The intent of any surgical training program, both for residents and for established surgeons acquiring a new procedural skill, is to enable the trainee to acquire the requisite skill sets necessary to perform the designated surgical procedure well and safely. To accomplish this mission, a clearly defined endpoint or set of skill proficiencies must be identified. Furthermore, it must be verified that mastery of those skill sets can accurately be measured during the trainee’s progress. It must also be confirmed that the acquisition of those skills is predictive of the ability to perform an effective surgical procedure. Many experienced surgeons who are “proficient” in the performance of a specific procedure and are able to perform it well are also able to identify and agree on the essential “steps” to be completed, as well as the “errors” to be avoided, for that procedure. The reader is referred to Table 1 for a glossary of terms used throughout the article. One challenge, however, in identifying those key features is that surgeons rarely think about the procedures they perform with this level of detail. Surgeons who are proficient in the performance of a specific operation will exhibit many if not all of the important “performance characteristics” (Table 1) that contribute to actually performing the procedure well. They may, however, have automated to many of these steps and how they are performed and, as a consequence, may be less cognizant of the details and more granular elements of the techniques they use.1-3
The units of performance that constitute skill can be elucidated with a “task analysis” (Table 1), or breakdown and detailed description of the steps or actions necessary to perform the procedure. In attempting to characterize specific skills, psychologists have subjected them to a detailed task analysis and then “operationally defined” (Table 1), rather than simply described, the resulting steps. A definition specifies the order, duration, and result of the specific action and provides precise parameters such that it can be unambiguously determined whether that specific event did or did not occur. A “description” (Table 1), on the other hand, only offers a general characterization of an event or behavior in qualitative terms. Definitions are the preferred foundations of measurement science. The definitions, or “metrics” (Table 1), for a specific...
procedure provide a quantitative standard of measurement that can be used to objectively assess performance. These metrics must then be validated with respect to whether their characterization fits with what is known about the skill being analyzed. The task analysis–derived characterizations, or “metric units” (Table 1), of skilled performance do not have to capture every aspect of performance but should at least allow for ordinal differentiation between different levels of performance as described by Dreyfus and Dreyfus.4 The metrics created from this analysis can serve as a tool to evaluate the effectiveness of different training protocols for a particular surgical procedure.

“Face validity” (Table 1) is verified by the opinion of an experienced panel that reviews the content of an assessment or tool to determine if it is appropriate and relevant to the concept it purports to measure. “Content validity” (Table 1) of a testing instrument is similarly obtained and based on the opinion of an experienced panel that performs a detailed examination of the contents of the test items. Thus the face validity and content validity of tools assessing procedural skill are not verified by statistical analysis but, rather, by the summary opinion of an experienced panel of surgeons. An additional question that relates to establishing the validity of the metric definitions for a particular procedure asks, “Do more skilled individuals perform better on the defined metrics than less skilled or experienced individuals, and do the specific metrics identify the quality, ability, or trait they were designed to measure (‘construct validity’ [Table 1])?” In contrast to face and content validity, the establishment of construct validity requires sufficient data and statistical analysis to prove that it exists.

Task analysis for a particular operation should be performed initially for a “reference procedure” (Table 1)5–7—one that is straightforward with a generally accepted or agreed on method that is uncomplicated under ideal circumstances. An optimal approach to learning should ensure that trainees are capable of performing a routine procedure before they have to deal with the technique variations necessary to address more complex pathology.

We sought to study the effectiveness of “proficiency-based progression” (Table 1) training plus simulation for the acquisition of surgical skills. Proficiency-based progression dictates that the learner must demonstrate the ability to meet specific performance benchmarks before he or she is permitted to progress in training. This investigation required the development and validation of specific tools to conduct the analysis. The first component needed was a “metric tool” (Table 1) that could objectively and accurately characterize an arthroscopic Bankart repair. The development of this tool is the focus of this study. Future investigations will report on the establishment of additional tools.

