A Proficiency-Based Progression Training Curriculum Coupled With a Model Simulator Results in the Acquisition of a Superior Arthroscopic Bankart Skill Set

Richard L. Angelo, M.D., Richard K. N. Ryu, M.D., Robert A. Pedowitz, M.D., Ph.D.,
William Beach, M.D., Joseph Burns, M.D., Julie Dodds, M.D., Larry Field, M.D.,
Mark Getelman, M.D., Rhett Hobgood, M.D., Louis McIntyre, M.D., and
Anthony G. Gallagher, Ph.D., D.Sc.

Purpose: To determine the effectiveness of proficiency-based progression (PBP) training using simulation both compared with the same training without proficiency requirements and compared with a traditional resident course for learning to perform an arthroscopic Bankart repair (ABR). **Methods:** In a prospective, randomized, blinded study, 44 postgraduate year 4 or 5 orthopaedic residents from 21 Accreditation Council for Graduate Medical Education–approved US orthopaedic residency programs were randomly assigned to 1 of 3 skills training protocols for learning to perform an ABR: group A, traditional (routine Arthroscopy Association of North America Resident Course) (control, n = 14); group B, simulator (modified curriculum adding a shoulder model simulator) (n = 14); or group C, PBP (PBP plus the simulator) (n = 16). At the completion of training, all subjects performed a 3 suture anchor ABR on a cadaveric shoulder, which was videotaped and scored in blinded fashion with the use of previously validated metrics. **Results:** The PBP-trained group (group C) made 56% fewer objectively assessed errors than the traditionally trained group (group A) (*P* = .011) and 41% fewer than group B (*P* = .049) (both comparisons were statistically significant). The proficiency benchmark was achieved on the final repair by 68.7% of participants in group C compared with 36.7% in group B and 28.6% in group A. When compared with group A, group B participants were 1.4 times, group C participants were 5.5 times, and group C^{PBP} participants (who met all intermediate proficiency benchmarks) were 7.5 times as likely to achieve the final proficiency benchmark. **Conclusions:** A PBP training curriculum and protocol coupled with the use of a shoulder model simulator and previously validated metrics produces a superior arthroscopic Bankart skill set when compared with traditional and simulator-enhanced training methods. **Clinical Relevance:** Surgical training combining PBP and a simulator is efficient and effective. Patient safety could be improved if surgical trainees participated in PBP training using a simulator before treating surgical patients.

Changing work patterns and a reduction in hours available for training[1,2] have forced the surgical community to consider new methods to augment and enhance training. Surgical simulation-based training, first proposed by Satava[3] in 1993 as a potential solution to this problem, has developed in sophistication and adoption among the medical education and training communities.[4,5] The first prospective, randomized, double-blind clinical trial of simulation-based training for the operating room showed that surgical residents trained to a "proficiency benchmark" (Table 1) on a virtual reality simulator made

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
Damage to non-target tissue	Iatrogenic damage to tissues not intended to be addressed in the specific step (e.g., articular cartilage damage)
Definition	A definite, distinct, and clear objective characterization providing an accurate and reliable identification of whether an event was or was not observed to occur
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 modified form, experts answer queries or 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
Likert-scale assessment	A method of assessing a range of attitudes; it ascribes a quantitative value to qualitative data (e.g., 1, extremely unlikely; 3, likely; and 5, highly likely)
Metric	A standard of measurement of quantitative assessments used for objective evaluations to make comparisons or to track performance
Operational definition	Terms used to define a variable or event in terms of a process (or set of validation tests) needed to determine its existence, quantity, and duration
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, constitute a complete operative procedure
Proficiency/proficiency benchmark	A specific level of performance defined by a quantitative score (benchmark) on a standardized test or other form of assessment
Proficiency-based progression	A training program that dictates that skill performance be demonstrated, to a predetermined benchmark level, by the trainee before advancement to more complex techniques
Sentinel error	An event or occurrence involving a serious deviation from optimal performance during a procedure that either (1) jeopardizes the success or desired result of the procedure or (2) creates significant 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

significantly fewer objectively assessed intraoperative errors when compared with the control group.[6] The reader is referred to Table 1 for a list of terms used throughout this article. Gallagher et al.[7] and Gallagher and O'Sullivan[8] have argued that simulation-based training is optimal when trainees are given precise feedback on their performance with specific recommendations for improvement, proximate to the performance. They have also suggested that trainees be provided a quantitative performance benchmark to work toward and that this benchmark should be a valid representation of a clinically important performance characteristic or task. Thus trainees must demonstrate the ability to meet specific performance benchmarks before they are permitted to progress in training (proficiency-based progression [PBP] training) (Table 1). The effectiveness of this methodology is well supported.[6,9,10]

We sought to study the effectiveness of PBP training plus simulation for the acquisition of surgical skills. For the patient with unidirectional anterior instability due primarily to a Bankart lesion (capsulolabral detachment from the anteroinferior glenoid) without significant bone loss, a suture anchor repair using 3 implants is a commonly accepted method used to obtain a successful patient outcome.[11-17] In addition, the essential components of the procedure are well outlined regardless of whether the patient is placed in the lateral decubitus or beach-chair orientation.[18,19] Thus an arthroscopic Bankart repair (ABR) was selected as the platform for this research.

The investigation into PBP training plus simulation required the development and validation of 3 separate, specific tools to conduct the analysis. The first component to be created was a "metric tool" that could objectively and accurately characterize an ABR by clearly defining the essential "steps," "errors," and "sentinel errors" (more serious errors) for a standard reference repair (Table 1). The metric tool created was shown to have face and content validity[20] using a

modified Delphi panel methodology (Table 1).[21] Second, a "training tool" (a shoulder model simulator coupled with the ABR metrics) was proved to have construct validity (Table 1), showing the ability to distinguish between novice and experienced surgeon performance. A proficiency benchmark for the use of the metrics with the simulator was established.[22] Lastly, an "assessment tool" (a cadaveric shoulder coupled

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

I. Portals
  1. Establish posterior portal
  2. View posterior humeral head and extent of the Hill-Sachs when present
  3. Introduce midanterior 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 cuff
  10. Probe anteroinferior glenoid/Bankart pathology including rim fracture, articular defect
  View from anterosuperior portal
  11. View or probe the midsubstance 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 (demonstrate restoring tension)
  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 suture anchor
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 glenoid equator
  28. Drill anchor hole oblique to the glenoid articular face
  29. Insert suture anchor
  30. Test anchor security by pulling on 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 tissue
  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 midanterior cannula
Failure to maintain intra-articular position of the anterosuperior cannula
Damage to the superior border of the subscapularis
Damage to the anterior border of the supraspinatus
Loss of intra-articular position of arthroscope/sheath or operating cannula (loss of each portal is scored
 only once for each phase (Roman numeral), i.e., up to a total of 3 for arthroscope + 2 portals)
Lacerate intact capsulolabral tissue (sentinel error)
Failure to maintain control of working instrument (sentinel error)
Guide is not located in the inferior region of the anteroinferior quadrant of the glenoid (for the first anchor position)
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
Implant breakage
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 the anchor hole for anchor positions 1 and 2 (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 with knot tying (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 (n = 77 total errors).

with the ABR metrics) was evaluated and also shown to have construct validity (Table 1).[23]

The purpose of this study was to determine the effectiveness of PBP training using simulation both compared with the same curriculum without the proficiency requirements and compared with a traditional Arthroscopy Association of North America (AANA) Resident Course for learning to perform an ABR. We hypothesized that a training protocol coupling PBP training with a shoulder model simulator would be superior to an identical curriculum using a simulator but without the need to show proficiency, as well as to a traditional curriculum with no simulator or proficiency requirements.

