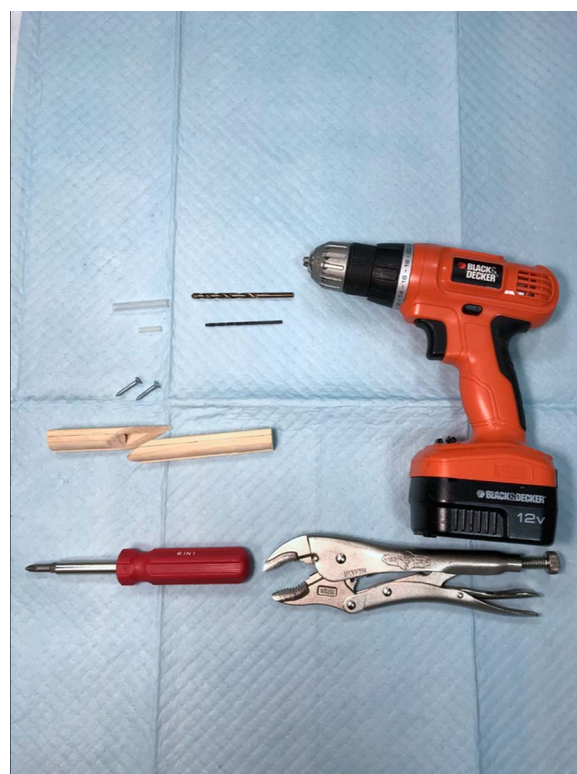

Fig. 1

Photograph of the bicortical lag screw of an oblique fracture simulation module.

Fig. 2
Photograph of the depth of plunge simulation module.

(PGY 4-5) participated. Junior residents had completed 0 to 1 hand surgery rotations while senior residents had completed 2 to 3 rotations. Participation was voluntary, and all residents in each program were encouraged to participate. Resident operative experience was quantified by ACGME case log volume, duration (months) in residency, and number of hand surgery rotations. The ASSH surgical simulation taskforce identified current procedural terminology codes for procedures related to each of the 8 surgical tasks (see Appendix). Case logs relevant to each task were summated for a task-specific total and together were summated for a grand total of all tasks.

Description of STEP, Individual Tasks, and Performance Criteria
The STEP program is a validated tool developed by the ASSH surgical simulation taskforce for the purpose of resident and fellow education. STEP consists of 8 modules: (1) lag screw fixation, (2) depth of plunge during bicortical drilling (3) flexor tendon repair, (4) scaphoid fixation, (5) phalangeal fracture pinning, (6) microsurgery, (7) full-thickness skin graft harvest, and (8) wrist arthroscopy. The overall start-up cost and cost per trainee assessment were ~$600 and $25, respectively. Assuming a 2-person assembly team with materials available, start-up time of 1 full 8-hour work day should be expected to construct the modules and test stations. Additional time (~1 hour) is needed to prepare for each test day (i.e., slicing and marking mangos, dissecting flexor tendons, etc.).

Lag Screw Fixation of an Oblique Fracture
This task simulated drilling and lag screw fixation of a bicortical wood model of an oblique fracture, constructed by gluing together pine corner and quarter-round molding (Fig. 1). A 4-cm segment of the quarter-round molding was chiseled out to simulate the intramedullary canal and bicortical model.

Depth of Plunge
This task used the same bicortical wood model, secured on top of a foam block (Fig. 2). The participant drilled 3 consecutive holes bicortical, limiting the depth of plunge into the underlying foam. The foam was sliced in half, and the depth of the plunge was measured.

Flexor Tendon Repair
Participants repaired a whole pig-foot flexor tendon laceration, simulating a zone II flexor tendon injury (Fig. 3). 3-0 Ethibond and 6-0 Prolene suture were provided for the repair technique of their choice. The repairs were then tested biomechanically, measuring strength of repair with a tensiometer (target goal 20 Newtons before gapping, simulating requirements for early active motion protocol).

