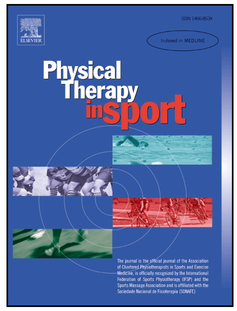
Accepted Manuscript



Reliability of a shoulder arm return to sport test battery

Margie Olds, Cade Coulter, Dan Marrant, Tim Uhl

PII: S1466-853X(19)30156-7

DOI: https://doi.org/10.1016/j.ptsp.2019.06.001

Reference: YPTSP 1063

To appear in: *Physical Therapy in Sport*

Received Date: 31 March 2019

Revised Date: 31 May 2019

Accepted Date: 1 June 2019

Please cite this article as: Olds, M., Coulter, C., Marrant, D., Uhl, T., Reliability of a shoulder arm return to sport test battery, *Physical Therapy in Sports* (2019), doi: https://doi.org/10.1016/j.ptsp.2019.06.001.

Margie Olds[1], Cade Coulter[2], Dan Marrant[2], Tim Uhl[2]

1- Flawless Motion, 7/88 Cook St, Auckland, New Zealand
2- Division of Athletic Training, Department of Rehabilitation Sciences, College of Health Sciences, University of Kentucky, Lexington, USA.

Corresponding Author:

Dr Margie Olds

Email: margie@flawlessmotion.com

Twitter: @margie_olds

Phone: +64-21573422

Permanent Address: 7/88 Cook St, Auckland, New Zealand

Address where work was done: Division of Athletic Training, Department of Rehabilitation Sciences, College of Health Sciences, University of Kentucky, Lexington, USA.

Funding Statement: This research was supported in part from University of Kentucky through grant UL1TR001998 to help with recruitment of subjects.

Reliability of a Shoulder Arm Return to Sport Test Battery

Abstract

Objectives: To establish the reliability and responsiveness of a clinical test battery developed to determine readiness to return to sport after an upper extremity injury. A second objective was to examine the limb symmetry in single limb tests.

Design: Methodological study

Participants: Forty healthy participants (20 male) were tested weekly on three occasions.

Main outcome measures: Learning effect, inter-rater and intra-rater reliability was calculated for each test in the Shoulder Arm Return to Sports (SARTS) battery with repeated measures ANOVA and intraclass correlation coefficient (ICC). Measurement error and responsiveness were determined using Standard Error of Measure (SEM) and Minimal Detectable Change (MDC).

Results

Drop Catches and Ball Taps showed a learning effect between Days 2-3. Intra-rater reliability for the remaining six tests between Days 2-3 ranged between 0.78(95%CI 0.63–0.88) and 0.96(95%CI 0.92–0.98) while inter-rater reliability on Day 2 ranged between ICC=0.96(95%CI 0.94–0.98) and ICC=0.99(95%CI 0.98-0.98). Two tests (BABER (91%) and Drop Catches (93%)) were significantly decreased on the non-dominant side (p=0.05).

Conclusions

Six of the eight tests in the SARTS test battery demonstrate good psychometric properties to evaluate both open and closed chain upper extremity activities indicating their readiness for clinical use.

Keywords: Athlete, Athletic Training, Physical Therapy, Recovery of Function, Return to Play, Sporting Activity Resumption

Word Count: 3999 of 4000 limit

1. Introduction

Decision-making regarding return to sport after a shoulder injury is difficult for clinicians and the lack of reliable and valid physical performance tests exacerbates this problem (Cools et al., 2014; Creighton, Shrier, Shultz, Meeuwisse, & Matheson, 2010). Shoulder injuries which result in lengthy time away from sport can be expensive for professional athletes and can jeopardize career opportunities (Headey, Brooks, & Kemp, 2007). Additionally, rates of recurrent shoulder instability after a shoulder dislocation can be higher than 50% following both surgical and non-surgical treatment (Torrance, Clarke, Monga, Funk, & Walton, 2018; Wheeler, Ryan, Arciero, & Milinari, 1989) indicating that improvements in decision-making are necessary to ensure athletes are safe to return to sport after a shoulder injury. One critical step in reducing recurrent shoulder injuries is the establishment of physical performance tests to facilitate decision-making regarding whether athletes are ready to return to sport.

Physical performance tests must be easy and inexpensive to perform in the clinic, demonstrate good psychometric properties, and have normative data and cut-off values generated before clinical use (Ardern, Bizzini, & Bahr, 2016; Goldbeck & Davies, 2000). Clinical tests should be representative of the demands of the sport to which the athlete is returning as there is wide variation in the demands of different sports (Cho, Hwang, & Rhee, 2006). Additionally, despite many sports requiring athletes to play for extended periods of time, current physical performance tests report isolated efforts (Ashworth, Hogben, Singh, Tulloch, & Cohen, 2018; Chmielewski et al., 2014; Gorman et al., 2012; Harris et al., 2011), and do not examine endurance capability. Thus, single one-off testing may not accurately measure an athlete's readiness to return to sport for the duration required, meaning performance tests which examine endurance are required. Finally, single tests which determine return to sport have limited clinical utility as they measure only one construct. Thus, a battery of tests which measures different constructs such as strength, power, range of motion and neuromuscular control, may improve the ability to determine safe return to sport.

