A Reliability Study of Measurement Techniques to Determine Static Scapular Position

Mark H. Gibson, MS, PT, ATC[1,6]
Gerald V. Goebel, MS, PT[2,6]
Terry M. Jordan, MS, PT, ATC[3,6]
Sam Kegerreis, MS, PT, ATC[4]
Teddy W. Worrell, EdD, PT, SCS, ATC, FACSM[5]

Current shoulder rehabilitation programs encourage scapular stabilization components although, to date, no scientific studies have evaluated changes in scapular position following such rehabilitation. Four different measurement methods of scapular position have been reported in the literature. The purpose of this study was to examine the intratester and intertester reliability of these four methods and to also examine if significant differences exist in scapular position between dominant and nondominant extremities. Thirty-two subjects volunteered for this study. Intraclass correlation coefficients (ICC) revealed acceptable intratester reliability (ICC = 0.81–0.95) for all measurement methods. However, while one method also proved to be acceptable (ICC = 0.91–.92) for intertester measurements, the other three methods were unacceptable (ICC = 0.18–0.69). One tester reported significant differences in scapular position of the dominant and nondominant extremities when using the most reliable method. The second tester found no significant differences with either method. Future research is recommended to reexamine reliability of these methods and measure subjects with shoulder pathology.

Key Words: *scapular position, reliability, measurement*

[1] *Assistant Professor, Department of Exercise and Sport Science; Director, Athletic Training Program; 134 Mitchell Hall, University of Wisconsin-La Crosse, La Crosse, WI 54601*
[2] *Staff Physical Therapist, Health Care Therapy Services, Indianapolis, IN*
[3] *Staff Physical Therapist, Rebound Physical Therapy, Albany, OR*
[4] *Associate Professor, Physical Therapy, Krannert School of Physical Therapy, University of Indianapolis, Indianapolis, IN*
[5] *Assistant Professor, Director of Research, Krannert School of Physical Therapy, University of Indianapolis, Indianapolis, IN*
[6] *At the time of this study, Mr. Gibson, Mr. Goebel, and Mr. Jordan were physical therapy students at the Krannert School of Physical Therapy, University of Indianapolis, Indianapolis, IN.*

The shoulder girdle is an arrangement of three bones, four articulations, and numerous muscles which must interact in a highly coordinated and complex manner in order to elevate the arm and position the hand for use. Because the glenohumeral joint is highly mobile and inherently unstable, it is essential that the scapula and corresponding musculature perform several crucial functions. While some authors (3,12) state that normal scapular orientation at rest maintains the glenoid surface in an upward tilt which enhances passive stability of the glenohumeral joint, others (11,16) state that the resting position of the glenoid fossa is downward. During elevation of the arm, the scapula must follow the humeral head with a lateral or upward rotation in order to facilitate congruency of the glenohumeral joint and, thus, maximize stability. It is also imperative that the acromion process elevate and rotate in relation to the greater tuberosity of the humerus in order to avoid impingement of soft tissues in the subacromial space. In addition to demands for mobility, the scapula must also be capable of dynamic stability during arm elevation in order to afford the movers of the glenohumeral joint a stable base with which to position the arm.

An understanding of principles of normal scapular motion during arm elevation coupled with skilled observation of scapular position may assist the clinician in detecting and monitoring shoulder girdle pathologies. For example, premature scapular motion during humeral abduction may signify a restricted glenohumeral joint characteristic of adhesive capsulitis. However, even with the development of observational skill, clinicians do not presently have a simple and reliable means to objectively quantify scapular position during the course of arm elevation. Standard goniometric measurements, while helpful in quantifying restrictions of glenohumeral motion, do not make distinctions as to the relative contributions of shoulder girdle components to this motion.

Shoulder girdle biomechanics, including the relationship of humeral motion to scapular motion during elevation of the arm, have been well studied. Inman et al (13) used radiographs to quantify the ratio of glenohumeral movement to scapulothoracic movement occurring during humeral abduction. Many researchers have subsequently undertaken similar studies that have reported variations in the original findings (1,9,11,16, 17). Others have attempted to define scapular instant centers of rotation during arm elevation, and their findings have proven valuable in determining the contributions of different muscles to scapular movement (1,10, 16). Electromyographic studies have also greatly increased our understanding of relative muscular contributions (2,6,13,17). However, these studies have all utilized equipment for data collection that clinicians typically do not have readily available for such analysis.

