Validation of an Online Assessment of Orthopedic Surgery Residents' Cognitive Skills and Preparedness for Carpal Tunnel Release Surgery

Janet Shanedling, PhD
Ann Van Heest, MD
Michael Rodriguez, PhD
Matthew Putnam, MD
Julie Agel, MA

Abstract

Background Surgical competence requires both cognitive and technical skills. Relatively little is found in the literature regarding the value of Web-based assessments to measure surgery residents' mastery of the competencies.

Objective To determine the validity and reliability of 2 online instruments for predicting the cognitive preparedness of residents for performing carpal tunnel release surgery.

Method Twenty-eight orthopedic residents and 2 medical school students responded to an online measure of their perception of preparedness and to an online cognitive skills assessment prior to an objective structured assessment of technical skills, in which they performed carpal tunnel release surgery on cadaveric specimens and received a pass/fail assessment. The 2 online assessments were analyzed for their internal reliability, external correlation with the pass/fail decision, and construct validity.

Results The internal consistency of the perception of preparedness measure was high ($\alpha = .92$) while the cognitive assessment was less strong ($\alpha = .65$). Both instruments demonstrated moderately strong correlations with the pass/fail decision, with Spearman correlation of .606 ($P = .000$) and .617 ($P = .000$), respectively. Using logistic regression to analyze the predictive strength of each instrument, the perception of preparedness measure demonstrated a 76% probability ($\eta^2 = .354$) and the cognitive skills assessment a 73% probability ($\eta^2 = .381$) of correctly predicting the pass/fail decision. Analysis of variance modeling resulted in significant differences between levels at $P < .005$, supporting good construct validity.

Conclusions The online perception of preparedness measure and the cognitive skills assessment both are valid and reliable predictors of readiness to successfully pass a cadaveric motor skills test of carpal tunnel release surgery.

Background

Surgical competence requires excellence in both cognitive and technical skills. Definitive metrics for assessing competence of surgical residents in one or both areas are being researched across medical education and within surgical specialties. A number of assessment methods have been developed and are being validated for cognitive and technical skills assessment. Simulation has been employed for learning assessment in laparoscopy and general surgery,[1–3] and positive, valid outcomes have been achieved using global rating scales and checklists in objective structured assessments of technical skills (OSATS) in obstetrics and gynecology, general surgery, laparoscopy, and other specialties.[2,4–6]

The validity of self-assessment as a means of ascertaining residents' surgical competence has not been robustly demonstrated to date. Some studies[7,8] suggest that the capacity of physicians and residents to accurately self-assess is limited, while others[5,9] suggest that residents' self-assessments can be as reliable and valid as faculty ratings. Self-assessment has also been effective in evaluating the 6 Accreditation Council for Graduate Medical Education competencies for otolaryngology residents[10] and for chart review skills of internal medicine residents.[11]

Relatively little is found in the literature regarding the value of Web-based assessments to measure the

All authors are at the University of Minnesota. **Janet Shanedling, PhD,** is Director of Academic Health Center Educational Development and Assistant Professor at the School of Nursing; **Ann Van Heest, MD,** is Professor at the Department of Orthopaedic Surgery; **Michael Rodriguez, PhD,** is Associate Professor of Quantitative Methods in Education at the College of Education & Human Development; **Matthew Putnam, MD,** is Professor at the Department of Orthopaedic Surgery; and **Julie Agel, MA,** is Statistician at the Department of Orthopaedic Surgery.

This study was supported by a TEL (Technology Enhanced Learning) Grant, University of Minnesota, Minneapolis, Minnesota.

Corresponding author: Janet Shanedling, PhD, AHC Office of Education, R664-A Children's Rehabilitation Center, MMC 501 Mayo, 420 Delaware St SE, Minneapolis, MN 55455, 612.626.3873, janet@umn.edu

Received December 15, 2009; revision received March 1, 2010; accepted March 29, 2010.

