The Bankart Performance Metrics Combined With a Shoulder Model Simulator Create a Precise and Accurate Training Tool for Measuring Surgeon Skill

Richard L. Angelo, M.D., Robert A. Pedowitz, M.D., Ph.D., Richard K. N. Ryu, M.D., and Anthony G. Gallagher, Ph.D., D.Sc.

Purpose: To determine if a dry shoulder model simulator coupled with previously validated performance metrics for an arthroscopic Bankart repair (ABR) would be a valid tool with the ability to discriminate between the performance of experienced and novice surgeons, and to establish a proficiency benchmark for an ABR using a model simulator. **Methods:** We compared an experienced group of arthroscopic shoulder surgeons (Arthroscopy Association of North America faculty) (n = 12) with a novice group (n = 7) (postgraduate year 4 or 5 orthopaedic residents). All surgeons were instructed to perform a diagnostic arthroscopy and a 3 suture anchor Bankart repair on a dry shoulder model. Each procedure was videotaped in its entirety and scored in blinded fashion independently by 2 trained reviewers. Scoring used previously validated metrics for an ABR and included steps, errors, and "sentinel" (more serious) errors. **Results:** The inter-rater reliability among pairs of raters averaged 0.93. The experienced group made 63% fewer errors, committed 79% fewer sentinel errors, and performed the procedure in 42% less time than the novice group (all significant differences). The greatest difference in errors between the groups involved anchor preparation and insertion, suture delivery and management, and knot tying. **Conclusions:** The tool comprised by validated ABR metrics coupled with a dry shoulder model simulator is able to accurately distinguish between the performance of experienced and novice orthopaedic surgeons. A performance benchmark based on the mean performance of the experienced group includes completion of a 3 anchor Bankart repair, enacting no more than 4 total errors and 1 sentinel error. **Clinical Relevance:** The combination of performance metrics and an arthroscopic shoulder model simulator can be used to improve the effectiveness of surgical skills training for an ABR. The methodology used may serve as a template for outcomes-based procedural skills training in general.

Some authors from professional bodies and health care training organizations around the world argue that the surgical trainee should acquire basic procedural skills outside of the surgical theater before operating on real patients.[1-3] Furthermore, evidence now clearly indicates that when performed to a quantitatively defined level, skills practiced and acquired outside the operating room are superior to skills acquired in a traditional apprenticeship manner primarily in the operating room.[4,5] Satava[6] first introduced the concept of simulation-based training in the early 1990s, with quantitative evidence from prospective, randomized, double-blinded clinical studies showing that simulation-based training is a powerful tool for the acquisition of surgical skills.[7-10] The simulator can be either a physical model or computer-generated video images[7,10] because both are equally effective if used as part of a "metric-based training curriculum."[11] Table 1 shows a glossary of terms used throughout this article.

An implicit assumption in a simulator-based training process is the use of validated metrics that appropriately characterize the procedure to be trained. Previously, Angelo et al.[12] reported on the development of a tool defining "performance metrics" ("steps" and "errors" as defined in Table 1) for a standard "reference approach"

From ProOrtho Clinic (R.L.A.), Kirkland, Washington; Professor Emeritus, University of California (R.A.P.), Los Angeles, California; The Ryu Hurvitz Orthopedic Clinic (R.K.N.R.), Santa Barbara, California, U.S.A.; and ASSERT, University College Cork (A.G.G.), Cork, Ireland.

The authors report the following potential conflict of interest or source of funding: R.L.A. receives support from DePuy Mitek. R.K.N.R. receives support from MedBridge and Mitek.

Received January 24, 2014; accepted April 15, 2015.

Address correspondence to Richard L. Angelo, M.D., ProOrtho Clinic, 12911 120th Ave NE, Ste H-210, Kirkland, WA 98072, U.S.A. E-mail: rlamdortho@comcast.net

0749-8063/1454/$36.00

http://dx.doi.org/10.1016/j.arthro.2015.04.092

Table 1. Glossary

	Definition
Construct validity	A type of evidence that supports that specific test items identify the quality, ability, or trait they were designed to measure
Content validity	An estimate (opinion) by experts of the validity of a testing instrument based on a detailed examination of the contents of the test items
Definition	A definite, distinct, and clear objective characterization providing an accurate and reliable identification of whether an event was or was not observed to have occurred
Damage to non-target tissue	Iatrogenic damage to tissues not intended to be addressed in the specific step (i.e., articular cartilage damage)
Delphi Panel (modified)	A structured communication technique originally developed as a systematic, interactive forecasting method that relies on the opinion of a panel of experts; in the modified form, experts answer queries/vote in 2 or more rounds (cycles) on the appropriateness of the metric-based operational definitions of detailed aspects of procedure performance with the goal of achieving consensus—voting is not anonymous
Error	A deviation from optimal performance
Face validity	An estimate (opinion) by experts who review the content of an assessment or tool to see if it seems appropriate and relevant to the concept it purports to measure
Inter-rater reliability	The extent of agreement between 2 raters on the occurrence of a series of observed events; it ranges between 0, no agreement, and 1.0, complete agreement
Metric	A standard of measurement of quantitative assessments used for objective evaluations to make comparisons or to track performance
Metric-based training curriculum	A procedural skills training program based on clear, specific, detailed definitions of the steps to be accomplished and errors avoided
Performance metric	The features determining the accomplishment of a given task measured against preset known standards of accuracy and completeness
Procedure phase	A group or series of integrally related events or actions that, when combined with other phases, make up or constitute a complete operative procedure
Proficiency/proficient	A specific level of performance defined by a quantitative score (benchmark) or scores on a standardized test or other form of assessment
Reference procedure	A straightforward operative procedure; an agreed on/accepted approach to the performance of an uncomplicated procedure under ideal circumstances
Sentinel error	An event or occurrence involving a serious deviation from optimal performance during a procedure that either (1) jeopardizes the success/desired result of the procedure or (2) creates iatrogenic insult to the patient's tissues
Step	A component task, the series aggregate of which constitutes the completion of a specific procedure
Task deconstruction	To break down a procedure into constituent tasks, steps, or components

to performing an arthroscopic Bankart repair.[13-17] This tool was derived from a careful "task deconstruction" (Table 1) using videos of complete Bankart procedures performed with patients in either the lateral decubitus or beach-chair orientation. The metrics were constructed so that they could be scored in an identical manner with the patient in either orientation. "Face validity" and "content validity" of the metrics were verified using a "modified Delphi Panel" methodology (Table 1).

The purpose of this study was to determine if a dry shoulder model simulator coupled with previously validated performance metrics for an arthroscopic Bankart repair would be a valid tool with the ability to discriminate between the performance of experienced and novice surgeons. We also sought to establish a "proficiency" (Table 1) benchmark for the arthroscopic Bankart procedure using the model simulator. The null hypothesis was that when using a shoulder model simulator, the Bankart metrics would fail to discriminate between experienced and novice surgeon performance.

Methods

No institutional review board (IRB) approval was obtained for this study investigating the validity of the Bankart metrics coupled with the model simulator. IRB approval was sought for the final Copernicus Study proper, which will compare 3 different training protocols. The Western IRB (No. 1-776362-1) opined that, as an educational curriculum study, this investigation was exempt from the need for full IRB approval [based on the criteria of 45 CFR 46.101(b)(1)]. The final study comparing the 3 training protocols was registered with the National Institutes of Health (ClinicalTrials.gov No. NCT01921621).

Study Groups

Two groups were compared in their performance of an arthroscopic Bankart procedure on a shoulder model simulator. The experienced group consisted of all faculty members who served as either a master or associate master instructor for a standard 3-day Arthroscopy Association of North America (AANA) Resident Course conducted at the Orthopedic

Learning Center (Rosemont, IL). The novice group was limited to postgraduate year (PGY) 4 and PGY 5 orthopaedic residents who had registered for a Resident Course and who volunteered to participate in the investigation.

