RESEARCH REPORT

KARYN HAITZ, BA1 • REBECCA SHULTZ, PhD2 • MELISSA HODGINS, MSPT, SCS, ATC3 • GORDON O. MATHESON, MD, PhD4

Test-Retest and Interrater Reliability of the Functional Lower Extremity Evaluation

etermining return-to-play status after injury can be difficult. This decision is often complex and subjective, combining input from physicians, athletes, physical therapists, athletic trainers, and coaches. In addition, the factors included in a returnto-play decision go beyond health and rehabilitation status, such as pressure to help the team or desire to "get back in the game."¹¹ Such

factors may influence return to sport participation before players have regained full function of the injured limb, which may increase the chance of reinjury.⁸

Reinjury rates can be 4 times the rates of initial injury.⁸ Reinjury in the

• **STUDY DESIGN:** Repeated-measures clinical measurement reliability study.

• **OBJECTIVES:** To establish the reliability and face validity of the Functional Lower Extremity Evaluation (FLEE).

BACKGROUND: The FLEE is a 45-minute battery of 8 standardized functional performance tests that measures 3 components of lower extremity function: control, power, and endurance. The reliability and normative values for the FLEE in healthy athletes are unknown.

• **METHODS:** A face validity survey for the FLEE was sent to sports medicine personnel to evaluate the level of importance and frequency of clinical usage of each test included in the FLEE. The FLEE was then administered and rated for 40 uninjured athletes. To assess test-retest reliability, each athlete was tested twice, 1 week apart, by the same rater. To assess interrater reliability, 3 raters scored each athlete during 1 of the testing sessions. Intraclass correlation coefficients were used to assess

lower extremities is even more frequent than overall reinjury rates and is most common in the knee or ankle.^{1,2,19} Arnason et al¹ reported that previous ankle or knee sprain resulted in a 5-fold risk of incurring an ankle or knee sprain dur-

the test-retest and interrater reliability of each of the FLEE tests.

RESULTS: In the face validity survey, the FLEE tests were rated as highly important by 58% to 71% of respondents but frequently used by only 26% to 45% of respondents. Interrater reliability intraclass correlation coefficients ranged from 0.83 to 1.00, and test-retest reliability ranged from 0.71 to 0.95.

• CONCLUSION: The FLEE tests are considered clinically important for assessing lower extremity function by sports medicine personnel but are underused. The FLEE also is a reliable assessment tool. Future studies are required to determine if use of the FLEE to make return-to-play decisions may reduce reinjury rates. J Orthop Sports Phys Ther 2014;44(12):947-954. Epub 13 November 2014. doi:10.2519/jospt.2014.4809

 KEY WORDS: ACL, anatomy/lower extremity, functional testing, hop test, injury, knee, return to play ing a soccer season. A volleyball study by Bahr and Bahr² showed similar results, with 79% of all ankle sprains during a season being repeat sprains. This susceptibility to reinjury may occur because full function has not returned to the injured limb or abnormal movement patterns exist that may contribute to the initial injury.

Ideally, clinicians should adopt standardized testing procedures to make informed return-to-play decisions. Standardized functional testing can be used throughout rehabilitation to assess limb function by comparing functional performance to preinjury data or to normative data from healthy uninjured athletes. Functional deficits can be revealed by functional performance testing that involves high-level exercise maneuvers that mimic the demands placed on the limb during athletic activities.^{3.6}

Existing Lower Extremity Functional Performance Tests

The single-leg hop, timed hop, triple hop, and crossover hop were first combined by Noyes and colleagues¹³ into a hop test sequence. However, this hop test sequence was found to have low sensitivity with respect to identifying lower extremity abnormalities, as it was only able to correctly identify half of the patients with anterior cruciate ligament deficiency in the population tested.¹³ Thus, it was sug-

¹Harvard Medical School, Boston, MA. ²Human Performance Laboratory, Department of Orthopaedic Surgery, Stanford University School of Medicine, Stanford, CA. ³Tahoe Center for Health and Sports Performance, Truckee, CA. ⁴Stanford Sports Medicine Center, Department of Orthopaedic Surgery, Stanford University School of Medicine, Stanford, CA. This study received funding from a UAR Major grant from Stanford University (proposal 4870). This research was approved by the Stanford University Institutional Review Board (protocol 21085). The authors certify that they have no affiliations with or financial involvement in any organization or entity with a direct financial interest in the subject matter or materials discussed in the article. Address correspondence to Karyn Haitz, 107 Avenue Louis Pasteur, Mailbox 320, Boston, MA 02115. E-mail: Karyn_Haitz@hms.harvard.edu @ Copyright ©2014 Journal of Orthopaedic & Sports Physical Therapy[®]