Table 4. Demographic and Baseline Perceptual, Visuospatial, and Psychomotor Assessment Data for Participants

Demographic Variable	Group A	Group B	Group C	P Value
Gender, n				
Male	13	13	14	
Female	1	1	2	
PGY of training, n				
PGY 4	6	8	7	
PGY 5	8	6	9	
Right hand dominant	82%	93%	94%	
Age, yr, mean (SD)	33 (4)	31 (2)	32 (3)	.54
Perceptual assessment (PicSOr), mean (SD)	0.92 (0.07)	0.93 (0.07)	0.93 (0.03)	.89
Visuospatial assessment, mean (SD)	25 (10)	26 (7)	24 (7)	.76
Psychomotor assessments, mean (SD)				
Dominant hand				
Correct incisions	12 (3)	11 (2)	10 (4)	.45
Incorrect incisions	0.3 (0.5)	0.9 (0.1)	0.1 (0.3)	.15
Nondominant hand				
Correct incisions	10 (4)	9 (5)	9 (5)	.66
Incorrect incisions	0.1 (0.5)	0.1 (0.3)	0.2 (0.6)	.6

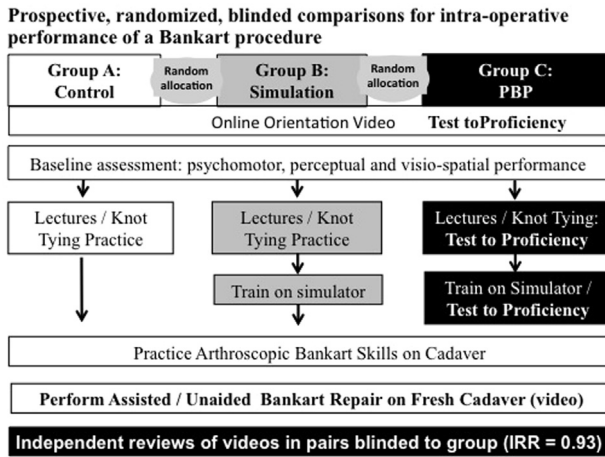 PGY, postgraduate year; PicSOr, pictorial surface orientation.

Fig 1. Study pathways for 3 separate training protocols: groups A (traditional training [control]), group B (simulation enhanced), and group C (proficiency-based progression [PBP] training plus simulation). All groups underwent baseline assessments followed by lectures on knot tying, but only group C was required to pass a cognitive examination on the metrics and to test to proficiency on knot tying. Groups B and C trained to perform an arthroscopic Bankart repair (ABR) on a model simulator, but only group C had to test to proficiency. All 3 groups trained to perform or practiced performing an ABR on a cadaver. All participants from each group performed an unaided ABR on a fresh cadaveric shoulder that was videotaped and scored in blinded fashion. (IRR, inter-rater reliability.)

Methods

Participants/Subjects

Forty-four postgraduate year (PGY) 4 or 5 residents from 21 Accreditation Council for Graduate Medical Education–approved orthopaedic residency training programs from across the United States participated. All subjects were assigned a unique identifying number that gave no indication of their PGY, residency program, or study group. The Western Institutional Review Board opined (No. 1-776362-1) that, as an educational curriculum study, this investigation was exempt from the need for full institutional review board approval [based on the criteria of 45 CFR 46.101(b)(1)]. The study protocol was registered with the National Institutes of Health (ClinicalTrials.gov No. NCT01921621) before initiation of the investigation.

Bankart Procedure Performance Metrics

The surgical residents were evaluated on their skill in performing an ABR on a cadaveric specimen. Previously validated "performance metrics" formed the basis of this evaluation and included 45 key steps with related steps grouped into 1 of 13 phases (Table 1) of the procedure[21] (Table 2). Seventy-seven potential errors to be avoided were specified (Table 3). Of these errors, 20 were designated as more serious, or sentinel, errors because either (1) the error's enactment had the potential to seriously compromise the success of the procedure or (2) the error had the potential to create significant iatrogenic damage to the shoulder. The metrics were clearly defined with beginning points and endpoints for each step, as well as precisely what did and did not constitute each potential error. All subject surgeons and faculty were provided a link on the AANA Web site to 2 full-length orientation videos, 1 each in the lateral decubitus and beach-chair orientations. Access was available 4 weeks before the course in which they were participating. Each video demonstrated all of the steps, in addition to either demonstrating or specifically identifying each of the potential
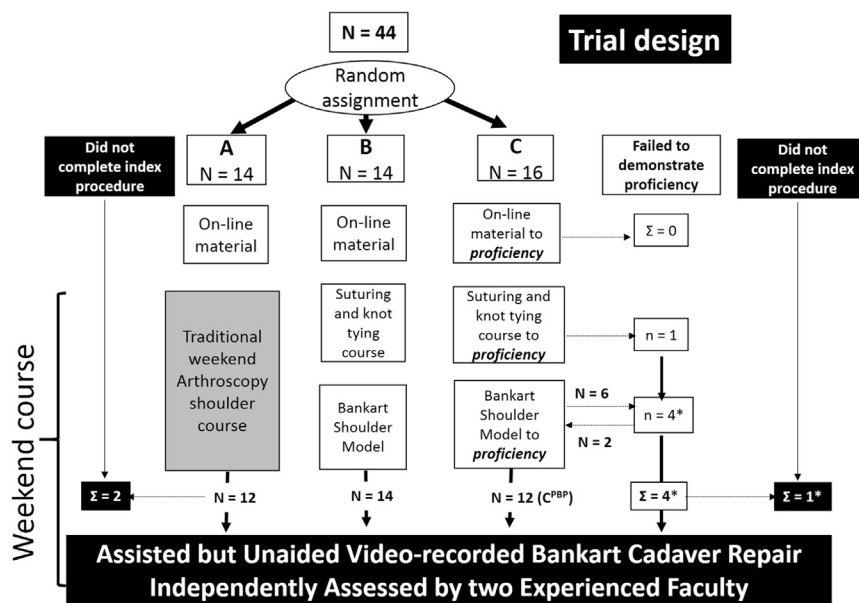
Fig 2. Participant flowchart detailing that 2 of 14 participants in group A were unable to complete the procedure, 1 group C subject could not complete knot tying to proficiency, and 4 group C residents were unable to show proficiency on the shoulder model simulator (failed 2 attempts). The group C subject who failed to show proficiency both on knot tying and on the model simulator also failed to complete the final Bankart repair. The remaining 12 subjects from group C who met all intermediate proficiency benchmarks, designated as group C[PBP], would be the only subjects allowed to progress per the PBP protocol. Asterisks indicate that the participant failed to demonstrate proficiency in one or more intermediate procedure components.

Fig 3. AANA Fundamentals of Arthroscopic Surgery Training workstation with multiple potential portal sites in an opaque dome covering a mandrel around which knots are tied through an arthroscopic cannula (inset on laptop screen). A USB camera (Ipevo, Sunnyvale, CA), directed through a window, projects an image of the inside of the dome onto a laptop computer screen.

errors (including sentinel errors) to be avoided in performing an ABR safely.