Therefore, it is necessary that a battery of safe, valid and reliable physical performance tests be devised. The primary aim of this study is to assess the inter-rater and intra-rater reliability of a battery of functional performance tests in a healthy population. It is hypothesized that there will be significant familiarization to the test battery between Day 1 and Day 2 while reliability of Day 2 and Day 3 will be good to excellent, exceeding 0.8 intraclass correlation coefficients. The secondary aim of this study is to examine side to side differences within tests that measure limbs independently to identify potential effects of dominance.

2. Materials

2.1 Participants

Potential participants were recruited to participate from Midwestern University, USA. Eighty-three people made inquiries into participation in the study, but only 45 participants ultimately agreed to participate, and 40 participants completed the three weeks of testing. Potential participants were excluded from participation if they had recent injury to the arm, shoulder, back or neck within the last 12 months, or if they had a history of shoulder or spine surgery, upper extremity fracture, ongoing neck pain, neurological symptoms in any extremity, low back, hip or knee pain. Participants were included if they were aged 18-55 years and were willing to consent to the study. The forty subjects: 20 males/ 20 females with an age 25 ± 6 years, height 175 ± 10 cm, and mass 84 ± 17 kg read and signed a university approved consent form prior to participation in the three week study. Arm length, history of contact and overhead sport was recorded, as well as PENN Scores and levels of upper extremity activity (Table 1). This study was approved by the university institutional review board (IRB).

All participants were evaluated in a test-retest design of three tests, one week apart. Inter-rater reliability was examined between three raters (one expert and two novice) on Day 2, while intra-rater of the expert rater was examined across three time points. The expert rater has 24 years of experience as a physiotherapist and developed the SARTS battery of tests based on her clinical experience. The two novice raters were certified athletic trainers with one year of experience and were enrolled in a post-professional Masters athletic training program during data collection.

	Age (yrs) mean(sd)	Height (cm) mean(sd)	Weight (kgs) (mean, sd)	Arm length (cm) (mean, sd)	Shoulder Activity Score	PENN score
Males	25.25 (6.86)	182.01 (5.80)	85.97 (15.27)	79.25 (3.03)	12.50 (3.85)	99.00 (1.59)
Females	23.75 (4.78)	168.03 (6.48)	66.63 (11.41)	74.60 (4.15)	11.75 (2.42)	98.15 (3.70)
(yrs=years, cm =centimetres, kgs = kilograms						

Table 1: Demographic data of the 40 participants

2.2 Procedure

The test battery was comprised of four open chain and four closed chain tests that were randomly assigned for a participant using Randomization.com. The participant alternated between open and closed chain tests for their test session. Each participant performed the tests in their assigned order on subsequent testing days. All tests were performed for one minute, with a rest period between tests of between one to two minutes. The total number of repetitions performed was recorded as the total score on an excel database (Microsoft, Redwood, WA, USA). Open chain tests included Ball Abduction External Rotation (BABER), Drop Catches, Ball Taps and Overhead Snatch. Closed chain tests included Push-Up Claps, Line Hops, Modified Closed Kinetic Chain Upper Extremity Stability Test (MCKCUEST), and Side Hold Rotations. Before each of the tests, the rater explained and demonstrated the tests. The participant had an opportunity to familiarize themselves by performing three to five repetitions of the tests before recording began.

These clinical tests were developed to replicate the shoulder demands of athletes engaged in sport and designed to assess readiness to return to sport. The International Olympic Committee has recommended that pre-participation clinical tests use minimal equipment, are inexpensive, reproducible and easy to perform (Ljungqvist et al., 2009). A fixed load independent of body weight has been used previously in upper extremity performance tests (Borms & Cools, 2018; Borms, Maenhout, & Cools, 2016; Negrete et al., 2010). The single load of 3 and 5kg and Swiss ball size of 55cm was chosen to increase the clinical utility of the tests, and enable comparison across sexes, levels of performance and different sports. *2.3 Technical equipment*

A 5kg medicine ball (Body Sport, Hudson, OH, USA) was used for the Overhead Snatch, and a 3kg medicine ball (Body Sport, Hudson, OH, USA) was used for the BABER. A 55cm Swiss Ball (Hygenic Corp, Akron, OH, USA) was used for the ball taps. Push-Up Claps were performed on a mat 2.5cm thick, while Line-hops and MCKCUEST were performed on a mat 0.5cm thick. All tests were performed for one minute, recorded with a stopwatch and a hand-held counter was used to count the number of repetitions. Tests were video-recorded by one rater, and the remaining two raters used the video recordings to independently count each test. Raters watched the video one time only at standard speed to simulate the live testing environment.

2.4 Open chain tests

2.4.1 Ball Abduction External Rotation

The participant stood holding a 3kg medicine ball at their shoulder with their elbow bent. (Figure 1a) They extended their elbow at 90 degrees of shoulder abduction until full extension (Figure 1b). The ball then returned to the shoulder (Figure 1c), before the arm was extended overhead to 180 degrees of flexion (Figure 1d) and then return to the shoulder (Figure 1e). Repetitions were not counted if the elbow was not fully extended, if the hand was not returned to the shoulder at the mid position, or if they dropped the ball. This test was repeated for each individual limb.