DOI: 10.4300/JGME-D-09-00103.1

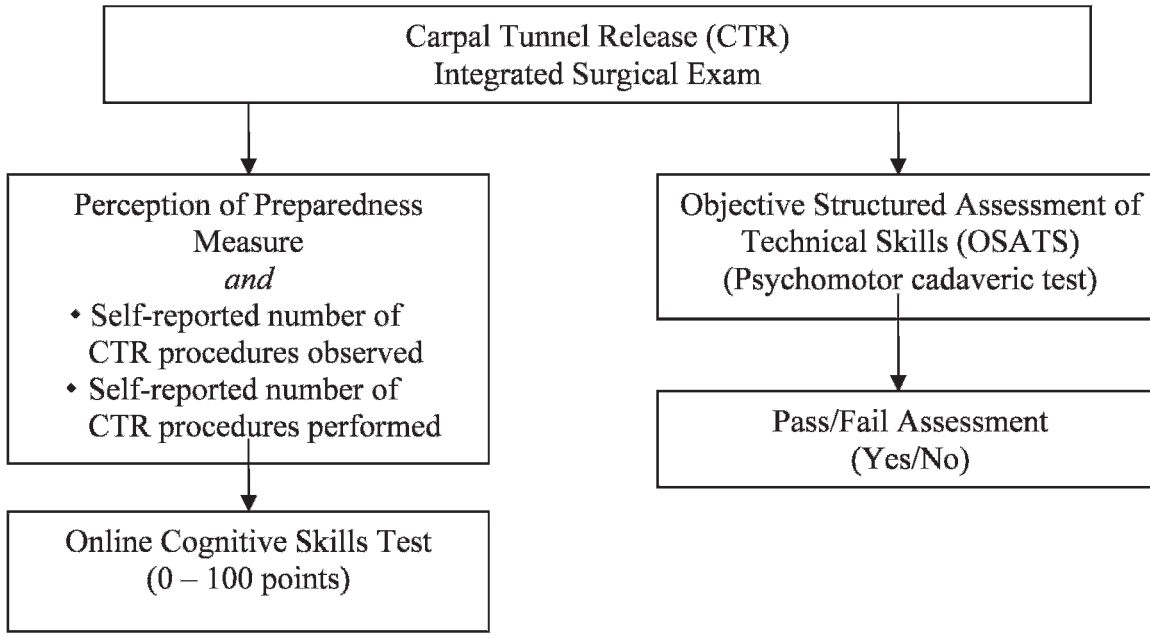


FIGURE 1 **Carpal Tunnel Release Surgery Examination Protocol**

competencies required by surgery residents. Schell and Flynn[12] report on a self-paced, online curriculum for minimally invasive surgery. The study used Web-based modules for formative learning rather than as a summative assessment tool.

Our study evaluated the validity of one measure of residents' self-reported perception of preparedness (P of P) for performing carpal tunnel release (CTR) surgery, and of an online cognitive skills test that assessed their knowledge regarding CTR surgery. We sought to prove that the P of P measure and the online cognitive skills test are reliable and valid means of assessing the readiness and competence of orthopedic surgery residents to successfully perform CTR surgery.

Methods

Our study was part of an examination sequence evaluating upper extremity knowledge and motor skills in 28 participants in the Orthopaedic Surgery Residency Program at the University of Minnesota; the study included residents ranging from postgraduate year (PGY)-1 to PGY-5, 2 PGY-6 hand fellows, and 2 medical students, all in good standing at the time of the examination. Learners consented to participate in the study. The complete cognitive and motor skills examination protocol is depicted in FIGURE 1. The study received Institutional Review Board approval.

Online Perception of Preparedness Measure

Participants completed a 7-item measure, via the WebCT Vista course management site (Blackboard, Inc, Washington, DC), of their self-reported level of confidence and readiness to perform the key components of CTR surgery using a 5-point Likert scale to rate their confidence in anatomy, physical exam and patient preparation, interpretation of literature, surgical steps, operative reporting, and management of complications or errors. Residents also indicated the number of CTR procedures they had observed and the number of procedures they had performed to date.

Online Cognitive Skills Test

Participants then completed a 100-point, 43-item test with dictated operative report via microphone in the same WebCT Vista site. The test content was developed by a board-certified orthopedic surgeon/Certificate of Added Qualifications–certified hand surgeon, and it underwent 3 iterations following review by 4 board-certified orthopedic surgeons/Certificate of Added Qualifications–certified hand surgeons.