Fig. 3
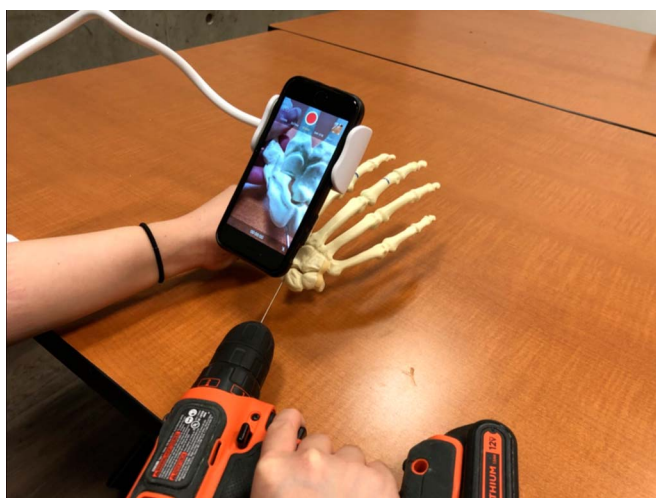Photograph of the flexor tendon repair simulation model and testing jig.

Fig. 4

Fig. 5

Fig. 4 Photograph of the central axis scaphoid pinning simulation model using 2D video camera to simulate fluoroscopy. **Fig. 5** Photograph of the phalangeal pinning simulation model using 2D video camera to simulate fluoroscopy.

Scaphoid Fracture Fixation
This task simulated placement of a central-axis scaphoid pin using a sawbones hand model with a mobile proximal carpal row (Fig. 4). K-wire pinning was performed under 2D viewing through a phone camera to simulate fluoroscopy.

Phalangeal Fracture Pinning
This task simulated cross-pinning of both proximal-third proximal phalanx and distal-third middle phalanx fracture of the index and long fingers, respectively (Fig. 5). K-wire pinning was similarly conducted under 2D viewing through a phone camera.

Microsurgical Suturing
The participant performed microsurgical suturing of a 2-cm laceration of a latex glove taped over a Petri dish (Fig. 6). Ten simple interrupted sutures with 8-0 nylon suture were performed under a table-top dissection microscope.

Full-Thickness Skin Graft
This task simulated harvest of a full-thickness skin graft using an unripened mango (Fig. 7). A 2-cm × 4-cm diamond was harvested using a #10 blade and iris scissors. The green peel represented the skin while the yellow flesh of the fruit represented fat.

Wrist Arthroscopy
This station was constructed using a 3 × 5 index card box with 1 to 2, 3 to 4, 4 to 5, and 6-U portals, cardstock paper with printed bullseye targets and triangles, 2 pegs, a rubber washer, and a Desptech endoscope, simulating a wrist arthroscopy (Fig. 8). Portal establishment was simulated by aiming and puncturing the red bullseye with a 18-gauge spinal needle passed from behind the index card box, through an area cut out, and replaced with foam gutter guard material. A free pass was given to orient the needle; the second pass was counted and scored. Probing a structure was simulated by tracing 1 triangle with dotted lines with a fine tip Sharpie pen. Object translation was simulated by transferring the rubber washer from one peg to the other, releasing it, then transferring it back using a 3.5-mm arthroscopic grasper.

All modules were constructed, administered, and scored by 6 trained examiners, per previously established protocol in the original validation study of STEP[20]. Scoring was modeled

Fig. 6

Photograph of the microsurgical suturing simulation model using a desktop microscope.
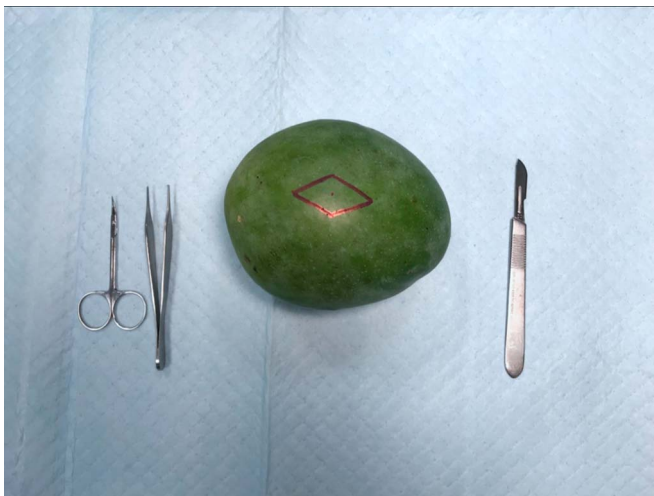
Fig. 7
Photograph of the full-thickness skin graft simulation module.