Arthroscopic Bankart Repair Metrics

Metrics have been previously defined for a standard reference arthroscopic Bankart repair.[12] Forty-five essential steps in 13 "phases" (Table 1) (Roman numerals) were defined with beginning points and endpoints (Table 2). Twenty-nine potential unique errors were specified (Table 3), 8 of which were designated as "sentinel" (Table 1). The more serious (sentinel) errors were defined as those expected to either (1) substantially compromise the outcome of the shoulder stabilization (e.g., "capsular penetration of the suture passing instrument is superior to the anchor hole," resulting in failure to achieve retentioning of the capsule and inferior glenohumeral ligaments) or (2) potentially lead to iatrogenic damage to the shoulder (e.g., "laceration of the intact labrum"). Some of the same errors could be enacted more than once during different phases of the procedure. Thus a total of 77 potential errors, 20 of which were sentinel errors, were specified for the complete procedure. In addition, events that led to less consequential "damage to non-target tissue" (DNTT) (Table 1) were recorded as a standard error (e.g., scuffing of the articular cartilage). A perfect score would indicate that all 45 steps were completed satisfactorily without committing any errors.

Dry Shoulder Model Simulator

The shoulder simulator used is a physical model composed of a dense foam plastic endoskeleton including a humerus, scapula, glenoid, coracoid, acromial spine, and acromion with proportions appropriate to the human skeleton (Sawbones; Pacific Research Laboratories, Vashon, WA) (Fig 1). The articulating surfaces of the humerus and glenoid are laminated with a softer, white layer designed to mimic articular cartilage. A Hill-Sachs lesion measuring 1 cm by 3.5 cm is oriented vertically on the posterior aspect of the humeral head and is represented by a red impaction trough. A rim of off-white, rubber-like material encircles and lightly adheres to the glenoid neck, simulating the labrum. Red staining in the region where the labrum is joined to the anteroinferior glenoid represents the Bankart lesion. The adhesive attachment of the labrum requires the operator to intentionally "liberate" the labrum from the glenoid to demonstrate mobilization of the capsulolabral tissues. A more medial and superficial pink layer of soft foam represents the subscapularis muscle. A tubular strand of rubber simulates the long head of the biceps tendon and courses from its anatomic attachment to the superior labrum, out of the shoulder joint, into the bicipital groove of the humerus. The capsule is replicated by a pliable, rubberized material containing the glenohumeral joint and has a molded imprint of the inferior glenohumeral ligaments on the articular surface. A separate band represents the superior border of the subscapularis tendon. Holes measuring 8 mm in diameter are created in the capsule during molding and enable cannulas for the posterior, mid-anterior, and anterosuperior portals to pass through the relatively tough capsular material. Beige-colored, soft, moderately dense foam represents the skin and soft tissues exterior to the glenohumeral joint and possesses a contour and bulk that mimic the shape of the human shoulder. The acromion, acromial spine, and coracoid landmarks are readily palpable through the "soft tissues" and assist in locating proper portal placement.

Arthroscopic Bankart Repair

During a single weekend AANA resident arthroscopy course, the surgeons from both groups were instructed to establish portals (posterior, anterosuperior, and mid-anterior), complete a thorough diagnostic arthroscopy, and perform a 3-anchor arthroscopic Bankart repair on the simulator model. Furthermore, they were instructed to demonstrate/complete all of the steps for the Bankart repair that they would normally perform in clinical practice on a real patient. The model was secured in either the lateral decubitus or beach-chair orientation according to surgeon preference. Equipment representatives from multiple different vendors served as surgical assistants and were randomly assigned to participant surgeons. The assistants were instructed to act only at the specific direction of the operating surgeon. Prompting and coaching (of technique) were prohibited (the procedures were proctored by staff from the Orthopedic Learning Center). A standard equipment tower with a 30° arthroscope, along with all instruments necessary to complete an arthroscopic Bankart repair, was provided (Table 4).

The surgeon created the required portals based on the palpable "bony" landmarks of the shoulder and then progressed to complete the diagnostic arthroscopy and Bankart repair. A continuous video recording was made beginning with the first arthroscopic view of the joint from the posterior portal and ending with the withdrawal of the arthroscope after the surgeon's examination of the completed repair with a hook probe. No time limit was imposed on the performance of the procedure on the simulator model.

Video Reviewer Training

Once the construction of the metrics for an arthroscopic Bankart repair was completed and face and

Table 2. Thirteen Phases of Bankart Procedure (in Roman Numerals) and Brief Summary of 45 Steps of Procedure

I. Portals
- 1. Posterior portal established
- 2. View posterior humeral head and extent of the Hill-Sachs when present
- 3. Introduce mid-anterior spinal needle immediately superior to the subscapularis and direct it toward the anteroinferior glenoid and labrum
- 4. Establish a cannula that abuts the superior border of the subscapularis near the lateral subscapularis insertion
- 5. Demonstrate instrument access to the anteroinferior glenoid/labrum
- 6. Introduce anterosuperior spinal needle at the superolateral aspect of the rotator interval and direct it toward the anterior glenoid
- 7. Establish an anterosuperior cannula, arthroscopic sheath, or switching stick

II. Arthroscopic instability assessment
- View from posterior portal
  - 8. View or probe the superior labral attachment onto the glenoid
  - 9. View or probe articular surface of the rotator cuff
  - 10. Probe anteroinferior glenoid/Bankart pathology including rim fracture, articular defect
- View from anterosuperior portal
  - 11. View or probe the mid-substance of the anterior-inferior glenohumeral ligaments
  - 12. View or probe the insertion of the anterior glenohumeral ligaments onto the anterior humeral neck

III. Capsulolabral mobilization/glenoid preparation
- 13. Elevate the capsulolabral tissue from the glenoid neck and articular margin
- 14. View the subscapularis muscle superficial to the mobilized capsule
- 15. With an instrument, grasp and perform an inferior to superior shift of the capsulolabral tissue (to show tension is restored)
- 16. Obtain a view of the anterior glenoid neck
- 17. Mechanically abrade the glenoid neck

IV. Inferior anchor preparation/insertion
- 18. Seat the guide for the most inferior anchor hole at the inferior region of the anteroinferior quadrant
- 19. Drill anchor hole oblique to the glenoid articular face
- 20. Insert anchor
- 21. Test the anchor security by pulling on the suture tails

V. Suture delivery/management
- 22. Pass a cannulated suture hook or suture retriever through the capsular tissue—inferior to the anchor
- 23. Pass anchor suture limb through the capsular tissue and deliver out the anterior cannula

VI. Knot tying
- 24. Deliver an arthroscopic sliding knot
- 25. Back up with 3 or 4 half-hitches
- 26. Cut suture tails

VII. Second anchor preparation/insertion
- 27. Seat the drill guide for the second anchor superior to the first anchor and inferior to the equator of the glenoid
- 28. Drill anchor hole oblique to the glenoid articular face
- 29. Insert suture anchor
- 30. Test anchor security by pulling on the suture tails

VIII. Suture delivery/management
- 31. Pass a cannulated suture hook or suture retriever through the capsular tissue inferior to the suture anchor
- 32. Pass anchor suture limb through the capsular tissue and deliver out the anterior cannula

IX. Knot tying
- 33. Deliver an arthroscopic sliding knot
- 34. Back up with 3 or 4 half-hitches
- 35. Cut suture tails

X. Third anchor preparation/insertion
- 36. Seat the drill guide for the third anchor at or superior to the equator
- 37. Drill anchor hole oblique to the glenoid articular face
- 38. Insert suture anchor
- 39. Test anchor security by pulling on suture tails

XI. Suture delivery/management
- 40. Pass a cannulated suture hook or suture retriever through the capsular at or inferior to the suture anchor
- 41. Pass anchor suture limb through the capsular tissue and deliver out the anterior cannula