The purpose of this study was to establish the metrics (operational definitions) necessary to characterize a reference arthroscopic Bankart procedure and to seek consensus from experienced shoulder arthroscopists on the appropriateness of the steps, as well as errors identified. The null hypothesis was that face and content validity for the step and error metrics derived from task deconstruction of an arthroscopic Bankart procedure would not be demonstrated.

### Methods

**Arthroscopic Bankart Metric Development**

Three experienced arthroscopic shoulder surgeons (R.L.A., R.K.N.R., R.A.P.), each with over 25 years of clinical practice, and an experimental psychologist (A.G.G.) formed the Metrics Group that characterized an arthroscopic Bankart repair. A detailed task analysis and deconstruction process (described in detail elsewhere)8 was used to identify the units of performance that are integral to the skilled performance of the instability repair. The goal was to characterize a “reference” arthroscopic Bankart repair and not one attempting to manage unusual or complex instability pathology. Procedure performance characterization (task deconstruction) was guided by (1) decades of practice and teaching experience by the Metrics Group, (2) published studies on arthroscopic Bankart repair,9,10 and (3) manufacturer guidelines on device use. Two 2.5-day face-to-face meetings and eight 1.5- to 2-hour online conferences were conducted, along with countless E-mail exchanges, to craft the procedural metrics. For the online sessions, the use of Skype videoconferencing (Microsoft, Redmond, WA; available at www.skype.com) enabled the investigators (R.L.A., R.K.N.R., R.A.P., A.G.G.; who reside in different geographic locations) to simultaneously review arthroscopic videos in real time with acceptable resolution. One investigator initiated a standard Skype video connection for a group call using a laptop computer. A second computer (desktop) with a high-resolution screen was used to play the arthroscopic video being studied. An independent USB camera (Ipevo, Sunnyvale, CA) was connected to the USB port of the laptop to which the Skype video input was directed, instead of the resident camera on the laptop screen (“settings” tab in Skype). Thus all of the members on the group Skype call viewed the arthroscopic image rather than the call initiator’s image.

Fourteen video recordings of a complete in vivo Bankart procedure, performed by surgeons with varying levels of experience (Table 2), were reviewed by the Metrics Group in detail to assist in the creation and stress testing of the metrics. The videos represented surgeons with practice experience ranging from 3 to 33 years. Both the lateral decubitus (n = 10) and
potential procedural step, metrics were also created to identify groups of related steps. In addition to specifying each step was further defined by identifying beginning points and endpoints during the procedure for that metric. The aim was that these detailed metric units would accurately capture the essence of procedure performance, as well as serve as a sound and comprehensive training guide for persons learning the procedure. The metrics included the specific operative steps, the general order in which they should be accomplished, and the instruments and the manner in which they should be used. “Procedural phases” (Table 1) were specified for groups of related steps. In addition to specifying each procedural step, metrics were also created to identify potential “errors” (Table 1), or actions that deviate from optimal performance and should not be done. The intent, again, was to create unambiguous operational definitions (rather than descriptions) for each metric error. A special designation was made for more serious, or “sentinel,” errors defined by events that, by themselves, could either (1) jeopardize the outcome of the procedure or (2) lead to significant iatrogenic damage to the shoulder joint. An additional error characterization was termed “damage to non-target tissue.” This occurrence defined an event that was injurious to tissues not intentionally being addressed during the defined task, such as “scuffing of articular cartilage by an instrument” or “lacerating the intact labrum.”

By agreed on convention, an event (step or error) must have been observed on the video to be scored. Thus inference that an event was “likely to have occurred” was eliminated. For example, if comparable views of the anterior humeral head showed relatively healthy or pristine articular cartilage early in the procedure, with scuffing and abrasion later during the repair, but the injurious event was not observed on the video, it was not scored as an error (or damage to non-target tissue).