Researchers recently have attempted to devise simple, practical, and reliable means of quantifying scapular position (7,14,15). DeVita et al (7) used a piece of string to measure linear distance between the spinous process of the third thoracic vertebrae and the inferior angle of the acromion process of the scapula. This study was limited to measures of the scapula when the arm was at rest at the side. Neiers and Worrell (15) replicated DeVita et al's measurement method and reported similar intratester reliability for scapular position but found the measures to normalize scapular abduction for subject size to be unreliable. Kibler (14) discussed measurement of "lateral slide" of the scapula in three different positions of humeral abduction. Kibler used a linear measurement from the nearest spinous process to the inferior angle of the scapula. No data regarding reliability of repeated measures using the Kibler methods were reported.

The DeVita et al and Kibler studies are the only studies which explain clinically practical measurement techniques to quantify scapular position. Before such measures can ever be considered valid in elucidating biomechanical abnormalities or pathology, they must first be proven reliable. Therefore, the purpose of this study was threefold: *1)* to determine the intratester reliability of the four aforementioned measurement methods, *2)* to determine the intertester reliability of these methods, and *3)* if acceptable reliability is established, to determine if differences exist between scapular position of the dominant vs. nondominant extremity.

METHODS

Subjects

Thirty-two healthy, nonathletic subjects (17 male, 15 female) without a history of previous pathology or surgery to either shoulder girdle participated in this study on a voluntary basis (age range = 22–34 years, $\bar{X}$ = 26.16 years; height range = 157.5–198.1 cm, $\bar{X}$ = 175.82 cm). Prior to participating in this study, all subjects read and signed an informed consent approved by the University of Indianapolis Committee on Research Involving Human Participants.

Instrumentation and Procedures

During the measurement procedure, five people were present in the examination room: two subjects, two testers, and one recorder. Subjects were required to have their vertebrae and scapulae exposed for visualization and palpation. Each tester palpated the cervical spinous processes of their subject to locate C7. This landmark was determined to be the first prominent spinous process while palpating caudally that remained consistently positioned when the subject extended the cervical spine. Next, the tester sequentially palpated the thoracic spinous processes to locate T3, and this landmark was marked with a ¼-inch adhesive square. The tester then continued to palpate caudally to locate the T8 spinous process and mark it in a similar manner.

Each tester conducted a measure of scapular position on the subject's dominant side as described by DeVita et al:

> "The subject then stood in a relaxed position for 30 seconds before the same examiner palpated the inferior angle of the acromion. The examiner then used an unmarked section of string to measure the distance from the inferior angle of the acromion to the spinous process of the third thoracic vertebrae. The outstretched piece of string was marked at the site of the palpated inferior angle of the acromion and at the tagged third thoracic spinous process. The linear distance from the third thoracic vertebra to the inferior angle of the acromion was defined as the total scapular distance" (7, p. 472).

The tester transferred the string to a metric rule fastened to a table, and the recorder read and recorded the measurement in cm so that the tester was blinded to the results. The tester then released the measure, returned to the subject, and repeated the measurement so that an average value could be calculated. DeVita et al also measured the length of the scapula from the inferior angle of the acromion to the medial border along the spine and then divided the scapular distance by this measure in order to normalize scapular abduction for subject size. DeVita et al reported fair reliability (ICC = .78) for this normalized scapular abduction, but Neiers and Worrell reported poor reliability (ICC = .34) for this measure in their study. Therefore, we chose to perform only the initial measure as described for scapular position.