after the FLS scoring system which used time, accuracy, and various performance metrics to score each task[21]. Scoring, performance criteria, and penalties for each task are detailed in Table I. The total time to test and score 1 individual is approximately 1.5 to 2 hours.

Statistical Analysis

Data were analyzed for normal distribution and linearity. Continuous and categorical data were reported as mean (SD), median (interquartile range), and frequency (percentage) where appropriate. Student t test and χ^2 analyses were performed on continuous and categorical data, respectively. Spearman correlations were performed between task scores and case log numbers, months in residency, and number of hand rotations given the nonlinear, monotonic relationship of the data being tested. Strength of correlation was categorized as none (rs = 0), weak (0 < rs ≤ 0.29), moderate (0.30 ≤ rs ≤ 0.59), strong (0.60 ≤ rs ≤ 0.79), or very strong (0.80 ≤ rs ≤ 1.00).

Results

Thirty residents from 2 orthopaedic residency programs completed all 8 surgical simulation tasks of the STEP. There were 15 senior residents and 15 junior residents. The 2 programs demonstrated similar number of senior/junior residents, mean months in training, and mean number of hand rotations (Table II). Total case logs and task-specific case log numbers were similar between the 2 programs, except for wrist arthroscopy. Detailed comparisons of senior and junior residents are demonstrated in Table III. Overall, seniors had significantly higher mean number of total case logs (mean difference 96.2, 95% confidence interval [CI] 67.5-124.8, p < 0.01). Seniors also had significantly higher number of task-specific cases except for skin graft and wrist arthroscopy (Table III).

Seniors achieved significantly higher mean scores in tendon repair (p = 0.03) and in total (p < 0.01), as well as higher mean scores approaching significance in scaphoid fixation (p = 0.06) and microsurgery (p = 0.06) (Table IV).

Moderate correlation was observed between case log numbers and scaphoid fixation score (rs = 0.423, 95% CI 0.07-0.69) and total score (rs = 0.584, 95% CI 0.25-0.79). Moderate correlation was observed between months in residency and scaphoid fixation (rs = 0.377, 95% CI 0.01-0.66) and microsurgery (rs = 0.483, 95% CI 0.13-0.73); strong correlation with total score (rs = 0.656, 95% CI 0.35-0.83). Moderate correlation was observed between number of hand rotations and tendon repair (rs = 0.362, 95% CI −0.01 to 0.65), skin graft (rs = 0.385, 95% CI 0.01-0.66), wrist arthroscopy (rs = 0.391, 95% CI 0.02-0.67), microsurgery (rs = 0.461, 95% CI 0.10-0.71), and scaphoid fixation (rs = 0.578, 95% CI 0.25-0.79); strong correlation with total score (rs = 0.670, 95% CI 0.37-0.84) (Table V).

Discussion

Simulation is becoming a critical part of learning in orthopaedic residency, including hand subspecialty[4,5,8,15,17]. Simulation provides an opportunity to build fundamental psychomotor skills outside of the operating room and to test for improvement throughout the course of residency training. To date, there exists only one widely accepted surgical simulation platform, FLS, developed by the Society of American Gastrointestinal Endoscopic Surgery[18]. It is a well-validated psychomotor skills assessment tool with objective, metric-based scoring used as both a learning tool throughout training and as prerequisite for board eligibility[18,22]. Recognizing the need for a similarly

Fig. 8
Photograph of the wrist arthroscopy module with a laptop screen showing real-time feed of the simulated arthoscopic camera.

