The Interrater Reliability of the Load and Shift Test for Anterior Shoulder Instability: A Technical Report

Morey J. Kolber, PT, Ph.D, OCS, Cert MDT, CSCS[1]
Melissa Corrao, BS, CSCS[2]

1. Assistant Professor
2. Doctoral Student

Department of Physical Therapy, Nova Southeaster University, Ft. Lauderdale, Florida.

United States

ABSTRACT

Background: Shoulder disorders affect up to 67% of the adult population at some point in their lifetime. The shoulder complex ranks third, trailing low back and neck pain, in musculoskeletal disorders for which individuals seek physical rehabilitation. The assessment of shoulder stability is an integral component of the patient examination, as it may assist the clinician in making a diagnosis, measuring improvement or deterioration, and determining functional impairments. Reliable tests and measurements are therefore essential to both the clinician and researcher desiring to objectively monitor disease progression, outcomes, and mobility impairments. Many of the currently used clinical tests for anterior instability are documented based on a dichotomous outcome, thus do not provide a means of quantifying or ranking the degree of anterior translation associated with the instability. The load and shift test has been described in textbooks and research investigations as a method to quantify (rank) anterior shoulder instability and/or translation at the glenohumeral joint. This test is attractive clinically, as it requires no special equipment or space, thus may be utilized in multiple settings and without cost; however, few studies have investigated the interrater reliability of this test. **Purpose:** The purpose of this investigation was to determine the interrater reliability of the seated load and shift test using a 4-level grading criteria. **Method:** Two investigators performed the load and shift test on the non-dominant shoulder of 29 asymptomatic female participants in a repeated measures intrasession design. Each investigator was blinded to the results and the order of testing was counterbalanced to minimize potential bias. **Results:** Results indicated good interrater reliability with an Intraclass Correlation Coefficient (2, 1) (95% CI) = 0.80 (0.61, 0.90). **Conclusion:** The load and shift test appears to be a reliable clinical test for detecting and quantifying anterior translation of the glenohumeral joint when using the grading criteria outlined in this investigation. Shoulder translation when excessive is associated with shoulder instability, thus this test may prove valuable as a measurement of anterior instability. **Recommendations:** Future investigations are needed to determine the reliability and validity of this test among a symptomatic population and including a male cohort.

INTRODUCTION

The lifetime prevalence of shoulder disorders in the general population has been reported to range from seven to sixty-seven percent.[1] The shoulder complex ranks third, trailing low back and neck pain, in musculoskeletal disorders for which individuals seek physical rehabilitation.[2] Numerous precipitating factors and biomechanical explanations have been offered to diagnose shoulder pain; however, the etiology is often multifactorial. Anterior glenohumeral joint instability (AI) is a common biomechanical impairment implicated in the etiology of shoulder pain in both the general and athletic population.[3-7] AI may be evaluated using a

range of criteria including the clinical history, special tests that rely on pain provocation, and manual assessment. In particular, the manual assessment of anterior humeral head translation is performed clinically as it is suggested to be pathognomonic of AI.[3,4,6-8]

The load and shift test is used by clinicians and researchers as a means to detect and quantify anterior glenohumeral joint translation and instability. This test may be performed in various abduction angles while seated or supine and requires the examiner to determine the amount of anterior humeral head translation on a fixed glenoid fossa.[8] The ranking advocated for this test is a 4-level grading criteria ranging from normal mobility (grade 0) to a frank dislocation (grade 3). AI is suggested when the test is graded on the upper limits of the scale, hyperlaxity when graded on the lower limits, and zero indicating normal mobility. Advantages of this test include portability and cost, as no special equipment or space is required. A limitation of the ranking system is its inability to account for hypomobility. A disadvantage of the technique is that it requires manual palpation, which may be difficult in cases where there is considerable muscle hypertrophy or guarding.