### Metric Stress Testing and Reliability of Identification

After the 4 members of the Metrics Group were satisfied that the entirety of the procedure had been well characterized, they “stress tested” (Table 1) the metrics by subjecting them to a robust assessment of how reliably they could be independently scored in blinded fashion. Eight video recordings of complete arthroscopic Bankart procedures that were performed by surgeons possessing a wide range of technical skill were independently reviewed and scored. Both the lateral decubitus and beach-chair orientations were represented by the videos studied. Each metric was scored in binary fashion as either yes or no (occurring or not occurring). After each video review, differences in the scoring of each metric by the reviewers (R.L.A., R.K.N.R., R.A.P.) were compared and discussed. Where necessary, operational definitions were clarified, modified, or dropped and new ones added to optimize the functionality of the characterizations as a whole. This process of independent viewing, scoring, and revising the step and error metrics was continued until the Metrics Group was satisfied that the metrics accurately and unambiguously characterized the specifics of an arthroscopic Bankart procedure and could be “reliably identified” (Table 1) by independent reviewers. The extent of agreement between 2 raters for the entire group of step and error Bankart metrics could potentially range between 0, no agreement, and 1.0, complete agreement.

### Face and Content Validation of Bankart Metrics by Modified Delphi Panel

The Delphi Panel method (Table 1) is a process that provides an interactive communication structure between researchers (i.e., the Metric Group authors) and an experienced panel (as described later) in a field or discipline to provide systematic feedback on a given topic (i.e., the accuracy of the metrics developed for a reference approach to a Bankart procedure). The Delphi method uses an “iterative process” (Table 1) for progressing toward a desired result by means of repeated cycles of deliberations. The iterative process should be convergent, that is, it should come closer to the desired result as the number of iterations or cycles of review increases. For the Bankart characterization, the desired result (consensus on the appropriateness of a particular metric) was obtained by means of repeated

### Table 2. Source Videos for Bankart Metric Creation and Stress Testing

<table>
<thead>
<tr>
<th>Video No.</th>
<th>Surgeon Time in Practice, yr</th>
<th>Patient Orientation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>LD</td>
</tr>
<tr>
<td>2</td>
<td>17</td>
<td>LD</td>
</tr>
<tr>
<td>3</td>
<td>25</td>
<td>LD</td>
</tr>
<tr>
<td>4</td>
<td>26</td>
<td>LD</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>LD</td>
</tr>
<tr>
<td>6</td>
<td>17</td>
<td>BC</td>
</tr>
<tr>
<td>7</td>
<td>18</td>
<td>BC</td>
</tr>
<tr>
<td>8</td>
<td>26</td>
<td>LD</td>
</tr>
<tr>
<td>9</td>
<td>28</td>
<td>LD</td>
</tr>
<tr>
<td>10</td>
<td>21</td>
<td>LD</td>
</tr>
<tr>
<td>11</td>
<td>24</td>
<td>LD</td>
</tr>
<tr>
<td>12</td>
<td>25</td>
<td>LD</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>BC</td>
</tr>
<tr>
<td>14</td>
<td>4</td>
<td>BC</td>
</tr>
</tbody>
</table>

BC, beach chair; LD, lateral decubitus.
cycles of questioning, deliberation, metric modification, and voting on the appropriateness of each refined metric definition. The methodology assumes that good-quality knowledge evolves from the process. The Delphi method was modified to the extent that the voting cycles, with each new iteration, were not anonymous.

The determination of face and content validity for the Bankart characterization was made by subjecting each metric to an appraisal by a group of surgeons who were very experienced in the performance of an arthroscopic Bankart repair. Twenty-seven board-certified orthopaedic surgeons (the 3 Metrics Group surgeons and 24 additional Arthroscopy Association of North America shoulder faculty instructors) with an average of over 23 years in clinical practice involving shoulder arthroscopy attended a Delphi Panel. Four of the panelists are full-time academicians, 9 are in private group practice and have direct involvement in teaching fellows, and 14 are in private practice with a clinical affiliation with a university orthopaedic department. Each member of the Delphi Panel is a master or associate master faculty member of the Arthroscopy Association of North America and has taught the technique for arthroscopic Bankart repair during shoulder courses conducted at the Orthopedic Learning Center (Rosemont, IL). An experimental psychologist facilitated the meeting.