TABLE I Performance and Scoring Criteria of Individual Surgical Training and Educational Platform Tasks

Task	Starting Score	Penalty—Time	Performance Criteria and Relevant Penalties			
Depth of plunge	120	Total time (s)	−5 point per 1 mm of plunge per drill hole (×3)			
Lag screw	1,000	Total time (s)	−5 point per 1 mm of fracture malreduction	−50 points per lag screw not erpendicular to fracture line	−50 points if gap between screws < 8 mm	−100 points per screw < 4 mm from fracture edge
Flexor tendon repair	1,200	Total time (s)	<4 strand repair = automatic task failure (task score= 0)		−120 points if no epitendinous repair performed	−1 point per 1 N of force below minimum 20 N
Phalangeal fracture pinning	750	Total time (s)	Penalty for pin holes too close to fracture (<10 mm): (−4 points per 1 mm less than 10 mm)		Penalty for redirection attempts (i.e., extra holes): proximal 1/3 fracture: −10 per hole; distal 1/3 fracture: −20 per hole	
Scaphoid pinning	180	Total time (s)	−5 point for each 1 mm away from target exit point		−20 point for penetration of volar cortex	
Skin graft	230	Total time (s)	−10 points for inefficient instrument handling	−20 points for clumsy or poor technique	Fat harvest penalty: 0 point for no residual fat (yellow mango pulp); −10 points for 25%-50% residual fat; −20 points for >50% residual fat	
Microsurgery	1,800	Total time (s)	−90 point per suture passed <1 mm from wound edge		−180 points per suture with <3 throws	−90 points per suture that is not square or appears loose
Wrist arthroscopy	1,000	Total time (s)	Distance from center of bullseye on 2nd pass (−20 point for every 1 mm away from center)		Sum of deviation from each line on triangle (5 point per 1 mm deviation × 3 sides)	

robust learning and assessment tool tailored to hand surgery trainees, the ASSH surgical simulation committee developed STEP, a platform focused on testing the critical psychomotor skills that translate across a wide variety of hand surgery operations. The purpose of this study was to further validate the platform through testing construct validity by correlating resident performance with case log numbers and level of training.

In 2001 the ACGME established the case log system, implementing minimum case number requirements to

TABLE II Baseline Comparisons of Residents from 2 Training Programs*†			
	Program 1 (n = 19)	Program 2 (n = 11)	p
PGY			**<0.01**
2	10 (53%)	0	
3		5 (46%)	
4	2 (11%)		
5	7 (36%)	6 (54%)	
Level of training			0.71
"Senior" resident	9 (47%)	6 (55%)	
"Junior" resident	10 (53%)	5 (45%)	
Months in training	33.8 (17.5)	40.1 (12.5)	0.26
Hand rotations	1.4 (1.2)	1.1 (1.0)	0.52
Total cases	89.6 (66.9)	74.9 (53.1)	0.54
Lag screw	28.5 (23.1)	23.4 (17.4)	0.53
Depth of plunge	49.0 (37.6)	41.1 (26.7)	0.55
Tendon repair	0.8 (1.2)	1.9 (3.0)	0.18
Scaphoid fixation	2.2 (1.9)	1.8 (1.9)	0.64
Phalangeal pinning	4.3 (4.7)	3.0 (3.7)	0.45
Microsurgery	2.6 (3.1)	2.7 (4.5)	0.95
Skin graft	0.2 (0.4)	0.5 (1.0)	0.22
Wrist arthroscopy	2.0 (2.1)	0.5 (1.2)	**0.03**

*PGY = postgrade year. †Data are presented as mean (SD) and number (%) for continuous and categorical data, respectively. Bold denotes significant p-values.

ensure competency is achieved in core surgical cases before residency graduation. Previous studies have called into question the effectiveness of the case log system on predicting readiness, confidence, and surgical competency of graduating residents[23,24]. Furthermore, a recent survey of orthopaedic surgery residencies indicated that case log reporting may