Baseline Assessments

To ensure homogeneity among the 44 subjects, all residents completed previously validated assessments of their visuospatial,[6,24] perceptual,[6,25] and psychomotor abilities[26-28] (Table 4). "Visuospatial ability" is one component of cognitive function that is related to the capacity to process and interpret visual information about where objects are in space. In this assessment, a pencil was used to create the shortest and most appropriate route between 2 specific points on a block-grid street map. A possibility of 20 correct routes between various points existed for each of two tests. The number of correct routes created in two 3-minute time periods was scored. Each registrant completed 2 tests (scores range from 0 to 40). "Perceptual ability" refers to the capacity to identify, organize, and interpret sensory information about visual depth of field. It was assessed with a computer-generated and -scored task requiring the subject to orient the axis of a spinning cone perpendicular to a designated face of a cube.[25] Each of 30 trials placed the cube in a different 3-dimensional orientation (scores range from 0.0 to 1.0). "Psychomotor ability" refers to the capacity for coordinated activity involving the arms, hands, and fingers (and, potentially, movement of the feet). Performance was assessed using a lighted endoscopic box trainer with a fixed overhead view projected onto a laptop screen. A 4 × 8–inch piece of paper had a series of 1-inch-long parallel lines drawn perpendicular to and along the long border of the sheet. Each of 30 parallel lines was separated by 10 mm. Instruments were passed through openings in the front of the box trainer. An endoscopic grasper was controlled by one hand and used to hold the paper within the box. Endoscopic scissors were controlled by the other hand and used to make cuts in the paper between the designated lines. The number of accurate paper cuts (between, but not touching, the parallel lines) able to be
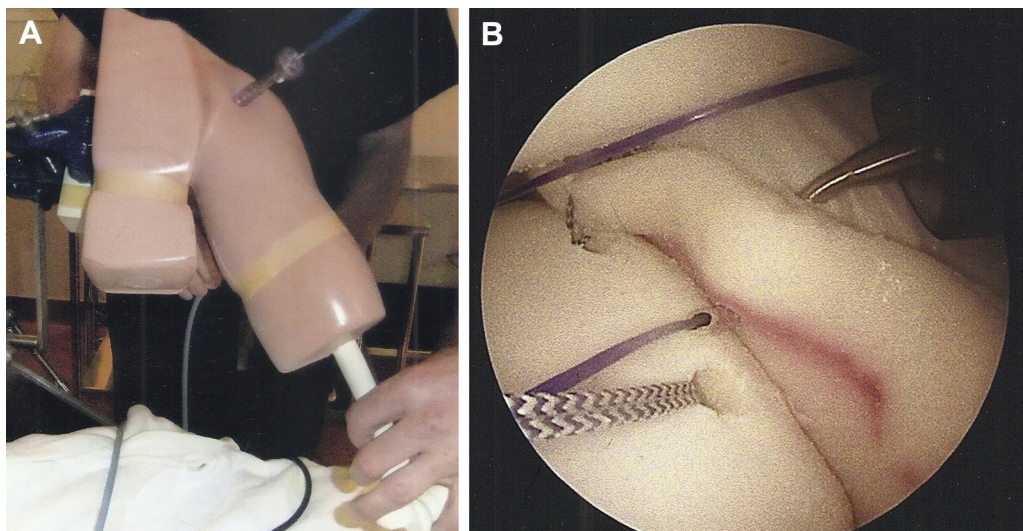
Fig 4. (A) Exterior view of the anterior aspect of a left shoulder model simulator supported in the beach-chair orientation; a hook probe is delivered through a midanterior portal. (B) Arthroscopic view from the anterosuperior portal of a left shoulder in the lateral decubitus orientation. The second anchor is in place with its sutures coursing toward the posterior portal; a suture hook device passes through the midanterior portal, pierces the labrum, and delivers a monofilament suture (to be used later for shuttling).

made in 60 seconds was tabulated. Two trials were run, one with the scissors in the dominant hand and the other with them in the nondominant hand.

Study Groups

During the weekend courses, all groups were provided similar background shoulder instability lectures that focused on indications, contraindications, and case-based examples. References to surgical technique were avoided in the lecture presentations. For each of the 3 groups, separate, dedicated, experienced Master and Associate Master AANA faculty members worked closely with that cohort of residents. The duration of training was similar for each of the 3 groups. The 3 training curricula are outlined in Figure 1. All training was conducted at the Orthopedic Learning Center (OLC) in Rosemont, Illinois.

Group A: Traditional (Control). Group A was derived from a cohort of PGY 4 and 5 residents who had independently registered for a 3-day AANA Resident Course at the OLC. The curriculum involved equal time spent on the practice of knee and shoulder procedures, which included an ABR. The registrants of the course were given the opportunity to be involved in the study but were not required to do so. A cohort of 14 residents (of a total of 48 registered residents), elected to participate in the research project, and completed an ABR at the end of their course of training.

The 14 subjects in group A served as the control group and followed an agenda typical for an AANA Resident Course (Fig 2). The curriculum included lectures on various topics including shoulder instability. Knot-tying skill was practiced under the direction of an experienced faculty member. Both sliding and non-sliding knots, as well as half-hitches to secure the primary knot, were practiced. Knot-tying boards provided the opportunity to tie knots around hooks using large cord and/or No. 2 FiberWire suture (Arthrex, Naples, FL). In addition, practice tying suture knots and delivering them down an 8.5-mm arthroscopic cannula was afforded. Finally, knots could be created and delivered through a cannula with the loop around a mandrel (smooth bar) using the AANA Fundamentals of Arthroscopic Surgery Training workstation (Fig 3). The opaque dome eliminated direct surgeon view of the knot being tied and required the use of triangulation skills. A small USB camera was directed through a window in the dome with the image viewable on a laptop computer screen. The field of view contained the knot being tied. The subject surgeon had the opportunity to spend as much time practicing knots as desired and until the participant believed he or she was proficient. A series of 5 knot trials were then tied and labeled in sequence for later testing.

Study in the laboratory using a cadaveric specimen then followed under the direct supervision of an experienced faculty member with knowledge of the Bankart metrics. Fresh-frozen specimens with a complete shoulder girdle from the scapula to the mid humerus with associated soft tissues were used. After appropriate thawing, the scapula was mounted with a clamp in the surgeon's orientation of preference (lateral decubitus *v* beach chair). Bony landmarks were identified and marked with a surgical pen, portals established, and a diagnostic arthroscopy performed. An ABR was generally the first intra-articular procedure studied. After a Bankart lesion was created, practice was conducted in the steps necessary to mobilize the capsulolabral tissue and complete a 3-anchor repair. Standard instruments for an ABR were made available. A 45° cannulated suture hook was the primary tool used to deliver sutures through the capsulolabral tissue. Single-loaded push-in anchors and a simple loop suture pattern were used. The residents were able to continue with guided study on the Bankart repair as long as desired and until they believed they were proficient. Subsequently, additional arthroscopic shoulder procedures could be electively studied as well.

Group B: Simulator. Multiple randomly selected residency program coordinators were notified of the opportunity for their residents to participate in this PBP training study. The first 32 PGY 4 and 5 residents to register became the study participants. Training for groups B and C was conducted concurrently at the OLC but during a different weekend from group A training. Of the 32 preregistered subjects, 16 were randomly assigned to 1 of 2 training protocols (group B or C) based on a computer-generated random allocation[29] (Fig 2). Of the 16 residents randomized to group B, 2 who were preregistered failed to show up for the weekend course; thus group B comprised 14 residents. The residents in group B engaged in knot-tying study and practice similar to that for group A until they believed they were proficient. A series of 5 knot trials were then tied and labeled in sequence for later testing. Group B participants were afforded the additional opportunity to train and practice an ABR using a dry shoulder model simulator (Fig 4) secured in the orientation of surgeon preference.[22] A standard equipment tower with a 30° arthroscope was provided along with all instruments necessary to complete an ABR. The simulator model was composed of a dense plastic endoskeleton palpable through simulated skin and soft tissues. Posterior, anterosuperior, and midanterior portals were created. A glenoid, humeral head, biceps, capsule, and labrum, in addition to Bankart and Hill-Sachs lesions, were present and provided the opportunity to complete all

of the steps demonstrated in the orientation videos for an ABR. Work on the simulator continued as long as the residents desired and until they believed they were proficient with the steps and sequences of the capsulolabral repair. Further study and guided practice of the steps for an ABR on a cadaveric specimen then followed and continued as long as participants desired and until they believed they were proficient.

Group C: PBP. All residents randomized to group C attended the course (n = 16) and were exposed to a protocol identical to that of group B with the additional requirement of showing proficiency at various stages of the training (Fig 2). Each of the individual proficiency benchmarks for the procedural components was established based on the mean performance of separate groups of experienced surgeons on the specific exercise.[6-10,22,23] After arriving at the OLC, if the resident had not yet taken and passed the cognitive test online, covering the validated metric steps and errors demonstrated in the orientation videos, he or she was required to do so on site. A minimum score of 84% was required to pass. Those who initially failed continued to study the material and were provided additional faculty instruction.

Knot-tying study progressed in a manner similar to groups A and B. Once residents believed they had mastered the knot-tying skills, they had the opportunity for the integrity of their knots to be tested using the Fundamentals of Arthroscopic Surgery Training workstation knot tensiometer[30] if desired. To pass, the loop-knot construct had to elongate less than 3 mm when subjected to a static load of 15 lb for 15 seconds. The benchmark was set at a minimum of 3 of 5 knots meeting this standard. Once the formal testing process began, the subject tied 5 knot trials, which were labeled in sequence. All 5 were then tested using the tensiometer. Subjects who failed continued to practice until confidence was gained, and then the testing sequence was repeated with a new series of 5 knots being tied. This process continued until the resident achieved proficiency or was unable to do so and failed to show progressive improvement in the knot-tying skill set.[30]

Work and practice then began with the same shoulder model simulator used by group B. The model was oriented according to physician preference. Landmarks were identified, and posterior, midanterior, and anterosuperior portals were established. After a diagnostic examination was performed, the steps for a 3-anchor Bankart repair were practiced. The faculty instructors provided proximate feedback and recommended corrections based on the previously defined step and error metrics demonstrated in the orientation videos. Practice and faculty feedback continued through a complete procedure until the subject and his or her faculty

instructor both believed the subject had adequately prepared for testing.