**Delphi Panel Procedure**

An overview of the project and meeting objectives was presented. Background information regarding proficiency-based progression training, prior literature showing the validity of this training approach for procedural specialties, and the specific objectives of the current Delphi Panel were reviewed. It was explained that the Bankart metrics had been developed by the Metrics Group for a reference approach to arthroscopic anterior shoulder stabilization for unidirectional anterior glenohumeral instability. It was acknowledged that the designated reference procedure might not accurately embodied the essential and key components of the procedure. An affirmative vote by a panel member indicated that the metric definition presented was accurate and acceptable as written but not necessarily that it was the manner in which that particular panelist might have chosen to complete the step. “Consensus” meant that there was unanimity in voting among the panelists and that a particular metric definition was “not wrong or inappropriate.” Each of the procedural steps and potential errors were evaluated individually. After each metric definition was presented, panel members voted on whether or not the metric was acceptable as written. If the panel could not achieve consensus because of lack of clarity or differences in opinion, the metric definition was revised accordingly and a new vote conducted on the acceptability of the modified metric. This process was repeated until the metric was accepted. If consensus could not be achieved through a series of modifications, the metric was deleted. When it was deemed necessary, a new metric was defined and added.

**Results**

**Bankart Procedure Metrics**

The step metrics resulting from task deconstruction were grouped into 13 separate phases of the procedure (in Roman numerals). Each phase (e.g., “arthroscopic instability assessment” or “inferior anchor preparation/insertion”) contains a series of related, unambiguously defined, observable procedure events (steps) with specific beginning and ending points. All potential errors identified had been noted to occur during the stress testing of the metrics. Some of the identical errors and sentinel errors could occur during different phases of the procedure that recurred during the 3-anchor repair (e.g., “uncorrected entanglement of shuttling device or suture”).

**Modified Delphi Panel**

All phases of the procedure were accepted as identified. Only a minority of procedure phases and their associated metrics were accepted without discussion. At the conclusion of the deliberations, consensus among the Delphi Panel was reached for 45 steps, 77 errors (29 unique), and 20 sentinel errors (8 unique) (Tables 3 and 4). During the panel deliberations, 2 metrics were deleted, 1 was added, and 5 were modified before consensus was achieved (Table 5).

**Summary of Points Raised and Voting Outcomes of Bankart Delphi Panel**

The minutes of the metric validation meeting (Copernicus Study/Delphi Panel), held November 18, 2011, are presented herein (recorded by Robert Pedowitz, M.D., Ph.D.). The meeting chair was Rick Angelo, M.D. The Project Leadership Team comprised Rick Angelo, Rick Ryu, Rob Pedowitz, and Tony Gallagher. The attendees comprised R. Angelo, R. Ryu, R. Pedowitz, J. Tokish, R. Bell, R. Hunter, K. Nord, V. Goradia, A. Barber, S. Snyder, B. Beach, J. Abrams, B. Shaffer, J. Tauro, L. Higgins, S. Weber, S. Koo, D. Richards, J. Esch, J. Dodds, J. Randle, J. Richmond, A. Curtis, J. Burns, N. Saggionlo, J. Kelly, and S. Powell (27 voting attendees), as well as T. Gallagher (meeting facilitator) The meeting overview is as follows:

1. Dr. Angelo presented a brief overview of the project and meeting objectives.
2. Dr. Gallagher presented the background of proficiency-based training, as well as some prior literature demonstrating the validity of this training
Table 3. Thirteen Phases of Arthroscopic 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 midsurface 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 anchor security by pulling on 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 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
approach for procedural specialties, and he explained the specific objectives of the current Delphi meeting.

3. Dr. Angelo presented each procedural step and explained the associated metrics that have been developed by the Metrics Group for a reference approach to anterior shoulder stabilization for glenohumeral instability. The comments and recommendations for each of the steps, with associated vote, are presented in Table 6.