The examination included: a cognitive skills assessment for the anatomy of the carpal tunnel area; a preoperative evaluation that comprised indications for surgery, interpretation of electromyography/nerve conduction velocity, and determination of whether CTR is indicated; an interpretation of the literature section that applied the literature to a patient's case; a surgical steps section that required participants to demonstrate knowledge of correct patient positioning, topical landmarks, and layers of dissection; a section on the surgical incision that required residents to use their computer mouse as a scalpel to draw the actual incision site; a section that required participants to dictate an operative report; and a final section, complications, which assessed participants' ability to recognize, respond, and avoid nerve or artery lacerations.

TABLE 1 **Internal Reliability of the Perception of Preparedness Measure and Cognitive Skills Test Scores**

Instrument	No. Respondents	No. Items or Sections	Coefficient α	Range of Spearman Correlations
Perception of preparedness measure	29	7 items	.922	.622–.968
Cognitive skills online test	24	7 sections	.662	−.130–.642

OSATS Testing

A week after the online cognitive skills test, the residents and medical students performed carpal tunnel surgery, unassisted, on cadaveric specimens in an OSATS exam. Two hand surgeons independently evaluated each resident and determined an independent pass/fail assessment. The inter-rater reliability of the pass/fail determinations was .759. Another study[13] used 4 measures to assess knowledge and technical skills: a Web-based knowledge test (one of the two measures addressed in this report), a detailed observation checklist, a global rating scale, and a pass/fail assessment applied in a cadaver testing lab. While the purpose of the former study was to evaluate the reliability and validity of the 4 testing measures used to assess motor skill competence in performing CTR, the focus of this study was to evaluate the reliability and validity of 2 online measures assessing cognitive knowledge and preparedness, a self-reported P of P measure, and the cognitive skills test since online learning.

Statistical Methods

Scores from the P of P measure and for the cognitive test were automatically calculated and stored by WebCT Vista (scores for the audio-dictated operative report were hand-entered into the online grade book). Item scores for each individual were aggregated into the 7 sections of the cognitive skills test and exported to a database. The scores of one individual who did not complete the P of P measure and portions of the cognitive skills test were not included in the analysis. The score of a second individual who did not complete the audio section of the cognitive skills test was dropped from analysis of that test only.

To evaluate the internal consistency of the 7-item P of P measure, we calculated a coefficient α to obtain overall score consistency and item discrimination indices. A coefficient α was also calculated to determine the internal reliability of sections and total scores on the cognitive skills test.

TABLE 2 **Spearman Correlations Across Measures**[a]

	Perception of Preparedness Total Score	Cognitive Skills Test Total Score	Number of Procedures Observed	Number of Procedures Performed
Cognitive skills test total score	*.549*			
P value	.002			
n	29			
Number of procedures observed	*.774*	*.477*		
P value	.000	.009		
n	29	29		
Number of procedures performed	*.733*	*.515*	*.834*	
P value	.000	.004	.000	
n	29	29	29	
Pass/fail	*.606*	*.589*	*.577*	*.736*
P value	.000	.000	.001	.000
n	29	30	29	29

[a] Correlations with $P < .01$ are italic.

TABLE 3 **Strength of Prediction (Logistic Regression)**

Measure	η^2	Correct Prediction of Pass/Fail Decision, %	*d*
Perception of preparedness measure	.354	76	1.2
Cognitive skills online test	.381	73	1.3
Number of procedures observed	.220	76	1.0
Number of procedures performed	.320	93	1.2

Abbreviations: *d*, mean difference (in standard deviations) on each measure between those who passed and those who failed; η^2, percentage of variance explained.

Spearman correlations were run to determine the external correlations of a number of measures: the P of P individual item and total scores, cognitive skills section and total scores, number of procedures observed, number of procedures performed, and pass/fail decisions. Statistical significance was noted at $P < .01$. For both the P of P and cognitive skills scores, standardized mean differences and variance explained (η^2) were analyzed at the $P < .01$ level to determine the degree to which the variation of scores on the 2 instruments related to the variance in pass/fail decisions. Coefficients of determination and logistic regression were used to determine the degree to which each of the 2 online instruments, as well as numbers of procedures observed and performed, could predict the pass/fail decisions.