A new simulator model was then oriented in the resident's position of preference. Equipment representatives from multiple different vendors served as surgical assistants and were randomly assigned to participating 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 OLC). The resident surgeon then proceeded to complete a diagnostic evaluation and perform a 3-anchor ABR attempting to mimic the key steps identified in the orientation videos. By use of the Bankart metric score sheet, 1 of 6 designated faculty members, intimately familiar with the Bankart metrics, scored the subject in real time during the arthroscopic repair on the simulator model. The simulator benchmark for a passing score on the shoulder model was established from a prior study[22] and included a 3-anchor repair with no more than 4 total errors and no more than 1 sentinel error. If a resident failed to meet the benchmark, the faculty who scored the model test, the assigned training instructor, and the resident all conferred to identify the specific deficiencies exhibited and the appropriate corrections. The subject then worked toward acquisition of the requisite skills with instructor guidance. When confident, the subject was given 1 additional opportunity to repeat the scored procedure on a new model. In a normal PBP protocol, residents who fail to meet each of the intermediate proficiency benchmarks would not be allowed to progress in training and would require additional practice until the deficiencies were corrected (and would not be allowed to progress to working with the cadaver). However, given the artificial finite time constraints of the study weekend, all group C participants, regardless of persistent deficiencies, were allowed to proceed to practice with a cadaveric specimen and guided instruction similar to groups A and B.

Final Videotaped Bankart Repair Assessment

At the completion of their respective courses, the subjects from each group performed an assisted, unaided arthroscopic diagnostic survey and a 3-anchor Bankart repair on a fresh cadaveric shoulder. The cadaveric specimens were considered acceptable if (1) arthroscopic visibility of the target tissues was obtainable; (2) the specimen permitted adequate access to the target tissues (flexibility); and (3) the integrity of the capsulolabral tissues was sufficient to permit mobilization, suture delivery, and knot tying. All necessary instrumentation and implants were made available. Residents participating in the course served as assistants for each other. They were instructed to act only at the request of the operating surgeon and were

prevented from coaching or prompting. The OLC staff proctored the procedures for compliance.

The procedure was videotaped in its entirety beginning with the initial view from the posterior portal. The resident surgeon mapped the bony landmarks and then created his or her preferred portals. All or a portion of the diagnostic examination was completed. The arthroscope was withdrawn, and a red card was videotaped for 5 seconds to signal that the subject surgeon was no longer operating. One of 4 designated faculty members then reintroduced the arthroscope and, using a sharp elevator from either the anterosuperior portal, midanterior portal, or both, created a standard Bankart lesion, 6 to 9 mm deep (medial from the bony rim) and from the 2- to 6-o'clock position along the glenoid. Once the Bankart lesion was created, care was taken to avoid additional mobilization of the capsulolabral tissue. The arthroscope was then withdrawn, and a green card was videotaped for 5 seconds, signaling that the subject surgeon was operating for the balance of the procedure. The arthroscope was reintroduced into the glenohumeral joint by the subject surgeon, and any remaining elements of the diagnostic survey were completed. The subject then performed a 3-anchor Bankart repair, attempting to mimic the steps demonstrated in the orientation videos and practiced in the simulation model.

All subject surgeons used identical implants: single-loaded (2.8-mm) Gryphon push-in anchors with a single No. 2 Orthocord (DePuy Mitek, Raynham, MA). A 45° cannulated suture hook was used to deliver a shuttling device with retrograde passage of the anchor sutures through the capsulolabral tissues. Before work was begun on the final scored Bankart repair, instructions were given to all residents regarding the protocol for anchor pullout from cadaveric bone. If an anchor failed prior to completion of the index sliding knot, the surgeon was permitted to remove the anchor and suture and to replace it with a metal 5.5-mm screw-in anchor (Smith & Nephew, Andover, MA). The procedure then continued with no penalty. The time required for the reintroduction of the screw-in anchor and re-passage of the anchor suture through the capsulolabral tissue was subtracted from the total procedure time. If the anchor failed subsequent to the initial sliding knot being completed (i.e., efforts to back up the primary knot with half-hitches), the surgeon was instructed to abandon the first anchor position and proceed to the second anchor position. No time limit was imposed on the performance of the Bankart repair, and participants were able to continue to work as long as they believed they were making progress. At the point in the procedure when the subject surgeons did not believe they could make further progress in the Bankart repair, they could electively choose to terminate the procedure.

Video Reviewer Training

Once the construction of the metrics for an ABR was completed and face and content validity verified,[21] a final version of a score sheet was formatted. Ten Master/Associate Master AANA faculty surgeons formed the panel of reviewers designated to score the videos. This group included the 3 arthroscopic surgeons (R.L.A., R.K.N.R., and R.A.P.) who, in conjunction with a consultant experimental psychologist (A.G.G.), developed the arthroscopic Bankart metric definitions (Table 1). The 10 reviewers were randomly assigned to form 5 fixed pairs, which remained constant throughout the scoring of all videos. 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 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 tabulated. During 2 subsequent 2-hour group phone conferences, the differences and discrepancies among all reviewers were compared and discussed, seeking conformity in scoring. In addition, each designated pair of reviewers conducted 1 to 3 additional phone conferences to analyze the specific instances in which the two of them scored particular events differently. Subsequently, all reviewers scored practice videos 3 and 4 (with each patient orientation again represented), and the results were tabulated. 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 discussed later) fall below an acceptable level[8] of 0.8, at 0.76. Thus, confidence was established that the future scoring of the videos generated by the participants of this study could be accomplished with a high IRR.

Video Scoring

The AANA research coordinator randomly assigned the 44 full-length study videos, each with only the designated unique identifying number attached, to a single pair of reviewers. Other than the research coordinator and the study consultant, all video reviewers remained blinded to the source of the video being reviewed. Each video was independently reviewed and scored by the 2 members of an assigned pair of reviewers. All scores were tabulated for each of the 13 phases of the procedure. Each step and error metric were scored as either yes or no, designating whether
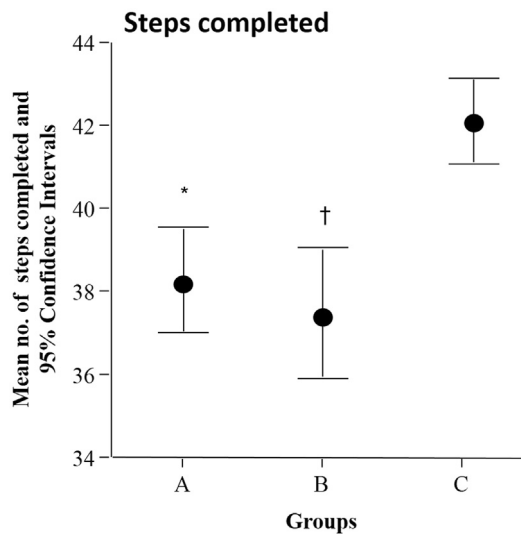
Steps completed

Fig 5. Mean number of Bankart procedure steps (and 95% confidence intervals) completed for each study group. The probability values observed for differences in performance were as follows: [*]P < .001 between group C and group A; [†]P < .001 between group C and group B.

the specific event was or was not observed to have occurred by the reviewer. In addition to scoring of steps and errors, each event characterized as "damage to non-target tissue" (Table 1) (e.g., gouging the articular cartilage or tearing of the capsule) was scored. There was no limit to the number of individual instances in which damage to non-target tissue could be scored, with each occurrence tallied as a single error event. The score sheet also contained a box for specific reviewer comments for each metric.

Score Tabulation

For each of the 13 separate phases of the procedure, the numbers of uncompleted steps and errors made were tabulated and the scores for the 2 reviewers averaged. Furthermore, for each subject, the step and error data were pooled for the 3 repetitive components of the procedure: (1) anchor preparation, (2) suture passage/management, and (3) knot tying. These data were used to determine which of the procedural phases showed the greatest differences in performance among the groups (1-factor analysis of variance analysis) (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 for the pair of reviewers. The subject's operative time was obtained by subtracting the faculty time to create the Bankart lesion from the total recording time for the procedure.