To our knowledge, only one investigation has reported the reliability of the seated load and shift test.[3] In that study, moderate interrater reliability (ICC = 0.53) was identified for the seated load and shift test in a cohort of 13 symptomatic patients.[3] No studies were identified that assessed the reliability of the load and shift test among asymptomatic participants despite use of the unaffected extremity as a comparison for many clinical tests. Moreover, hyperlaxity and instability has been reported among individuals without shoulder complaints, thus it is necessary to identify the reproducibility of this test among an asymptomatic cohort.[9] Based on the limited research available for this test, it is apparent that further investigations are necessary to determine the reliability of the load and shift test. The purpose of this study therefore was to determine the interrater reliability of the seated load and shift test for detecting anterior humeral head translation when using an ordinal 4-level classification system. The results of this investigation should be of value to clinicians who must recognize the merits and limitations of their measurements and to researchers who may use the data as the basis for future comparison in a symptomatic cohort.

METHODS
Participants
Twenty-nine participants were consecutively recruited from a local university setting. Participants who met study requirements were provided with an informed consent document approved by the Institutional Review Board at Nova Southeastern University and all questions were answered to their satisfaction prior to commencing data collection.

Participants completed a questionnaire to report age, weight (body mass), height, and arm dominance. Exclusion criteria consisted of recent shoulder surgery for which the participant was still receiving care or current participation in a rehabilitation or active mobility program for the non-dominant shoulder. Nineteen of the participants reported performing recreational weight-training at least twice a week, whereas the remaining 10 acknowledged no exercise or sport participation. Testing was done on the non-dominant arm to account for confounding variables such as compulsory use of the dominant extremity. The left arm was the non-dominant arm of 27 out of 29 participants.

Procedures
Testing procedures were explained to participants and all questions were answered prior to commencing data collection. Following completion of demographic paperwork and consent to participate, individuals were brought to a private testing laboratory for data collection. Once in appropriate attire exposing the glenohumeral joint, the participants performed a standardized warm-up that was supervised by the raters. One of the raters (Rater 2) is a Physical Therapist with 15 years of experience and board certification in orthopaedics through the American Board of Physical Therapy Specialties, and the other rater (Rater 1) was a doctoral physical therapy student at the time of data collection. The doctoral student-rater had approximately 1-year of experience performing the load and shift test. The warm-up required approximately three minutes total time and consisted of two exercises including arm pendulums and standing shoulder blade squeezes, neither of which we believed would offer a mobilization effect.

Following the warm-up, participants proceeded with testing. The procedure for all 29 participants was similar with the only exception that testing order was alternated to avoid systematic error that may be induced from a mobilization effect from the first tester.

Participants were required to sit on a treatment table without back support and the raters ensured that individuals were sitting upright with an appropriate posture and their tested arm relaxed on the thigh. The rater stood posterior to the participant and stabilized the shoulder with one hand over the clavicle and scapular spine. The other hand grasped the humeral head with the

fingers anteriorly and the thumb posteriorly to determine first if the humeral head was seated centrally in the glenoid fossa (Figure 1).

Figure 1: The load and shift test is illustrated. Arrow on superior shoulder denotes direction of manual translation.

If the humeral head was positioned anterior or posterior in the fossa, the examiner compressed the humeral head medially toward the fossa and applied a gentile anterior or posterior force to centrally seat the humeral head in the glenoid fossa as described by Magee.[8] This step is the "load" component of the test. Next, with the participant relaxed and sitting upright, the hand contacting the humeral head was translated anteriorly and slightly medially to reflect the natural orientation of the glenoid fossa as illustrated by the arrow in Figure 1. This step is the "shift" component of the test. Careful attention is made during the test to avoid scapular protraction as this may cause the humeral head to translate anteriorly in the glenoid fossa and narrow the subacromial space.[8] The examiner at this point determines the amount of humeral head translation respective to the glenoid fossa and assigns a grade based on the criteria illustrated in table 1. Once this procedure was completed, the second grader repeated the process within a 60-second duration and documented their grade on a data collection sheet blinded to the results of the first rater. Each rater performed the test once and documented their outcome using a modified version of the grading criteria defined by Altchek (Table 1).[10,11]