Construct validity was provided by several layers of analysis. First, the iterative design process provided significant support for the content coverage and meaning of each measure. The reliability evidence supported score interpretation, providing evidence of internal consistency and score consistency. Internal and external correlations for each of the measures also provided criterion-related validity evidence. Construct-related validity evidence was obtained by correlating the mean scores of residents in each year of the program on both the P of P measure and the cognitive skills test with the pass/fail decision.

Results

Reliability

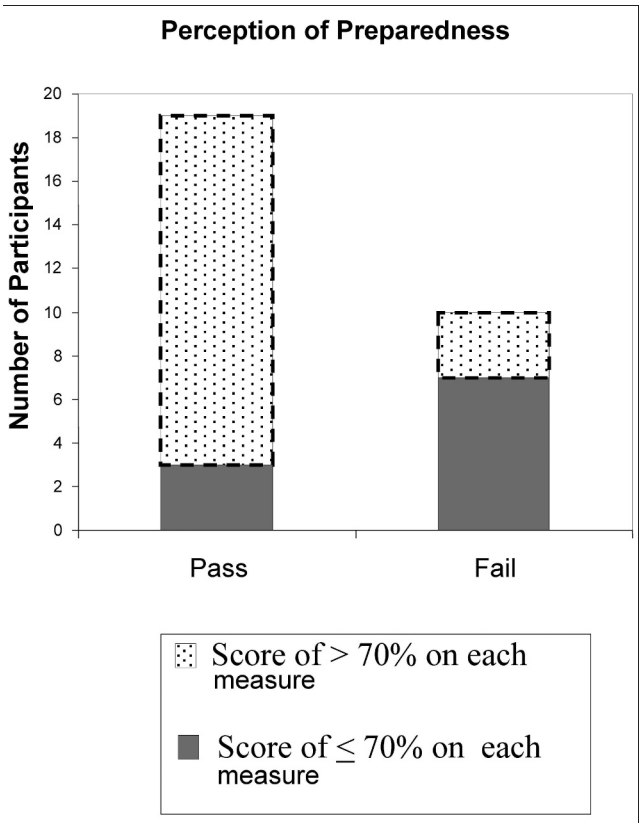
Reliability supports the validity argument by providing evidence of internal consistency and score consistency. The internal consistency of the 7-item P of P measure and of the 7-section cognitive skills measure were analyzed by calculation of coefficient α, by item-total correlations (item discrimination), and by Spearman correlations between items or sections on the instruments. TABLE 1 presents internal reliability data for the P of P measure and for the cognitive skills test.

Reliability of P of P Measure

The coefficient α for the composite measure is very high, $\alpha = .92$. Reliability greater than .80 is generally accepted for determining competency levels for groups; this level would be appropriately high for individual-level decisions. The correlations among the individual items on the

TABLE 4 **Average Scores on Each Measure by Postgraduate Year (PGY)**

PGY	N	Perception of Preparedness (0–35 points), mean (SD)	Cognitive Skills Online Test (0–100 points), mean (SD)	Number of Procedures Observed, mean (SD)	Number of Procedures Performed, mean (SD)	Pass/Fail Decision (Pass = 1; Fail = 0), mean (SD)
0	2	19.00 (7.07)	75.50 (13.44)	2.00 (1.41)	.00 (.00)	.00 (.00)
1	2	21.50 (4.95)	70.50 (3.54)	3.50 (2.12)	.00 (.00)	.00 (.00)
2	8	23.25 (3.24)	81.75 (7.69)	2.50 (2.00)	.25 (.46)	.25 (.46)
3	5	28.20 (3.63)	86.59 (5.43)	7.30 (5.12)	10.70 (5.47)	1.00 (.00)
4	6	31.67 (3.61)	95.57 (3.91)	20.00 (10.95)	24.17 (16.86)	1.00 (.00)
5	4	30.00 (3.37)	85.88 (4.97)	32.50 (25.98)	39.25 (25.86)	1.00 (.00)
6	2	35.00 (.00)	88.00 (11.31)	35.00 (21.21)	32.50 (10.61)	1.00 (.00)
η^2		.667	.594	.569	.629	.775