XII. Knot tying
- 42. Deliver an arthroscopic sliding knot
- 43. Back up with 3 or 4 half-hitches
- 44. Cut suture tails

XIII. Procedure review
- 45. View and/or probe final completed repair

Table 3. Summary of 29 Different Bankart Procedure Metric Errors

Failure to maintain intra-articular position of the posterior cannula
Failure to maintain intra-articular position of the mid-anterior cannula
Failure to maintain intra-articular position of the anterosuperior cannula
Damage to the superior border of the subscapularis during creation of the mid-anterior portal
Damage to the anterior border of the supraspinatus during creation of the anterosuperior portal
Loss of intra-articular position of arthroscope/sheath or operating cannula (loss of each portal is scored only once for each Roman numeral, i.e., up to a total of 3 for arthroscope + 2 portals)
Laceration of intact capsulolabral tissue (sentinel error)
Failure to maintain control of a working instrument (sentinel error)
Guide is not located in the inferior region of the anteroinferior quadrant of the glenoid for the most inferior anchor
Entry of the completed tunnel lies outside safe zone of 0 to 3 mm from the bony glenoid rim (sentinel error)
Shallow undermining and deformation of articular cartilage (sentinel error)
Failure to maintain secure seating of the drill guide during anchor insertion
Breakage of the implant
Implant remains visibly proud (sentinel error)
Failure to insert the anchor with the inserter laser line (when present) to or beyond the laser line on the drill guide
Anchor fails to remain securely fixed within bone at the appropriate depth
Capsular penetration is at or superior to anchor hole (sentinel error)
Capsular penetration is not at or peripheral to the capsulolabral junction
Instrument breakage
Tearing of capsulolabral tissue
Uncorrected entanglement of shuttling device or suture
Off-loading of suture anchor
Breakage of suturing device
Failure to create and maintain indentation of the capsule or labral tissue on knot completion (sentinel error)
Visible void is present between throws of the completed primary knot (sentinel error)
Completed knot abuts articular cartilage
Visible void is present between throws of the completed half-hitches
Suture breakage
Guide is inferior to the equator of the glenoid (for the third and final anchor)

NOTE. Metric errors can be associated with multiple phases and steps of the procedure (77 total errors).

content validity were verified,[12] a final version of a score sheet was formatted. Ten AANA master/associate master faculty surgeons (none belonging to the experienced group from this study) formed the panel of reviewers designated to score the videos. This group included the 3 members (R.L.A., R.K.N.R., R.A.P.) who developed the arthroscopic Bankart metric definitions (Table 1) in conjunction with a consultant experimental psychologist (A.G.G.). The 10 reviewers were assigned by the AANA research coordinator to 1 of 5 fixed pairs, which remained constant throughout the scoring of all videos. Assignments were made based on similar time zones of the reviewers' residence/practice. Reviewer training was initiated with an 8-hour in-person meeting during which time each metric was studied in detail. Multiple video examples of live patient cases were shown to illustrate each particular metric. Videos of the patients in both the lateral decubitus and beach-chair orientations were represented. Discussion helped to clarify how each step and error were to be scored, including the nuances and conventions to be used. Several weeks later, full-length practice videos 1 and 2 (one each in the lateral decubitus and beach-chair orientation) were sent to and independently scored by each of the 10 reviewers, and the scores were then tabulated. During 2 subsequent 2-hour group phone conferences, the differences and discrepancies among all reviewers were compared and discussed, seeking conformity in scoring. Each of the designated pairs of reviewers also participated in 1 to 3 additional phone conferences to analyze the specific instances in which the 2 of them scored particular events differently. Subsequently, all reviewers scored practice videos 3 and 4, and the results were tabulated (each patient orientation was again represented). The scores for each of the 5 designated pairs of reviewers were compared for the second set of practice videos. In only 1 of 10 comparisons (2 videos × 5 reviewer pairs) did the inter-rater reliability (IRR) (Table 1) calculation (as described later) fall below an acceptable level of 0.8,[18] with an IRR value of 0.76.

Video Scoring

The AANA research coordinator randomly assigned each of the 19 full-length study videos of experienced and novice surgeons performing an arthroscopic Bankart procedure on the shoulder simulator model to a single pair of reviewers. Other than the research coordinator and the study consultant, all reviewers remained blinded to the source of all videos. Each of the 19 videos were independently reviewed and scored by the 2 members of an assigned pair of reviewers. All scores were tabulated for each of 13 phases of the procedure

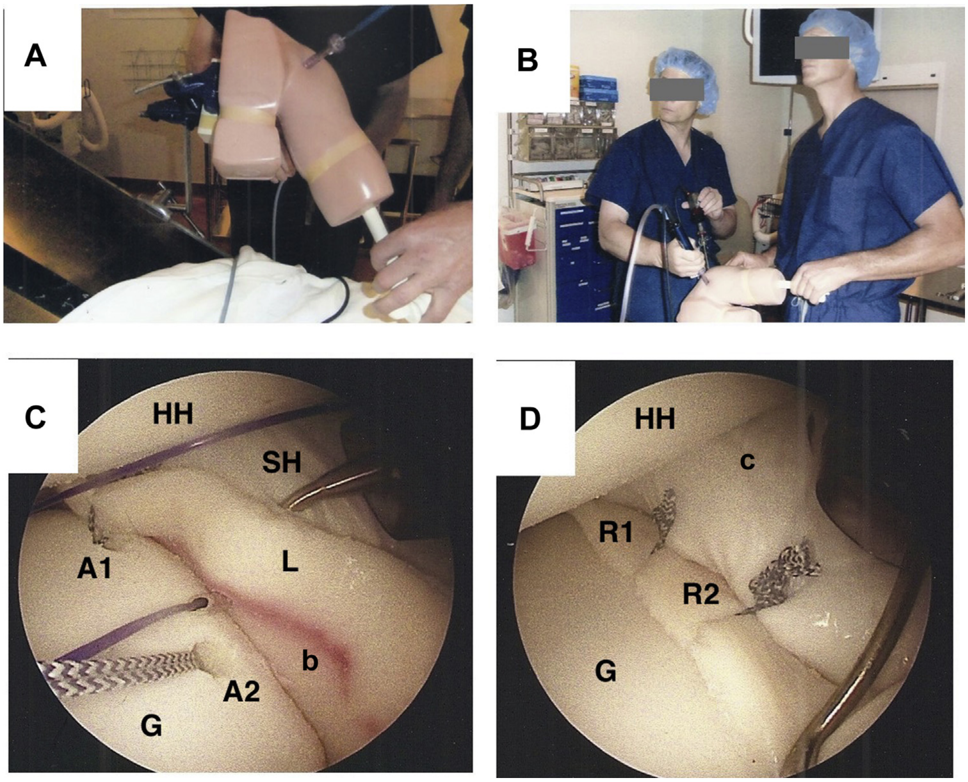

Fig 1. Dry shoulder simulator model used in study (left shoulder). (A) Anterior view of shoulder simulator oriented in beach-chair position. (B) An operator and assistant performing arthroscopic surgery on the simulator model oriented in the lateral decubitus position. (C, D) Arthroscopic view from anterosuperior portal. (C) Placement of the inferior-most anchor and sutures has been completed. The second anchor has been inserted, and the sutures have been retrieved out of the posterior cannula. A cannulated suture hook (SH) enters through the midanterior portal and is passed through the capsule and labrum (L) inferior to the exit of the suture anchor hole; a monofilament shuttle suture is then delivered. (A1, anchor position 1; A2, anchor position 2; b, Bankart lesion; G, glenoid; HH, humeral head.) (D) A hook probe is used to examine the completed repair (with the third anchor just out of view beneath the hook probe); the capsule (c) has been retensioned and the labrum secured to the glenoid rim. (G, glenoid; HH, humeral head; R1, repair position 1; R2, repair position 2.)

(Tables 5 and 6). Each step and error metric was scored as either yes or no, designating whether the specific event was or was not observed to have occurred by the reviewer. In addition to scoring steps and errors, each event characterized as DNTT was scored. There was no limit to the number of individual instances DNTT could be scored, with each occurrence simply tallied as a single error event. The score sheet also contained a box for specific reviewer comments for each metric. The 2 individual scores from a pair of reviewers were averaged to obtain the overall score for each step, error, or DNTT event. In addition, the score agreement or disagreement between the specific pair of reviewers was tabulated for each individual event (step, errors, and DNTT events) and used to calculate IRR correlations (as described in the "Statistical Methods" section). The total time in minutes was documented for each video, beginning with the first view of the arthroscope from the posterior portal and ending with withdrawal of the arthroscope after examination of the completed repair.