2.4.2 Drop Catches

The participant began holding a tennis ball with the shoulder abducted to 90 degrees and elbow flexed to 90 degrees (Figure 1f). They then drop the tennis ball and quickly catch the ball by twisting their shoulder from an externally rotated position to an internally rotated position. If the participant dropped the ball, did not return to the start position, did not keep their elbow at shoulder height, or did not keep their elbow flexed at 90 degrees, the repetition was not counted. This test was repeated for each individual limb.

2.4.3 Ball Taps

The participant began with the ball on the wall at 180 degrees of shoulder abduction. They bounced the ball on the wall to 90 degrees of abduction and returned the ball to the overhead position (Figure 1g). Repetitions were not counted if they dropped the ball, or did not complete the movement to 90 degrees abduction and back to the vertical line. This test was repeated for each individual limb.

2.4.4 Overhead Snatch

The participant began in a squat position holding onto a 5kg medicine ball (Figure 1h). They lifted the ball overhead to maximum hip, knee and elbow extension (Figure 1i), before returning the ball to the floor between their feet. Repetitions were not counted if the participant did not fully extend their elbows, if they let go of the ball or did not touch the ball onto the floor between repetitions.

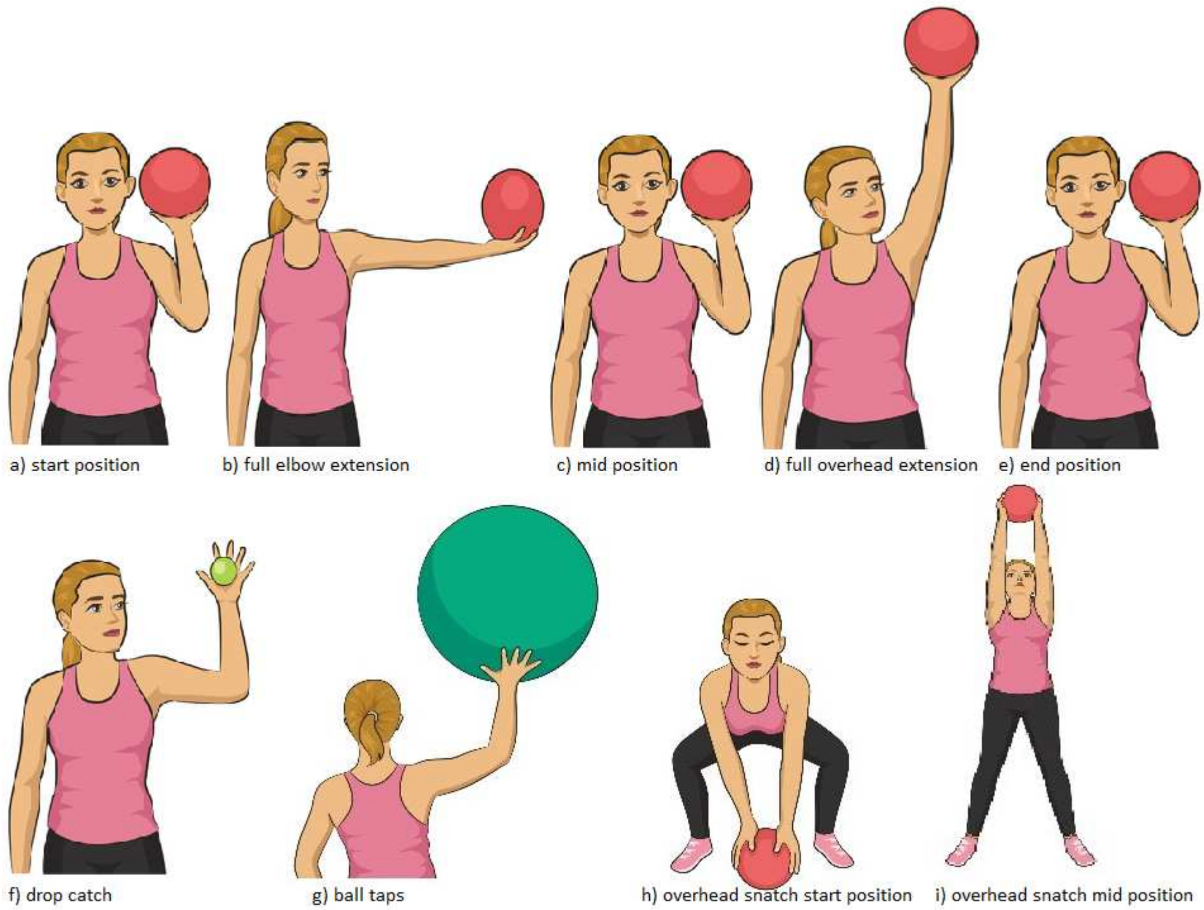


Figure 1: Open chain physical performance tests

2.5 Closed chain tests

2.5.1 Push-Up Claps

The participant began in a push-up position (Figure 2a) and performed an explosive push-up to clap their hands while in flight (Figure 2b) and returned their hands to the mat (Figure 2c). Repetitions were not counted if the participant did not clap their hands, or was unable to return their hands to the mat before their body.

