EDU

Training less-experienced faculty improves reliability of skills assessment in cardiac surgery

Xiaoying Lou, BS,[a] Richard Lee, MD,[b] Richard H. Feins, MD,[c] Daniel Enter, MD,[a] George L. Hicks, Jr, MD,[d] Edward D. Verrier, MD,[e] and James I. Fann, MD[f]

Objective: Previous work has demonstrated high inter-rater reliability in the objective assessment of simulated anastomoses among experienced educators. We evaluated the inter-rater reliability of less-experienced educators and the impact of focused training with a video-embedded coronary anastomosis assessment tool.

Methods: Nine less-experienced cardiothoracic surgery faculty members from different institutions evaluated 2 videos of simulated coronary anastomoses (1 by a medical student and 1 by a resident) at the Thoracic Surgery Directors Association Boot Camp. They then underwent a 30-minute training session using an assessment tool with embedded videos to anchor rating scores for 10 components of coronary artery anastomosis. Afterward, they evaluated 2 videos of a different student and resident performing the task. Components were scored on a 1 to 5 Likert scale, yielding an average composite score. Inter-rater reliabilities of component and composite scores were assessed using intraclass correlation coefficients (ICCs) and overall pass/fail ratings with kappa.

Results: All components of the assessment tool exhibited improvement in reliability, with 4 (bite, needle holder use, needle angles, and hand mechanics) improving the most from poor (ICC range, 0.09-0.48) to strong (ICC range, 0.80-0.90) agreement. After training, inter-rater reliabilities for composite scores improved from moderate (ICC, 0.76) to strong (ICC, 0.90) agreement, and for overall pass/fail ratings, from poor (kappa = 0.20) to moderate (kappa = 0.78) agreement.

Conclusions: Focused, video-based anchor training facilitates greater inter-rater reliability in the objective assessment of simulated coronary anastomoses. Among raters with less teaching experience, such training may be needed before objective evaluation of technical skills. (J Thorac Cardiovasc Surg 2014;148:2491-6)

See related commentary on pages 2497-8.

Supplemental material is available online.

From Northwestern University,[a] Chicago, Ill; St Louis University,[b] St Louis, Mo; University of North Carolina at Chapel Hill,[c] Chapel Hill, NC; University of Rochester Medical Center,[d] Rochester, NY; University of Washington Medical Center,[e] Seattle, Wash; and Stanford University,[f] Stanford, Calif.

This study was supported by the Thoracic Surgery Directors Association and the Joint Council for Thoracic Surgery Education, the Thoracic Surgery Foundation for Research and Education Simulation Grant (R.H.F. and J.I.F.), and the Western Thoracic Surgical Association Doty Award (J.I.F.). Corporate sponsors of this session include the Chamberlain Group, Sorin Group, St Jude Medical Inc, CryoLife Inc, Scanlan International, LifeLike BioTissue Inc, Teleflex Medical, and Ethicon Inc.

Disclosures: James I. Fann reports equity ownership in Apica. Richard H. Feins reports equity ownership in KindHeart Inc. The other authors have nothing to disclose with regard to commercial support.

Read at the 40th Annual Meeting of The Western Thoracic Surgical Association, Dana Point, California, June 25-28, 2014.

Received for publication June 23, 2014; revisions received Aug 25, 2014; accepted for publication Sept 4, 2014; available ahead of print Oct 11, 2014.

Address for reprints: James I. Fann, MD, Department of Cardiothoracic Surgery, Stanford University, 300 Pasteur Dr, Stanford, CA 94305 (E-mail: jfann@stanford.edu).

0022-5223/$0.00

Published by Elsevier Inc. on behalf of The American Association for Thoracic Surgery

http://dx.doi.org/10.1016/j.jtcvs.2014.09.017

Technical skill is a key component of surgical competence and a core component of cardiothoracic (CT) surgery training. For the last 2 decades, the use of surgical simulators has evolved as a way for trainees to learn and practice technical skills in a safe, cost-effective, and low-stress environment.[1] Simulation also affords opportunities for direct observation for formative and summative assessment. For such assessments to accurately reflect a trainee's level of technical skill, however, they must be standardized. As the role of simulation expands with the potential for incorporation in high-stakes settings, such as those used for promotion and certification, it is paramount that assessment tools demonstrate high inter-rater reliability and ease of execution.[2]

In CT surgery, the Joint Council on Thoracic Surgery Education (JCTSE) and the Thoracic Surgery Directors Association (TSDA) have developed instruments to evaluate trainee competence in common operative procedures.[3-6] For the JCTSE coronary artery anastomosis assessment tool, high inter-rater reliability among experienced educators and senior faculty members, even without rater training, has been demonstrated.[7] Because junior faculty members with less experience as educators are often charged with evaluating trainee competence, it is requisite that they achieve similar levels of inter-rater reliability.