Table 4. Arthroscopic Instruments Used to Perform Bankart Procedure on Cadaveric Shoulder

5.5- and 8.5-mm obturator cannulas
Switching sticks
Hook probe
Regular and looped graspers
Liberator/elevator
Shaver
Drill guide/drill
Push-in anchor loaded with single suture
Mallet
Cannulated suture hook
Penetrator
Monofilament suture
Knot pusher
Arthroscopic scissors

Performance Benchmark

Prior research has used the metric-based mean performance of a group of experienced or expert operators to objectively define "proficiency."[7,9-11,19] Before initiating this study, the 4 primary investigators specified

Table 5. Copernicus Resident/Master Model Metric Validation: Novices

	64A	64B	64 Ave	94A	94B	94 Ave	24A	24B	24 Ave	34A	34B	34 Ave	14A	14B	14 Ave	44A	44B	44 Ave	84A	84B	84 Ave
Video																					
I. Portals																					
Steps uncompleted	0	0	0	0	0	0	0	0	0	3	2	2.5	2	3	2.5	1	2	1.5	1	2	1.5
Errors made	0	2	1	3	3	3	0	1	0.5	3	0	1.5	1	1	1	1	1	1	2	0	1
II. Instabl assess																					
Steps uncompleted	2	1	1.5	1	0	0.5	1	1	1	1	1	1	3	3	3	0	1	0.5	2	1	1.5
Errors made	0	0	0	1	1	1	1	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0
III. Caps/gen prep																					
Steps uncompleted	2	2	2	3	3	3	1	1	1	3	4	3.5	5	5	5	3	3	3	3	4	3.5
Errors made	1	0	0.5	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IV. First inf anch prep																					
Steps uncompleted	0	0	0	0	1	0.5	0	1	0.5	0	0	0	1	2	1.5	1	3	2	1	1	1
Errors made	1	0	0.5	2	1	1.5	3	3	3	1	4	2.5	1	2	1.5	0	1	0.5	0	0	0
V. First sut del/mgmt																					
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	0	0	0	0	0	0	1	1	1	3	3	3	0	0	0	0	0	0	1	0	0.5
VI. First knot tying																					
Steps uncompleted	0	0	0	1	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	0	0	0	0	1	0.5	0	0	0	0	0	0	2	2	2	1	1	1	1	0	0.5
VII. Second anch prep																					
Steps uncompleted	0	0	0	0	1	0.5	1	0	0.5	0	0	0	1	1	1	1	1	1	1	0	0.5
Errors made	0	0	0	2	1	1.5	0	1	0.5	2	2	2	1	0	0.5	0	0	0	0	0	0
VIII. Second sut del/mgmt																					
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	2	1	1.5	0	0	0	1	1	1	3	3	3	0	0	0	0	0	0	0	0	0
IX. Second knot tying																					
Steps uncompleted	1	0	0.5	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Errors made	2	1	1.5	0	1	0.5	0	0	0	1	1	1	0	0	0	0	0	0	1	0	0.5
X. Third anch prep																					
Steps uncompleted	0	0	0	0	1	0.5	1	1	1	0	0	0	1	1	1	1	1	1	0	0	0
Errors made	0	0	0	0	1	0.5	0	0	0	2	2	2	0	0	0	1	0	0.5	3	3	3
XI. Third sut del/mgmt																					
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	1	2	1.5	0	0	0	1	1	1	2	2	2	1	1	1	1	1	1	2	0	1
XII. Third knot tying																					
Steps uncompleted	0	1	0.5	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
Errors made	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	3	0	1.5
XIII. Eval repair																					
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total time (Port, Dx, and Rx), min			105			93			67			76			62			67			39
Rating pairs																					
Steps completed (45)	40	41	40.5	40	39	39.5	41	41	41	36	36	36	32	31	31.5	38	34	36	37	37	37
Errors made (77)	7	6	6.5	10	11	10.5	8	6	7	17	17	17	7	6	6.5	5	5	5	13	3	8

(continued)

Table 5. Continued

	64A	64B	64 Ave	94A	94B	94 Ave	24A	24B	24 Ave	34A	34B	34 Ave	14A	14B	14 Ave	44A	44B	44 Ave	84A	84B	84 Ave
Sentinel errors	3	2	2.5	1	4	2.5	3	3	3	3	4	3.5	1	3	2	0	1	0.5	4	1	2.5
Steps completed																					
Agreement			42			40			43			43			41			41			39
Disagreement			3			5			2			2			4			4			6
Step IRR			0.93			0.89			0.95			0.95			0.91			0.91			0.87
Errors made																					
Agreement			73			70			72			69			70			73			65
Disagreement			4			7			5			8			7			4			12
Error IRR			0.95			0.91			0.93			0.9			0.91			0.95			0.84
Total score (S + E)																					
Agreement			115			110			115			112			111			113			104
Disagreement			7			12			7			10			11			9			18
Total IRR			0.94			0.9			0.94			0.92			0.91			0.93			0.85

anch, anchor; caps, capsule; Dx, diagnostic; E, errors; eval, evaluate; glen, glenoid; inf, inferior; instab, instability; IRR, inter-rater reliability; mgmt, management; min, minutes; port, portals; Rx, treatment; S, steps; sut, suture; sut del, suture delivery.

that any participant from the experienced group who was performing more than 2 SDs better or worse than the group as a whole would be deemed an "outlier" and not representative of the experienced group. If any such performance by a participant in this group occurred, the participant's scores would be removed from the analysis so as not to skew the establishment of the reference benchmark.

Statistical Methods

For each of the 13 separate phases of the procedure, the numbers of "uncompleted steps" (steps that were not performed) and "errors made" were tabulated and the scores for the 2 reviewers averaged (Tables 5 and 6). These data were used to determine which of the procedural phases showed the greatest differences in performance when comparing the experienced and novice surgeons (1-factor analysis of variance) (IBM SPSS statistical software program; IBM, Armonk, NY). Furthermore, for the entire procedure, the total numbers of steps "completed," errors made, and sentinel errors enacted were also averaged and tabulated for the pair of reviewers.

The 2 raw score sheets were compared for each of the individual steps (45 steps in total) and the number of "agreements" tabulated (either both reviewers documented that a step was performed or both scored the step as not being completed). In addition, the number of "disagreements" in scoring steps was tabulated (one of the reviewers indicated that the step had been completed and the other scored that the step had not been completed). The IRR for the steps was calculated according to the following formula: Number of agreements/Number of agreements + Number of disagreements.

In a similar manner, there was either agreement or disagreement in the 2 scores for each of the potential errors (77 errors in total). The IRR for error scoring was calculated in the same manner as that for the steps. Finally, the IRR for scoring the metrics for the complete procedure was calculated using both the step and error agreements/disagreements for the entire procedure (122 in total). An acceptable IRR is equal to or greater than 0.80.[18]

Results

Participants

The entire group of 12 master and associate master instructors serving as faculty for an AANA Resident Course chose to participate and comprised the experienced group. The faculty, all fellowship trained in arthroscopy or sports medicine, averaged over 17 years in clinical practice, with each having routine experience in arthroscopic shoulder techniques. All faculty members have been recognized nationally by

Table 6. Copernicus Resident/Master Model Metric Validation: Experienced

	Excluded																																		
	15A	15B	15 Ave	25A	25B	25 Ave	35A	35B	35 Ave	65A	65B	65 Ave	75A	75B	75 Ave	45A	45B	45 Ave	55A	55B	55 Ave	85A	85B	85 Ave	95A	95B	95 Ave	105A	105B	105 Ave	115A	115B	115 Ave	125A	125 B	125 Ave
Video																																				
I. Portals																																				
Steps uncompleted	3	2	2.5	1	0	0.5	2	1	1.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1.5	2	1	1.5
Errors made	0	1	1.5	0	0	0	1	0	0.5	1	0	0.5	0	1	0.5	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	3	2	2.5	0	0	0
II. Instabl assess																																				
Steps uncompleted	3	2	2.5	1	1	1	1	2	1.5	1	0	0.5	2	1	1.5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.5	2	0	1	2	1	1.5
Errors made	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
III. Caps/gen prep																																				
Steps uncompleted	2	3	2.5	2	1	1.5	4	4	4	1	3	2	2	3	2.5	2	2	2	1	3	2	1	2	1.5	1	3	2	1	3	2	4	3	3.5	3	3	3
Errors made	0	0	0	0	0	0	0	1	0.5	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
IV. First inf anch prep																																				
Steps uncompleted	1	1	1	0	0	0	0	1	0.5	0	1	0.5	0	0	0	1	0	0.5	0	0	0	1	1	1	0	0	0	0	0	0	1	1	1	1	1	1
Errors made	2	2	2	0	1	0.5	2	2	2	1	0	0.5	0	0	0	0	0	0	0	0	0	1	0	0.5	0	0	0	2	2	2	0	0	0	0	0	0
V. First sut del/ mgmt																																				
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	0	0	0	2	0	1	1	0	0.5	0	0	0	0	0	0	1	0	0.5	0	0	0	0	0	0	0	1	0.5	0	1	0.5	1	0	0.5	0	0	0
VI. First knot tying																																				
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	1	0.5	0	0	0	1	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0
VII. Second anch prep																																				
Steps uncompleted	1	1	1	0	0	0	1	1	1	0	1	0.5	0	0	0	1	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0	2	1	1	1	1
Errors made	0	0	0	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	4	3	1	0	0.5
VIII. Second sut del/mgmt																																				
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.5	0	0	0	0	0	0	0	0	0
Errors made	1	0	0.5	0	0	0	1	1	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
IX. Second knot tying																																				
Steps uncompleted	1	0	0.5	0	0	0	0	1	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0.5	0	0	0	0	0	0
Errors made	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0.5	1	2	1.5	0	0	0	0	0	0	0	0	0	0	0	0	2	1	1.5	0	1	0.5
X. Third anch prep																																				
Steps uncompleted	1	1	1	2	0	1	1	1	1	1	1	1	0	0	0	1	3	2	1	1	1	1	1	1	0	0	0	0	0	0	0	1	0.5	1	1	1
Errors made	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XI. Third sut del/mgmt																																				
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0.5	0	0	0	0	0	0	0	0	0
Errors made	1	0	0.5	0	0	0	1	0	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
XII. Third knot tying																																				
Steps uncompleted	1	1	1	0	0	0	0	1	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	1	0	0.5	0	0	0	1	1	1	0	0	0	0	0	0	0	1	0.5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
XIII. Eval repair																																				