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

In a similar manner, there was either agreement or disagreement in the 2 scores for each of the potential errors (n = 77). The IRR for error scoring was calculated in the same manner as that for the steps. Finally, the IRR for scoring the entire procedure was calculated using both the step and error agreements or disagreements for the complete procedure (n = 122). The acceptable IRR was defined as 0.80 or greater.[8]

Statistical Methods

The analysis was conducted as a series of multiple regressions. The exogenous variables (covariates) were the 3 intervention conditions, that is, PBP plus simulator (condition C), simulator (condition B), and traditional training (condition A). Group C was used as the reference condition within the analysis. As a check on the veracity and stability of the results, all of the analyses were also conducted using Poisson regression. The substantive interpretation remained unchanged regardless of the model used. All of the reported results are based on the analyses from the multiple regressions. Furthermore, a logistic regression analysis was performed to estimate the probability of those trainees from the different training curricula being able to attain the proficiency benchmark for the final ABR.

A secondary analysis was conducted to evaluate the subset of group C subjects who successfully met all of the intermediate benchmarks throughout training,

Table 5. Regression Analysis of Steps Completed

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	SE	β	t	P Value
Reference group: group C (constant)	42.167	0.618		68.225	< .001
Group A	−3.833	0.927	−0.559	−4.135	< .001
Group B	−4.702	0.890	−0.715	−5.286	< .001

B, beta; β, standardized beta; SE, standard error; t, test statistic.
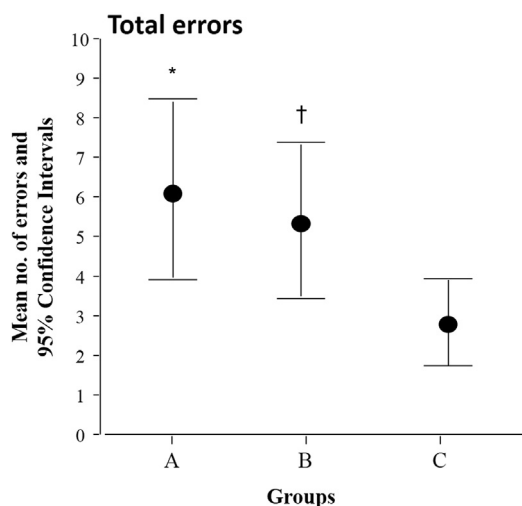
Total errors

Fig 6. Mean number of procedure errors enacted for each study group. The probability values observed for differences in performance were as follows: $^*P = .01$ between group C and group A; $^\dagger P = .049$ between group C and group B.

designated as group C^{PBP}. Group C^{PBP} was evaluated for the same performance metrics as the other groups, which included steps, errors, sentinel errors, and time, as well as the probability of attaining the benchmark on the final repair. All of the participants in group C followed the PBP training "curriculum." The PBP "protocol," in distinction, would only permit those individuals who meet each intermediate proficiency benchmark to progress in training (group C^{PBP}) (Fig 2).

Results

The mean and standard deviation scores on the baseline assessments of perceptual, visuospatial, and psychomotor performance are shown in Table 4. Although group A performed somewhat better on the psychomotor test than groups B and C, these differences were not statistically significant.

Intermediate Proficiency Training Benchmarks for Group C

All 16 participants in group C were able to obtain a passing score on the cognitive examination, although several required additional instruction after failing to achieve a passing score of 84% on their initial test. One subject from this group was unable to show proficiency in knot tying despite repeated training and practice. Six group C participants failed their first attempt to meet the benchmark for a Bankart repair on the simulator model. After additional guided training and practice, 2 of 6 were able to show proficiency on their second attempt with the shoulder model. Of the 4 participants who were unable to show proficiency on the model, one of whom was also the participant who failed to show proficiency at knot tying. Thus 12 of 16 group C subjects met all of the intermediate proficiency benchmarks during training. On the basis of the PBP protocol, these 12 (designated group C^{PBP}) would have been the only participants from group C allowed to progress to working on the cadaver.

Final Cadaveric Bankart Assessment

Two cadavers, 1 each from groups B and C, failed to meet the acceptability criteria and were replaced with better specimens. The video recording was restarted with the onset of work on the replacement specimen. For the 44 videos scored, the mean IRR for the total number of steps performed and errors made was 0.93 (range, 0.84 to 0.99; SD, 0.04).

Incomplete Final Procedures

Of all 44 subjects, only 3 failed to complete their final Bankart repair on a cadaveric shoulder. Two individuals from group A were only able to finish the first anchor with an average of 16.25 steps completed, 7 errors made, and 0.5 sentinel errors enacted. They worked for an average of 99 minutes. In group C, 1 subject had the first anchor pullout during efforts to deliver and secure the primary suture knot, and this subject elected not to replace that anchor (although the subject could have done so according to the anchor pullout protocol). The subject completed all of the second and third anchor components, thus performing only a 2-anchor final repair, which was deemed incomplete. During this procedure, an average of 37 steps were completed, along with 4 errors made and 0 sentinel errors enacted. The operative time was 92 minutes. This subject was the one who had previously failed to show proficiency both on knot tying and on the shoulder model repair components of the training curriculum and would not

Table 6. Regression Analysis of Errors Made

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	SE	β	t	P Value
Reference group: group C (constant)	2.600	0.819		3.173	.003
Group A	3.275	1.229	0.444	2.665	.011
Group B	2.400	1.179	0.339	2.035	.049

B, beta; β, standardized beta; SE, standard error; t, test statistic.
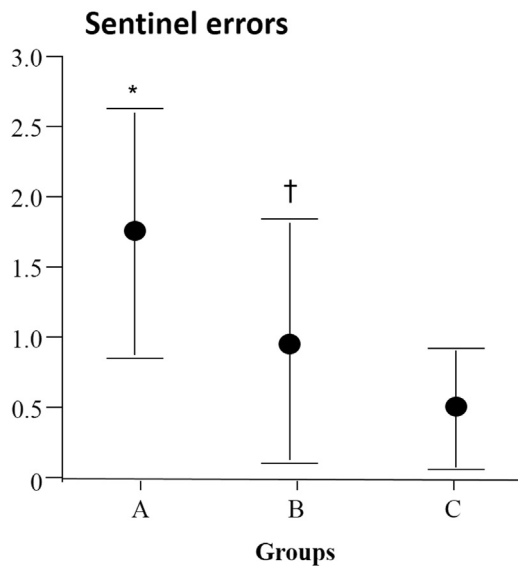
Sentinel errors

Fig 7. Mean number of sentinel errors enacted for each study group. The probability values observed for differences in performance were as follows: [*]$P = .023$ between group C and group A; [†]$P = .351$ between group C and group B.

normally have been allowed to progress to training with the cadaver. It was not possible to estimate or accurately extrapolate the number of errors, number of sentinel errors, or time for the 3 incomplete procedures. Thus, for the comparative analysis of the 3 groups for steps, errors, sentinel errors, and time, group A comprised 12 subjects, group B comprised 14, and group C comprised 15.

Steps Completed

Figure 5 shows the mean values and 95% confidence intervals (CIs) of procedure steps completed by groups A, B, and C. Groups A and B completed a similar number of procedure steps, whereas group C, on average, completed 4 more steps. The differences between the groups' performances using the regression model with group C as the reference group are summarized in Table 5. The results showed that group C completed, on average, 42.2 procedure steps. Subjects in group A completed 3.8 fewer steps, whereas those in group B completed 4.7 fewer steps. Both of these differences were statistically significant ($P < .001$ for group C v group A and $P < .001$ for group C v group B).

Procedure Errors

The mean number of errors and 95% CIs for each group are shown in Figure 6. The mean number of errors among subjects in group C was 2.6 (Table 6). On average, the subjects in group A made 3.3 more errors whereas those in group B made 2.4 more errors than those in group C. Both of these differences were statistically significant ($P = .011$ for group C v group A and $P = .049$ for group C v group B). Overall, the subjects in group C showed a 56% reduction in the mean number of errors over group A and a 41% reduction over the number of errors made by those in group B. The participants in group C were also more consistent, with the range in number of errors much smaller when compared with groups A and B.