Currently, inter-rater reliability among less-experienced educators has not been established. Moreover, although rater

EDU

Abbreviations and Acronyms

CT	= cardiothoracic
ICC	= intraclass correlation coefficient
JCTSE	= Joint Council on Thoracic Surgery Education
P/F	= pass/fail
TSDA	= Thoracic Surgery Directors Association

training has been recognized to improve inter-rater reliability, its effects have not been assessed in CT surgery. To address these needs, a skills assessment session was held at the JCTSE Educate the Educators program at the TSDA Boot Camp in 2013. The session included rater training for the JCTSE coronary artery anastomosis assessment tool. Rater training aims to improve rater performance by developing the necessary knowledge, skills, and attitudes to accurately evaluate skills and competencies.[8,9] The type of training used in this session can be characterized as performance dimension training with elements of frame of reference training. Performance dimension training teaches raters to recognize appropriate behaviors associated with each dimension targeted for evaluation using written or visual depictions. Examples representing expert consensus are provided to raters so that they associate similar behavioral cues with the dimension being evaluated. Frame of reference training involves recognition and expert-facilitated discussion of discrepancies between raters to provide feedback that improves rater performance.[9,10]

Although no standardized rater training techniques currently exist, it is generally agreed that jointly examining the sources of inter-rater variability and establishing a consensus to address any uncertainties enhances rater reliability.[11] In this study, we thus evaluated inter-rater reliability of less-experienced educators and the impact of focused training with a video-embedded coronary anastomosis assessment tool on improvement in inter-rater reliability.

MATERIALS AND METHODS

Nine CT surgery faculty members from different academic institutions participated as raters in the JCTSE Educate the Educators session on assessment at the TSDA Boot Camp at University of North Carolina, Chapel Hill. During coronary anastomoses training sessions, 4 individuals (2 medical students and 2 CT surgical residents) were recruited to perform a coronary artery anastomosis using a simulator; the individuals had a level of experience with coronary anastomoses consistent with their level of training. Approval for the study was obtained from the institutional review board at the University of North Carolina, Chapel Hill.

Model for Coronary Artery Anastomoses and Video Recordings

Coronary vessel anastomoses were performed using a synthetic graft task station and video recorded.[4] The medical students and residents anastomosed a 3-mm synthetic vein graft onto a 3-mm synthetic target vessel mounted in a portable chest model (HeartCase; Chamberlain Group, Great Barrington, Mass) using 6-0 polypropylene sutures and surgical instruments (Figure 1). The video recordings were edited to approximately 5 to 6 minutes, which included representative clips for subsequent evaluation of the assessment components. All video recordings were de-identified and limited to views of the simulation model and the participant's hands.

Joint Council on Thoracic Surgery Education Assessment Tool for Coronary Artery Anastomosis

The JCTSE assessment tool consists of 13 assessment components: arteriotomy, graft orientation, bite, spacing, needle holder use, use of forceps, needle angles, needle transfer, suture management, knot tying, hand mechanics, use of both hands, and economy of time and motion. Because of the limitations of the simulation model and the varying degree of aid of an assistant surgeon, 3 assessment components (arteriotomy, graft orientation, and economy of time and motion) could not be evaluated. The other components are scored on a Likert scale from 1 (poor) to 5 (excellent), with anchoring of 1, 3, and 5 ratings with behavioral descriptors (Appendix Table E1).