Immediately following the measurements using the DeVita et al technique, the tester continued on the same subject's dominant side with the three measures described by Kibler. Kibler established the nearest spinous process as the medial point

		Tester 1			Tester 2		
		X (cm)	SD (cm)	Range (cm)	X (cm)	SD (cm)	Range (cm)
DeVita	D	21.64	2.21	17.2–25.65	21.82	2.07	17.35–25.85
	ND	21.18	2.36	15.95–27.2	21.18	2.00	16.8–25.8
K-1	D	10.08	1.91	6.7–15.3	8.97	1.80	5.4–14.8
	ND	9.78	1.85	6.25–14.1	8.87	1.81	5.4–13.65
K-2	D	10.81	1.84	7.75–15.8	9.60	1.61	6.5–14.3
	ND	10.55	1.84	6.65–15.1	9.55	1.57	6.0–13.35
K-3	D	12.94	1.97	8.6–17.9	10.77	1.81	6.45–15.85
	ND	12.86	1.85	8.3–17.0	10.66	1.96	5.5–16.85

DeVita = DeVita et al measurement.
K-1 = Kibler measurement #1.
K-2 = Kibler measurement #2.
K-3 = Kibler measurement #3.
D = Dominant.
ND = Nondominant.

TABLE 1. *Descriptive data for determining scapular position (mean of four measures for each method).*

of the measurement. For the purpose of this study, we established T8 for standardization as the nearest spinous process and measured to the inferior angle of the scapula. These measures were taken with string with the subject's arm in three different positions. The first position was the arm relaxed at the side, the second was the subject's hand on the hip with the web space between the thumb and second finger placed on the lateral iliac crest, and the third was with the arm abducted to 90° as measured with a standard goniometer. Two measurements were taken at each position to allow calculation of an average. Each measurement was read and recorded in the same manner as described for the DeVita et al technique.

After completing the measurements for the single DeVita et al and the three Kibler positions on the subject's dominant side, the entire procedure was repeated on the nondominant side. When these measurements were completed, the tags marking the spinous processes were removed; the testers switched subjects, located, and remarked the spinous processes; and the entire measurement procedure was repeated. The subject then returned to his/her original tester for remeasurement using both measurement methods on the dominant and nondominant sides and, finally, returned to the second tester for repeat of the same process. Thus, each tester used both the DeVita et al and the Kibler measurement techniques twice for each subject.

Data Analysis

Means, standard deviations, and ranges of scapular distance as measured by each tester were calculated for the single DeVita et al and three Kibler positions. In addition, means and mean differences were calculated to compare the subjects' dominant vs. nondominant sides.

Intraclass correlation coefficients (ICC) (2,1) were calculated to determine reliability (18). The ICC was based upon a repeated measures analysis of variance (ANOVA) ($N = 32$) which compared the means of the two measures taken at each position. To calculate intratester reliability, the mean values of the two measures taken during each of the two trials by the same tester on one subject were compared. To calculate intertester reliability, the mean values of the two measures constituting the first trial for one tester on a given subject were compared with those from the first trial of the second tester on the same subject. For the purpose of this study, the ICCs were classified as follows: <.69, poor correlation; .70–.79, fair correlation; .80–.89, good correlation; and .90–1.00, high correlation (5).

The standard error of measurement (SEM) (SD $\sqrt{1\text{-ICC}}$) was used to determine the measurement error associated with repeated scapular measurements (4). The SEM is expressed in the same units as the original measurement. Therefore, the SEM is meaningful because the actual measurement error is known for a given tester and measurement method. The SEM is plus or minus the value of interest. For example, for tester 1, using the DeVita et al technique to measure the dominant side, the mean was 21.64 cm (Table 1) ± .49 cm (SEM) (Table 2), which means that 68% of the time the measure will be within .49 cm above or below the mean value.

Probability values (p) are reported from the repeated measures ANOVA. The p value indicates whether significant differences exist between repeated measures. For reliability purposes, the p values should be greater than .05, which indicates no significant difference existed between measurement sessions.

		Tester 1			Tester 2		
		ICC	SEM (cm)	p	ICC	SEM (cm)	p
DeVita	D	0.95	0.49	0.54	0.92	0.59	0.76
	ND	0.95	0.53	0.39	0.94	0.53	0.57
K-1	D	0.92	0.54	0.007*	0.94	0.44	0.99
	ND	0.93	0.49	0.19	0.91	0.54	0.48
K-2	D	0.88	0.64	0.03*	0.92	0.45	0.43
	ND	0.89	0.61	0.07	0.88	0.54	0.49
K-3	D	0.89	0.65	0.85	0.81	0.79	0.52
	ND	0.91	0.56	0.52	0.90	0.62	0.35

** $p < 0.05$.*
DeVita = DeVita et al measurement.
K-1 = Kibler measurement #1.
K-2 = Kibler measurement #2.
K-3 = Kibler measurement #3.
D = Dominant.
ND = Nondominant.