Table 1: Grading scale used to describe anterior humeral head translation during the load and shift test

Grade	Description used to quantify grade
0	Normal translation = the humeral head moves anterior but not to the glenoid rim
1	Excessive translation up to glenoid rim, but no subluxation
2	Humeral head translates over glenoid rim, but spontaneously reduces
3	Frank dislocation of humeral head over glenoid rim with locking

DATA ANALYSIS

Data analysis was performed with SPSS version 15.0 for Windows. Descriptive data including the mean and standard deviation for participant age, height, body mass, and BMI was calculated. Frequency counts for the results from rater 1 and 2 were tabulated and are presented in figure 2.
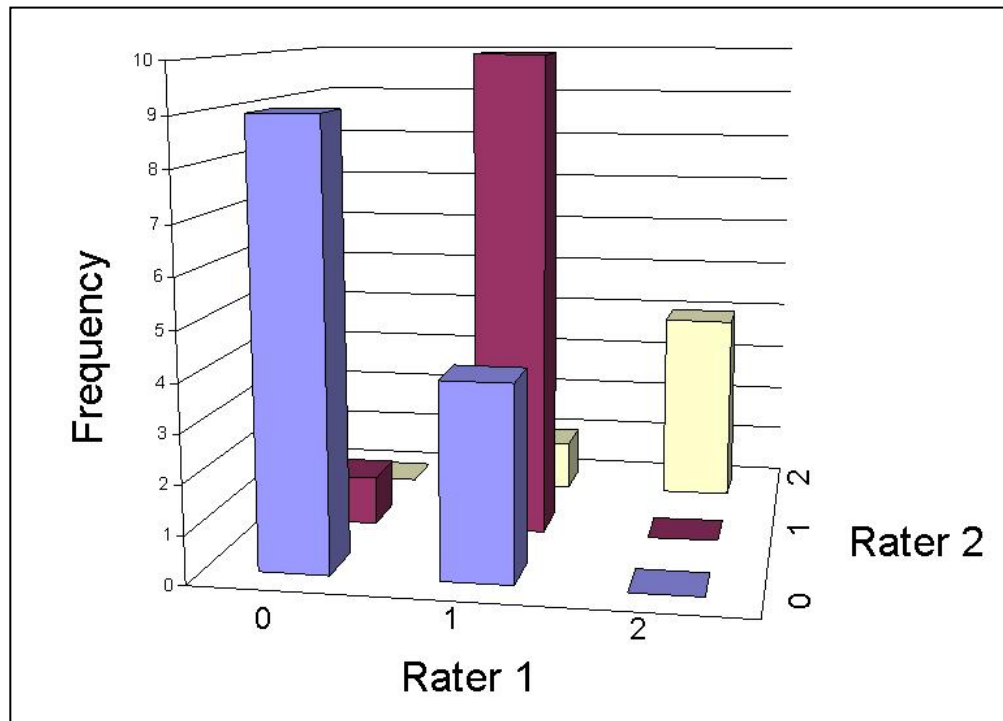

Figure 2: 3-dimensional frequency comparison of grades on the load and shift test between raters.