2.5.2 Line Hops

Pilot testing revealed that it was impossible for many athletes to perform line hops balanced on their feet as in the side hold rotations, push-up claps, and MCKUEST tests. Therefore, to increase the clinical utility of this test, and ensure safety, this test is performed with the weight balanced on the knees and hands. The participant began by balancing on their knees with their right hand on the right side of a 3inch strip of athletic tape (Figure 2d). With hips extended, they hopped their hand over the tape (Figure 2e) and back (Figure 2f) to perform one repetition. Repetitions were not counted if the participant's hand landed on the athletic tape, if the participant was unable to complete the task across the line and back or if the participant was unable to maintain their balance during the task. This test was repeated for each individual limb.

2.5.3 Side Hold Rotations

The participant began in a side hold position, their weight balanced on one hand and both feet (Figure 2g). The participant then rolled their body onto their toes with pelvis parallel with the floor, crossed the top hand over the line between the placed hand and the feet (Figure 2h), returned to the start position to complete the repetition (Figure 2i). Repetitions were not counted if the moving hand supported their weight on the ground, if their hips did not rotate to parallel to the floor, if they did not roll onto their toes during the movement, or they did not return to the start position in full horizontal extension between repetitions. This test was repeated for each individual limb.

2.5.4 Modified Closed Kinetic Chain Upper Extremity Stability Test

The participant started with hands on the outside of a tape 36 inches apart (Figure 2j), weight supported on their feet. Maintaining this plank position, they were instructed to reach one hand across to tap the back of the supporting hand (Figure 2k), and return the moving hand to the start position (Figure 2l). Repetitions were not counted if the participant was unable to complete the movement, or if their hand landed on, or within the tape. This test has been previously described as recorded for 15 seconds with 45 seconds rest between three testing sessions (Goldbeck & Davies, 2000). The test was modified to perform the maximum number of repetitions over a one minute period in order to examine the endurance aspect of this test.