RESEARCH REPORT]

gested that the hop test sequence be used in conjunction with other clinical assessments to better identify underlying deficiencies.¹³ Test-retest reliability for the hop test sequence has been shown to be high in various studies involving different populations.^{3,12,13,15,17}

Functional performance tests involving endurance also have been developed and studied. The square hop, for which acceptable test-retest reliability has been shown,9 involves hopping in and out of a square while moving clockwise. This test measures dynamic postural balance, coordination, and strength of the thigh and calf muscles.^{5,7,9,14,22} The lower extremity functional test (LEFT) is a multidirectional test composed of 8 different tasks.20 The LEFT addresses multidirectional movement and has been established as a reliable assessment tool. However, its developers have suggested that it should not be used as a standalone test but in conjunction with other clinical measurements and functional performance tests for a more comprehensive return-to-play assessment.20

The Functional Lower Extremity Evaluation

The Functional Lower Extremity Evaluation (FLEE) was created by a collaboration of sports medicine personnel (physicians, physical therapists, and athletic trainers) at Stanford University. The goal of developing the FLEE was to compile a comprehensive battery of tests to assess the rehabilitation status of athletes who sustain various lower extremity injuries. The essential categories measured during functional performance testing have been described by Reiman and Manske¹⁶: balance and proprioception; speed and agility; anaerobic and aerobic conditioning; as well as muscle flexibility, strength, power, and endurance. More specifically, the following 10 clinical components have been identified in the literature as important variables with respect to normal lower extremity function: hip/ knee/foot alignment,6,9,19 balance,4,8,15 accuracy of foot placement,^{8,15} strength,^{3,8,12}

Control sequence:

- 1. Timed lateral step-down
- 2. Timed leap and catch
- Hop sequence:
- 3. Single-leg hop for distance
- 4. Single-leg timed hop
- 5. Single-leg triple hop for distance
- 6. Crossover hop for distance
- Endurance sequence:
- 7. Square hop test
- 8. LEFT

FIGURE 1. The FLEE 8-test battery. The tests listed in this box represent the 8 tests of the FLEE battery, subdivided into their respective focus sequences. Abbreviations: FLEE, Functional Lower Extremity Evaluation; LEFT, lower extremity functional test.

endurance,²⁰ coordination,^{5,8} agility,⁵ control in multiple planes of direction,³ landing technique,³ and deceleration control.¹⁶ With these components in mind, the FLEE was designed as a comprehensive battery of 8 tasks.

The 8 tasks that comprise the FLEE (in order) are the timed lateral stepdown, lateral leap and catch, single-leg hop, timed hop, triple hop, crossover hop, square hop, and LEFT (FIGURE 1). The FLEE tasks were ordered to increase in complexity from the least to the most functionally demanding. This was done for safety reasons, due to the higher demands imposed on the musculoskeletal system by the later tests. For example, rehabilitating athletes who experience pain during single-leg squats (test 1) would not be progressed to hopping tests, which involve movements that require greater skill. The square hop and LEFT add multiple planes of direction at a rapid pace, making these the most demanding on the lower extremities. The FLEE battery of tests takes 45 minutes to complete, which more closely resembles the amount of time an athlete would spend in an actual sport activity. Therefore, the inclusion of the 8 tests in the FLEE creates a comprehensive lower extremity functional assessment that builds in complexity and mimics the time spent and/or endurance required during a sporting event.

The purpose of the current study was to establish the face validity of the FLEE, as well as to determine test-retest and interrater reliability of the FLEE. A secondary purpose was to report FLEE scores for healthy individuals. This study is an important step in creating a standardized lower extremity functional performance test to determine return-to-play status following lower extremity injury and/or surgery.

METHODS

Face Validity Survey

OLLOWING THE CREATION OF THE FLEE, a face validity survey was sent to California sports physical therapists, Pacific-12 Conference athletic trainers, and sports medicine physicians (approximately 200 surveys total). Respondents (n = 73) had an average of 13 years of experience in sports medicine settings. The input of these individuals was deemed important, because sports medicine personnel typically make functional performance assessments for rehabilitating athletes. Respondents were asked to evaluate the level of importance of each FLEE test as being "not at all important," "minimally important," "somewhat important," "important," or "very important." Participants also reported the frequency of use of each test in evaluating athletes as 0% to 10%, 10% to 25%, 25% to 40%, 40% to 55%, 55% to 70%, 70% to 85%, or 85% to 100% of the time. In our assessment of the survey responses, we considered high importance to be scores of "important" or "very important" and frequent use to be usage ranging from 70% to 100% of the time. The Stanford University Institutional Review Board approved distribution of this survey, and responses were deidentified.