Training Protocol and Data Collection

After a brief introduction to the use of simulation of coronary artery anastomosis, raters were provided paper copies of the assessment tool and allowed 5 minutes to review the tool and behavioral anchors. No further explanation of the tool or its anchors was provided. Raters then consecutively viewed and evaluated 2 video recordings of 1 medical student and 1 resident performance of a coronary anastomosis on the simulator. For each anastomosis, a rating from 1 to 5 was assigned for 10 assessment components, yielding an average composite score. Each performance also received an overall pass/fail (P/F) rating. All evaluations were completed on paper, independently, and without knowledge of the subject's level of training. Assessment took place concurrently with video viewing. Afterward, raters used audience response clickers to input their ratings, which were captured by live polling software (TurningPoint 5.2.1; Turning Technologies, Youngstown, Ohio). This setup provided raters with immediate visual feedback that compared their ratings with those by the rest of the group.

Training consisted of 30 minutes of expert-facilitated discussion of the behavioral descriptors used to anchor the assessment tool. Raters were asked to review a series of 10- to 15-second video clips embedded into the assessment tool depicting the levels of skill corresponding to 1, 3, and 5 ratings for each of the 10 assessment components (Figure 2). The embedded video clips had been collected before the rating session and had been deemed to be representative samples of these anchors by the group of experienced raters involved in the development of the assessment tool.[7] All questions posed by raters were also answered, and areas of discrepancy were discussed. Immediately after the training session, all raters evaluated the remaining 2 videos of a different medical student and resident performing the task using the same procedure as outlined previously.

Statistical Analysis

Data are expressed as mean $\pm$ standard deviation. Inter-rater reliability of composite scores as continuous variables and assessment component scores as ordinal variables were assessed using intraclass correlation coefficients (ICCs), and overall P/F ratings as dichotomous variables using Fleiss' kappa of concordance (κ). Internal consistency reliability among assessment components was assessed with Cronbach's α. Reliability is an index ranging from 0 to 1. Although no consensus on index levels currently exists, it is generally accepted that tools with reliabilities in the 0.0 to 0.5 range are imprecise and those in the 0.5 to 0.8 range are moderately reliable. Tools with reliability indices greater than 0.8 exhibit

FIGURE 1. The coronary artery anastomosis simulator (HeartCase; Chamberlain Group, Great Barrington, Mass) is a moderate-fidelity simulator with a synthetic vessel mounted on an adjustable stand. The end-to-side anastomosis is performed using a 3-mm target vessel and a 3-mm graft vessel.

strong reliability and can be used with confidence for high-stakes examinations.[12,13]

RESULTS

The characteristics of the CT surgery faculty (66.7% are male, mean years post-CT surgical training, 4.5 ± 3.5) are listed in Table 1. Seven faculty (77.8%) identified themselves as general thoracic surgeons, and 6 faculty (66.7%) were involved in teaching both cardiac and thoracic surgery trainees. Eight faculty (88.9%) taught coronary anastomoses in the simulated or clinical setting, and the majority (77.8%) had no experience with using surgical rating tools in the past. Reliability data, including (1) inter-rater reliability among raters for composite scores, individual assessment component scores, and overall P/F ratings; and (2) internal consistency reliability among the assessment components are listed in Table 2.

Inter-Rater Reliability

Composite scores. Inter-rater reliability for composite score increased from moderate (ICC = 0.76) agreement before training to strong (ICC = 0.90) agreement after training.

Individual assessment component scores. Before training, 6 assessment components (spacing, use of forceps, needle transfer, suture management, knot tying, use of both hands) demonstrated moderate agreement (ICC range, 0.53-0.71), 3 assessment components (needle holder use, needle angles, hand mechanics) demonstrated poor agreement (ICC range, 0.39-0.48), and 1 assessment component (bite) exhibited very poor agreement (ICC, 0.09). The mean ICC across the 10 components was 0.51. After training, all components of the assessment tool exhibited improvement, with 4 (bite, needle holder use, needle angles, and hand mechanics) improving the most from poor (ICC range, 0.09-0.48) to strong (ICC range, 0.80-0.90) agreement. The mean ICC across components increased to 0.84 after training.

Overall pass/fail ratings. Inter-rater reliability for overall P/F ratings increased from poor (kappa = 0.20) agreement before training to moderate (kappa = 0.78) agreement after training.

Internal consistency reliability. Internal consistency remained stable with Cronbach's α greater than 0.98 both before and after training. The relationship between composite scores and overall P/F ratings is presented in Figure 3. Across all performances, a passing score (a failing score) was associated with a mean composite score of 3.8 ± 0.61 (2.1 ± 0.57), ranging from 2.6 to 4.9 (1.0-2.7). Two composite scores less than 3.0 were given a "pass" rating; however, both of these instances occurred before training. In general, regardless of whether ratings were completed before or after training, a performance with a composite score greater than 3.0 was given a pass, whereas less than 3.0 was a fail.