TABLE 2. *Intratester reliability.*

For comparison of the scapular position (mean of four measures for each tester and method) of the dominant vs. nondominant extremity, an independent t test was used. A Bonferroni adjustment was made to correct for alpha level inflation resulting from multiple t tests. The adjusted alpha level for these data were, therefore, calculated to be $p < 0.006$ (0.05/8).

RESULTS

The descriptive statistics of scapular measurements are illustrated in Table 1. The means (four measures for each tester and method) obtained by tester 2 for the three Kibler positions were consistently smaller than those obtained by tester 1.

The ICC, SEM, and p values for intratester reliability are shown in Table 2. The ICC ranged from .92 to .95 for the DeVita et al method and from .81 to .94 for the Kibler methods. The SEM ranged from .49 to .59 for the DeVita et al method and from .44 to .79 for the Kibler methods. Significant differences occurred between measures of tester 1 for the dominant arm with both Kibler measures 1 and 2.

The ICC, SEM, and p values for intertester reliability are presented in Table 3. The ICC of the DeVita et al method exceeded that of the Kibler methods, and the ICC of the three Kibler measures declined sequentially from position 1 to position 3. The SEM values ranged from .60 to 1.65, the smaller value occurring with the DeVita et al technique. The SEM then increased progressively from Kibler measures 1 to 3. Probability values were nonsignificant for the DeVita et al comparison, but significant differences were noted between testers on all Kibler measures.

The means, mean differences, and p values for the intratester comparison of dominant vs. nondominant sides are illustrated in Table 4. Means were consistently greater for the dominant side than the nondominant side. However, using the adjusted alpha level, these differences were only significant for the DeVita et al measures of tester 2.

		ICC	SEM (cm)	p
DeVita	D	0.91	0.65	0.28
	ND	0.92	0.60	0.99
K-1	D	0.67	1.04	0.0001*
	ND	0.69	1.02	0.0002*
K-2	D	0.52	1.20	0.0001*
	ND	0.53	1.17	0.0004*
K-3	D	0.28	1.59	0.0001*
	ND	0.18	1.65	0.0001*

** $p < 0.05$.*
DeVita = DeVita et al measurement.
K-1 = Kibler measurement #1.
K-2 = Kibler measurement #2.
K-3 = Kibler measurement #3.
D = Dominant.
ND = Nondominant.

TABLE 3. *Intertester reliability.*

	Tester 1				Tester 2			
	D $\bar{X}$ (cm)	ND $\bar{X}$ (cm)	$\bar{X}$ Diff (cm)	*p*	D $\bar{X}$ (cm)	ND $\bar{X}$ (cm)	$\bar{X}$ Diff (cm)	*p*
DeVita	21.64	21.18	0.46	0.01	21.82	21.18	0.64	0.005*
K-1	10.08	9.78	0.30	0.08	8.97	8.87	0.11	0.53
K-2	10.81	10.55	0.25	0.27	9.60	9.55	0.05	0.77
K-3	12.94	12.86	0.08	0.72	10.77	10.66	0.12	0.57

** $p < 0.006$.*
DeVita = DeVita et al measurement.
K-1 = Kibler measurement #1.
K-2 = Kibler measurement #2.
K-3 = Kibler measurement #3.
D = Dominant.
ND = Nondominant.

TABLE 4. *Dominant vs. nondominant extremity comparisons using independent t tests with corrected alpha level (p = 0.006) (mean of four measures for each tester and method).*

DISCUSSION

Intratester Reliability

Intratester reliability of each tester using the DeVita et al measurement method was high. These findings are in agreement with both DeVita et al (ICC = .94) and Neiers and Worrell (ICC = .80). Intratester reliability of the Kibler measurement method was also high for position 1 and ranged from good to high for positions 2 and 3. However, despite acceptable ICC values, significant differences existed using the Kibler 1 and 2 measures of the dominant arm for tester 1. The fact that significant differences existed for Kibler measures 1 and 2 indicates that the larger measurement error was associated with these techniques for one tester. Therefore, caution should be used when using the Kibler technique as performed in this study.