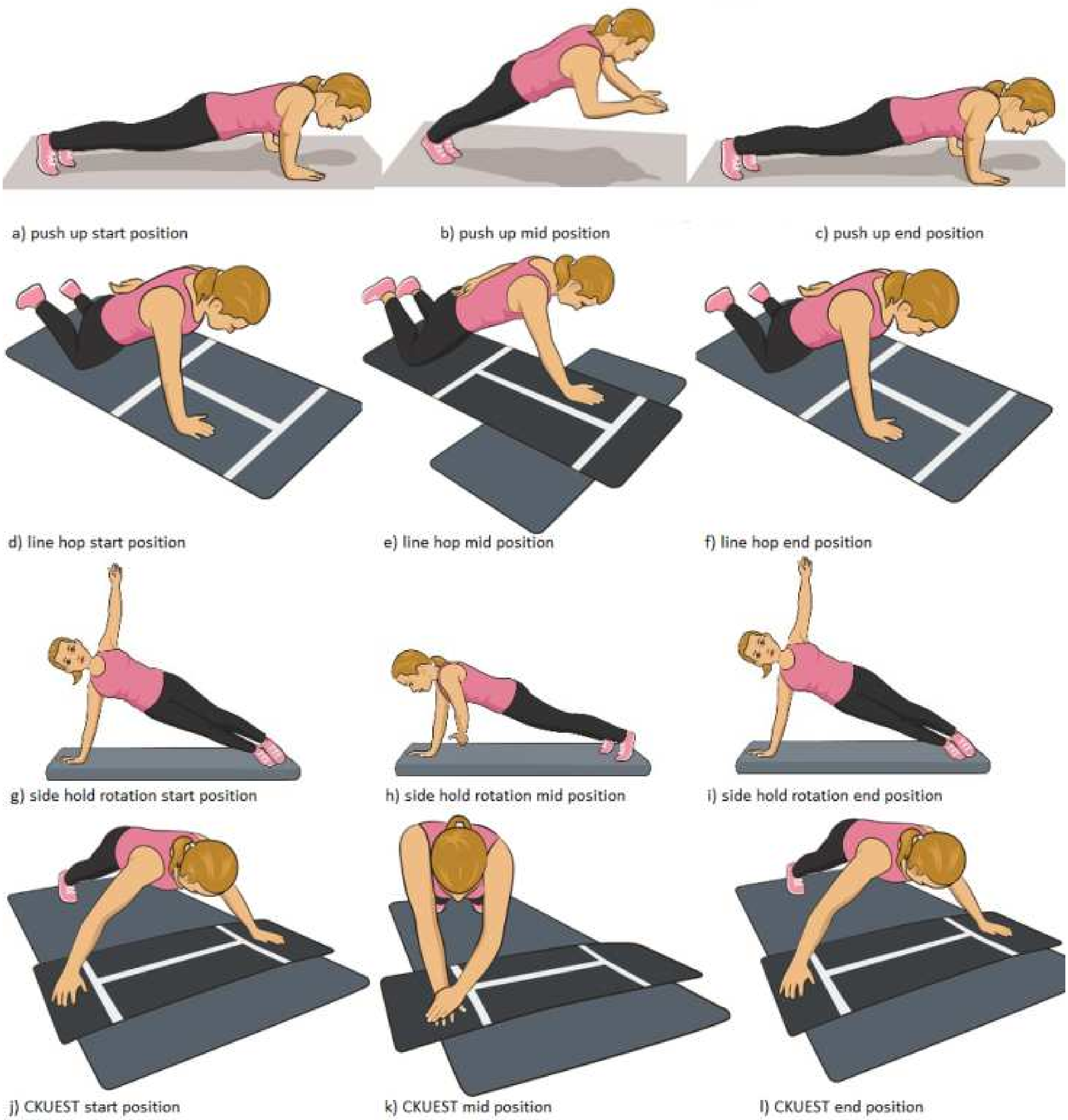


Figure 2: Closed chain physical performance tests

2.6 Statistical analyses

More than thirty participants were required to achieve a fair level of reliability (Mokkink et al., 2010). Analysis for a learning effect was examined using the expert rater data only across three days of testing using a repeated measure analysis of variance (ANOVA) for each of the 13 tests. A Bonferroni post-hoc analysis was used to compare for difference between days if necessary. To analyze intra-rater reliability, the intraclass correlation coefficient ($ICC_{2,1}$) of each rater across three days were calculated. Measurement error and responsiveness of the tests was determined using standard error of measure (SEM) and minimal detectable change (MDC) with a 90% confidence interval (Denegar & Ball, 1993). Inter-rater reliability was assessed using the same approach as above with all three raters counts compared on the second day of testing. Limb symmetry was calculated from mean scores across participants on Day 2 as recorded by the expert rater. Differences between limbs were compared with a paired t-test with alpha set at 0.05. All calculations were done using SPSS Statistics (version 23.0 for Windows, IBM Corp, Armonk, NY).

3. Results

All 40 participants were right hand dominant and all subjects reported high PENN shoulder scores (Male mean (sd)=99.00 (1.59) and female mean (sd)=98.15 (3.70)) indicating normal function of the upper extremity. Four tests (MCKCUEST, Overhead Snatch, and dominant and non-dominant Line Hops) showed no significant difference across the three days of testing (Table 2). However, post-hoc pairwise comparisons showed all other components of the SARTS required at least one testing session for familiarization (Table 2). Excellent reliability between raters (Table 3) and within raters on Day 2-3 (Table 4) was found for all tasks. Two of the five tests which assessed limb symmetry (BABER (91%) and Drop Catches (93%)) showed significantly decreased scores on the non-dominant side (p<0.01). Ball Taps were decreased on the non-dominant side (93%), but this difference was not significant (p=0.06).

Test	Day 1 (mean (SD))	Day 2 (mean (SD))	Day 3 (mean (SD))	p value
SIDEHOLDROTN_NONDOM	16 (6)	18 (5)	18 (5)	p=0.02 ɣ
SIDEHOLDROTN_DOM	16 (6)	18 (5)	18 (5)	p<0.01 ɣ
LINEHOP_NONDOM	22 (9)	23 (9)	24 (9)	P=0.43
LINEARMHOP_DOM	22 (8)	23 (8)	23 (8)	p=0.41¥
MCKCUEST	22 (17)	23 (17)	24 (16)	p=0.23¥
PUSHUPCLAPS	13 (9)	14 (9)	15 (10)	p<0.01¥*ɣ
BABER_NONDOM	13 (5)	14 (4)	15 (4)	p<0.01* ɣ
BABER_DOM	15 (5)	16 (5)	16 (5)	p<0.01ɣ
DROPCATCH_NONDOM	44 (10)	48 (13)	54 (11)	p<0.01*□ ɣ
DROPCATCH_DOM	46 (10)	52 (12)	56 (13)	p<0.01¥*□ ɣ
BALLTAPS_NONDOM	5 (3)	6 (3)	7 (4)	p<0.01¥* ɣ
BALLTAPS_DOM	6 (3)	7 (3)	8 (4)	p<0.01¥*□ ɣ
OHSNATCH	24 (6)	24 (5)	24 (5)	p=0.54¥

*significant learning effect between days 1-2, □significant learning effect between days 2-3, ɣsignificant learning effect between days 1-3, ¥indicates Greenhouse-Geisser correction applied. NONDOM=non-dominant limb, DOM = dominant limb