DISCUSSION

Standardizing assessment tools for competency evaluation must parallel the development of surgical simulation-based learning if the latter is to gain widespread implementation. Such efforts necessitate that reliability be established among raters with varying levels of clinical and teaching experience. It has been demonstrated that high inter-rater reliability can be achieved among experienced educators and program directors using the JCTSE coronary anastomosis assessment tool without prior rater training.[7] In this study, focused, video-based anchor training of less-experienced CT surgery faculty led to greater inter-rater reliability in the objective assessment of simulated coronary anastomoses.

The ability of training to strengthen the reliability of observational performance ratings has been met with mixed results.[9] Cook and colleagues[14] examined the impact of a half-day rater-training workshop on improvement in reliability of mini-Clinical Evaluation Exercise scores among internal medicine preceptors. Training techniques included rater error training, performance dimension training, behavioral observation training, and frame of reference training using lecture, video, and facilitated discussion. Despite the range of these techniques, training did not significantly improve inter-rater reliability or accuracy of mini-Clinical Evaluation Exercise scores, with ICCs ranging from 0.40 to 0.43 before training and 0.43 to 0.53 after training. Likewise, Newble and colleagues[15] examined the role of training among 18 raters using an objective checklist to evaluate medical student performance of physical examinations on simulated patients. Although moderate inter-rater reliability was achieved initially, training yielded no