While the ICCs provide information about relative reliability or the relationship between repeated measures, the SEM provides information about absolute reliability or the variability of scores with repeated measures (8). We believe that the SEM values are acceptable in comparison to the mean values. As an example, an SEM of .5 cm with a mean of 10 cm indicates that the actual scapular distance would range between 9 to 11 cm 96% of the time.

Intertester Reliability

The intertester reliability value for the DeVita et al measurement method was high, and the *p* value revealed no significant differences obtained by the two testers. Neither DeVita et al nor Neiers and Worrell examined intertester reliability, and we could locate no other studies for comparison. Intertester reliability values for the Kibler measurement methods were all poor, and the *p* values demonstrated significance in differences between testers for all Kibler measures. We believe that this was due largely to differences in measuring to the inferior angle of the scapula by the two testers. While the landmark used in the DeVita et al technique (the inferior angle of the acromion process) is relatively discrete, the inferior angle of the scapula used in the Kibler measures is not. This landmark assumes different contours on different subjects and typically describes an arc rather than a point. Because tester 2's Kibler values were consistently lower than tester 1's, it is postulated that perhaps tester 2 measured to the vertebral border of the inferior angle while perhaps tester 1 measured to the most inferior point on the scapula. Intraclass correlation coefficients measure the extent to which multiple measures agree or are equal rather than simply the extent to which they are associated (18), which could explain good to high intratester reliability but poor intertester reliability using the Kibler measurement method. The SEM values increased in comparison to the intratester values for all measurement techniques and also increased progressively within the Kibler technique from position 1 to position 3. This would be expected as the SEM is calculated from both the SD and ICC and, thus, a decrease in the ICC will increase the SEM.

Dominant vs. Nondominant Comparison

We found consistently larger means for the dominant extremity in comparison to the nondominant one using both the DeVita et al and the three Kibler measurement methods. Using the adjusted alpha level (0.006) to correct for multiple *t* tests, these differences were only significant as measured by one tester using the DeVita et al method. The other tester's results were nonsignificant, and no significant differences were found by either tester using the three Kibler measures. Perhaps these findings should not be surprising in a sample which is neither athletic nor engaged in heavy or repetitive manual work favoring the dominant extremity.

Limitations

Our subject sample consisted of normal, nonathletic individuals. Therefore, caution must be used in extrapolating these findings to other populations.

Despite the testers each having several years of clinical experience (tester 1 = 11 years, tester 2 = 6 years), we found palpation of the landmarks to be difficult at times, primarily for the Kibler measures. For Kibler measures 2 and 3, the lateral rotation of the scapula was accompanied by increasing approximation of the inferior angle against the rib cage in contrast to its relative prominence when using the DeVita et al and Kibler 1 measure. For Kibler measure 3, the dorsal muscle contraction produced by the active humeral abduction also made the inferior angle of the scapula more difficult to palpate accurately. As described earlier, the lack of a discrete, standardized point on the inferior angle of the scapula for these measures was believed to contribute to measurement variation between testers. Also, while none of our subjects would be classified as obese, it was relatively easy to locate spinous processes and scapular landmarks on the very lean subjects in comparison to more muscular subjects or those with increased subcutaneous fat. These difficulties became magnified when using the Kibler position 3 for reasons previously described.

Clinical Implications

The DeVita et al method holds promise for quantification of scapular position. Their original findings of high reliability using this measurement method have now been supported by two other studies, the Neiers and Worrell study (15) (good reliability) and this study.

The Kibler measurement methods hold promise and are worthy of further study, as they measure scapular distance in three different arm positions, not solely with the arm at the side. However, these methods are more difficult for the clinician and may not prove to be reliable when the data of multiple examiners are compared.

The validity of these static measures of a highly dynamic complex has not been established. The relationship of static positions and function remains to be seen. While it would be hoped that such measures would eventually prove useful in evaluation and rehabilitation of pathology, this remains unanswered at this time.

