

Original Research

Reliability of Upper Extremity Functional Performance Tests for Overhead Sports Activities

Bryan L Riemann^a, Kevin E Wilk, George J Davies

Keywords: Shoulder, Return to Sport, Limb Symmetry Index, Seated Single Arm Shot Put Test

<https://doi.org/10.26603/001c.74368>

International Journal of Sports Physical Therapy

Vol. 18, Issue 3, 2023

Background

There is lack of consensus on which tests, particularly upper extremity functional performance tests (FPT) that should be used for clinical decision making to progress a patient through a rehabilitation program or criteria for return to sport (RTS). Consequently, there is a need for tests with good psychometric properties that can be administered with minimal equipment and time.

Purpose

(1) To establish the intersession reliability of several open kinetic chain FPT in healthy young adults with a history of overhead sport participation. (2) To examine the intersession reliability of the limb symmetry indices (LSI) from each test.

Study Design

Test-retest reliability, single cohort study.

Methods

Forty adults (20 males, 20 females) completed four upper extremity FPT during two data collection sessions three to seven days apart: 1) prone medicine ball drop test 90°shoulder abduction (PMBDT 90°), 2) prone medicine ball drop test 90°shoulder abduction/90° elbow flexion (PMBDT 90°-90°), 3) half-kneeling medicine ball rebound test (HKMBRT), 4) seated single arm shot put test (SSASPT). Measures of systematic bias, absolute reliability and relative reliability were computed between the sessions for both the original test scores and LSI.

Results

Except for the SSASPT, all tests demonstrated significant ($p \leq 0.030$) improvements in performance during the second session. Generally, for the medicine ball drop/rebound tests, the absolute reliability was the highest (less random error) for the HKMBRT, next the PMBDT 90° followed by PMBDT 90°-90°. Excellent relative reliability existed for the PMBDT 90°, HKMBRT, and SSASPT, whereas fair to excellent relative reliability for the PMBDT 90°-90°. The SSASPT LSI revealed the highest relative and absolute reliability.

Conclusion

Two tests, HKMBRT and SSASPT demonstrated sufficient reliability; therefore, the authors' recommend those tests can be used for serial assessments to advance a patient through a rehabilitation program as well as criteria for progression to RTS.

^a Corresponding Author:
Bryan L Riemann, PhD, ATC
Biodynamics and Human Performance Center
Georgia Southern University- Armstrong Campus
11935 Abercorn Street
Savannah, GA 31419
briemann@georgiasouthern.edu

Level of Evidence

3

INTRODUCTION

Despite comparable injury rates between the shoulder and knee, particularly in overhead sports, much less research attention has been paid to the return to sport (RTS) clinical decision-making process following shoulder injury and surgery. For example, there are over twice as many references indexed by [PubMed.gov](https://pubmed.ncbi.nlm.nih.gov/) (searches conducted 11/26/22) concerning RTS for the knee ([return to sport] AND [knee]=1988 results) compared to the shoulder ([return to sport] AND [shoulder]=914 results). Based upon several recent systematic reviews and meta-analyses,¹⁻⁴ the success of RTS and return to previous level of play following shoulder surgery is highly variable between investigations. While overall RTS ranges between 62.7% to 100%,¹⁻⁵ those seeking to return to competitive overhead sports demonstrate less success at returning to the same or higher level of play.^{1,3,4} In addition to level of play, the specific injury and management (i.e., conservative versus surgical), age of athlete, and duration of follow up also likely contribute to the highly variable RTS estimates.

While it is certainly multi-factorial as to why and how an athlete returns to sport following shoulder injury or surgery, the complexity of RTS clinical decision-making is complicated by the lack of consensus or agreed upon metrics regarding the criteria that should be used.⁶⁻⁹ Systematic reviews and meta-analyses consistently report time following surgery as the most commonly used criteria for RTS.^{3,10-12} A few studies have used strength or range of motion, either in isolation, or in conjunction with time as RTS criteria.¹⁰⁻¹² While time post-surgery would account for biological tissue healing with strength and range of motion being important markers of preliminary functional recovery, there appears to be a gap between the RTS criteria being used and whether a patient truly has the necessary recovery prerequisites to being ready for more rigorous sport specific activities and return to prior levels of performance.

The most recent consensus statement regarding RTS following shoulder injury recommends sport specific upper extremity tests be used with RTS decision making.⁹ Interestingly several of the tests suggested are bilateral tasks, while others do not replicate common movement patterns or demands imposed by many sport activities (e.g., the Y-Balance test) to provide insight about the readiness to RTS. In addition to many sports involving unilateral activities, the use of bilateral tasks may permit compensation by the unaffected limb, thereby masking underlying performance deficiencies. Specific to overhead athletes is the function of both the anterior and posterior musculature with the shoulder in an abducted position. First suggested by Wilk et al. for assessments with overhead athletes,¹³ and included in the Bern Consensus,⁹ are the prone medicine ball drop test (PMBDT) and wall throws test. Despite the apparent ecological validity of these assessments to the overhead athlete, reliability metrics remain unknown.