Table 2: Learning effect across all three days for expert rater

TASK	Rater 1 Mean (SD)	Rater 2 Mean (SD)	Rater 3 Mean (SD)	ICC	95% CI	SEM	MDC90
SIDEHOLDROTN_NONDOM	18 (5)	18 (6)	17 (6)	0.99	0.98, 0.99	1	1
SIDEHOLDROTN_DOM	18 (5)	18 (5)	18 (5)	0.97	0.94, 0.98	1	1
LINEHOP_NONDOM	23 (9)	23 (8)	22 (8)	0.99	0.98, 0.99	1	1
LINEHOP_DOM	23 (8)	23 (7)	23 (8)	0.98	0.96, 0.99	1	2
MCKCUEST	23 (17)	24 (17)	23 (17)	0.99	0.98, 1.00	2	2
PUSHUPCLAPS	15 (9)	15 (9)	15 (9)	0.98	0.96, 0.99	1	1
BALLABER_NONDOM	14 (4)	14 (5)	14 (5)	0.97	0.96, 0.98	1	1
BALLABER_DOM	16 (5)	16 (4)	16 (4)	0.94	0.89, 0.96	1	2
DROPCATCH_NONDOM	48 (13)	47 (13)	47 (13)	0.97	0.95, 0.98	2	3
DROPCATCH_DOM	52 (12)	51 (13)	51 (12)	0.97	0.94, 0.98	2	3
BALLTAPS_NONDOM	6 (3)	6 (3)	6 (3)	0.93	0.89, 0.96	1	1
BALLTAPS_DOM	7 (3)	7 (3)	7 (3)	0.96	0.94, 0.98	1	1
OHSNATCH	24 (5)	23 (5)	24 (5)	0.98	0.97, 0.99	1	1

NONDOM=non-dominant limb, DOM = dominant limb

Table 3: Inter-rater reliability between three raters on Day 2 of testing

TASK	DAY2 Mean(sd)	DAY3 Mean(sd)	ICC	95%CI	SEM	MDC	MDC90
SIDEHOLDROTN_NONDOM	18(6)	18(5)	0.80	0.65, 0.89	2	3	5
SIDEHOLDROTN_DOM	18(5)	19(5)	0.80	0.65, 0.89	2	3	5
LINEHOP_NONDOM	23(8)	24(8)	0.78	0.63, 0.88	4	6	9
LINEHOP_DOM	23(7)	23(8)	0.90	0.82, 0.95	3	4	6
MCKCUEST	23(17)	24(16)	0.95	0.90, 0.97	4	5	0
PUSHUPCLAPS	15(9)	15(10)	0.96	0.92, 0.98	2	3	4
BALLABER_NONDOM	14(4)	15(4)	0.88	0.80, 0.94	2	2	4
BALLABER_DOM	16(5)	16(5)	0.92	0.86, 0.96	1	2	3
DROPCATCH_NONDOM	48(13)	54(11)	0.87	0.77, 0.93	4	6	10
DROPCATCH_DOM	52(12)	56(13)	0.94	0.88, 0.97	3	4	7
BALLTAPS_NONDOM	6(3)	7(4)	0.80	0.66, 0.89	2	2	4
BALLTAPS_RIGHT	7(3)	8(4)	0.90	0.82, 0.95	1	1	2
OHSNATCH	24(5)	24(5)	0.90	0.83, 0.95	2	2	4

NONDOM=non-dominant limb, DOM = dominant limb

TABLE 4: Intra-rater reliability for rater 1 between Days 2 and 3

TEST	Dominant Side Mean	Non-Dominant Mean	Percentage Difference (nondom/dom)	p value
SIDEHOLDROTN	18 (5)	18 (5)	100%	p=0.90
LINEHOP	23 (8)	23 (9)	100%	p=1.00
BALLABER	16 (5)	14 (4)	91%	p<0.01
DROPCATCH	52 (12)	48 (13)	93%	p<0.01
BALLTAPS	7 (3)	6 (3)	93%	p=0.06

Table 5: Limb symmetry of unilateral tests

4. Discussion

The SARTS test battery contains reliable physical performance tests that may assist clinicians with clinical decision making regarding return to sport. Repeated measures analysis of variance indicated that Overhead Snatch, Line Hops and MCKCUEST demonstrated no learning effect across the three days (Table 2). Therefore, these tests can be used to benchmark athletes without prior practice. Other tests showed significant differences between Days 1-2. 2-3 and 1-3 (Table 2). Side Hold Rotations, BABER and Push-Up Claps demonstrated high levels of reliability (ICC>0.80) between days 2-3 which indicates that these three tests can be used to benchmark athletic performance after one familiarization session. While the SARTS test battery showed excellent inter-rater reliability, two tests (Drop Catches and Ball Taps) demonstrated a significant difference between all three days of testing indicating that learning is still acquired across the three sessions. Thus Drop Catches and Ball Taps require more than three sessions of practice before they are stable to use in clinical practice. While they may be useful clinical exercises, we do not recommend using these two tests to benchmark athletic performance. Further work is required to establish tests which measure coordination and neuromuscular factors which are stable across time.

The use of unilateral tests enables clinicians to compare the affected limb with the unaffected limb. Knowledge that BABER scores in the non-dominant limb are 91% of the dominant limb will enable clinicians to confidently report differences between dominant and non-dominant limbs. Adaption of the Overhead Snatch to examine limbs independently may be advantageous to establish limb symmetry in the overhead position. Other tests which evaluate limb symmetry in positions which replicate throwing and overhead sports are required.