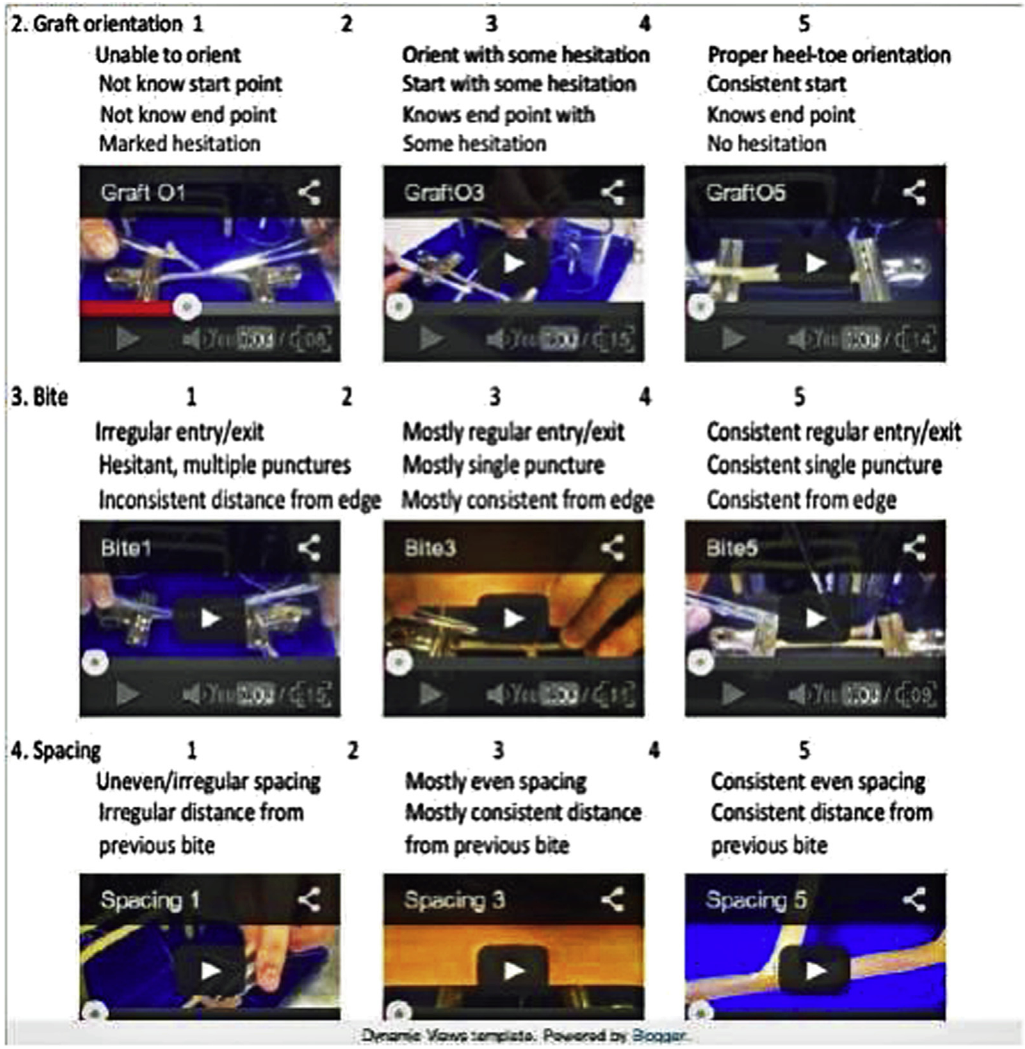

FIGURE 2. The assessment training tool consists of a series of 10- to 15-second video clips depicting the levels of skill corresponding to 1, 3, and 5 ratings for each of the assessment components. The clips were deemed to be representative samples of these anchors by the group of experienced raters involved in the development of the assessment tool.

significant improvement in rater agreement.[16] Yule and colleagues[17] found that even after training, novice ratings in behavioral assessment matched those by expert raters in just 50% of cases, particularly for ratings in the middle range. In contrast to the nonsurgical domain, technical skills assessment in surgery may be more amenable to improvement with training and calibration. In this study, components of the assessment tool demonstrated improvement in inter-rater reliability after rater training using a novel video-based anchoring assessment, with 4 components (bite, needle holder use, needle angles, and hand mechanics) improving the most. Inter-rater reliability for composite scores and overall P/F ratings also improved.

Compared with the raters who were predominantly cardiac surgeons in a previous study,[7] the majority of raters in this study were general thoracic surgeons. The type of clinical practice may influence the differences in inter-rater reliability between these cohorts. Although the majority of participants in this study have been involved in teaching coronary anastomoses in the simulated or operative setting, most can be considered to be less experienced at the time of the study, as evidenced by the number of years since completion of training, the number of years of experience with surgical simulators, and their experience with surgical rating tools in the past. Thus, even less-experienced surgeons who have variable contact with trainees performing coronary anastomosis can achieve a high level of inter-rater reliability in using the assessment tool after focused, video-based anchor training.

To date, few studies have measured internal consistency. Those that have tend to report internal consistency as inter-station reliability between multiple examination stations

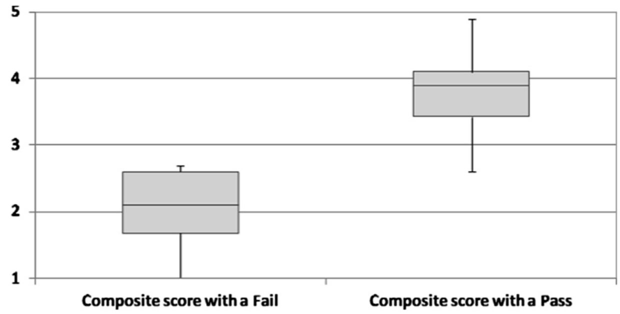

FIGURE 3. The relationship between composite scores and overall P/F ratings.

and not between individual assessment components. High internal consistency among assessment components across multiple simulation models was achieved for experienced raters in the previous study.[7] Likewise, high internal consistencies (Cronbach's α >0.98) were achieved in the present study both before and after training. These findings, independent of raters' level of experience, are likely the result of the assessment components representing a broad base of fundamental technical skills; it is reasonable to conclude that these skills are not acquired independently, that is, when a trainee is good at spacing bites evenly, he/she is also good at managing suture tension.

The training strategies used in this study consisted of performance dimension training with elements of frame of reference training. Performance dimension training instructs raters to recognize and use the appropriate dimensions targeted for evaluation by associating similar behavioral cues with the dimension.[8] We used 2 components of performance dimension training: 1. The assessment tool itself contains behavioral descriptors of the level of competence for scores of 1, 3, and 5 for each component. 2. Each of these written descriptions was supplemented with a short, video clip embedded in the assessment tool that represented an expert consensus rating of 1, 3, and 5. Also, 2 components of frame of reference training were used: 1. After raters independently completed initial ratings on paper, they entered their ratings using audience response clickers. These ratings were immediately analyzed, providing raters with visual feedback depicting how their ratings compared with those provided by the other raters. 2. The facilitators at the session were part of the original group of educators who participated in the assessment tool development. During training, the facilitators identified and mediated sources of confusion among raters. The combination of these strategies may explain why even brief training led to significant improvement in inter-rater reliability among less-experienced educators.