The seated single arm shot put test (SSASPT) is an additional unilateral upper extremity functional performance test that has been studied in both healthy individuals¹⁴⁻¹⁸ and patients with shoulder pain or injury.¹⁹⁻²¹ Similar to other unilateral functional performance tests, in circumstances of unilateral pathology, a limb symmetry index (LSI) can be computed to provide a unitless metric to interpret a patient's performance thereby avoiding many of the confounding issues with normative data comparisons. Previous literature has reported SSASPT LSI in healthy persons ranges between 103 to 111%,^{14-18,20} favoring the dominant limb. Additionally, a previous investigation¹⁷ demonstrated that, except for greater release velocities for the dominant limb, which were attributable to underlying strength and power differences, healthy persons perform the SSASPT with similar underlying projection mechanics between the dominant and nondominant limbs. Most recently,¹⁹ the SSASPT LSI was demonstrated to have minimal association with a patient reported outcome measure (QuickDASH) in patients being discharged from physical therapy following shoulder injury or surgery. This finding suggests that both subjective and objective measures are needed to attain a complete assessment of a patient's functional and perceived status. Furthermore, patients with nondominant shoulder involvement being discharged from rehabilitation exhibited lower SSASPT LSI than patients with dominant shoulder involvement.²¹ Surprisingly, despite all the attention the SSASPT LSI has received in the literature, to date, there have been no investigations regarding the reliability of the LSI metric.

The purpose of this investigation was to establish the intersession reliability of several open kinetic chain (OKC) functional performance tests (FPT) in healthy young adults with a history of overhead sport participation. Specifically studied were two variations of the PMBDT (shoulder abducted 90°, shoulder abducted 90°/elbow flexed 90°) and a variation of the wall throws test (half-kneeling medicine ball rebound test [HKMBRT]). To provide context for the reliability results of the three assessments, the SSASPT was also included. A secondary purpose was to examine the intersession reliability of the LSI computed from the SSASPT, HKMBRT and two PMBDT. The hypotheses were that all tests would demonstrate moderate to good reliability, with the reliability of LSI expected to be slightly less than the original scores.

METHODS

PARTICIPANTS

Forty healthy young adults (20 males, 20 females) with a history of overhead sport participation were recruited for the study. Prior to study participation, participants completed a demographic, injury and physical activity history form and the 2019 Physical Activity Readiness-Questionnaire. All forms were reviewed by a member of the inves-

tigative team to verify that each participant was appropriate for study participation. Inclusion criteria included being between 18 and 35 yrs old, meeting the American College of Sports Medicine criteria for being physically active,²² and participation in an overhead recreational or competitive sport for a minimum of one year. Participants were excluded if they had a previous history of cervical spine or upper extremity injury or surgery within a year prior of data collection, were deficient in the range of motion needed to adopt the upper extremity testing positions required to perform the tests or were unable to complete the tests as prescribed. This investigation was approved by a university institutional review board and all participants read and signed an approved informed consent form.

RESEARCH DESIGN

This study utilized a randomized repeated measures research design. Study participation required two testing sessions three to seven days apart. Testing sessions were scheduled at near identical times each day and participants were asked to avoid vigorous physical activity (e.g., upper extremity resistance or exercise training), 24 hours prior to each session. At each session, participants completed identical protocols. Each participant was randomly allocated a test and limb (dominant, nondominant) order; the order used for the first session was replicated for the participant's second session. At the beginning of each session, all participants completed a warm-up which consisted of arm circles forwards, backwards, and arm crosses. Each warm-up activity was completed for thirty seconds, for a total of three sets. After completion of the warm-up, participants were shown a pre-recorded demonstration video illustrating the four tests to be performed. Prior to the two PMBDT and HKMBRT tests, participants were given time to practice each task and demonstrate proficiency. Proficiency was defined as being able to repetitively and continuously perform the two PMBDT and HKMBRT tasks without hesitation between catches. Familiarization to the SSASPT procedures was a component of the four-trial gradient warm up described below. Additional time and cuing were provided to participants as needed. Between each test, participants completed two-minute rest periods.

TEST PROCEDURES

PRONE MEDICINE BALL DROP TEST AT 90° SHOULDER ABDUCTION (PMBDT 90°)

The PMBDT 90° ([Figure 1](#)) was performed with the participants lying prone on a treatment plinth with their testing shoulder abducted to 90°, elbow straight, and the forearm supinated (palm to floor).¹³ The non-testing arm was supported over the opposite side of the table. Participants were instructed to not grasp the table during testing. Participants dropped and caught a .91 kg medicine ball as many times as possible for thirty seconds.¹³ A mobile stool was placed under the hand to catch the medicine ball in circumstances in which the participants missed a catch. During the test participants were verbally instructed to maintain



Figure 1. Positioning for the performance of the Prone Medicine Ball Drop Test at 90° shoulder abduction.

the 90° shoulder abducted position. One trial was performed on each arm, with a 30s rest between each testing trial. The total number of catches recorded during the 30s trial served as the performance outcome metric.

PRONE MEDICINE BALL DROP TEST AT 90° SHOULDER ABDUCTION/90° ELBOW FLEXION (PMBDT 90°-90°)

The PMBDT 90°-90° ([Figure 2](#)) was performed with the participants lying prone on a treatment plinth with their testing shoulder abducted to 90°, the elbow flexed to 90°, and forearm supinated (palm to floor).²³ The non-testing arm was supported over the opposite side of the table. Participants were instructed to not grasp the table during testing. Participants dropped and caught a .91 kg medicine ball as many times as possible for thirty seconds.¹³ A mobile stool was placed under the hand to catch the medicine ball in circumstances in which the participants missed a catch. During testing, participants were verbally instructed to maintain the 90° shoulder abducted position and 90° elbow flexed position. One trial was performed on each arm, with a 30s rest between each testing trial. The total number of catches recorded during the 30s trial served as the performance outcome metric.