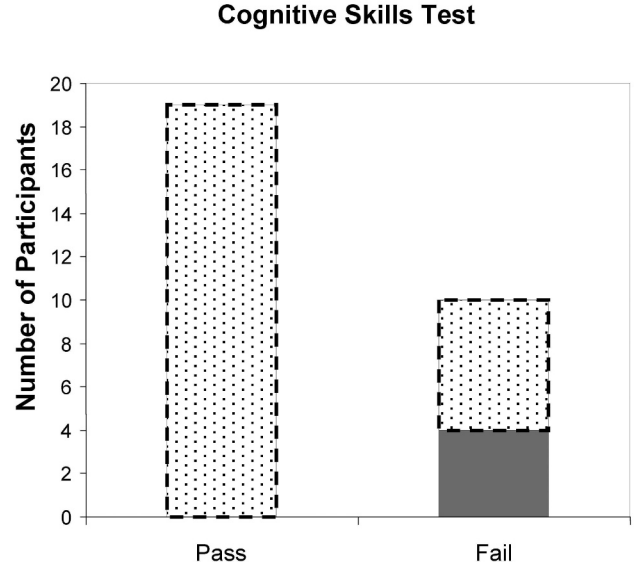


FIGURE 2 **Correlation of Each Online Measure With the Pass/Fail Assessment of the Objective Structured Assessments of Technical Skills Cadaveric Exam**

measure, all significant at the $P < .001$ level, indicate a very high level of internal consistency.

Reliability of Cognitive Skills Test

The reliability coefficient for the total cognitive skills online test is not strong (.66), being just below the threshold typically considered appropriate for research purposes (.70), and well below the level required for group-level decisions. The item discrimination indices provide information about the contribution of each item to the total cognitive skills test score, indicating inconsistencies in section scores and total scores. Three sections are inconsistent with the other sections, with item-total correlations of .024, .029, and .140. All other sections correlated between .470 and .698, indicating that the capacity of these 3 sections to differentiate among residents is not consistent with that of the overall test measure. Two of those sections contained only 1 question each. TABLE 1 summarizes the internal reliability measures for both the P of P and the cognitive skills scores, including ranges of Spearman correlations used to correlate sections for a measure of internal consistency. The Spearman correlations among the sections of the cognitive skills test, excluding the 3 sections of interpretation of the literature, surgical incision, and complications/errors, represented a smaller range than that indicated in TABLE 1, from .478 to .643. Each correlation was significant, with $P < .01$.

Validity

In addition to determining whether the items and sections of each of the 2 assessment instruments consistently measure a unified construct, it is important to also ascertain whether each instrument actually measures residents' preparedness to successfully pass an examination of CTR surgical skills. Each instrument was therefore correlated with the other, and, in addition, with number of procedures observed, number of procedures performed, and the ultimate pass/fail decision.

Spearman correlations across measures suggest intercorrelation of all measures, with $P < .01$, as shown in TABLE 2. Correlations suggest that each measure provides a similar, but distinct, contribution regarding the residents' readiness for CTR surgery. The P of P total scores correlate more strongly with the number of procedures observed and number of procedures performed than do the cognitive skills test total scores. However, the difference in correlations between P of P and cognitive skills with the pass/fail decision is much smaller (.606 and .589, respectively).