Implications for Future Research

We are yet to be convinced that string is necessarily the best measurement tool for these techniques. Error may exist when, on one trial, the string is pulled taut to yield the shortest distance between landmarks, and, on another trial, it is allowed less tension in order to follow skin contours. For this reason, experimentation with calipers might yield better results.

Normalized data for scapular abduction may prove helpful if used with understanding of its limitations. Therefore, future studies using these measurement techniques should assess the reliability of association of measures of scapular distance to another anatomical parameter.

We were unable to locate any other reliability studies using the Kibler technique. The discrepancy existing between our intratester and intertester reliability values for this technique warrants additional study. Because this technique is inherently more difficult, a lengthy familiarization period is recommended prior to data collection.

Future studies should use these measurement techniques on pathological populations in order to determine if significant differences can be quantified. Finally, if differences should be established, the effects of therapeutic intervention can then be monitored in order to prove or disprove real effects.

SUMMARY

We measured scapular position using methods reported by DeVita et al and Kibler in order to evaluate intratester and intertester reliability and also to determine if positional differences existed between the dominant and nondominant extremities. Both intratester and intertester reliability of the DeVita et al measures proved high. Intratester reliability of the Kibler methods ranged from good to high, but the intertester reliability was poor. Despite consistent, unidirectional differences in the scapular position of the dominant and nondominant extremities, only one tester using the DeVita et al method found significant differences between sides. All other differences proved to be nonsignificant. JOSPT

REFERENCES

1. Bagg SD, Forrest WJ: A biomechanical analysis of scapular rotation during arm abduction in the scapular plane. Am J Phys Med Rehabil 67:238–245, 1988
2. Bagg SD, Forrest WJ: Electromyographic study of the scapular rotators during arm abduction in the scapular plane. Am J Phys Med 65:111–124, 1986
3. Basmajian JV, Bazant FJ: Factors preventing downward dislocation of the adducted shoulder joint. J Bone Joint Surg 41A:1182–1186, 1959
4. Baumgartner TA: Norm-referenced measurement: Reliability. In: Safrit M, Woods T, (eds), Measurements Concepts in Physical Education and Exercise Science, pp 42–72. Champaign, IL: Human Kinetics Publishers, 1989
5. Blesh TE: Measurement in Physical Education (2nd Ed), New York: The Ronald Press Company, 1974
6. DeLuca CJ, Forrest WJ: Force analysis of individual muscles acting simultaneously on the shoulder joint during isometric abduction. J Biomech 6:385–393, 1973
7. DeVita J, Walker ML, Skibinski B: Relationship between performance of selected scapular muscles and scapular

abduction in standing subjects. *Phys Ther* 70:470–476, 1990

8. Domholdt E: *Physical Therapy Research: Principles and Applications*, Philadelphia, PA: W.B. Saunders Company, 1992
9. Doody SG, Waterland WJ, Freedman L: Scapulohumeral goniometer. *Arch Phys Med Rehabil* 51:711–713, 1970
10. Dvir Z, Berme N: The shoulder complex in elevation of the arm: A mechanism approach. *J Biomech* 11:219–225, 1978
11. Freedman L, Munro RR: Abduction of the arm in the scapular plane: Scapular and glenohumeral movements. *J Bone Joint Surg* 48A:1503–1510, 1966
12. Hertling D, Kessler RM: *Management of Common Musculoskeletal Disorders* (2nd Ed), p 177. Philadelphia, PA: J.B. Lippincott Company, 1990
13. Inman VT, Saunders JB, Abbott LC: Observations on the function of the shoulder joint. *J Bone Joint Surg* 26A:1–30, 1944
14. Kibler WB: Role of the scapula in the overhead throwing motion. *Conte Orthop* 22:525–532, 1991
15. Neiers L, Worrell TW: Assessmen scapular position. *J Sports Reh* 2:20–25, 1993
16. Poppen NK, Walker PS: Normal a abnormal motion of the shoulder *Bone Joint Surg* 58A:195–201, 197
17. Saha AK: *Theory of Shoulder Mec nism: Descriptive and Applied*, Sprir field, IL: Charles C. Thomas, 1961
18. Shrout PE, Fleiss JL: Intraclass corre tions: Uses in assessing rater reliabili *Psychol Bull* 86:420–428, 1979