HALF-KNEELING MEDICINE BALL REBOUND TEST (HKMBRT)

The HKMBRT was slightly modified from the original description.¹³ Instead of being performed in a standing position, the participants assumed a half-kneeling position in a doorway to eliminate contributions from the legs ([Figure 3](#)). The testing shoulder was abducted to 90°, the elbow flexed to 90°, and forearm supinated (palm toward wall). The non-testing hand was placed on the ipsilateral knee.



Figure 2. Positioning for the performance of the Prone Medicine Ball Drop Test at 90° shoulder abduction/90° Elbow Flexion.



Figure 4. Starting position for the seated single arm shot put test.



Figure 3. Positioning for the performance of the Half-Kneeling Medicine Ball Rebound Test.

Participants threw and caught a .91kg medicine ball against the wall as many times as possible for 30s.¹³ One trial was performed on each arm, with a 30s rest between each testing trial. The total number of catches recorded during the 30s trial served as the performance outcome metric.

SEATED SINGLE ARM SHOT PUT TEST (SSASPT)

Like previous studies,¹⁶⁻¹⁸ the SSASPT began with the participants assuming a long sitting position on the floor against a wall with the non-testing hand in their lap. The test began with the participant holding a 2.0kg medicine ball in the palm of their hand while keeping their elbow adjacent to their torso.^{16,17} Participants were instructed to “put or press the medicine ball as hard as they could for greatest distance” while maintaining their back against the wall and avoid the test limb crossing torso midline. Furthermore, participants were cued to perform the movement with a pure concentric action without any preloading (stretch-shortening). Prior to performing three maximal effort trials, participants performed four gradient sub-maximal to maximal warm-up practice trials at 25%, 50%, 75% and 100% effort. The distance from the wall to the first location of medicine ball ground contact was measured and averaged across the three test trials.

STATISTICAL ANALYSIS

All statistical procedures were conducted using IBM SPSS Statistics for Windows, version 27 (IBM Corp., Armonk, NY, USA) and Microsoft Excel, version 16 (Microsoft Corporation, Redmond, WA, USA). Separate statistical analyses were conducted for the males and females. Exploratory analyses were conducted on the data from each session to identify potential erroneous data entry errors. Normality of the between session difference scores was examined using Q-Q plots and Shapiro-Wilks tests. Heteroscedasticity between sessions was examined by using the Bland-Altman method.²⁴ Systematic bias between testing sessions were

Table 1. Participant demographics and overhead sport participation.

	Females	Males
Age (yrs)	23.2 ± 1.7	24.5 ± 2.6
Height (m)	1.65 ± .08	1.81 ± .07
Mass (kg)	67.0 ± 8.6	85.0 ± 13.9
Sport		
Softball	11	4
Baseball	0	7
Swimming	1	1
Tennis	3	0
Volleyball	5	5
Football (quarterback)	0	3

evaluated using dependent t tests. Absolute reliability was determined by computing standard error of measurement (SEM),²⁵ 90% minimal detectable difference (MDD_{90%}), and coefficient of variation (CV).²⁶ Relative reliability was computed using intraclass correlation coefficients (ICC, model: 2,1). Additionally, limb symmetry indices (LSI) were computed ([dominant/nondominant] *100) for each test and subjected to similar systematic bias, absolute reliability, and relative reliability analyses. Coefficients of variation were considered acceptable when values were below 10%. The magnitude for the ICC were interpreted as follows: less than 0.40: poor, between 0.40 and 0.59: fair, between 0.60 and 0.74: good, and between 0.75 and 1.00: excellent.²⁷

RESULTS

The demographics of the subjects are presented in [Table 1](#).

The results of the exploratory analysis for normality revealed no significant ($p = 0.071$ to 0.927) departures for the difference scores. Results of the heteroscedasticity analysis yielded no significant relationships; with the exception of dominant PMBDT-90° for both the females (Kendall's tau=.202, $p=0.226$) and males (Kendall's tau=.304, $p=0.067$), all Kendall's Tau values were between -.146 to .145.

Descriptive statistics for test performance across the two sessions are provided in [Table 2](#). Except for the SSASPT for both sexes and limbs ([Table 3](#)), all tests demonstrated significant ($p \leq 0.030$) improvements in performance during the second session compared to the first session. Generally, for the three-medicine ball drop and rebound tests, the absolute reliability, whether expressed as catches (SEM, MDD_{90%}) or CV, was the highest (less random error) for the HKMBRT, followed by the PMBDT 90°, which in turn was followed by PMBDT 90°-90°. The ICC results demonstrated excellent relative reliability for the PMBDT 90°, HKMBRT, and SSASPT, whereas the relative reliability for the PMBDT 90°-90° ranged from fair to excellent.

Descriptive statistics for the LSI across the two sessions are provided in [Table 4](#). None of the LSI demonstrated significant systematic bias ($p > 0.202$) between sessions ([Table 5](#)). The absolute reliability (less random error) was the highest for the SSASPT LSI and the lowest for the

PMBDT 90°-90° LSI. Relative reliability for the SSASPT LSI ranged from good to excellent whereas the other three tests ranged from poor to fair.