Study Limitations

One limitation is that simulators may not reproduce the tissue response seen clinically. Thus, it can be argued that assessment of technical performance using this tool may be relevant only in the laboratory setting. But as with any assessment tool, reliability does not reside within the tool itself. Instead, the data generated in any setting would need to be tested and confirmed for reliability. Another limitation is that the improvement in inter-rater reliability may have been the result of the familiarity gained from using the assessment tool and observing and discussing the variability of scores among the participants, rather than the improvement being attributed to the video-based anchoring approach per se. It is likely that comprehensive rater training (including familiarity with the tool and in-depth discussion) combine to improve inter-rater reliability using this assessment tool; future studies with a control group without the use of the video-based assessment could determine the degree to which this anchoring approach contributes to improved inter-rater reliability. One important issue not addressed in this study is whether the rater training

TABLE 1. Rater demographics

Gender	Male	Female			
	66.7% (6)	33.3% (3)			
Years since completion of CT surgical training	1	2-5	6-10	11+	Average
	22.2% (2)	44.4% (4)	22.2% (2)	11.1% (1)	4.5
Type of practice	Adult thoracic only	Adult cardiac and thoracic			
	77.8% (7)	22.2% (2)			
Type of trainees	Adult thoracic only	Adult cardiac and thoracic	Other		
	22.2% (2)	66.7% (6)	11.1% (1)		
Years of experience with surgical simulators	<5	5-<10	10+		
	77.8% (7)	11.1% (1)	11.1% (1)		
Estimate the no. of coronary anastomoses you have completed since completion of CT surgical training	≤100	101-≤500	501-≤1000	1001+	
	44.4 (4)	22.2% (2)	22.2% (2)	11.1% (1)	
Have you been directly involved with teaching coronary anastomoses in the simulated or operative setting	Yes	No			
	88.9% (8)	11.1% (1)			
Have you had any experience with using surgical rating tools in the past?	Yes	No			
	22.2% (2)	77.8% (7)			

CT, Cardiothoracic.

TABLE 2. Inter-rater reliability and internal consistency of technical skills assessment

	Before training	After training
Inter-rater reliability		
Composite score, ICC	0.761	0.903
Assessment components, ICC		
1. Arteriotomy	—	—
2. Graft orientation	—	—
3. Bite	0.089	0.800
4. Spacing	0.711	0.797
5. Needle holder use	0.389	0.837
6. Use of forceps	0.668	0.870
7. Needle angles	0.408	0.843
8. Needle transfer	0.526	0.842
9. Suture management	0.611	0.839
10. Knot tying	0.539	0.760
11. Hand mechanics	0.480	0.902
12. Use of both hands	0.625	0.877
13. Economy of time and motion	—	—
Mean ICC of assessment components	0.505	0.837
Overall P/F, kappa	0.196	0.775
Internal consistency		
Cronbach's α	0.989	0.999

ICC, Intraclass correlation coefficient; *P/F*, pass/fail.

leading to improved inter-rater reliability is durable; such evaluation will need to be addressed in future studies. Also, 3 components of the assessment form were eliminated from analysis (arteriotomy, graft orientation, and economy of time and motion), because they were dependent on the assisting surgeon and not well represented in the video recordings. Improving the simulator setup and the method of video acquisition should provide sufficient information for evaluation of these 3 components.

CONCLUSIONS

Standardization of assessment in CT surgery should involve evaluation of inter-rater reliability. Focused, video-based anchor training facilitates greater inter-rater reliability in the objective assessment of coronary artery anastomosis, particularly for components such as needle angles, needle transfer, and suture management, in the simulated environment. Among raters with less teaching experience, such training may be needed before formative and summative evaluation of technical skills.

The authors thank Ara A. Vaporciyan, MD, and Stephen C. Yang, MD, for assistance in the organization of the JCTSE Educate the Educators Program; and Rebecca Mark and Beth Winer for support and contributions in organizing the TSDA Boot Camp. This study would not have been possible without the active participation of the CT surgery faculty, surgery residents, and medical students.