Sentinel Errors

The mean values and 95% CIs for sentinel errors are shown in Figure 7. On average, the subjects in group C made 0.53 sentinel errors, whereas those in group A made 1.175 more sentinel errors and those in group B made 0.43 more sentinel errors (Table 7). The difference between group A and group C for sentinel errors was statistically significant ($P = .017$), but the difference between group C and group B was not. Overall, the subjects in group C made 69% fewer sentinel errors than those in group A and 44% fewer than group B.

Bankart Performance Time

The mean values and 95% CIs for time taken by the groups to perform the index procedure are shown in Figure 8. Groups A, B, and C took a similar amount of time to complete the procedure, with no significant differences observed (Table 8).

Analysis of Group C[PBP]

A secondary analysis was conducted to determine the performance of the 12 group C[PBP] subjects. The differences between the mean scores of group C[PBP] and those of group C were calculated. The difference in steps completed was marginal (42.46 for group C[PBP] v 42.2 for group C). The error analysis showed that 12% fewer errors were made (2.29 for group C[PBP] v 2.6 for group C), 6% fewer sentinel errors were made (0.5 for group C[PBP] v 0.53 for group C), and 4% less time was required to complete the procedure (77.17 minutes for group C[PBP] v 80.44 minutes for group C).

Table 7. Regression Analysis of Sentinel Errors Made

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	SE	β	t	P Value
Reference group: group C (constant)	0.533	0.313		1.706	.096
Group A	1.175	0.469	0.425	2.506	.017
Group B	0.431	0.450	0.162	0.958	.344

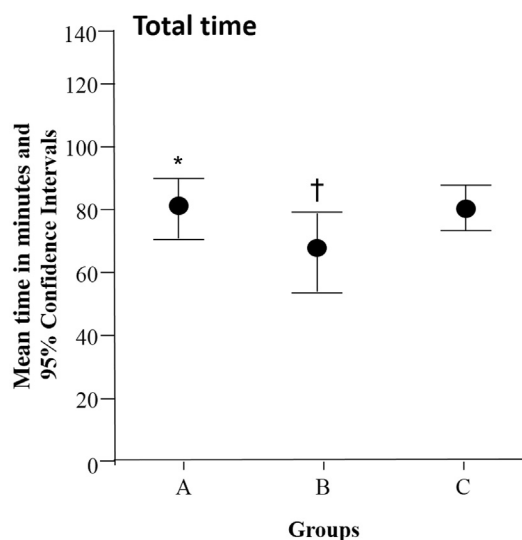B, beta; β, standardized beta; SE, standard error; t, test statistic.

Fig 8. Mean procedure duration in minutes for each study group. There were no significant differences in the probability values between groups C and A (asterisk) or between groups C and B (dagger).

Final Bankart Proficiency Benchmark

The proficiency benchmark (set as the mean performance of an experienced group of surgeons) had previously been established for a cadaveric shoulder[23] as no more than 3 total errors (1 less error than the simulator model benchmark) and no more than 1 sentinel error. In addition, a 3-anchor Bankart repair must have been completed. Overall, 28.6% of group A subjects (4 of 14), 36.7% of group B (5 of 14), 68.7% of group C (11 of 16), and 75% of group C^{PBP} (9 of 12) were able to achieve the final proficiency benchmark.

Logistic regression analysis for the relative differences between the control condition (group A, traditionally trained group) and group B (simulator), group C (simulator + BPB curriculum), and group C^{PBP} (simulator + PBP protocol) was performed and used to determine the odds ratios for the comparisons. Relative to group A, group B subjects were 1.4 times ($P = .121$), group C subjects were 5.5 times ($P = .033$), and group C^{PBP} subjects were 7.5 times ($P = .024$) as likely to achieve the final quantitatively defined proficiency benchmark. Only the comparisons of proficiency between group A and group C, as well as between group A and group C^{PBP}, were statistically significant (Fig 9). Trainees in group C^{PBP}

had a 36.4% greater probability of achieving the final benchmark than those in the entire group C.

Discussion

PBP Paradigm

Two primary conclusions can be drawn from the data in this study. First, the performance of the entire group C on the final Bankart evaluation shows that the PBP curriculum using simulation is superior both to the traditional curriculum (group A) and to the curriculum identical to that in group C (including the use of the simulator) but without the requirement to show proficiency (group B). Second, the performance of group C^{PBP} shows the superiority of the PBP protocol itself, in which only those trainees who meet each sequential intermediate proficiency benchmark during training are permitted to progress in the curriculum. The most important and revealing comparison of the 3 training protocols, therefore, compares groups A, B, and C^{PBP}. The subjects in group C^{PBP} performed more of the operative steps but did not take significantly longer to do so. They also made significantly fewer objectively assessed intraoperative errors and were over 7 times more likely to achieve the final benchmark than those in group A, who followed a traditional training pathway.

The performance of group B was only marginally better than that of group A, and this finding suggests that it is not simply access to working with the simulator that is important. Rather, it is the metric-dependent PBP curriculum coupled with the simulator that optimizes the effectiveness of the training. The findings of this investigation strongly support the "outcomes" (objective assessments)−based approach rather than the "process" (time spent/exposure gained)−based approach to graduate medical education advocated by the Institute of Medicine.[31]

Because of the artificial constraint of a limited period for training (a weekend course), all members of group C were allowed to progress and practice with the cadaver and participate in the final Bankart assessment whether they met all of the intermediate benchmarks or not. During the training process, 25% of the subjects in group C failed to show proficiency during the knot-tying phase or on the shoulder simulation model (1 subject failed both). These individual performances

Table 8. Regression Analysis of Time Taken

Model	Unstandardized Coefficients		Standardized Coefficients		
	B	SE	β	t	P Value
Reference group: group C (constant)	80.438	4.455		18.055	< .001
Group A	0.634	6.522	0.016	0.097	.923
Group B	−13.009	6.522	−0.331	−1.995	.053

B, beta; β, standardized beta; SE, standard error; t, test statistic.

are, thus, not representative of the PBP protocol and diminish the performance of group C as a whole. Had these 4 subjects had additional time and further opportunities to show proficiency, it is likely that they may have been able to do so.

Overall, confidence in the observed effects of the training methods was high. There were no statistically significant differences between the groups on pre-course, baseline visuospatial, perceptual, or psychomotor assessments. Furthermore, the blinded and objectively assessed videotaped performances of the subjects' final Bankart repairs were scored with a consistently high IRR (>90% agreement between the raters for all assessments, with none falling below 80%).

Performance Outliers

For the 2 group A subjects who effectively completed only a 1-anchor Bankart repair, it is probable that they would have enacted substantially more errors during a complete 3-anchor procedure. It was not possible to accurately estimate the total number of errors that would have been enacted for a full repair. Consequently, all of their data (including steps completed and errors enacted) had to be excluded from the statistical analysis. The necessary exclusion of these 2 subjects' data results in an overestimation of the performance of group A as a whole. The 1 group C subject whose repair was considered incomplete abandoned the effort on the first anchor repair because this subject was unable to complete the knot-tying steps. This subject did, however, complete the second and third anchor components of the repair and, along with performing 37 steps, made only 4 errors and no sentinel errors. Thus the subject's overall performance was not substantially different from group C as a whole.