DISCUSSION

This study sought to primarily examine the intersession reliability of four OKC FPT which could be used as criteria for rehabilitation program progression and RTS clinical decision making. The PMBDT 90° and PMBDT 90°-90° largely focused on the posterior glenohumeral musculature whereas HKMBRT and SSASPT were more anterior musculature focused. The results support the incorporation of the SSASPT and HKMBRT as potential tools into clinical decision making because of their excellent reliability. While the relative reliability of the PMBDT 90° was excellent, the absolute reliability bordered just beyond acceptability thresholds. The reliability of PMBDT 90°-90° was below acceptable thresholds. Thus, this investigation provides greater support for the two anterior musculature focused tests and less support for the current versions of the two posterior musculature centered tests. The secondary purpose of the investigation, examination of the intersession LSI reliability of the four OKC FPT, supports the SSASPT LSI as being a reliable metric. Compared to the SSASPT, larger changes in HKMBRT and PMBDT 90° LSI scores would be needed to be confident the changes are beyond measurement error.

Table 2. Descriptive statistics for each of the functional performance tests. The units associated with PMBDT 90°-90°, PMBDT 90°, and HKMBRT are medicine ball catches, whereas the units for the SSASPT are meters.

Test	Limb	Sex	Session 1 Mean ± SD	Session 2 Mean ± SD
PMBDT 90°-90°	D	F	50.4 ± 12.0	55.7 ± 13.5
		M	61.5 ± 9.9	68.3 ± 12.8
	ND	F	44.4 ± 12.8	51.0 ± 11.1
		M	58.1 ± 10.1	66.5 ± 10.4
PMBDT 90°	D	F	64.7 ± 13.4	71.6 ± 13.9
		M	74.7 ± 9.9	79.8 ± 12.3
	ND	F	58.9 ± 13.4	63.6 ± 14.3
		M	70.4 ± 9.9	76.7 ± 9.3
HKMBRT	D	F	69.0 ± 16.5	76.9 ± 14.4
		M	74.9 ± 12.6	82.7 ± 13.5
	ND	F	64.0 ± 12.9	71.0 ± 13.6
		M	70.1 ± 11.9	77.4 ± 13.0
SSASPT	D	F	3.35 ± 0.47	3.33 ± 0.39
		M	4.91 ± 0.67	4.83 ± 0.66
	ND	F	3.22 ± 0.55	3.19 ± 0.47
		M	4.63 ± 0.69	4.60 ± 0.61

PMBDT 90°-90°: prone medicine ball drop test in 90°-90° position; PMBDT 90°: prone medicine ball drop test in 90° position; HKMBRT: half-kneeling medicine ball rebound test; SSASPT: seated single arm shot put test; SD: standard deviation; D: dominant; ND: nondominant; F: female; M: male

Table 3. Results of the systematic bias, absolute reliability, and relative reliability analysis conducted on the functional performance tests. The units associated with PMBDT 90°-90°, PMBDT 90°, and HKMBRT are medicine ball catches, whereas the units for the SSASPT are meters.

Test	Limb	Sex	Systematic Bias		Absolute Reliability			Relative Reliability ICC (95% CI)
			Mean ± SD	P	SEM	CV %	MDD _{90%}	
PMBDT 90°-90°	D	F	5.3 ± 10.1	.030	7.1	17.6	16.7	.688 (.363-.863)
		M	7.0 ± 8.7	.002	6.1	10.3	14.3	.713 (.405-.876)
	ND	F	6.6 ± 7.4	<.001	5.2	14.3	12.2	.809 (.580-.920)
		M	8.4 ± 11.0	.003	7.8	13.0	18.2	.424 (-.011-.724)
PMBDT 90°	D	F	7.1 ± 9.1	.003	6.5	10.0	15.1	.775 (.515-.904)
		M	5.1 ± 7.9	.010	5.6	7.3	13.1	.748 (.466-.892)
	ND	F	4.7 ± 8.4	.021	5.9	10.7	13.8	.818 (.597-.924)
		M	6.3 ± 9.0	.005	6.4	9.1	13.1	.748 (.466-.892)
HKMBRT	D	F	7.9 ± 8.4	.001	5.7	9.4	13.9	.851 (.661-.938)
		M	7.8 ± 6.4	<.001	4.5	6.1	10.5	.881 (.724-.951)
	ND	F	7.0 ± 5.4	<.001	3.8	5.5	8.9	.917 (.803-.966)
		M	7.3 ± 6.5	<.001	4.6	6.5	10.7	.865 (.691-.944)
SSASPT	D	F	-.026 ± .238	.637	.170	5.3	.397	.848 (.657-.937)
		M	-.089 ± .199	.060	.141	3.0	.330	.956 (.892-.982)
	ND	F	-.034 ± .187	.423	.130	4.2	.304	.934 (.841-.973)
		M	-.037 ± .202	.426	.141	3.2	.330	.952 (.883-.981)

PMBDT 90°-90°: prone medicine ball drop test in 90°-90° position; PMBDT 90°: prone medicine ball drop test in 90° position; HKMBRT: half-kneeling medicine ball rebound test; SSASPT: seated single arm shot put test; SD: standard deviation; SEM: standard error of the measurement; CV: coefficient of variation; MDD_{90%}: 90% minimal detectable difference; ICC: intraclass correlation coefficient; CI: confidence interval; D: dominant; ND: nondominant; F: female; M: male

Table 4. Descriptive statistics for each of the limb symmetry indices (%) from the functional performance tests.