The reliability of the SARTS battery compares favorably with the psychometrics of other shoulder physical performance measures (Ashworth et al., 2018; Goldbeck & Davies, 2000; Negrete et al., 2010). However, many other studies examine the tests across two days and therefore are unable to show the presence of a learning effect (Ashworth et al., 2018; Falsone, Gross, Guskiewicz, & Schneider, 2002; Negrete et al., 2010), or only examine intra-rater and not inter-rater-reliability. The number of participants in our study is similar to that of Negrete et al.(2010) who had the necessary numbers to establish a fair level of reliability (Mokkink et al., 2010). The intra-rater reliability results from this study, show excellent reliability (ICC=0.78) and are similar to the intra-rater reliability of the Upper Quarter Y Balance test (Gorman et al., 2012). However, they are less than that published by Ashworth et al. (2018) (ICC=0.96 to 0.98) which measured strength using a force platform and Harris et al.(2011) (ICC=0.99) which measured power of older adults using distance of a medicine ball throw.

Negrete et al.(2010) reported the reliability and MDC for three tests of upper extremity function and power. The reliability of a seated shot put test ranged between 0.98 and 0.97, while modified pull-ups and push-ups performed over 15 seconds were 0.99 and 0.96 respectively. The lower levels of reliability reported in the current study compared with Negrete et al.(2010) may be due to increased variation from performing the tests over 60 seconds rather than 15 seconds. Large variation was also seen in the MCKCUEST scores, and the average values reported in this study over 60 seconds were similar to that of Goldbeck and Davies (2000), who performed the test over 15 seconds. The discrepancy in tests scores over 15 and 60 seconds and the variation of the MCKCUEST scores may be due to sex differences, as women had more difficulty performing this test. While the primary purpose of this study was not to examine sex difference between the SARTS battery of tests, it was noted that women had more difficulty performing the MCKCUEST compared with men. Women in this study had shorter arm length compared with men. However, the distance for the MCKCUEST was not adapted accordingly. Further examination of the sex differences in the MCKCUEST is required and may result in establishing a shorter distance for women due to their decreased arm length.

One of the strengths of this study is the utilization of tests which examine physical performance over an extended time period. Many sports involve athletes to participants for extended periods of time, yet previous clinical tests are single one-off tests to test max power/strength or distance (Ashworth et al., 2018; Harris et al., 2011; Negrete et al., 2010). The utilization of a test battery which examines the endurance capability of athletes may more closely replicate sporting activity and provide a clearer indication of readiness to return to sport. Other strengths of this study include testing over three weeks, and a sample size of 40 participants.

Limitations of the current study include testing reliability in a healthy unaffected population. While the reliability may have been different if tested in pathological populations, increased heterogeneity of the sample due to pathological differences may have also increased the measurement variability. For this reason, use of a population with healthy shoulders was deemed to be necessary. However, it may be that the values demonstrated in this study, may not hold true.

While the current study shows reliability of the SARTS test battery, further studies are required to investigate the kinetics, kinematics and muscle activity involved in performing these tests. Additionally, while the means and standard deviations presented in this study represent healthy active college students in the USA, it may be that different sporting populations present with different normative data. Further studies which provide bench-marked levels of ability may be required in respective sports. Future studies to improve the reliability of the Ball Taps may include examination of the effect of ball size, arm length and horizontal extension angle. Finally, inclusion of a test which examines forced shoulder external rotation may improve the SARTS battery as this is a common position of anterior shoulder dislocation (Maki et al., 2017).

5. Conclusion

Six of the eight tests in the SARTS battery of tests are reliable measures which can be used by clinicians to guide return to sport. Some tests (BABER, Side Hold Rotations and Push-Up Claps) require practice before clinical testing, while two tests (Drop Catches and Ball Taps) continued to show learning across the three days of testing. Further examination of the tests in sport-specific populations is required to develop normative guidelines. Relationships between the clinical tests, and known valid measures or strength, power and neuromuscular control are required.

References:

Ardern, C. L., Bizzini, M., & Bahr, R. (2016). It is time for consensus on return to play after injury: Five key questions. *British Journal of Sports Medicine*, *50*(9), 506–508. https://doi.org/10.1136/bjsports-2015-095475

Ashworth, B., Hogben, P., Singh, N., Tulloch, L., & Cohen, D. D. (2018). The Athletic Shoulder (ASH) test: Reliability of a novel upper body isometric strength test in elite rugby players. *BMJ Open Sport and Exercise Medicine*, *4*(1), 1–6. https://doi.org/10.1136/bmjsem-2018-000365

Borms, D., & Cools, A. (2018). Upper-Extremity Functional Performance Tests: Reference Values for Overhead Athletes. *International Journal of Sports Medicine*, *39*(06), 433–441. https://doi.org/10.1055/a-0573-1388

Borms, D., Maenhout, A., & Cools, A. (2016). Upper quadrant field tests and isokinetic upper limb strength in overhead athletes. *Journal of Athletic Training*, *51*(10), 789–796. https://doi.org/10.4085/1062-6050-51.12.06

Chmielewski, T. L., Martin, C., Lentz, T. A., Tillman, S. M., Moser, M. W., Farmer, K. W., & Jaric, S. (2014). Normalization considerations for using the unilateral seated shot put test in rehabilitation. *Journal of Orthopaedic & Sports Physical Therapy*, *44*(7), 518–524. https://doi.org/10.2519/jospt.2014.5004

Cho, N. S., Hwang, J. C., & Rhee, Y. G. (2006). Arthroscopic Stabilization in Anterior Shoulder Instability : Collision Athletes Versus Noncollision Athletes, *22*(9), 947–953. https://doi.org/10.1016/j.arthro.2006.05.015

Cools, A., Struyf, F., De Mey, K., Maenhout, A., Castelein, B., & Cagnie, B. (2014). Rehabilitation of scapular dyskinesis: From the office worker to the elite overhead athlete.