### Discussion

The principal findings of this study are that (1) an arthroscopic Bankart procedure can be deconstructed

---

**Table 5. Delphi Panel Metric Changes**

<table>
<thead>
<tr>
<th>Modification</th>
<th>Issue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deleted (2)</td>
<td>Should failure of anchor purchase remain an error? Alternating posts for knot tying</td>
</tr>
<tr>
<td>Added (1)</td>
<td>Completed knot position</td>
</tr>
<tr>
<td>Modified (5)</td>
<td>Diagnostic steps—probe or view (how long?) Adequacy of capsular mobilization Whether to ascribe “critical” to “laceration of labrum” Consider deleting the term “sliding” from knot description Should diagnostic steps be included in the procedure metrics?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Deliberation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cadaveric bone variability too great to score accurately Arthroscopic views not consistent enough to score reliably May cause iatrogenic damage if it abuts articular cartilage Does “looking” equal “ascertaining”? Should demonstrate effort at capsular mobility Labral variability too great in cadavers; can still see violation of “hoop” integrity Sliding knot would be acceptable for a reference procedure May skew results if there is excessive influence of diagnostic steps in procedure</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resolution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Delete error Delete error Add error to each of the 3 knots tied Add error to each of the 3 knots tied View must be held long enough to determine pathology Must take instrument (grasper) to demonstrate Retain as a sentinel error Retain the term “sliding” Include only steps directly related to instability assessment</td>
</tr>
</tbody>
</table>
into the essential steps necessary for the effective completion of the repair, (2) the potential errors related to the procedure are able to be identified and characterized, and (3) face and content validity for the resulting step and error metrics can be obtained through use of the modified Delphi Panel technique. Traditionally, surgeons have been trained using the “apprenticeship” model, which is related to “process” or time, based (a certain variety of rotations, exposure to numbers of specific cases, etc.). A paradigm shift toward proficiency-based progression training, which is outcomes based, is occurring and mandates that the trainee be able to demonstrate the ability to meet specific skills benchmarks to progress in training. These benchmarks must have specific, clear, objective, and fair standards of performance. Validated metrics will be essential in defining these standards. In addition, as the move toward including surgical skills credentialing and procedural competency occurs for licensing, the same validated standards will be needed. The methodology used in this study provides a framework for the development of those metrics and standards.

An arthroscopic Bankart repair (index procedure) was selected as the reference surgical procedure to
study for several reasons. For the patient with unidi-
rectional anterior instability due primarily to a Bankart
lesion without significant bone loss, a suture anchor
repair with 3 implants is a commonly accepted method
used to obtain a successful patient outcome.13-16 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.17-21

The task analysis stage of metric development is
crucial because metrics are the fundamental building
blocks of a good training program. Metrics, thus, not
only define how the training should be characterized
and the procedure performed by the trainee but also
must afford the opportunity for meaningful assess-
ment of the trainee’s performance and progress. The
total process of metric development should be as
transparent, objective, and unambiguous as possible.
Metric definitions should be characterized in such a
way that they are sufficiently complete and detailed
for an individual—not associated with the initial
development process—to use them to score perfor-
ance reliably. Metric definitions should include
behavioral markers that indicate the beginning points
and endpoints of the performance characteristics
(steps) to be assessed. These parameters will become
particularly important in the future as the procedural
metrics are used with higher-fidelity simulators. The
details of the metric definitions will be necessary for
the simulator to be appropriately programmed to
provide the trainee with performance assessments and
accurate feedback.

Other approaches to the measurement of surgical
performance use qualitative descriptions of perfor-
ance and require the user to rate items on a gradu-
ated Likert-type scale (Table 1), which ascribes a
quantitative value to qualitative data to make them
amenable to statistical analysis. Likert scales (often
with a range from 1 to 5 or from 1 to 7) are typically
constructed with responses (opinion) around a neutral
option (e.g., “suture delivery was 1, awkward, . . . 3,
effective, . . . 5, highly efficient”) and were originally
designed to assess a range of attitudes. Because of the
component of subjectivity, this method of attempting to
rate objective performance can render it difficult to
obtain acceptable levels of inter-rater reliability
(>80%) in the scoring of events. In contrast, the
approach to the assessment of performance in our
study uses precise definitions of performance and
simply requires the reviewer to report whether the
specific event occurred or not. This binary approach to
the measurement of individual events has been shown
to facilitate the reliable scoring of metric-based per-
formance units across a variety of functions during
skills training22-25 of individuals with different experi-
ence levels.26,27 This approach has also been shown to
be more reliable than Likert-scale scoring.28