Logistic regression, a logistic curve used to predict the probability of the occurrence of an event, was also used to analyze the predictive strength of each of the 2 online and 2 experiential measures relative to the pass/fail decision (TABLE 3). This analysis indicates that, even if used alone, each of the 2 online measures offers a valid predictive capacity for the OSATS exam pass/fail decision. The difference in the means of those measures based on the pass/

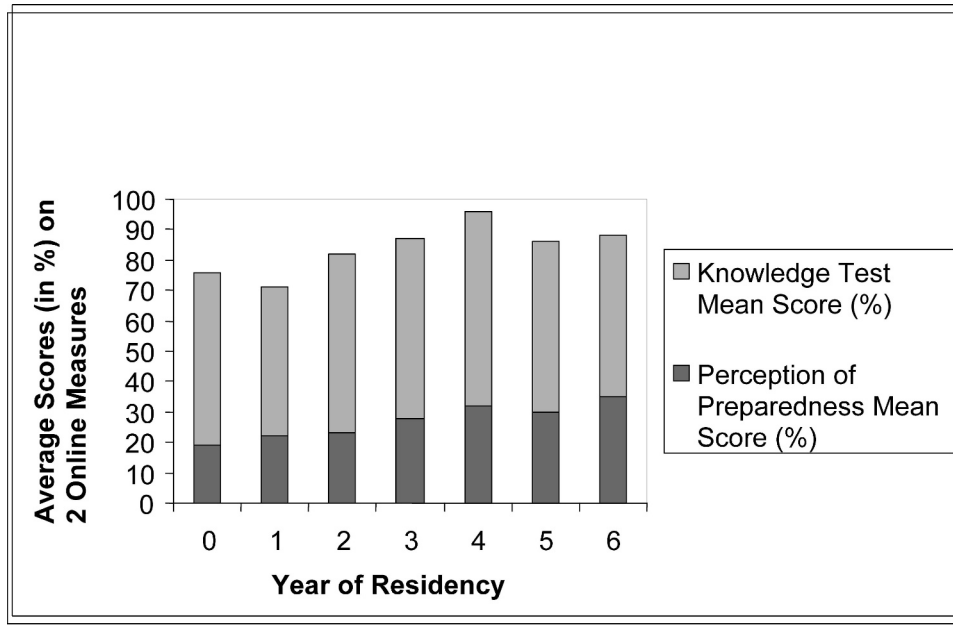


FIGURE 3 **Correlation of Average Scores (in %) by Postgraduate Year on Each Online Measure With Pass/Fail Assessment on Objective Structured Assessments of Technical Skills Cadaveric Exam**

fail decision is summarized in a standardized mean difference metric (*d*). This provides an index of the mean difference (in standard deviations) on each measure between residents who passed versus those who failed on the overall performance decision. The differences between passing and failing groups were very large, more than one standard deviation different.

Finally, construct-related validity is supported by evidence from performance on each measure based on year in the program. As shown in TABLE 4, scores on each measure increase fairly consistently as the number of years in the program increases, particularly from PGY-1 to PGY-4. Residents begin to observe and perform procedures primarily in PGY-3 and PGY-4; prior to year 3, it is unlikely that residents can obtain a pass decision on the motor skills performance measure. For each of the measures in TABLE 4, the analysis of variance model resulted in significant differences between PGY levels at $P < .005$, indicating that PGY level explained a significant amount of the variation in each measure; this was also indicated by η^2(percent of variance explained).

Discussion

This study evaluated the reliability and validity of 2 online instruments for assessing readiness and competence of residents to successfully perform carpal tunnel release surgery. The perception of preparedness measure and the cognitive skills test were each determined to provide valid assessments of the cognitive skills that are prerequisite for performing CTR surgery, and each significantly and positively correlated with the ultimate pass/fail decision made in a cadaveric laboratory (FIGURE 2). Each measure demonstrated construct validity through generally increasing mean scores according to increasing years of residency (FIGURE 3).

The 7-item P of P measure provides highly reliable and internally consistent scores as a self-reporting measure. Its strong correlations with the reported number of procedures observed and performed suggest that the instrument measures confidence levels, and that such confidence may be related to the extent of operating room experience, either in an observer or surgeon role.

Overall, the cognitive skills online test also provides reliable scores of knowledge required for carpal tunnel release surgery. Three sections of the test—interpretation of literature, surgical incision, and complications/errors—did not significantly correlate with the other 4; 2 of those sections are composed of only 1 question. In the complications/errors section, the number of residents responding correctly to most or all of the 8 questions was high. That lack of variability may indicate that the skills being assessed reflect basic cognitive skills of surgery, rather than decision-making skills distinct to carpal tunnel surgery.