PBP Superiority

An important finding from this study is that the training process must be more than an educational experience. Simple knowledge of the metrics (steps and errors) and the opportunity to practice with expert feedback (group A, traditional AANA Resident Course) resulted in an inferior demonstration of arthroscopic Bankart skills. Furthermore, the addition of the opportunity to work with the simulator (group B) resulted in a modest improvement in performance over the control (group A). The use of the metric-dependent PBP curriculum coupled with the simulator (group C) resulted in the acquisition of a statistically superior ABR skill set. Multiple potential reasons exist for the superiority of this protocol:

1. The requirement to obtain a passing score on the cognitive examination at the outset ensures that the trainee is very familiar with the steps to be completed and errors to be avoided for the reference repair.
2. Proximate feedback linked to established and validated metrics facilitates the prompt, specific, and effective correction of errors.
3. Deliberate practice[32] in attempting to mimic the specific skills demonstrated in the orientation videos ensures uniformity in acquiring the essential skills needed to perform the reference procedure.
4. The medium-fidelity simulator provides the opportunity for practice and repetition of the important skills necessary for effective performance of a Bankart repair.
5. The validated performance benchmark for the simulator serves as an intermediate assessment tool and helps identify individual trainee deficiencies requiring correction.
6. The trainee's knowledge of the requirement to show proficiency by meeting each of the intermediate benchmarks to be able to progress in training helps the trainee focus on acquiring the necessary skills.
7. Trainee performance at a quality-assured performance level based on validated metrics must be shown.[7,8,33,34]

Fig 9. Odds ratios and statistical significance of differences between group A and groups B, C, and C^PBP for final Bankart proficiency demonstration. (PBP, proficiency-based progression.)
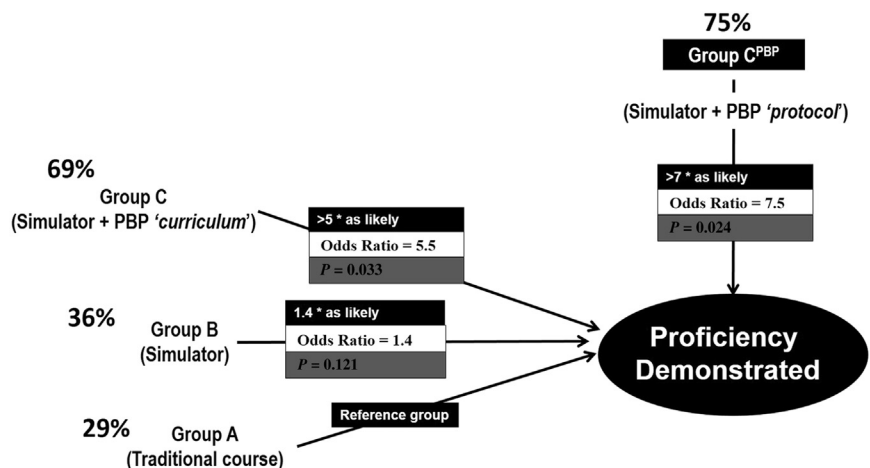
Metric-Based Curriculum

There are a number of unique aspects to this study, and they primarily relate to the development and use of the procedural metrics. This is the first simulation-based study, to our knowledge, in which the metric characterization and validation of a complete procedure have been carried out. This effort sought to investigate the merits of the emerging paradigm shift in surgical skills training from the apprenticeship model to a PBP format and, consequently, became known as the AANA Copernicus Initiative (Nicolaus Copernicus is credited with the paradigm shift from the earth to the sun being considered the center of the universe). In the approach reported on in this study, the simulation was simply one of the vehicles for the delivery of a metric-based training curriculum. Much of the effort focused on the development and validation of metric-based performance characteristics that appropriately captured a reference approach to the performance of an ABR.[21] Face validity and content validity for the Bankart metrics were verified using a modified Delphi panel meeting with Master and Associate Master AANA shoulder faculty.[21] The construct validity of the metrics coupled with the shoulder simulation model[22] and separately with a cadaveric shoulder[23] was confirmed. On the basis of these results, specific performance benchmarks were established separately for the shoulder model and for the cadaveric specimens.

The approach to the assessment of performance in this study uses precise metric definitions of performance and simply requires the scorer to determine whether a specific event did or did not occur. This binary approach to the measurement of performance has been shown to facilitate the reliable scoring of metric-based performance units across a variety of functions during skills training[10,24,35,36] with different experience levels.[37,38] In contrast, Likert-scale assessments (Table 1) result in a less focused approach to minimizing errors because the deviations from optimal performance are less clearly defined.[39] A Likert-type scale is a method of ascribing a quantitative value to qualitative data to make it amenable to statistical analysis and was originally designed to assess a range of respondent attitudes.[40] Because of the inherent subjectivity in this method of attempting to rate objective performance, it can be difficult to obtain acceptable levels of IRR (>80%) in the scoring of events.[18] It has been shown that Likert-scale scoring may be less reliable than metric-based assessments[41] and simply gives the trainee feedback information on the global aspects of his or her performance.

Simulation Platforms

The detailed metrics enabled a simulation platform that already existed (an anatomically accurate shoulder model) to be used for training and assessment. This medium-fidelity platform is relatively inexpensive, is readily available, and serves as an accurate representation of the human shoulder joint. One shortcoming of a physical simulator, however, is that it is unable to capture any performance data or provide feedback to the trainee. A significant investment in time and effort for instructional faculty and video reviewers was required to obtain detailed data and formulate accurate performance assessments. The approach used for simulation-based training in this study, nevertheless, holds considerable promise in the short-term because the vast majority of surgical procedures (particularly for traditional open surgery) have no virtual reality platform. Relative to higher-fidelity computer-based simulators, physical simulation models are much easier to develop in an expedient manner. This capability affords the surgical community the opportunity to develop PBP simulation-based training programs in a reasonable amount of time for traditional and new surgical procedures. The crucial element in terms of the effectiveness of any simulator will be its coupling with appropriate and accurate metric-based characterizations and the "operational definitions" (Table 1) for those metrics.

Studies assessing the value of simulation in surgical skills education and training have begun to emerge. We were unable to find any studies in the literature with which to compare our investigation. Frank et al.[42] performed a systematic review of the published literature (19 studies) on modern arthroscopic simulator training models. The analysis suggested that practice on arthroscopic simulators improves performance on the simulators, but evidence that skills obtained during simulator training are transferred to the operating room is lacking. Cannon et al.[43] studied the impact on transfer of training using the ArthroSim (TolTech, Aurora, CO) virtual reality arthroscopic knee simulator, which has previously been shown to have construct validity.[44] The PGY 3 orthopaedic residents who trained on the simulator (for an average of 11 hours) showed greater proficiency on a live diagnostic knee arthroscopy than the control group trained in a traditional fashion. They performed significantly better on the procedural checklist and assessment of probing skills but not on the assessment of visualization skills.

Implications

For well over a century, the apprenticeship model has been the predominant method used to assist surgical trainees in skill acquisition and preparation for the practice of surgery. A paradigm of repeated observation in addition to graded, enhanced responsibility and independence during operations of increasing technical complexity has been used. Although reasonably effective, this approach is inefficient and produces

considerable variability in the skill sets obtained by trainees with equivalent time exposure and experience.[5] Alternatively, PBP training using simulation enables trainees to focus on the acquisition of specific procedural skills, measure their progress, and correct deficiencies. It is through the process of deliberate practice[6] that they learn not only what to do but, perhaps more importantly, what not to do. The trainee is thus able to enact errors and learn to correct them in an inconsequential manner and without risk to patients. In addition, this structured approach promotes the acquisition of a more homogeneous skill set at the completion of training.[33]

One of the concerns of the medical community prior to this study was the generalizability of simulation-based training. Even simulation enthusiasts harbor the concern that simulation-based training effectiveness may be, in part, a function of the effort that enthusiasts put into the training initiatives and the reported science. The results of this study showed that simulation-based training is very effective, even when applied across a large number of residents from training programs throughout the United States using faculty equally dispersed. One of the reasons for this success was our deliberate choice of a reference approach to a particular procedure. This method of standardization means that, at the outset of their learning a particular procedure, trainees do not have a myriad of approaches and techniques to master. They can develop their own surgical style once they have acquired safe operative skills for the reference approach.

Limitations

Other than their year in training, no additional information regarding the participants was available. The extent of an individual resident's arthroscopic surgery exposure and experience, particularly shoulder arthroscopy, was likely variable. The diverse group of residency programs represented (21) and the fact that residents from an individual program were randomized to different training protocols should have minimized any selection bias. Although the residents in group A performed the best overall on the baseline visuospatial, perceptual, and psychomotor assessments, the between-group differences were not statistically significant.

The number of participants enrolled was based, in part, on previous studies (e.g., Seymour et al.[6]) and secondarily on the logistic challenges of having more than 12 simultaneous recording stations at the OLC during the final Bankart repair. Because there were more than 12 participants in each of the groups, the residents were randomly assigned to flights to complete their videotaped repair. The 14 group A residents were volunteers from a normal AANA Resident Course, and their Bankart recording was performed over 2 flights.

Training for the 14 group B and 16 group C participants was conducted on a separate study weekend. The final Bankart repair for these 2 groups required 3 flights of surgical procedures, some of which lasted 2 hours. With turnover, 8 hours were required to record the 3 flights (30 total recordings).