(continued)

Table 6. Continued

																			Excluded																	
			15			25			35			65			75			45			55			85			95			105			115			125
	15A	15B	Ave	25A	25B	Ave	35A	35B	Ave	65A	65B	Ave	75A	75B	Ave	45A	45B	Ave	55A	55B	Ave	85A	85B	Ave	95A	95B	Ave	105A	105B	Ave	115A	115B	Ave	125A	125 B	Ave
Steps uncompleted	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Errors made	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Total time (Port, Dx, and Rx), min			38			47			37			37			29			43			43			52			42			31			35			27
Rating pairs																																				
Steps completed (45)	33	34	33.5	41	43	42	36	33	34.5	42	39	40.5	42	41	41.5	40	42	41	39	42	40.5	41	40	40.5	42	42	42	43	39	41	36	35	35.5	37	37	37
Errors made (77)	5	3	4	3	1	2	7	6	6.5	3	2	2.5	4	6	5	3	4	3.5	1	1	1	3	1	2	3	4	3.5	3	4	3.5	8	7	7.5	1	1	1
Sentinel errors	1	0	0.5	0	0	0	3	3	3	0	0	0	1	1	1	0	3	1.5	0	0	0	0	0	0	0	0	0	0	0	0	2	2	2	1	0	0.5
Steps completed																																				
Agreement			42			39			38			40			43			42			43			44			41			41			38			41
Disagreement			3			6			7			5			2			3			2			1			4			4			7			4
Step IRR			0.93			0.87			0.84			0.89			0.95			0.93			0.95			0.98			0.91			0.91			0.84			0.91
Errors made																																				
Agreement			73			73			73			75			75			73			75			75			76			76			72			75
Disagreement			4			4			4			2			2			4			2			2			1			1			5			2
Error IRR			0.95			0.95			0.95			0.97			0.97			0.95			0.97			0.97			0.99			0.99			0.93			0.97
Total score (S + E)																																				
Agreement			115			112			111			115			118			115			118			119			117			117			111			116
Disagreement			7			10			11			7			4			7			4			3			5			5			11			6
Total IRR			0.94			0.92			0.91			0.94			0.97			0.94			0.97			0.97			0.96			0.96			0.91			0.95

anch, anchor; caps, capsule; Dx, diagnostic; E, errors; eval, evaluate; glen, glenoid; inf, inferior; instab, instability; IRR, inter-rater reliability; min, minutes; mgmt, management; port, portals; Rx, treatment; S, steps; sut, suture; sut del, suture delivery.

AANA for their talent and ability to teach and communicate shoulder arthroscopy skills to trainees. The novice group was made up of 7 volunteers (from a total of 46 orthopaedic residents registered for a weekend course) who elected to participate in the study and perform an arthroscopic Bankart repair on a shoulder model simulator.

IRR Assessments

The IRR calculations across each of the assessments were strong. The mean IRR for the paired scoring of all 19 videos was 0.92 (range, 0.84 to 0.98) for procedural steps and 0.94 (range, 0.84 to 0.99) for errors (includes DNTT events). The mean IRR for the total of steps and errors was 0.93 (range, 0.85 to 0.97). In no instance did any of the 3 IRR calculations for the 19 scored videos fall below 0.80.

Outlier Experienced Surgeon Performance

Before analysis of the data, score profiles of participants in the experienced group were examined for atypical performance. One subject in the experienced group was found to have enacted 3 sentinel errors in comparison with a mean experienced-group sentinel error rate equal to 0.71 (SD, 0.98). The mean and standard deviation were used to convert the data to *Z* scores, with a mean of 0 and an SD of 1. The outlier subject's *Z* score for sentinel errors was 2.31 (>2 SDs from the mean of the experienced group) ($P = .01$). When this subject's score is removed from the sentinel error data, the new mean equals 0.5 (SD, 0.71). Figure 2 shows that the exclusion of this 1 experienced surgeon had little impact on the overall experienced-group scores. In subsequent statistical analysis, all of this outlier's data were excluded.

Overall Performance Comparisons

On average, the experienced surgeons completed more steps than participants in the novice group (39.54 *v* 37.36), but this difference was not statistically significant (Fig 2A). The experienced group took significantly less time to perform the procedure on the shoulder model in comparison with the novice surgeons (39 minutes *v* 66 minutes, $P < .001$) (Fig 2B). The experienced surgeons also made significantly fewer errors than the novice surgeons (3.23 *v* 8.64, $P = .001$) (Fig 2C) and significantly fewer sentinel errors (0.5 *v* 2.36, $P < .001$) (Fig 2D). Overall, experienced arthroscopists made 63% fewer errors, made 79% fewer sentinel errors, and performed the procedure in 42% less time than novice surgeons. The procedural phases that exhibited the greatest differences in performance between the groups included anchor preparation and insertion, suture delivery and management, and knot tying. The experienced group also performed better than the novices on the phases of portal placement, arthroscopic instability