Behaviors that deviate from optimal performance (er-
rors) can be characterized, including those of a “more
serious nature.” The issue of whether those more serious
errors should be termed “critical errors” or some alter-
native label was raised at the outset during the metric
definition process. It was agreed on by the Metrics Group
that use of the term “critical error” could imply that the
event was life-threatening or might have serious medi-
colegal implications. It was elected, instead, to use the
term “sentinel” (Table 1) to connote an error that should
be carefully “watched for and to avoid.” Sentinel errors
involve a serious deviation from optimal performance
during a procedure because they can either jeopardize
the success/desired result of the procedure or create
iatrogenic insult to the patient’s tissues. A single specific
sentinel error may not always lead to a poor outcome but
should stringently be avoided.6 The underlying philoso-
phy of this approach to errors is that suboptimal out-
comes do not happen by accident but usually result from
the coalescence of deviations from optimal procedure
performance.

The face and content validity of the metric-based
procedure characterization by subject specialists can
be verified using the modified Delphi Panel method-
ology reported in this study. The metrics developed
were informed by research studies, professional
guidelines, clinical experience, and manufacturers’
guidelines.6,7 Although the surgeons in the Metrics
Group are very experienced in the performance of the
Bankart procedure, the Delphi process provided an
excellent method to ensure that the procedure char-
acterization is appropriate, represents best practice,
and is acceptable to a larger group of experienced
master and associate master Bankart faculty. As
anticipated, many surgeons pointed out that they
might perform a specific step in a different manner but
that the approach outlined by the Metrics Group was
“not incorrect” or inadvisable. The members of the
panel made very helpful suggestions for improving the
definitions.

Assuming that the Bankart metric identifications and
definitions represent a real-world surgical procedure,
these performance characteristics should be able to
distinguish between experienced (skilled) surgeons and
novices, that is, provide construct validity. Future
studies regarding construct validity will seek to provide
information about which metrics best distinguish be-
tween experienced and novice surgeon performance.
This information will facilitate the establishment of a
benchmark to define the “proficiency level” that
trainees should acquire before progressing to in vivo
practice.6,29

**Limitations**

A limitation of this study resides in the fact that
every potential error, regardless of how rare the
occurrence, might not have been included. The Delphi Panel, however, confirmed that the errors listed were those most likely to occur and that should be avoided in the safe performance of an arthroscopic Bankart repair. Although common errors may be relatively easy to agree on, it is somewhat more challenging to decide which errors should be designated as “sentinel,” without a specific weighting methodology. Although the issue of using this designation for events that cause iatrogenic damage is more straightforward, the concept of also using the term for events that might “potentially lead to a suboptimal outcome” is more subject to the opinion of the Metrics Group and the Delphi Panel.

Furthermore, data are not available to confirm that the specific steps identified by the Metrics Group and the Delphi Panel directly correlate with a successful surgical outcome for patients with unidirectional shoulder instability. Therefore the metrics created remain predominantly based on the opinion of experienced surgeons and instructors. An outcomes study will be needed to fully establish the predictive validity for the Bankart metrics as authored. In addition, the surgeons comprising the Metrics Group and Delphi Panel were all North American surgeons. International arthroscopists may have created somewhat different metrics for an arthroscopic Bankart repair.

Conclusions

This study rejects the null hypothesis and confirms that a core group of experienced arthroscopic surgeons is able to perform task deconstruction of an arthroscopic Bankart repair and create unambiguous step and error definitions (metrics) that accurately characterize the essential components of the procedure. Analysis and revision by a larger panel of experienced arthroscopists were able to validate the Bankart metrics.

References