British Journal of Sports Medicine, *48*(8), 692–697. https://doi.org/10.1136/bjsports-2013-092148

Creighton, D. W., Shrier, I., Shultz, R., Meeuwisse, W. H., & Matheson, G. O. (2010). Return-to-Play in Sport : A Decision-based Model. *Clin J Sport Med*, *20*(5), 379–385.

Denegar, C. R., & Ball, D. W. (1993). Assessing Reliability and Precision of Measurement: An Introduction to Intraclass Correlation and Standard Error of Measurement Intraclass Correlation and Standard Error of Measurement. *Journal of Sport Rehabilitation*, *2*, 35–42.

Falsone, S. A., Gross, M. T., Guskiewicz, K. M., & Schneider, R. A. (2002). One-Arm Hop Test: Reliability and Effects of Arm Dominance. *Journal of Orthopaedic & Sports Physical Therapy*, *32*(3), 98–103. https://doi.org/10.2519/jospt.2002.32.3.98

Goldbeck, T. G., & Davies, G. J. (2000). Test-Retest Reliability of the Closed Kinetic Chain Upper Extremity Stability Test: A Clinical Field Test. *Journal of Sport Rehabilitation*, *9*(1), 35–45. https://doi.org/10.1123/jsr.9.1.35

Gorman, P. P., Butler, R. J., Plisky, P. J., Orman, P. G., Utler, R. J. B., Lisky, P. J. P., … 1, 3 AND KYLE B. KIESEL1, 3. (2012). Upper quarter Y balance test: Reliability and performance comparison between genders in active adults. *Journal of Strength and Conditioning Research*, *26*(11), 3043–3048. https://doi.org/10.1519/JSC.0b013e3182472fdb

Harris, C., Wattles, A., Debliso, M., Sevene-Adams, P., Berning, J., & Adams, K. (2011). The seated medicine ball throw as a test of upper body power in older adults. *Journal of Strength & Conditioning Research*, *25*(8), 2344–2348.

Headey, J., Brooks, J. H. M., & Kemp, S. P. T. (2007). The Epidemiology of Shoulder Injuries in English Professional Rugby Union. *The American Journal of Sports Medicine*, *35*(9), 1537–

1543. https://doi.org/10.1177/0363546507300691

Ljungqvist, A., Jenoure, P., Engebretsen, L., Alonso, J. M., Bahr, R., Clough, A., … Thill, C. (2009). The International Olympic Committee (IOC) Consensus Statement on periodic health evaluation of elite athletes March 2009. *British Journal of Sports Medicine*, *43*(9), 631–643. https://doi.org/10.1136/bjsm.2009.064394

Maki, N., Kawasaki, T., Mochizuki, T., Ota, C., Yoneda, T., Urayama, S., & Kaneko, K. (2017). Video Analysis of Primary Shoulder Dislocations in Rugby Tackles. *Orthopaedic Journal of Sports Medicine*, *5*(6), 1–7. https://doi.org/10.1177/2325967117712951

Mokkink, L. B., Terwee, C. B., Knol, D. L., Stratford, P. W., Alonso, J., Patrick, D. L., … Vet, H. C. W. De. (2010). The COSMIN checklist for evaluating the methodological quality of studies on measurement properties : A clarification of its content. *BMC Medical Research Methodology*, *10*(22), 1–8.

Negrete, R. J., Hanney, W. J., Kolber, M. J., Davies, G. J., Ansley, M. K., McBride, A. B., & Overstreeet, A. L. (2010). Reliability, Minimal Detectable Change, and Normative Values for Tests of Upper Extremity Function and Power. *Journal of Strength & Conditioning Research*, *24*(12), 3318–3325.

Torrance, E., Clarke, C. J., Monga, P., Funk, L., & Walton, M. J. (2018). Recurrence After Arthroscopic Labral Repair for Traumatic Anterior Instability in Adolescent Rugby and Contact Athletes. *The American Journal of Sports Medicine*, 036354651879467. https://doi.org/10.1177/0363546518794673

Wheeler, J. H., Ryan, J., Arciero, R. A., & Milinari, R. N. (1989). Arthroscopic versus nonoperative treatment of acute shoulder dislocations in young athletes. *Arthroscopy - Journal of Arthroscopic and Related Surgery*, *5*(3), 213–217.

FUNDING: This work was supported by University of XX [Grant Number UL1TR001998]

Highlights

- 6/8 tests in the SARTS test battery are reliable and can be used by clinicians to guide return to sport
- 3 tests (BABER, Side Hold Rotations &Push-Up Claps) require a familiarisation session before use
- 2 tests (Drop Catches and Ball Taps) show a learning effect across the 3 testing sessions