Test	Sex	Session 1	Session 2
		Mean ± SD	Mean ± SD
PMBDT 90°-90°	F	119.2 ± 32.0	110.2 ± 16.0
	M	106.2 ± 11.0	102.8 ± 12.4
PMBDT 90°	F	111.5 ± 14.2	115.5 ± 18.0
	M	106.4 ± 6.3	104.1 ± 10.6
HKMBRT	F	107.3 ± 12.1	108.6 ± 9.7
	M	107.3 ± 9.5	107.4 ± 11.2
SSASPT	F	104.7 ± 6.9	105.0 ± 7.6
	M	106.4 ± 8.0	105.3 ± 9.4

PMBDT 90°-90°: prone medicine ball drop test in 90°-90° position; PMBDT 90°: prone medicine ball drop test in 90° position; HKMBRT: half-kneeling medicine ball rebound test; SSASPT: seated single arm shot put test; SD: standard deviation; F: female; M: male

Intraclass correlation coefficients are the most frequently reported reliability statistics reported in the literature and reflect the stability of the rank or position of an individual relative to the group across a repeated assessment. Using the arbitrary threshold of .75 as having clinical meaningfulness, all tests across both sexes and limbs except for the PMBDT 90°-90° met the criteria. Also critical to fully interpreting ICC is consideration of the confidence interval precision and range, particularly the lower bound, around each ICC point estimate. Consistent with previous

literature,^{15,20,28,29} the ICC for the SSASPT were all >.85. Additionally, except for the dominant limb for females, the confidence interval lower bounds were above .75, further supporting the relative reliability of the SSASPT. The ICC values for the HKMBRT and PMBDT 90° also met the clinical meaningfulness threshold, although many of the lower confidence interval bounds reside in the fair to good range. Except for nondominant limb for females, the PMBDT 90°-90° ICC values were below clinical meaningfulness threshold, and the confidence intervals were very wide. Thus, likely because of the larger random error that was reflected by the absolute reliability statistics, the relative reliability for PMBDT 90°-90° was reduced.

Of the previous studies^{15,28,29} that considered relative and absolute reliability of the SSASPT, none appear to have considered systematic bias. Remarkably, the SSASPT was the only assessment in the current investigation void of significant repeated exposure changes. For the two PMBDT, practitioners can expect an average improvement between five to seven catches and for the HKMBRT between seven to eight catches that are attributable to completion of the test a second time.

Absolute reliability has the most pertinence to clinicians by providing an estimate of the expected random error associated with a test.³⁰ For example, when conducting serial assessments during rehabilitation to monitor patient progress, a patient's performance change must exceed the mean systematic bias plus an absolute reliability estimate (e.g., SEM, MDD) to definitively declare improvement. With the two PMBDT and the HKMBRT, patients must attain

Table 5. Results of the systematic bias, absolute reliability, and relative reliability analysis conducted on the limb symmetry indices (%) for each functional performance test.

Test	Sex	Systematic Bias		Absolute Reliability			Relative Reliability ICC (95% CI)
		Mean $\pm$ SD	P	SEM	CV %	MDD _{90%}	
PMBDT 90°-90°	F	-9.0 $\pm$ 30.5	.202	21.6	19.7	50.4	.271 (-.183-.630)
	M	-3.4 $\pm$ 14.6	.312	10.4	11.0	24.1	.219 (-.236-.596)
PMBDT 90°	F	4.0 $\pm$ 17.1	.307	12.1	10.6	28.2	.446 (.017-.737)
	M	-2.4 $\pm$ 11.7	.375	8.3	8.3	19.3	.106 (-.342-.515)
HKMBRT	F	1.3 $\pm$ 11.4	.619	8.1	7.8	18.8	.462 (.036-.746)
	M	0.1 $\pm$ 9.7	.960	6.9	6.0	16.1	.563 (.172-.801)
SSASPT	F	0.3 $\pm$ 5.7	.233	4.0	3.7	9.4	.697 (.378-.868)
	M	-1.2 $\pm$ 5.1	.323	3.6	3.3	8.3	.833 (.625-.930)

PMBDT 90°-90°: prone medicine ball drop test in 90°-90° position; PMBDT 90°: prone medicine ball drop test in 90° position; HKMBRT: half-kneeling medicine ball rebound test; SSASPT: seated single arm shot put test; SD: standard deviation; SEM: standard error of the measurement; CV: coefficient of variation; MDD_{90%}: 90% minimal detectable difference; ICC: intraclass correlation coefficient; CI: confidence interval; F: female; M: male

~seven additional catches (systematic bias) plus ~five catches (SEM) on a subsequent assessment to have some confidence that their improvement is not attributable to learning effects or measurement error. In contrast, for the SSASPT, because there was no significant systematic bias, patients would just need to exhibit ~.15m (SEM) improvement. The MDD is a more conservative estimate of measurement error and when changes exceed the magnitude of the MDD, clinicians can be extremely confident that the patient has experienced improvement. The MDD for the two PMBDT and HKBRT are rather large, ranging between nine and 17 catches. As a result, it is likely that patients may experience clinically meaningful changes that may not be reflected in test performance exceeding the MDD.³⁰

Coefficient of variation as an index of variability is particularly useful when data demonstrate heteroscedasticity. It is noteworthy that none of the tests assessed, for either limb or sex, in the current investigation demonstrated heteroscedasticity. An additional advantage of the CV is that because it is unitless, comparisons can be made between various instruments. It is important that such CV comparisons be made between CV derived from similar populations. For example, in the current study, for both the dominant and nondominant limbs, the males demonstrated higher SSASPT values than the females. If the variability between sessions was the same for both sexes, the CV would be higher for the females compared to the males because the two sessions mean (denominator) would be lower. Thus, it is most prudent to compare the CV between tests within each sex. Coefficient of variation value comparisons within each sex reveal similar results, with SSASPT having the lowest CV, followed by the HKBRT. With regards to the 10% CV acceptability threshold, only the PMBDT 90°-90° (both sexes and limbs) largely exceeded. The CV results, in conjunction with the large MDD, further challenges to the clinical utility of the current version of the PMBDT 90°-90° as a reliable test of functional performance.