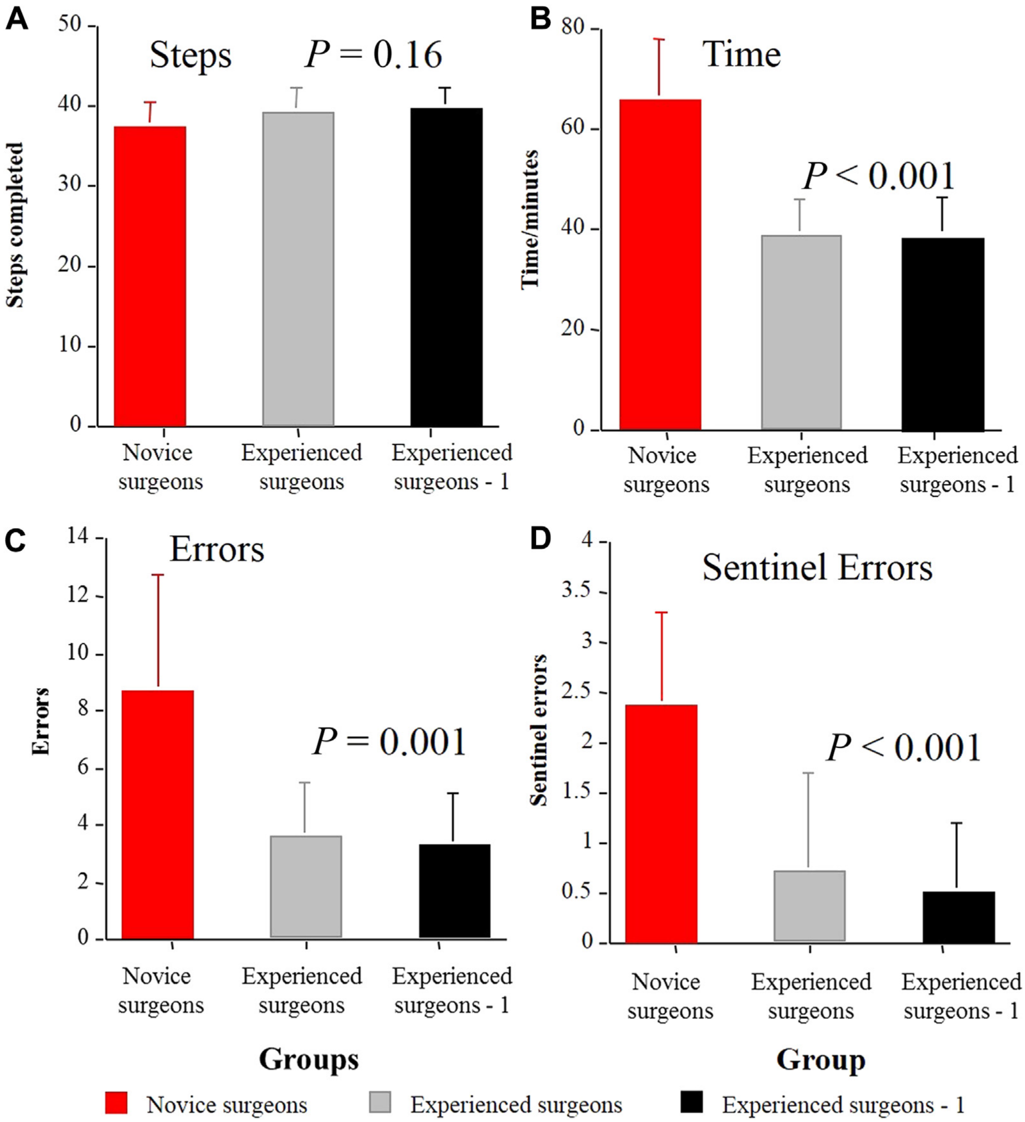

Fig 2. Summary performance data for the shoulder model simulator showing the mean and standard deviation scores for (A) steps completed, (B) time taken, (C) errors enacted, and (D) sentinel errors made by novice and experienced operators. Also shown are the mean scores of the experienced group with the 1 outlier's dataset excluded.

assessment, and capsulolabral mobilization/glenoid preparation, but the differences were not significant.

Anchor Preparation and Insertion Steps

Figure 3A shows the mean number of steps not completed by both groups during the anchor preparation/insertion phase of the procedure. Few steps were not completed or were omitted by either group. Experienced surgeons performed better on anchor 1, but surgeons in the novice group performed marginally better on anchors 2 and 3. These differences were not statistically significant.

Anchor Preparation and Insertion Errors

Figure 3B shows the mean (and standard deviation) number of errors made by the experienced group and the novice group during the anchor preparation and insertion phase of the procedure. Across all 3 anchors, the experienced group made fewer errors than the novices and showed more consistent performance as indicated by the smaller standard deviation scores. Experienced arthroscopists made significantly fewer errors than novices on the preparation and insertion of anchor 1 (0.32 *v* 1.36, $P = .02$). Although experienced surgeons also made considerably fewer errors than novices on the preparation and insertion of anchor 3, this did not reach statistical significance (0.14 *v* 0.86, $P = .07$).

Suture Delivery and Management Steps

The number of uncompleted steps during suture delivery and management for anchors 1, 2, and 3 was few, and there were almost no differences between the groups (Fig 3C).

Suture Delivery and Management Errors

Figure 3D shows the mean (and standard deviation) number of errors made by experienced and novice surgeons on the suturing steps of the procedure. The number of errors made by experienced arthroscopists was small

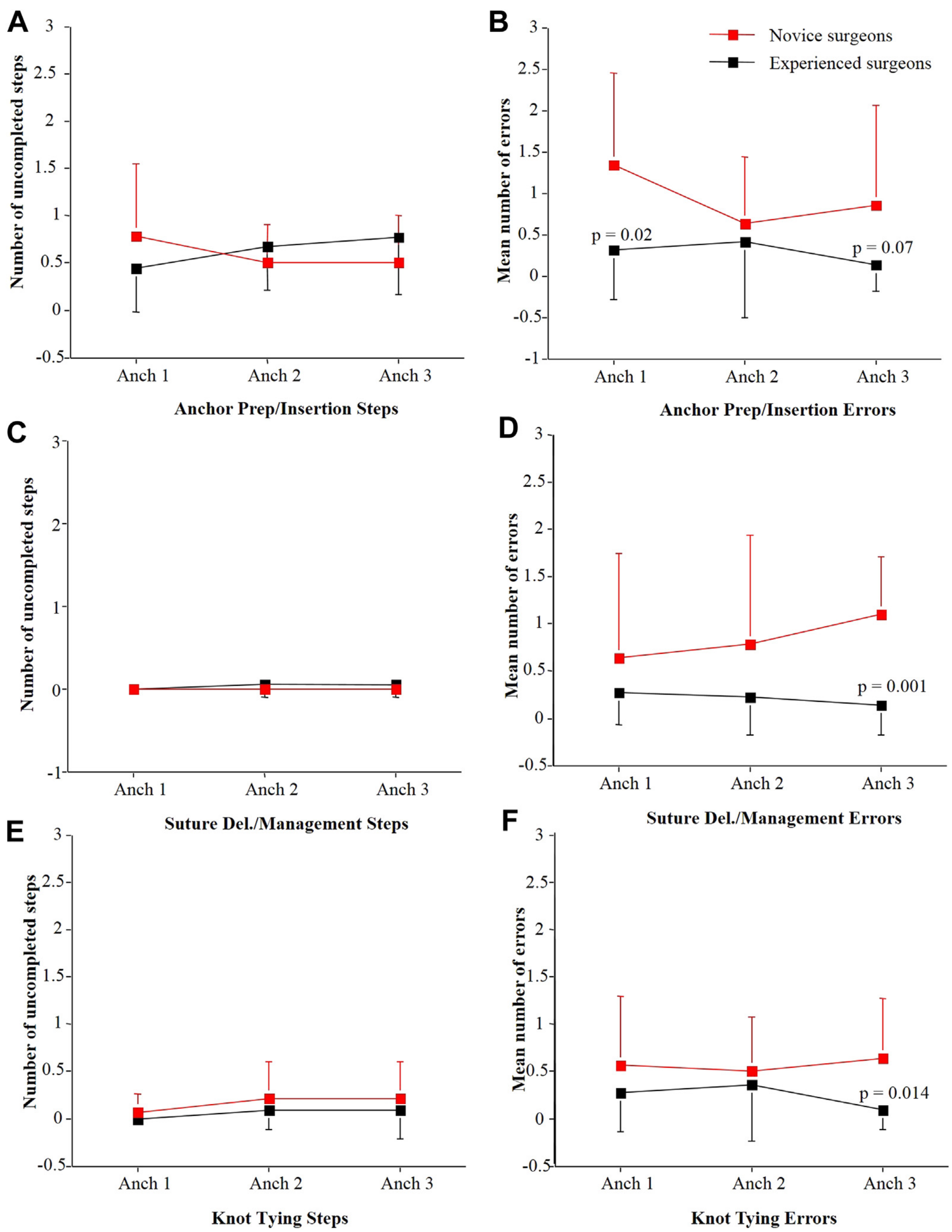

Fig 3. Mean and standard deviation data for novice and experienced surgeons for (A) anchor preparation (Prep) and insertion steps and (B) errors (phases IV, VII, and X); (C) suture delivery (Del) and management steps and (D) errors (phases V, VIII, and XI); and (E) knot-tying steps and (F) errors (phases VI, IX, and XII) during Bankart procedure. (Anch, anchor.)

across all 3 anchors and showed substantial consistency as indicated by the small standard deviation scores. In addition, their performance showed slight improvement across the anchors. In contrast, the novices showed considerable performance variability and performance deterioration across the 3 anchors. Only the differences in suture management and delivery on anchor 3, however, were found to be statistically significant (0.14 *v* 1.1, $P = .001$).

Knot-Tying Steps

Similar to the results for the suturing steps, only a small number of uncompleted or omitted steps during knot tying for anchors 1, 2, and 3 was observed for both groups (Fig 3E) and the small differences that did exist were not statistically significant.