The interrater reliability was determined by the Intraclass Correlation Coefficient (ICC) model 2,1.[12,13] The single value from each testing session was used for the analysis and the agreement option was selected. Model 2, 1 is recommended for interrater analysis as it may be used to determine if a particular test or instrument can be used with confidence among randomly chosen raters that represent the population of raters at large.[12,13] Moreover, the results can be generalized to other raters within the range of experience possessed by the current raters. Our interpretation of the ICC value was based on guidelines offered by Portney and Watkins, where a value of above 0.75 was classified as good, 0.50-0.75 was considered moderate and, < .50 was fair.[12]

RESULTS

Participants were female and between the ages of 18 to 41 years. The mean and standard deviation (SD) for the participants age, body mass index (BMI), body mass, and height were 24.7 years (4.9), 23.6 kg/m^2(3.2), 63.6 kg (10.2), and 164.0 cm (9.7) respectively. The interrater reliability analysis indicated good reliability; ICC (95% CI) = 0.80 (0.61, 0.90). Frequency counts indicate that rater 1 identified normal anterior translation (Grade 0) among 13(44.8%) of participants. Grade 1, 2, and 3 was identified at 11(37.9%), 5(17.2%) and 0(0%) respectively. Rater 2 identified 10(34.5%) of participants as normal with 15(51.7%) as having grade 1 anterior translation, 4(13.8%) as having grade 2 translation. Similar to rater, 1 none of the participants were graded as having a grade 3 load and shift test by rater 2 as expected in an asymptomatic cohort. Figure 2 provides an illustration of the grading percentage scored by each of the two raters. Although the participants were asymptomatic, twenty were graded as having excessive anterior translation on the modified Altchek scale. The raters only agreed on 9 participants as having normal translation.

DISCUSSION

The assessment of shoulder stability is an integral component of the patient examination. According to the *Guide to Physical Therapist Practice*, the examination of joint integrity is necessary to select appropriate interventions.[14] Moreover, recognizing a loss of joint stability may assist the clinician in making a differential diagnosis, measuring improvement or deterioration, determining functional impairment, and planning therapeutic interventions. Accurate and reliable measurement instruments are therefore essential to both the clinician and researcher desiring to objectively monitor disease progression, outcomes, and mobility impairment.

The load and shift test when performed as outlined in this investigation possesses good interrater reliability that exceeds previous values reported in the literature. Tzannes et al assessed 13 patients with a history suggesting shoulder instability and reported an interrater reliability coefficient of ICC = 0.53 for the load and shift test when performed according to the procedures outlined in the current investigation.[3] The lower value may be the result of using a symptomatic cohort compared to the present investigation which was asymptomatic.

Of interest in the current investigation was the degree of anterior translation identified in an asymptomatic cohort. Of the 29 participants in this investigation rater 1 graded 13 participants as normal whereas rater 2 graded 10 participants as normal. When looking at agreement between the 2 raters only 9 participants were graded as having normal anterior translation. Moreover, 4 participants were graded a 2 by both raters which is suggestive of anterior humeral head translation over the glenoid rim.[10,11] None of the participants had grade 3 anterior translation which may suggest that this grade may be more representative of symptomatic AI than grades 1 and 2. This finding would need to be explored further in a validity study among a symptomatic cohort. From a predictive perspective, it is not currently known whether asymptomatic individuals with abnormal findings are more likely to become symptomatic at some future point; thus, these finding may have predictive relevance.

In regard to activity participation, 19 of the 20 participants graded as having some degree of anterior translation reported participating in weight-training, which has previously been reported as a activity associated with anterior instability.[5,9] Future investigations are needed to further confirm this association. These findings have direct clinical and research implications. From a clinical perspective, it should be recognized that anterior translation classified as greater than normal on the load and shift test may be identified in an asymptomatic population, thus diagnosis should be supported by the clinical history and other clinical findings. From a research and clinical perspective, the results from this investigation should be considered when characterizing impairments among symptomatic individuals as only 31% of participants were considered normal by both raters.

In regard to testing position, the load and shift test may be performed in sitting or supine and can be assessed at varying levels of abduction.[3,8] In the current study, we chose to use the seated position with the arm in zero degrees of abduction. Our rational for choosing this position is its broad application in future investigations and based on efficiency as mobility restrictions or the inability to lie flat would not be a barrier to performing the test. Additionally, the test is more easily performed with the arm at the side as compared to an abducted position which would require the clinician to stabilize the joint. One may not generalize our finding during the seated version of the test to supine testing or testing that requires arm abduction.