References

1. McGaghie WC, Siddall VJ, Mazmanian PE, Myers J. Lessons for continuing medical education from simulation research in undergraduate and graduate medical education. *Chest*. 2009;135:62S-8S.
2. Levine AI, Schwartz AD, Bryson EO, DeMaria S. Role of simulation in US physician licensure and certification. *Mt Sinai J Med*. 2012;79:140-53.
3. Fann JI, Baker CJ, Calhoon JH, Carpenter AJ, Colson YL, Feins RH, Grossi EA, et al. Cardiothoracic surgery technical skills modular curriculum. Available at: http://www.jctse.org/wp-content/uploads/2013/02/Joint_Council_Technical_Skills_Curriculum-2-27-12_websiteready.pdf. Accessed February 10, 2014.
4. Fann JI, Calhoon JH, Carpenter AJ, Merrill WH, Brown JW, Poston RS, et al. Simulation in coronary artery anastomosis early in cardiothoracic surgical residency training: the Boot Camp experience. *J Thorac Cardiovasc Surg*. 2010;139:1275-81.
5. Hicks GL Jr, Gangemi J, Angona RE Jr, Ramphal PS, Feins RH, Fann JI. Cardiopulmonary bypass simulation at the Boot Camp. *J Thorac Cardiovasc Surg*. 2011;141:284-92.
6. Fann JI, Feins RH, Hicks GL Jr, Nesbitt JC, Hammon JW, Crawford FA Jr, Senior Tour in Cardiothoracic Surgery. Evaluation of simulation training in cardiothoracic surgery: the Senior Tour perspective. *J Thorac Cardiovasc Surg*. 2012;143:264-72.
7. Lee R, Enter D, Lou X, Feins RH, Hicks GL, Gasparri M, et al. The Joint Council on Thoracic Surgery Education coronary artery assessment tool has high interrater reliability. *Ann Thorac Surg*. 2013;95:2064-9.
8. Feldman M, Lazzara EH, Vanderbilt AA, DiazGranados D. Rater training to support high-stakes simulation-based assessments. *J Contin Educ Health Prof*. 2012;32:279-86.
9. Woehr DJ, Huffcutt AI. Rater training for performance appraisal: a quantitative review. *J Occup Organ Psychol*. 1994;67:189-205.
10. Holmboe ES, Hawkins RE, Huot SJ. Effects of training in direct observation of medical residents' clinical competence: a randomized trial. *Ann Intern Med*. 2004;140:874-81.
11. Azuma H, Hori S, Nakanishi M, Fujimoto S, Ichikawa N, Furukawa TA. An intervention to improve the interrater reliability of clinical EEG interpretations. *Psychiatry Clin Neurosci*. 2003;57:485-9.
12. Reznick R, Regehr G, MacRae H, Martin J, McCullock W. Testing technical skill via an innovative ''bench station'' examination. *Am J Surg*. 1997;173:226-30.
13. Hutchinson L, Aitken P, Hayes T. Are medical postgraduate certification processes valid? A systematic review of the published evidence. *Med Educ*. 2002;36:73-91.
14. Cook DA, Dupras DM, Beckman TJ, Thomas KG, Pankratz VS. Effect of rater training on reliability and accuracy of mini-CEX scores: a randomized, controlled trial. *J Gen Intern Med*. 2008;24:74-9.
15. Newble DI, Hoare J, Sheldrake PF. The selection and training of examiners for clinical examinations. *Med Educ*. 1980;14:345-9.
16. Noel GL, Herbers JE Jr, Caplow MP, Cooper GS, Pangaro LN, Harvey J. How well do internal medicine faculty members evaluate the clinical skills of residents. *Ann Intern Med*. 1992;117:757-65.
17. Yule S, Rowley D, Flin R, Maran N, Youngson G, Duncan J, et al. Experience matters: comparing novice and expert ratings of non-technical skills using the NOTSS system. *ANZ J Surg*. 2009;79:154-60.