Knot-Tying Errors

The experienced group consistently made fewer errors during the knot-tying phase of the model procedure (Fig 3F). They performed best on anchor 3 as indicated by their mean score and very small standard deviation. The novice group made, on average, less than 1 error per anchor. A significant difference between the groups was, however,

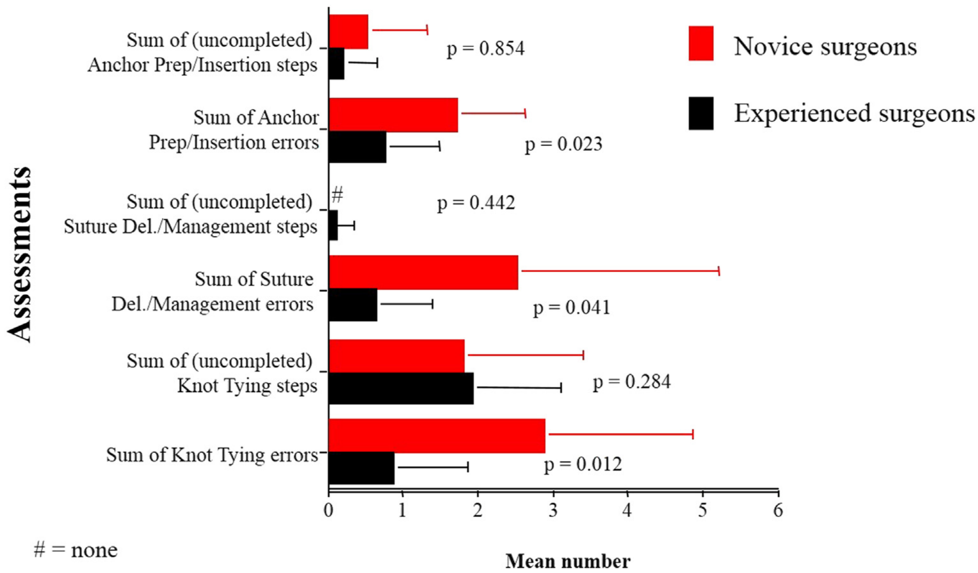

Fig 4. Mean and standard deviation data for novices and experienced surgeons across 6 groups of metrics: anchor preparation (Prep) and insertion steps and errors (phases IV, VII, and X); suture delivery (Del) and management steps and errors (phases V, VIII, and XI); and knot-tying steps and errors (phases VI, IX, and XII).

observed for knot-tying errors on anchor 3 (0.09 *v* 0.64, *P* = .014).

Performance Summary Assessments

The performances of both groups across the 6 measures presented in Figure 3 were summed to give an indicator of each group's overall performance on the 3-anchor repair. These data are presented in Figure 4. Although the experienced group completed more of the procedure steps than the novice group for the anchor preparation and insertion, suture management, and knot-tying phases, none of these differences in steps completed were statistically significant. In contrast, all of the error variables did show large and statistically significant differences between the groups. The experienced group made significantly fewer (i.e., 70%) anchor preparation and insertion and suture management errors (0.86 *v* 2.9, *P* = .012). The experienced group also made 74% fewer suture delivery and management errors in comparison with the novice group; this difference was found to be statistically significant (0.6 *v* 2.5, *P* = .041). The smallest difference between the error performances of the 2 groups was in the knot-tying phase. The experienced group still made 57% fewer knot-tying errors than the novice group; this difference was also found to be statistically significant (0.73 *v* 1.7, *P* = .023).

Procedure Review

Both groups completed the final examination of the repair, and there were no errors committed during this phase of the procedure by either group.

Proficiency Benchmark

The experienced group had a mean total error rate of 3.23. A surgeon could not make a portion of an error, so for practical purposes, the error benchmark was rounded to the next greater whole number, 4. The experienced group also created a mean number of 0.5 sentinel errors (SD, 0.71), rounded to 1. Thus the overall benchmark is set at completing a 3-anchor Bankart repair with no more than 4 total errors and no more than 1 sentinel error.

Discussion

The most important findings of this study include verification that the arthroscopic Bankart metrics coupled with a shoulder model simulator are able to accurately discriminate between the performance of experienced and novice orthopaedic surgeons and show "construct validity" (Table 1). In addition, a performance benchmark was able to be established based on the mean performance of the experienced group and included completion of a 3-anchor Bankart repair, enacting no more than 4 total errors and 1 sentinel error.

Bankart Metrics

The primary intent of the study was to determine whether construct validity could be demonstrated for the previously established arthroscopic Bankart metrics coupled with the use of a shoulder model simulator. For construct validity to be demonstrated, the combination of ABR metrics and simulator tools must be able to discriminate between the performance of experienced and novice surgeons. The differences between the 2 groups were significant, and those that best distinguished between experienced and novice surgeons in the performance of an arthroscopic Bankart procedure included (1) the errors enacted in the performance of the procedure, (2) the number of sentinel errors made, and (3) the time it took to perform the procedure.

In this study, error scores were a very powerful and accurate[20] discriminator between the groups, with the novices making more than twice as many errors as the experienced group, with a difference in the standard deviation scores of a similar magnitude. The goal of surgical education should be to help trainees perform well with as few errors as possible. The trainee, however, should be afforded the opportunity to create errors in an inconsequential manner (e.g., on a simulator without associated patient morbidity) and learn from them. Effective progression in training should be demonstrable with a concomitant reduction in errors. For the individual trainee, the identification of specific errors facilitates a focused correction of deficiencies. Performance errors can also be used as a powerful metric tool to shape and configure the related educational curriculums and to establish benchmarks that trainees must meet and demonstrate before progressing.[11,19,21] Lastly, defined errors can serve to guide the development of simulators, that is, not only what they should emulate but also what they should measure.

Within surgical and procedural disciplines, there is unanimous agreement that certain types of technique errors are so egregious and pose such a threat to either the success of the procedure or patient safety that they should constitute their own performance category. We have elected to term these more serious deviations from optimal performance "sentinel errors." Similar designations have not been formally made in other studies that have objectively assessed surgical performance and evaluated the construct validity of specific metrics.[7-10,22,23] The use of such a special metric classification could have profound implications for "high-stakes" assessments (i.e., determining whether a trainee is allowed to progress in the specific educational/residency program) and proficiency-based progression approaches to training.

Although the number of steps completed did not distinguish between the groups in this study, it is nonetheless important to include all of the essential steps in a training program. The procedure cannot be completed without knowing and performing all of the correct steps in the proper order.[11,19,21] The fact that this performance unit is being assessed also increases the probability that the steps will be learned.[24] However, the proper steps and sequences should be communicated and learned outside of the skills training proper, such as in an online educational module, because its inclusion is sensible rather than essential. A thorough diagnostic evaluation of pathology potentially related to shoulder instability is necessary for the comprehensive and appropriate treatment of the unstable shoulder and, thus, its inclusion in the steps of the procedure.

Simulator Model

Gallagher and O'Sullivan[11] have proposed that to be effective, a simulation model should provide the learner with the span of appropriate sensory responses to physical actions that are behaviorally consistent with what would be experienced in the real situation, including the opportunity to enact both appropriate actions (steps) and inappropriate actions (errors). The simulator should also afford the opportunity to execute the procedure in the same order and with the same tools and devices with which the procedure would normally be performed.[21] The simulator model used in this study was sufficiently realistic to provide the learner the opportunity to perform each of the 45 steps in a realistic fashion using the same tools, implants, and techniques used for an anterior stabilization in a real patient.

Benchmarking

The second purpose of this study was to establish a performance benchmark for the arthroscopic Bankart metrics coupled with a shoulder model simulator. The definition of "proficiency," in distinction to qualitatively described "competency," is based on objectively defined performance metrics. Proficiency-based progression training requires the establishment of a benchmark that trainees must reach to be able to progress. We sought to establish an objective, reliable, transparent, and fair performance benchmark for an anterior Bankart repair on the shoulder model simulator. Because the benchmark was to be established based on the mean performance of the group of experienced surgeons, it was important that the performance of the members be representative of that group. On the basis of a pre-study stipulation, the scores of 1 member of the experienced group were removed from the analysis because the performance was 2.3 SDs worse than that of this participant's peers. This policy was established so as not to skew the creation of the reference benchmark and was shown to have little impact on the overall scores of the experienced group.