The use of logistic regression presents solid evidence that each of the online instruments can provide valuable information about a resident's preparedness for surgery. While the number of procedures performed provides the

strongest single predictor of passing or failing the surgical skills test, the P of P measure alone is as strong a predictor as the number of procedures observed. Revising the measure from the use of a 5-point rating scale to one employing a range from novice to proficient (or master) may provide even more informative self-reporting.[14] In conjunction with the cognitive skills test, this measure can serve as an integrated approach for both self-assessment and objective assessment of cognitive readiness to demonstrate technical skill. Additional testing of each instrument is required to determine the predictive efficacy of each measure when administered alone. Consideration is being given to identifying a control group of residents who will not complete the online assessments in order to compare their results in the OSATS exam to those of residents who do complete the online measures.

Analysis of variance results suggest good construct validity for both instruments. Based on the data, assessment at the end of PGY-2 may provide the optimal point of differentiation between those prepared and those less prepared to successfully perform CTR surgery.

Conclusions

Determining valid and reliable methods for testing operative skills in orthopedic surgery is a primary concern at a time when documentation of competence and cost-effective assessment metrics are possible means for minimizing medical errors and adverse events. Cadaver labs and computerized simulators offer excellent learning and assessment tools, but at significant cost.

This study has demonstrated that 2 online measures, the perception of preparedness measure and the cognitive skills test, have the potential to provide valid evidence of the readiness of orthopedic surgery residents to perform carpal tunnel release surgery.

References

1 Aggarwal R, Grantcharaov T, Moorthy K, et al. An evaluation of the feasibility, validity, and reliability of laparoscopic skills assessment in the operating room. *Ann Surg*. 2007;245(6):992–999.

2 Martin JA, Regehr G, Reznick R, et al. Objective structured assessment of technical skill (OSATS) for surgical residents. *Br J Surg*. 1977;84(2):273–278.

3 Reznick RK, MacRae H. Teaching surgical skills—changes in the wind. *N Engl J Med*. 2006;355(25):2664–2669.

4 Chou B, Bowen CW, Handa VL. Evaluating the competency of gynecology residents in the operating room: validation of a new assessment instrument. *Am J Obstet Gynecol*. 2008;199(5):571.e1–571.e5.

5 Mandel LS, Goff BA, Lentz GM. Self-assessment of resident surgical skills: is it feasible? *Am J Obstet Gynecol*. 2005;193(5):1817–1822.

6 Goff B, Mandel L, Lentz G, et al. Assessment of resident surgical skills: is testing feasible? *Am J Obstet Gynecol*. 2005;192(4):1331–1340.

7 Davis DA, Mazmanian PE, Fordis M, Van Harrison RV, Thorpe KE, Perrier L. Accuracy of physician self-assessment compared with observed measures of competence: a systematic review. *JAMA*. 2006;296(9):1094–1101.

8 Brewster LP, Risucci DA, Joehl RJ, et al. Comparison of resident self-assessments with trained faculty and standardized patient assessments of clinical and technical skills in a structured educational module. *Am J Surg*. 2008;195(1):1–4.

9 Ward M, MacRae H, Schlachta C, et al. Resident self-assessment of operative performance. *Am J Surg*. 2003;185(6):521–524.

10 Roark RM, Schaefer SD, Guo-Pei Y, Branovan DI, Peterson SJ, Wei-Nchih L. Assessing and documenting general competencies in otolaryngology resident training programs. *Laryngoscope*. 2006;116(5):682–695.

11 Hildebrand C, Trowbridge E, Roach M, Sullivan AG, Broman AT, Vogelman B. Resident self-assessment and self-reflection: University of Wisconsin-Madison's five-year study. *J Gen Intern Med*. 2009;24(3):361–365.

12 Schell SR, Flynn TC. Web-based minimally invasive surgery training: competency assessment in PGY 1-2 surgical residents. *Curr Surg*. 2004;61(1):120–124.

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

14 Leach D. Unlearning: it is time. *ACGME Bull*. 2005;4:2–3.