TABLE III Baseline Comparisons of Senior and Junior Resident Participants*†				
	Total (n = 30)	Senior (n = 15)	Junior (n = 15)	p
PGY				**<0.01**
2	10 (33%)	0	10 (67%)	
3	5 (17%)	0	5 (33%)	
4	2 (6%)	2 (13%)	0	
5	13 (44%)	13 (87%)	0	
Months in training	36.1 (15.9)	51.2 (4.0)	21.0 (4.4)	**<0.01**
Hand rotations	1.2 (1.1)	2.2 (0.7)	0.3 (0.5)	**<0.01**
Total cases	84.2 (61.7)	132.3 (51.8)	36.1 (15.7)	**<0.01**
Lag screw	26.6 (21.0)	42.5 (18.6)	10.7 (4.9)	**<0.01**
Depth of plunge	46.1 (33.7)	71.3 (30.4)	20.9 (8.8)	**<0.01**
Tendon repair	1.2 (2.1)	2.2 (2.6)	0.3 (0.6)	**0.01**
Scaphoid fixation	2.0 (1.9)	3.1 (1.9)	1.0 (1.2)	**<0.01**
Phalangeal pinning	3.8 (4.3)	6.7 (4.4)	0.9 (1.1)	**<0.01**
Microsurgery	2.7 (3.6)	3.1 (1.9)	1.4 (1.9)	**0.04**
Skin graft	0.3 (0.7)	0.6 (0.9)	0.1 (0.3)	0.06
Wrist arthroscopy	1.4 (1.9)	2.0 (2.3)	0.9 (1.3)	0.11

*PGY = postgrade year. †Data are presented as mean (SD) and number (%) for continuous and categorical data, respectively. Bold denotes significant p-values.

TABLE IV T-test Comparison of Scores, Juniors Vs. Seniors*				
	Score (Junior)	Score (Senior)	Mean Difference	p
Lag screw	485	530	45	0.56
Depth of plunge	−54	−40	14	0.58
Tendon repair	109	404	294	**0.03**
Phalangeal pinning	265	326	61	0.45
Scaphoid fixation	−133	31	165	0.06
Skin graft	42	76	34	0.26
Microsurgery	−1,194	−448	746	0.06
Wrist arthroscopy	604	678	74	0.16
Total	124	1,559	1,435	**<0.01**

*Data are presented as mean scores. Bold denotes significant p-value.

underestimate the actual number of cases performed during residency by as much as a 24%[25]. Yet, this is currently the only objective metric of surgical performance consistently available for evaluation. In our study, we found performance on most individual STEP tasks correlated weakly with task-specific case log numbers. One possible explanation for the limited correlation may actually be the small sample size of cases. The 2018 to 2019 ACGME National Data Report for orthopaedic surgery cases logs found the median number of cases logged among all residents nationwide was 2 for skin graft, nerve repair, and wrist arthroscopy and 9 for microsurgery. Such a small sample size makes it difficult to correlate scores[26]. However, we found a strong correlation between residents' total scores and total case log numbers. This supports our hypothesis that performance on a simulated psychomotor task would improve with increased case log numbers.

We hypothesized that longer duration of training and greater number of hand surgery rotations would predict better performance on individual tasks and overall. For specific tasks, we found moderate correlation between task-specific score and months of training for scaphoid fixation, microsurgery, skin graft, and tendon repair tasks. Predictably, there was strong correlation between overall score and number of months in residency. We found moderate correlations between score and number of rotations for tendon repair, skin graft, wrist arthroscopy, scaphoid fixation, and microsurgery, and strong correlation with total score.