Currently, there is a void of FPT focusing upon the posterior glenohumeral musculature to use during clinical decision making. Unfortunately, in the current investigation,

the two tests emphasizing glenohumeral posterior musculature revealed the lowest reliability, with the PMBDT 90°-90° demonstrating slightly worse reliability than the PMBDT 90°. Both tests involved a 30s effort of catching a medicine ball while maintaining a prone position. The movement pattern differences between the two tests likely explains the slight difference in reliability. The PMBDT 90° involved a horizontal abduction movement whereas the PMBDT 90°-90° was slightly more complex and required maintaining 90° abduction while performing external rotation movements to repeatedly catch and drop the medicine ball. Based on the current results, both tests could benefit from some revision aimed towards improving their reliability. Based on the authors' experience with the tests during the investigation, as well as qualitative comments made by some of the participants, exploring the use of two 15s trials rather than a single 30s trial is recommended. Some participants reported fatigue of their hand muscles occurring during the last ~10s affected their ability to catch and maintain the transitory grip of the medicine ball required for each repetition. As the goal of the two PMBDT is to concentrate on the posterior shoulder musculature, it is important that grip fatigue as a potential performance limitation be reduced. The authors expect that modifying the two PMBDT to be based upon the average number of catches across two 15s trials will greatly improve both relative and absolute reliability metrics.

Because LSI incorporates performance of the dominant and nondominant limbs simultaneously, it was expected that the LSI reliability metrics would be weaker than the individual dominant and nondominant limb metrics. Statistically, there were no LSI differences (i.e., systematic bias) between the two testing sessions across the four assessments for either the males or females; however, except for the SSASPT, there appeared to be a lot of between participant variability based on the standard deviations. Additionally, aside from the SSASPT, attaining LSI changes that exceed the MDD would likely be very difficult, particularly for the PMBDT 90° and PMBDT 90°-90°. Given the expectation for lower LSI reliability as described earlier, using the LSI

SEM as a threshold for identifying change exceeding measurement error in a patient's performance may be more appropriate. Based on this suggestion, achieving LSI changes in patients that exceed the LSI SEM for the HKMBRT and PMBDT 90° (males) seems reasonable. Like the individual limb reliability metrics, the authors expect the PMBDT 90° and PMBDT 90°-90° LSI reliability metrics to improve with the modification of the tests to include two 15s trials.

The LSI for the SSASPT were the smallest (i.e., closest to 100%) of the four tests indicating more similar performance between the dominant and nondominant limbs. Furthermore, the SSASPT LSI scores in the current study were within, albeit towards the lower end, of the SSASPT LSI range (103 to 111% favoring dominant limb) previously reported in healthy persons.^{14-18,20} The LSI reliability metrics provided by the current investigation, coupled with six separate investigations^{14-18,20} reporting healthy persons demonstrate SSASPT LSI within the above range, supports clinicians using the SSASPT LSI for rehabilitation program progression evaluation and RTS clinical decision making. When using the SSASPT LSI to assess RTS readiness in a patient with UE pathology, there may be a need to consider which limb is the involved based upon a recent investigation reporting SSASPT LSI differences between patients with dominant versus nondominant limb involvement at the time of discharge from rehabilitation following shoulder injury or surgery.²¹ Specifically, the odds of a patient with a nondominant involved limb being below the normative LSI range (i.e., <89%) were two times higher than the odds of a patient with a dominant involved limb being below the normative range (i.e., <103%). With only one report examining SSASPT LSI differences between dominant/nondominant involved limbs at time of rehabilitation discharge, further research is clearly needed.

There are a few limitations to the current study that should be recognized. First, although the overall sample size is justifiable for a reliability study, preliminary analyses revealed sex differences to be present in some of the reliability metrics. As a result, all analyses were conducted separately for the males and females, thereby effectively lowering sample sizes within each sex to 20. With all else remaining constant (e.g., within subject variability across sessions), sample size directly influences confidence interval provision. Subsequently, this may explain some of the large ICC confidence interval widths. Furthermore, part of the inclusion criteria required participants to have been participating in an overhead recreational or competitive sport for a minimum of one year. The resulting sample was overwhelmingly (95%) involved with unilateral sports. Surprisingly, there were not large differences in the reliability metrics between the dominant and nondominant limbs within each sex, nor did the LSI for the SSASPT or HKMBRT reveal large average bilateral asymmetry. Unfortunately, there is no data regarding the level of play (recreational versus competitive), nor is there data regarding participation history. While collecting current level of play is straightforward, simultaneously trying to quantify history of participation and level of play simultaneously would be difficult. Additionally, it is likely the participants have his-