EDU

APPENDIX TABLE E1. The Joint Council on Thoracic Surgery Education assessment tool

	Poor				Excellent
1. Arteriotomy	1	2	3	4	5
	Not identify artery		Partial artery exposure		Full artery exposure
	Off-midline		Mainly midline		Consistent midline
	Multiple "tracks"		Thick single "track"		Thin single "track"
	Injury to back wall		Close to back wall		No injury to back wall
	Marked irregular edge		Mild irregular edge		Smooth edge
Additional comments:					
2. Graft orientation	1	2	3	4	5
	Unable to orient		Orient with some hesitation		Proper heel-toe orientation
	Not know start point		Start with some hesitation		Consistent start
	Not know end point		Knows end point with		Knows end point
	Marked hesitation		Some hesitation		No hesitation
Additional comments:					
3. Bite	1	2	3	4	5
	Irregular entry/exit		Mostly regular entry/exit		Consistent regular entry/exit
	Hesitant, multiple punctures		Mostly single puncture		Consistent single puncture
	Inconsistent distance from edge		Mostly consistent from edge		Consistent from edge
Additional comments:					
4. Spacing	1	2	3	4	5
	Uneven/irregular spacing		Mostly even spacing		Consistent even spacing
	Irregular distance from previous bite		Mostly consistent distance from previous bite		Consistent distance from previous bite
Additional comments:					
5. Needle holder use	1	2	3	4	5
	Awkward finger placement		Functional finger placement		Comfortable, smooth finger placement
	Unable to rotate instrument		Hesitant when rotating		Smooth rotation
	Awkward and not facile		Moderate facility		High facility
	Inconsistent needle placement		Generally good placement		Consistent proper placement
Additional comments:					
6. Use of forceps	1	2	3	4	5
	Awkward or no traction		Moderate proper traction		Consistent proper traction
	Unable to expose		Able to assist in exposure		Consistent proper exposure
	Not use to stabilize needle		Able to stabilize but rough		Knows when to stabilize, gentle
Additional comments:					
7. Needle angles	1	2	3	4	5
	Not aware of angles		Understand angles, not consistent		Consistent correct angles
	Not compensate for depth		Partial compensation for depth		Compensate for depth
	Does not consider subsequent angles		Partial consideration of subsequent angles		Consistent adjustment for subsequent angles
Additional comments:					
8. Needle transfer	1	2	3	4	5
	Marked hesitation in mounting needle		Able to mount needle with hand and partial manipulation		Able to mount needle and manipulate needle easily
Additional comments:					
9. Suture management	1	2	3	4	5
	Not use tension		Tension use inconsistent		Proper use of tension
	Suture entangled		Sutures occasionally get in way		Suture consistently not in way

(Continued)

APPENDIX TABLE E1. Continued

	Poor				Excellent
Additional comments:					
10. Knot tying	1	2	3	4	5
	Marked hesitancy, slow speed		Moderate facility, moderate speed		Consistent facility, no hesitancy
	No follow through		Intermittent follow through		Consistent follow through
	Not able to tie, breakage		Able to tie and tension,		Consistent tension and tight
	Loose or "air" knot		intermittently loose		
Additional comments:					
11. Hand Mechanics	1	2	3	4	5
	No pronation or supination		Incomplete pronation or supination		Able to modulate pronation/ supination
	Awkward finger/hand motion		Hesitant finger/hand motion		Smooth, comfortable motion
	No wrist motion		Incomplete wrist motion		Smooth, appropriate wrist motion
Additional comments:					
12. Use of both hands	1	2	3	4	5
	Awkward/not coordinated use		Moderately coordinated use		Smooth, seamless coordination
	Nondominant hand neglect		Moderate use of nondominant hand to assist/expose		Full use of nondominant hand to assist/expose
Additional comments:					
13. Economy of time and motion	1	2	3	4	5
	Marked hesitation		Some hesitation		No hesitation
	Not aware of goal		Some awareness of goal		Fully aware of goal
	Unable to do task		Able to do task but discontinuous		Able to do task smoothly
Additional comments:					
Overall	Pass	Fail			
General definitions:					
5. Excellent, able to accomplish goal without hesitation, showing excellent progress and flow					
4. Good, able to accomplish goal deliberately, with minimal hesitation, showing good progress and flow					
3. Average, able to accomplish goal with hesitation, discontinuous progress and flow					
2. Below average, able to partially accomplish goal with hesitation					
1. Poor, unable to accomplish goal; marked hesitation					