We acknowledge limitations in this study. First, correlation of individual task scores with task-specific case log numbers was weak. However, total score correlated well with total case logs and correlations improved when duration of training or number of hand rotations was considered. The explanation for this is likely multifactorial. First, scoring is based on an established scoring protocol for FLS which may not be optimal for the purpose of STEP. Most notably, time to complete individual tasks is weighed heavily in individual task

TABLE V Spearman Correlation of Scores*†									
	rs (Cases)‡	95% CI	p	rs (mo)§	95% CI	p	rs (Rotation)#	95% CI	p
Lag screw	0.04	−0.33 to 0.39	0.81	0.063	−0.30 to 0.41	0.74	0.164	−0.21 to 0.50	0.39
Depth of plunge	0.096	−0.27 to 0.44	0.62	0.130	−0.24 to 0.47	0.49	0.136	−0.24 to 0.47	0.48
Tendon repair	0.144	−0.23 to 0.48	0.45	0.312	−0.06 to 0.61	0.09	**0.362**	**−0.01 to 0.65**	**0.05**
Phalangeal pinning	−0.051	−0.40 to 0.32	0.79	0.176	−0.20 to 0.51	0.35	0.104	−0.27 to 0.45	0.59
Scaphoid fixation	**0.423**	**0.07 to 0.69**	**0.02**	**0.377**	**0.01 to 0.66**	**0.04**	0.578	0.25 to 0.79	<0.01
Skin graft	0.153	−0.22 to 0.49	0.42	0.341	−0.03 to 0.63	0.07	**0.385**	**0.01 to 0.66**	**0.04**
Microsurgery	0.009	−0.35 to 0.37	0.96	**0.483**	**0.13 to 0.73**	**<0.01**	0.461	0.10 to 0.71	0.01
Wrist arthroscopy	0.134	−0.24 to 0.47	0.48	0.159	−0.22 to 0.49	0.40	**0.391**	**0.02 to 0.67**	**0.03**
Total	0.584	0.25 to 0.79	<0.01	0.656	0.35 to 0.83	<0.01	0.670	0.37 to 0.84	<0.01

*CI = confidence interval. †Bold indicates moderate or greater correlation. ‡rs (cases): Spearman coefficient correlating scores to case log numbers. §rs (months): Spearman coefficient correlating scores to months in residency. #rs (rotations): Spearman coefficient correlating scores to number of hand rotations.

scores. The data suggest that for certain tasks (i.e., flexor tendon repair), a more experienced trainee might spend more time on this task. With experience, a trainee might know enough to invest time to use a more complex tendon repair technique, perhaps involving more strands or incorporating an epitendinous repair. A less experienced trainee, lacking knowledge of these techniques, might perform a quicker, more haphazard repair but score better because of the time factor. The scoring system proved sufficient to detect differences in skill level between CAQ hand surgeons and novice trainees (interns)[20]. However, it seems the scoring system in its current format is less effective in detecting more subtle differences in skill among intermediate trainees. This project serves to further gather and compile data for each of the psychomotor tasks and provides insight as to how the scoring system might be modified in the future with regards to weighing time and penalizing errors in performance. Given the substantial time and resource investment it took to complete this project, we did not feel it was feasible nor reasonable to retest our subjects under a modified scoring system. Score modification remains a future goal in the plan to improve and refine STEP. Second, low case log numbers for certain tasks limits the ability to correlate with those task-specific scores. Although case numbers were more robust for tasks related to bony fixation (depth of plunge and lag screw drilling), hand surgery–specific task-related case numbers (tendon repair, scaphoid and phalangeal, skin graft, microsurgery, and wrist arthroscopy) were low. Residents are likely more exposed to trauma-related cases than subspecialized cases throughout residency. With greater exposure (time in residency and hand surgical rotations), number of hand-specific cases improved modestly. Extrapolating, a comparison group of hand fellowship trainees may demonstrate higher number of task-related cases and improved correlation with scores. Three tasks (lag screw fixation, depth of plunge drilling, and phalangeal pinning) failed to demonstrate any correlation between score and case log volume or duration of training or number of hand surgery rotations. These tasks test basic orthopaedic psychomotor skills learned early during residency training thus may offer the least opportunity to differentiate skill among trainees at different points in their career.