It is acknowledged that conducting an analysis of PBP training during the finite time period of a weekend course imposes an artificial constraint. PBP training dictates that the trainee continue to study and practice as long as it takes to master the requisite skills. It is probable that most of the participants in group C who did not reach all of the intermediate benchmarks during the weekend course would likely have been able to do so with additional training and practice. Furthermore, it was not the intent of this investigation to study the efficiency of individual resident's skill acquisition or residents' efficiency in performing the index procedure. The efficiency of the PBP training protocol, however, is clearly implied. In essentially the same time frame as the other groups, a substantially greater percentage of residents trained using the PBP protocol coupled with simulation achieved the final benchmark compared with those participating in the other training methods.

The final proficiency assessment was performed using cadaveric shoulders. The specimen size, tissue compliance, and extent of pre-existing glenohumeral pathology likely introduced some inherent variability in the cadaveric shoulders. The impact of these differences was minimized by using the previously described acceptability criteria: Those specimens deemed unsuitable for an ABR were replaced, and a new video recording was initiated.

Finally, all participants and faculty for each of the 3 study groups were provided a link to 2 full-length videos (lateral decubitus and beach chair) demonstrating each of the 45 steps and showing or identifying all of the errors and sentinel errors. Although the link was provided to all participants and faculty 1 month before the course in which they were involved, we have no record of how often they viewed or studied the videos.

Conclusions

A PBP training curriculum and protocol coupled with the use of a shoulder model simulator and previously validated metrics produces a superior arthroscopic Bankart skill set when compared with traditional and simulator-enhanced training methods.

References

1. Barden CB, Specht MC, McCarter MD, et al. Effects of limited work hours on surgical training. *J Am Coll Surg* 2002;195:531-538.
2. Pickersgill T. The European working time directive for doctors in training. *BMJ* 2001;323:1266.

3. Satava RM. Virtual reality surgical simulator. The first steps. *Surg Endosc* 1993;7:203-205.

4. 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.

5. Pellegrini CA, Sachdeva AK, Johnson KA. Accreditation of education institutes by the American College of Surgeons: A new program following an old tradition. *Bull Am Coll Surg* 2006;91:8-12.

6. 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;236:458-463.

7. 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.

8. Gallagher AG, O'Sullivan GC. *Fundamentals of surgical simulation—Principles and practices*. London: Springer Verlag, 2011.

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. 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.

12. Petrera M, Dwyer T, Tsuji MR, Theodoropoulos JS. Outcomes of arthroscopic Bankart repair in collision versus noncollision athletes. *Orthopedics* 2013;36:e621-e626.

13. Harris JD, Bupta AK, Mall NA, et al. Long-term outcomes after Bankart shoulder stabilization. *Arthroscopy* 2013;29:920-933.

14. Archetti Netto N, Tamaoki MJ, Lenza M, et al. Treatment of Bankart lesions in traumatic anterior instability of the shoulder: A randomized controlled trial comparing arthroscopy and open techniques. *Arthroscopy* 2012;28:900-908.

15. Chahal J, Marks PH, Macdonald PB, et al. Anatomic Bankart repair compared with nonoperative treatment and/or arthroscopic lavage for first-time traumatic shoulder dislocation. *Arthroscopy* 2012;28:565-575.

16. Stein T, Linke RD, Buckup J, et al. Shoulder sport-specific impairments after arthroscopic Bankart repair: A prospective longitudinal assessment. *Am J Sports Med* 2011;39:2404-2414.

17. Ee GW, Mohamed S, Tan AH. Long term results of arthroscopic Bankart repair for traumatic anterior shoulder instability. *J Orthop Surg Res* 2011;6:28.

18. Angelo R. Arthroscopic Bankart repair for unidirectional shoulder instability. *Instr Course Lect* 2009;58:305-313.

19. Ryu RK. Arthroscopic approach to traumatic anterior shoulder instability. *Arthroscopy* 2003;19:94-101.

20. Smith R. All changed, changed utterly. *BMJ* 1998;316:1917-1918.

21. 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.

22. Angelo RL, Pedowitz RA, Ryu RKN, Gallagher AG. The Bankart performance metrics combined with a shoulder model simulator create a precise and accurate training tool for measuring surgeon skill. *Arthroscopy* 2015;31:1639-1654.

23. Angelo RL, Pedowitz RA, Ryu RKN, Gallagher AG. The Bankart performance metrics combined with a cadaveric shoulder create a precise and accurate assessment tool for measuring surgeon skill. *Arthroscopy* 2015;31:1655-1670.

24. McClusky DA III, Ritter EM, Lederman AB, et al. Correlation between perceptual, visuo-spatial, and psychomotor aptitude to duration of training required to reach performance goals on the MIST-VR surgical simulator. *Am Surg* 2005;71:13-20.

25. Gallagher AG, Cowie R, Crothers I, et al. PicSOr: An objective test of perceptual skill that predicts laparoscopic technical skill in three initial studies of laparoscopic performance. *Surg Endosc* 2003;17:1468-1471.

26. Gallagher AG, McClure N, McGuigan J, et al. An ergonomic analysis of the fulcrum effect in the acquisition of endoscopic skills. *Endoscopy* 1998;30:617-620.

27. Crothers IR, Gallagher AG, McClure N, et al. Experienced laparoscopic surgeons are automated to the "fulcrum effect": An ergonomic demonstration. *Endoscopy* 1999;31:365-369.

28. Gallagher AG, Smith CD, Bowers SP, et al. Psychomotor skills assessment in practicing surgeons experienced in performing advanced laparoscopic procedures. *J Am Coll Surg* 2003;197:479-488.

29. Randomization.com Web site. Available at http://www.randomization.com. Accessed November 23, 2013.

30. Pedowitz RA, Nicandri GT, Angelo RL, et al. Objective assessment of knot-tying proficiency with the fundamentals of arthroscopic surgery training program workstation and knot tester. *Arthroscopy* 2015;31:1872-1879

31. Asch DA, Weinstein DA. Innovation in medical education. *N Engl J Med* 2014;371:794-795.

32. Ericsson KA, Krampe RT, Tesch-Römer C. The role of deliberate practice in the acquisition of expert performance. *Psychol Rev* 1993;100:363-406.

33. 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.

34. 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.

35. Van Sickle K, Smith B, McClusky DA III, et al. Evaluation of a tensiometer to provide objective feedback in knot-tying performance. *Am Surg* 2005;71:1018-1023.

36. Van Sickle K, Gallagher AG, Smith CD. The effect of escalating feedback on the acquisition of psychomotor skills for laparoscopy. *Surg Endosc* 2007;21:220-224.

37. Neary PC, Boyle E, Delaney CP, et al. Construct validation of a novel hybrid virtual-reality simulator for training and assessing laparoscopic colectomy; results from the first course for experienced senior laparoscopic surgeons. *Surg Endosc* 2008;22:2301-2309.

38. Nicholson WJ, Cates CU, Patel AD, et al. Face and content validation of virtual reality simulation for carotid angiography: Results from the first 100 physicians attending the Emory NeuroAnatomy Carotid Training (ENACT) program. *Simul Healthc* 2006;1:147-150.

39. 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.

40. Likert R. A technique for the measurement of attitudes. *Arch Psychol* 1932;140:44-53.

41. Gallagher AG, O'Sullivan GC, Leonard G, Bunting BP, McGlade KJ. Objective structured assessment of technical skills and checklist scales reliability compared for high stakes assessments. *ANZ J Surg* 2014;84:568-573.

42. Frank RM, Erickson B, Frank JM, et al. Utility of modern simulator training models. *Arthroscopy* 2014;30: 121-133.

43. Cannon WD, Garrett WE Jr, Hunter RE, et al. Improving residency training in arthroscopic knee surgery with use of a virtual-reality simulator. A randomized blinded study. *J Bone Joint Surg Am* 2014;96:1798-1806.

44. Cannon WD, Nicandri GT, Reinig K, et al. Evaluation of skill level between trainees and community orthopaedic surgeons using a virtual reality arthroscopic knee simulator. *J Bone Joint Surg Am* 2014;96:e57.

Keep in Touch With Your Journal...

Arthroscopyjournal

Like us on Facebook

@ArthroscopyJ

Follow us on Twitter