An IRR equal to or greater than 0.80 is considered acceptable.[18,25] The very high IRR for the scores from reviewer pairs for the entire group of metrics (0.93) is reflective of the clarity and precision of the arthroscopic Bankart metrics drafted, as well as the thorough training of the 10 reviewers. The ability to score the steps and errors consistently is essential in obtaining a reliable measure of the surgeon's performance and skill level for a particular procedure.

Limitations

A limitation of this study is that there was no confirmation that the participants identified as master/associate master surgeons and representative of the experienced group possessed a specified level of expert skill in performing an arthroscopic Bankart procedure. Nevertheless, the individual surgeons so identified have been recognized by the AANA as skilled and effective educators either from lecture presentations with video exhibiting skilled shoulder arthroscopic techniques or from their performance in an arthroscopic laboratory setting in which they demonstrated the ability to teach each of the key components of a Bankart procedure. Therefore this group was defined as "experienced" rather than "expert." Similarly, other than identification of the year in training, no information was obtained to determine the extent of the novice group's (resident) experience with arthroscopic shoulder surgery, that is, the number of arthroscopy/sports medicine rotations previously completed or the number of shoulder arthroscopic procedures in which the participant served as an assistant surgeon. Even with those pieces of information, it would not be possible to gain any reliable measure of an individual resident's level of first-hand "experience" or skill with arthroscopic shoulder surgery. Furthermore, the structure of residency rotations and level of independence permitted vary a great deal among training programs. As a result, the residents' knowledge and skill sets are unlikely to be uniform but the PGY 4 and 5 levels provide a general measure of their training experience consistent with the designation "novice."

The numbers of surgeons in both groups were small, but one of the strengths of the detailed metric-based procedure characterization method that we used is the sensitivity to detect differences when in fact they exist. We obtained 123 data points (metrics) including the duration of the procedure in minutes for each scored video. Thus small numbers of subjects can still produce statistically powerful differences, assuming that performance has been reliably measured. An a priori power analysis was not performed because we found no previous studies of arthroscopic shoulder repair that used a similar detailed assessment methodology. This is the first study of its kind in this field, and our results will afford other researchers the opportunity to develop their sample sizes based on the reported mean and standard deviation scores. The only published scientific reports we could draw on were similar types of studies published in the laparoscopic surgical literature. Those reports could only give an indication of the possible sample sizes required.

Although not specifically a limitation of the study or design, the option to use either the lateral decubitus or beach-chair position could potentially introduce some variability. Both patient positions are in common use among practicing surgeons. The metrics were carefully constructed to facilitate unbiased scoring for the model simulator/patient in either orientation with no penalty. Several metrics require that the arthroscope be placed in the anterosuperior portal to adequately complete the step; however, this is true for both orientations (e.g., step 12, "view or probe the insertion of the anterior glenohumeral ligaments onto the anterior humeral neck," and step 16, "obtain a view of the anterior glenoid neck"). The challenge in completing these steps relates more to the position of the arthroscope (posterior *v* anterosuperior portal) than to the patient orientation. Although we believe that the lateral decubitus orientation makes some steps easier to perform (e.g., appropriate seating of the drill guide in the anterior/posterior dimension relative to the bony rim, as well as accurate passage of the suturing device through the capsulolabral tissue inferior to the anchor site), no inherent bias is introduced in scoring the metrics for procedures performed with either patient orientation.

On the basis of the data from this study, the null hypothesis is rejected. The shoulder model used, coupled with previously validated arthroscopic Bankart metrics, is able to accurately distinguish between experienced and novice operators. Construct validity is demonstrated for the simulator model coupled with Bankart performance metrics.

Conclusions

The tool comprising validated arthroscopic Bankart repair metrics coupled with a dry shoulder model simulator is able to accurately distinguish between the performance of experienced and novice orthopaedic surgeons. A performance benchmark based on the experienced group includes completion of a 3-anchor Bankart repair, enacting no more than 4 total errors and 1 sentinel error.

References

1. Healy GB. The college should be instrumental in adapting simulators to education. *Bull Am Coll Surg* 2002;87:10-11.
2. Pellegrini CA. Surgical education in the United States: Navigating the white waters. *Ann Surg* 2006;244:335-342.
3. Sachdeva AK, Pellegrini CA, Johnson KA. Support for simulation-based surgical education through American College of Surgeons—accredited education institutes. *World J Surg* 2008;32:196-207.
4. Selzer DJ, Dunnington GL. Surgical skills simulation: A shift in the conversation. *Ann Surg* 2013;257:594-595.
5. Zendejas B, Brydges R, Hamstra SJ, et al. State of the evidence on simulation-based training for laparoscopic surgery: A systematic review. *Ann Surg* 2013;257:586-593.
6. Satava RM. Virtual reality surgical simulator. The first steps. *Surg Endosc* 1993;7:203-205.
7. Seymour NE, Gallagher AG, Roman SA, et al. Virtual reality training improves operating room performance: Results of a randomized, double-blinded study. *Ann Surg* 2002;236:458-464.
8. Grantcharov TP, Kristiansen VB, Bendix J, et al. Randomized clinical trial of virtual reality simulation for laparoscopic skills training. *Br J Surg* 2004;91:146-150.
9. Ahlberg G, Enochsson L, Gallagher AG, et al. Proficiency-based virtual reality training significantly reduces the

error rate for residents during their first 10 laparoscopic cholecystectomies. *Am J Surg* 2007;193:797-804.
10. Van Sickle K, Ritter EM, Baghai M, et al. Prospective, randomized, double-blind trial of curriculum-based training for intracorporeal suturing and knot tying. *J Am Coll Surg* 2008;207:560-568.
11. Gallagher AG, O'Sullivan GC. *Fundamentals of surgical simulation; principles & practices*. London: Springer Verlag, 2011.
12. Angelo RL, Ryu RKN, Pedowitz RA, Gallagher AG. Metric development for an arthroscopic Bankart procedure: Assessment of face and content validity. *Arthroscopy* 2015;31:1430-1440.
13. Morgan CD, Bodenstab AB. Arthroscopic Bankart suture repair: Technique and early results. *Arthroscopy* 2010;26:819-820.
14. Streubel PN, Krych AJ, Simone JP, et al. Anterior glenohumeral instability: A pathology-based surgical treatment strategy. *J Am Acad Orthop Surg* 2014;22:283-294.
15. Waterman BR, Burns TC, McCriskin B, et al. Outcomes after Bankart repair in a military population: Predictors for surgical revision and long-term disability. *Arthroscopy* 2014;30:172-177.
16. Shibata H, Gotoh M, Mitsui Y, et al. Risk factors for shoulder re-dislocation after arthroscopic Bankart repair. *J Orthop Surg Res* 2014;9:53.
17. Ryu RK. Arthroscopic approach to traumatic anterior shoulder instability. *Arthroscopy* 2003;19:94-101.
18. American Educational Research Association. Standards for educational and psychological testing, 2014 ed. Available at www.apa.org/science/programs/testing/standards.aspx. Accessed January 12, 2014.
19. Gallagher AG, Ritter EM, Champion H, et al. Virtual reality simulation for the operating room: Proficiency-based training as a paradigm shift in surgical skills training. *Ann Surg* 2005;241:364-372.
20. Rossi MJ, Lubowitz JH, Provencher MT, Poehling GG. Precision versus accuracy: A case for common sense. *Arthroscopy* 2012;28:1043-1044.
21. Gallagher AG. Metric-based simulation training to proficiency in medical education—What it is and how to do it. *Ulster Med J* 2012;81:107-113.
22. Seymour NE, Gallagher AG, Roman SA, et al. Analysis of errors in laparoscopic surgical procedures. *Surg Endosc* 2004;18:592-595.
23. Larsen CR, Soerensen JL, Grantcharov TP, et al. Effect of virtual reality training on laparoscopic surgery: Randomized controlled trial. *BMJ* 2009;338:b1802.
24. Reinhardt-Rutland AH, Gallagher AG. Visual depth perception in minimally invasive surgery. In: Robertson SA, ed. *Contemporary ergonomics*. London: Taylor and Francis, 1995;531-536.
25. Gallagher AG, Ritter EM, Satava RM. Fundamental principles of validation, and reliability: Rigorous science for the assessment of surgical education and training. *Surg Endosc* 2003;17:1525-1529.