This study further establishes construct validity of STEP and also highlights some of the limitations of the testing in its current format. The psychomotor tasks have face and content validity and construct validity. However, correlation between STEP scoring and performance in the operating room still needs to be established. Furthermore, in its current form, STEP is not a feasible tool for widespread application. Although start-up cost is minimal (~$600), the time and human resource investment for initial construction of the stations, testing and scoring applicants, is significant and requires adequate support staff to complete. In addition, it requires time away from clinical obligations for trainees to be tested. In this pilot, a very small number of observers were trained to administer and score the tests, and even so, it was sometimes difficult to reliably agree on the scores awarded. We foresee problems with inter-rater reliability and the length of time allotted to testing as barriers to wider adoption. Additional work is required to establish STEP as a validated tool based on the Messick validity framework.

Conclusion

STEP is a cost-effective simulation for the assessment of fundamental psychomotor skills in hand surgery. Previous study demonstrated construct validity between interns and CAQ hand surgeons. This study supports that STEP can differentiate between the proficiency of junior and senior residents. Furthermore, it shows that the total STEP score correlated with the total task-related case volumes, months in training, and number of hand rotations. Scoring could be modified to improve the fidelity of assessing surgical performance. Although it is both time and human resource consuming to perform, administer, and score, this study demonstrates construct validity of STEP in assessing the progression of surgical skill through residency.

Appendix

eA Supporting material provided by the authors is posted with the online version of this article as a data supplement at jbjs.org (http://links.lww.com/JBJSOA/A252). ∎

Jeffrey J. Olson, MD[1]
Bo Zhang, MD[2]
Diana Zhu, MD[2]
Evan T. Zheng, BA[3]
George S.M. Dyer, MD, FAOA[4]
Tamara D. Rozental, MD, FAOA[5]
Dawn M. LaPorte, MD, FAOA[6]

[1]Harvard Combined Orthopaedic Residency Program, Boston, Massachusetts

[2]Johns Hopkins Orthopaedic Surgery Residency Program, Baltimore, Maryland

[3]Harvard Medical School, Boston, Massachusetts

[4]Harvard Medical School, Department of Orthopaedic Surgery, Brigham and Women's Hospital, Boston Massachusetts

[5]Harvard Medical School, Department of Orthopaedic Surgery, Beth Israel Deaconess Medical Center, Boston Massachusetts

[6]Johns Hopkins Department of Orthopaedic Surgery, Baltimore, Maryland

E-mail address for D. LaPorte: dlaport1@jhmi.edu

ORCID iD for J.J. Olson: 0000-0003-2176-1522
ORCID iD for B. Zhang: 0000-0003-2162-7446
ORCID iD for D. Zhu: 0000-0002-5757-8298
ORCID iD for E.T. Zheng: 0000-0002-9793-1877
ORCID iD for G.S.M. Dyer: 0000-0002-6542-4467
ORCID iD for T.D. Rozental: 0000-0002-5003-0924
ORCID iD for D.M. LaPorte: 0000-0001-6564-2707

References

1. Mauser NS, Michelson JD, Gissel H, Henderson C, Mauffrey C. Work-hour restrictions and orthopaedic resident education: a systematic review. Int Orthop. 2016;40(5):865-73.

2. Weatherby BA, Rudd JN, Ervin TB, Stafford PR, Norris BL. The effect of resident work hour regulations on orthopaedic surgical education. J Surg Orthop Adv. 2007; 16(1):19-22.

3. Wilson T, Sahu A, Johnson DS, Turner PG. The effect of trainee involvement on procedure and list times: a statistical analysis with discussion of current issues affecting orthopaedic training in UK. Surg J R Coll Surg Edinb Irel. 2010;8(1):15-9.

4. 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;97(8):659-66.

5. Lopez G, Martin DF, Wright R, Jung J, Hahn P, Jain N, Bracey DN, Gupta R. Construct validity for a cost-effective arthroscopic surgery simulator for resident education: J Am Acad Orthop Surg. 2016;24(12):886-94.

6. Weber EL, Leland HA, Azadgoli B, Minneti M, Carey JN. Preoperative surgical rehearsal using cadaveric fresh tissue surgical simulation increases resident operative confidence. Ann Transl Med. 2017;5(15):302.