LIMITATIONS AND FUTURE INVESTIGATIONS

The results of this study were new as reliability values for the load and shift test have not previously been reported in an asymptomatic cohort. While the results suggested good reliability, there was a distinct difference in the level of experience between the two raters. Thus, it is possible that the reliability coefficient may have been higher if the raters had similar experience.

As with any study using asymptomatic participant, one may not generalize the findings to a symptomatic population. Moreover, participants in this study were all female, thus it is not known whether or not this test could be performed on males with similar reproducibility due to varied morphology. Future investigations must be performed on a symptomatic population comprised of both males and females. Lastly, one must recognize that although the load and shift maneuver is advocated as a diagnostic tool for anterior instability, this study was a reliability investigation and made no attempt to test the diagnostic validity. Future investigations should assess both the diagnostic and predictive validity of this test.

CONCLUSION

The load and shift test possesses good reliability for testing anterior shoulder translation among an asymptomatic cohort when performed according to the procedures outlined in this investigation. The test requires no special equipment and is reproducible among clinicians with varying levels of experience, thus clinical application may not be limited by space or practitioner

experience. No study is meant to be the terminal point in research; therefore, future investigations are necessary to build upon these findings so that they may be considered with absolute confidence in the clinical environment.

REFERENCES
1. Luime JJ, Koes BW, Hendriksen IJ, Burdorf A, Verhagen AP, Miedema HS, Verhaar JA. Prevalence and incidence of shoulder pain in the general population; a systematic review. *Scand J Rheumatol.* 2004; 33:73-81.\
2. Di Fabio RP, Boissonnault W. Physical therapy and health-related outcomes for patients with common orthopaedic diagnoses. *J Orthop Sports PhysTher*. 1998 27:219-230.
3. Tzannes A, Paxinos A, Callanan M, Murrell GA. An assessment of the interexaminer reliability of tests for shoulder instability. *J Shoulder Elbow Surg*. 2004;13:18-23.
4. Matsen FA, Fu FH, Hawkins RJ. The shoulder: a balance of mobility and stability. *AAOS*. 1992:3-5.
5. Gross ML, Brenner SL, Esformes I, Sonzogni JJ. Anterior shoulder instability in weight lifters. *Am J Sports Med*. 1993;21:599-603.
6. Bigliani LU, Codd TP, Connor PM, Levine WN, Littlefield MA, Hershon SJ. Shoulder motion and laxity in the professional baseball player. *Am J Sports Med*. 1997;25:609-613.
7. Rupp S, Berni ger K, Hopf T. Shoulder problems in high level swimmers-impingement, anterior instability, muscular imbalance? *Int J Sports Med.* 1995;16:557-62.
8. Magee DJ. *Orthopedic Physical Assessment. 5th ed*. Saunders-Elsevier,St louis, Missouri, 2009.
9. Kolber, M. Shoulder Joint and Muscle Characteristics in the Recreational Weight Training Population [dissertation]. Ft. Lauderdale, FL. 2007.
10. Altchek DW, Warren RF, Skyhar MJ, et al. A technique for treating multidirectional instability in the athlete. *Orthop Trans*. 1989;13:560-561.
11. Altchek DW, Warren RF, Skyhar MJ, Ortiz G. T-plasty modification of the Bankart procedure for multidirectional instability of the anterior and inferior types. *J Bone Joint Surg Am*. 1991;73:105-112.
12. Portney LG, Watkins MP. *Foundations of clinical research. Applications to practice.3rd ed*. Upper Saddle River, Pearson Prentice Hall. Upper Saddle River, NJ. 2009.
13. Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychol Bull.* 1979; 86:420-428.
14. American Physical Therapy Association (APTA) .*Guide to Physical Therapist Practice 2nd ed*. Alexandria, VA: American Physical Therapy Association. 2003.