tories of participating in non-overhead athletic activities (e.g., resistance training). As a result, the variability of participation history, level of play, and other sporting activities may be responsible for the average SSASPT and HKMBRT LSI scores being <10% with the associated standard deviations ranging between 6.9 to 12.1%. Finally, with regard to the participant characteristics, it is worth noting that 80% of the women participants were involved with either softball or volleyball, whereas the men had more dispersion among participation in four sports. Although overall there were not large sex differences in the reliability metrics, the sport participation differences may help explain the trend for the men to have better dominant limb absolute reliability compared to the women, whereas the women demonstrated better nondominant limb absolute reliability compared to the men. Lastly, it is important to note that the current investigation used a three-to-seven-day test-retest interval. This interval is likely shorter than the typical serial testing intervals used for monitoring rehabilitation progression in patients. Critical to test-retest research design is the assumption that the underlying characteristic is not changing.³⁰ Using a longer than seven day interval to estimate random measurement error in physically active participants could be confounded by actual changes in functional performance. Hence, the rationale for using a three-to-seven-day test-retest interval in the current study. Using shorter test-retest intervals has the benefit that improvements associated with learning during novel test performance (e.g., systematic bias) are likely to be greater when shorter time intervals are used compared to longer intervals. Thus, we speculate the performance improvements revealed in the current study for the PMBDT and HKMBRT are liberal estimates compared to typical longer test-retest intervals used in monitoring rehabilitation progression.

CONCLUSION

There is a lack of consensus on which battery of tests, particularly UEFPT that should be used for clinical decision making to progress a patient through a rehabilitation program or criteria for RTS. Based on the results of this study, the authors recommend using the SSASPT and HKMBRT tests because they demonstrated moderate to good reliability, and interestingly, focus on the anterior musculature which is commonly used in overhead throwing sports. However, the posterior musculature, which is obviously critical in the throwing motion as part of the eccentric deceleration phase, still requires additional research to determine an ecologically valid test with good psychometric properties.

CONFLICTS OF INTEREST

The authors report no conflicts of interest.

ACKNOWLEDGEMENTS

Submitted: December 15, 2022 CDT, Accepted: March 26, 2023
CDT

We would like to thank Nicole Ebel, Morgan Taylor, Ben DeLoach and Luke Thayer for their assistance with participant recruitment and data collection.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-NC-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-nc/4.0> and legal code at <https://creativecommons.org/licenses/by-nc/4.0/legalcode> for more information.

REFERENCES

1. Altintas B, Anderson N, Dornan GJ, Boykin RE, Logan C, Millett PJ. Return to sport after arthroscopic rotator cuff repair: Is there a difference between the recreational and the competitive athlete? *Am J Sports Med.* 2020;48(1):252-261. [doi:10.1177/0363546519825624](https://doi.org/10.1177/0363546519825624)
2. Frantz TL, Shacklett AG, Martin AS, et al. Biceps tenodesis for superior labrum anterior-posterior tear in the overhead athlete: A systematic review. *Am J Sports Med.* 2021;49(2):522-528. [doi:10.1177/0363546520921177](https://doi.org/10.1177/0363546520921177)
3. Fried JW, Hurley ET, Duenes ML, et al. Return to play after arthroscopic stabilization for posterior shoulder instability-A systematic review. *Arthrosc Sports Med Rehabil.* 2021;3(1):e249-e256. [doi:10.1016/j.asmr.2020.08.007](https://doi.org/10.1016/j.asmr.2020.08.007)
4. Thayaparan A, Yu J, Horner NS, Leroux T, Alolabi B, Khan M. Return to sport after arthroscopic superior labral anterior-posterior repair: A systematic review. *Sports Health.* 2019;11(6):520-527. [doi:10.1177/1941738119873892](https://doi.org/10.1177/1941738119873892)
5. Ialenti MN, Mulvihill JD, Feinstein M, Zhang AL, Feeley BT. Return to play following shoulder stabilization: A systematic review and meta-analysis. *Orthop J Sports Med.* 2017;5(9):2325967117726055. [doi:10.1177/2325967117726055](https://doi.org/10.1177/2325967117726055)
6. Owens BD. We can do better. *Am J Sports Med.* 2019;47(12):2793-2794. [doi:10.1177/0363546519878451](https://doi.org/10.1177/0363546519878451)
7. Watson S, Allen B, Grant JA. A clinical review of return-to-play considerations after anterior shoulder dislocation. *Sports Health.* 2016;8(4):336-341. [doi:10.1177/1941738116651956](https://doi.org/10.1177/1941738116651956)
8. Wilson KW, Popchak A, Li RT, Kane G, Lin A. Return to sport testing at 6 months after arthroscopic shoulder stabilization reveals residual strength and functional deficits. *J Shoulder Elbow Surg.* 2020;29(7s):S107-S114. [doi:10.1016/j.jse.2020.04.035](https://doi.org/10.1016/j.jse.2020.04.035)
9. Schwank A, Blazey P, Asker M, et al. 2022 Bern consensus statement on shoulder injury prevention, rehabilitation, and return to sport for athletes at all participation levels. *J Orthop Sports Phys Ther.* 2022;52(1):11-28. [doi:10.2519/jospt.2022.10952](https://doi.org/10.2519/jospt.2022.10952)
10. Bravi M, Fossati C, Giombini A, et al. Criteria for Return-to-Play (RTP) after rotator cuff surgery: A systematic review of literature. *J Clin Med.* 2022;11(8):2244. [doi:10.3390/jcm11082244](https://doi.org/10.3390/jcm11082244)
11. Ciccotti MC, Syed U, Hoffman R, Abboud JA, Ciccotti MG, Freedman KB. Return to play criteria following surgical stabilization for traumatic anterior shoulder instability: A systematic review. *Arthroscopy.* 2018;34(3):903-913. [doi:10.1016/j.arthro.2017.08.293](https://doi.org/10.1016/j.arthro.2017.08.293)
12. Griffith R, Fretes N, Bolia IK, et al. Return-to-sport criteria after upper extremity surgery in athletes-A scoping review, Part 1: Rotator cuff and shoulder stabilization procedures. *Orthop J Sports Med.* 2021;9(8):23259671211021827. [doi:10.1177/23259671211021827](https://doi.org/10.1177/23259671211021827)
13. Wilk KE, Bagwell MS, Davies GJ, Arrigo CA. Return to sport participation criteria following shoulder injury: A clinical commentary. *Int J Sports Phys Ther.* 2020;15(4):624-642. [doi:10.26603/ijsp.20200624](https://doi.org/10.26603/ijsp.20200624)
14. Chmielewski TL, Martin C, Lentz TA, et al. Normalization considerations for using the unilateral seated shot put test in rehabilitation. *J Orthop Sports Phys Ther.* 2014;44(7):518-524. [doi:10.2519/jospt.2014.5004](https://doi.org/10.2519/jospt.2014.5004)
15. Negrete RJ, Hanney WJ, Kolber MJ, et al. Reliability, minimal detectable change, and normative values for tests of upper extremity function and power. *J Strength Cond Res.* 2010;24(12):3318-3325. [doi:10.1519/jsc.0b013e3181e7259c](https://doi.org/10.1519/jsc.0b013e3181e7259c)
16. Riemann BL, Davies GJ. Association between the seated single-arm shot-put test with Isokinetic pushing Force. *J Sport Rehabil.* 2020;29(5):689-692. [doi:10.1123/jsr.2019-0140](https://doi.org/10.1123/jsr.2019-0140)
17. Riemann BL, Johnson W, Murphy T, Davies GJ. A bilateral comparison of the underlying mechanics contributing to the seated single-arm shot-put functional performance test. *J Athl Train.* 2018;53(10):976-982. [doi:10.4085/1062-6050-388-17](https://doi.org/10.4085/1062-6050-388-17)
18. Watson MD, Davies GJ, Riemann BL. Relationship between seated single-arm shot put and isokinetic shoulder flexion and elbow extension strength. *J Sport Rehab.* 2021;30(3):521-524. [doi:10.1123/jsr.2020-0069](https://doi.org/10.1123/jsr.2020-0069)
19. Davies GJ, Riemann BL. Relationships between the seated single arm shot put test and Quick-DASH in patients being discharged from physical therapy. *Musculoskelet Sci Pract.* 2022;60:102568. [doi:10.1016/j.msksp.2022.102568](https://doi.org/10.1016/j.msksp.2022.102568)