7. Ruder JA, Turvey B, Hsu JR, Scannell BP. Effectiveness of a low-cost drilling module in orthopaedic surgical simulation. J Surg Educ. 2017;74(3):471-6.

8. Coelho G, Defino HLA. The role of mixed reality simulation for surgical training in spine: phase 1 validation. Spine (Phila Pa 1976). 2018;43(22):1609-16.

9. Dumestre D, Yeung JK, Temple-Oberle C. Evidence-based microsurgical skill-acquisition series Part 1: validated microsurgical models—a systematic review. J Surg Educ. 2014;71(3):329-38.

10. Dumestre D, Yeung JK, Temple-Oberle C. Evidence-based microsurgical skills acquisition series Part 2: validated assessment instruments—a systematic review. J Surg Educ. 2015;72(1):80-9

11. Evgeniou E, Walker H, Gujral S. The role of simulation in microsurgical training. J Surg Educ. 2018;75(1):171-81.

12. Ramachandran S, Ghanem AM, Myers SR. Assessment of microsurgery competency-where are we now? Microsurgery. 2013;33(5):406-15.

13. 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;73(6):1020-5.

14. 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;29(5):906-12.

15. Kalun P, Wagner N, Yan J, Nousiainen MT, Sonnadara RR. Surgical simulation training in orthopedics: current insights. Adv Med Educ Pract. 2018;9:125-31.

16. 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;474(4): 874-81.

17. Nousiainen MT, McQueen SA, Ferguson P, Alman B, Kraemer W, Safir O, Reznick R, Sonnadara R. Simulation for teaching orthopaedic residents in a competency-based curriculum: do the benefits justify the increased costs? Clin Orthop Relat Res. 2016;474(4):935-44.

18. Peters JH, Fried GM, Swanstrom LL, Soper NJ, Sillin LF, Schirmer B, Hoffman K; SAGES FLS Committee. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. Surgery. 2004;135(1):21-7.

19. Borgersen NJ, Naur TMH, Sørensen SMD, Bjerrum F, Konge L, Subhi Y, Thomsen ASS. Gathering validity evidence for surgical simulation: a systematic review. Ann Surg. 2018;267(6):1063-8.

20. ASSH Surgical Simulation Taskforce, Wright DJ, Uong J. Establishing Validity of a Comprehensive Hand Surgical Training and Educational Platform (STEP). J Hand Surg Am. 2020;45(12):1105-14.

21. Society of American Gastrointestinal and Endoscopic Surgeons. FLS Manual Skills Written Instructions and Performance Guidelines. Foundations of Laparoscopic Surgery; 2014. Available at: https://www.flsprogram.org/wp-content/uploads/2014/ 03/Revised-Manual-Skills-Guidelines-February-2014.pdf. Accessed April 1, 2020.

22. ABOG announces new eligibility requirement for board certification. 2018. Available at: https://www.flsprogram.org/news/abog-announces-new-eligibility-requirement-board-certification/. Accessed April 1, 2019.

23. Shah D, Haisch CE, Noland SL. Case reporting, competence, and confidence: a discrepancy in the numbers. J Surg Educ. 2018;75(2):304-12.

24. Jeray KJ, Frick SL. A survey of resident perspectives on surgical case minimums and the impact on milestones, graduation, credentialing, and preparation for practice: AOA critical issues. J Bone Joint Surg Am. 2014;96(23):e195.

25. Okike K, Berger PZ, Schoonover C, O Toole RV. Do orthopaedic resident and fellow case logs accurately reflect surgical case volume? J Surg Educ. 2018;75(4): 1052-7.

26. ACGME: Orthopaedic Surgery Case Logs National Data Report 2018-2019. Accreditation Council for Graduate Medical Education; 2019. Available at: https:// www.acgme.org/Portals/0/PDFs/260_National_Report_Program_Version_2018-2019.pdf. Accessed April 1, 2020.