20. Degot M, Blache Y, Vigne G, Franger G, Neyton L, Rogowski I. Intra- and intersession reliability and agreement of the unilateral seated shot-put test outcome measures in healthy male athletes. *BMC Sports Sci Med Rehabil.* 2021;13(1):72. [doi:10.1186/s13102-021-00301-4](https://doi.org/10.1186/s13102-021-00301-4)
21. Riemann BL, Davies GJ. Limb dominance effects on seated single arm shot put limb symmetry indices following shoulder rehabilitation. *J Sport Rehab.*
22. Reiebe D, Ehrman J, Liguori G, Magal M. *ACSM's Guidelines for Exercise Testing and Prescription.* Wolters Kluwer Health; 2018.
23. Ellenbecker TS, Sueyoshi T, Bailie DS. Muscular activation during plyometric exercises in 90 degrees of glenohumeral joint abduction. *Sports Health.* 2015;7(1):75-79. [doi:10.1177/1941738114553165](https://doi.org/10.1177/1941738114553165)
24. Bland JM, Altman DG. Measurement error proportional to the mean. *BMJ.* 1996;313(7049):106. [doi:10.1136/bmj.313.7049.106](https://doi.org/10.1136/bmj.313.7049.106)
25. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res.* 2005;19(1):231-240. [doi:10.1519/15184.1](https://doi.org/10.1519/15184.1)
26. Hopkins WG. Measures of reliability in sports medicine and science. *Sports Med.* 2000;30(1):1-15. [doi:10.2165/00007256-200030010-00001](https://doi.org/10.2165/00007256-200030010-00001)
27. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess.* 1994;6(4):284-290. [doi:10.1037/1040-3590.6.4.284](https://doi.org/10.1037/1040-3590.6.4.284)
28. Pinheiro JS, Monteiro OLS, Pinheiro CAB, et al. Seated single-arm shot-put test to measure the functional performance of the upper limbs in exercise practitioners with chronic shoulder pain: A Reliability Study. *J Chiropr Med.* 2020;19(3):153-158. [doi:10.1016/j.jcm.2020.01.001](https://doi.org/10.1016/j.jcm.2020.01.001)
29. Popchak A, Poploski K, Patterson-Lynch B, Nigolian J, Lin A. Reliability and validity of a return to sports testing battery for the shoulder. *Phys Ther Sport.* 2021;48:1-11. [doi:10.1016/j.ptsp.2020.12.003](https://doi.org/10.1016/j.ptsp.2020.12.003)
30. Riemann BL, Lininger MR. Statistical primer for athletic trainers: The essentials of understanding measures of reliability and minimal important change. *J Athl Train.* 2018;53(1):98-103. [doi:10.4085/1062-6050-503-16](https://doi.org/10.4085/1062-6050-503-16)