Subjects

Study participants consisted of 49 uninjured collegiate athletes, who were recruited from Stanford University's field, court, or running sports. Informed consent was obtained from all participants, per the study protocol approved by the Stanford University Institutional Review Board, and the rights of the study participants were protected. Nine participants dropped out between the first and second testing sessions due to reasons unrelated to the study (eg, illness, pain from workouts, or scheduling conflicts); thus, 40 subjects completed the study protocol (**TABLE 1**).

Raters

The present study utilized 3 raters: a physical therapist, an athletic trainer, and a research assistant. The physical therapist had 11 years of experience as an athletic trainer and 9 years as a physical therapist, and the athletic trainer had 8 years of sports medicine experience. Both raters were full-time clinicians who had several months to several years of experience with each of the FLEE tests. The research assistant practiced administering and rating the FLEE for 3 months and was thoroughly trained and evaluated by the director of physical therapy prior to evaluating study participants. All 3 raters were provided with a manual of procedures detailing how to administer and rate the FLEE.

The research assistant administered and scored both testing sessions for each athlete, and the other 2 raters scored each athlete during either the first or second testing session. Before each test on both testing days, the research assistant provided subjects with verbal instructions. The other raters did not give any instructions to subjects, and independently graded each test. All raters were blinded to each other's scores.

Procedures

Test sessions began with a 5-minute warm-up on a stationary bike. Each subject was tested on 2 occasions separated by 1 week. On day 1, subjects performed the first 7 FLEE tests and a submaximal trial of test 8 (LEFT). The LEFT was not performed and scored on day 1 due to time constraints, pre-established reliability, and the test's expected learning

TABLE 1	Subject Characteristics by Sex		
	Men (n = 22)	Women (n = 18)	Total (n = 40)
Age, y*	20.7 ± 1.8	19.9 ± 1.4	20.4 ± 1.6
Age range, y	18-24	18-22	18-24
Height, cm*	181.9 ± 5.3	165.6 ± 6.1	174.6 ± 9.9
Limb dominance, n			
Right	19	14	33
Left	3	4	7
Sport, n			
Field hockey	0	3	3
Gymnastics	0	2	2
Rugby	6	5	11
Soccer	11	2	13
Track	1	0	1
Triathlon team	3	3	6
Volleyball	0	3	3
Wrestling	1	0	1
*Values are mean $\pm S$	D.		



FIGURE 2. Timed lateral step-down. (A) Starting position with the squatting foot on the step and the other beside it. (B) Squat position with the nonsquatting heel lightly tapping the foam mat (or ground). The athlete alternates between positions A and B each time the metronome clicks (at 80 bpm).

curve.²⁰ The LEFT's complexity makes motor-learning effects more likely after several test exposures.²⁰

For the tests performed on a single leg, subjects began testing on their dominant limb, identified as the side used to kick

RESEARCH REPORT





a ball. Limb symmetry indices were recorded as the score of the nondominant compared to the dominant leg. All FLEE testing took place in the Stanford University Human Performance Laboratory.

Control Sequence (Tests 1 to 2)

Test 1 The timed lateral step-down involves continuous single-leg squats on a step with hands on hips, and requires that subjects gently tap the heel to the ground with the toe pointed upward during each squat (FIGURE 2). Step height was adjusted such that a 60° to 70° knee flexion angle was achieved when the heel touched the ground. A metronome was used to keep an 80-bpm pace, with each click signaling the subject to flex or extend the knee. Subjects were instructed to maintain neutral limb alignment during the test, which continued until 3 faulty movement-pattern strikes were made, the athlete chose to stop for reasons such as pain or inability to continue, or 180 seconds had passed. Strikes were given for the presence of knee valgus, loss of balance, falling off pace, or the hands coming off the hips. The recorded measure was the total time (seconds) prior to obtaining 3 strikes, as recorded by each rater using a stopwatch. During pilot testing, we defined knee valgus as the center of the patella moving medial to the first toe. When multiple raters were present, they were physically shielded

from each other and held up clipboards with large X's on the back to indicate when they had recorded 3 strikes. Subjects were instructed to stop before the 180-second time limit if all 3 testers recorded 3 strikes.

Test 2 The lateral leap and catch involved continuous unilateral jumping from one foot to the other over lines set at a distance of 60% body height. Subjects were allowed to practice prior to testing. The test lasted 60 seconds, with a metronome keeping a 40-bpm pace (each click indicating 1 jump). During each landing, body weight had to be transferred to the landing limb such that proper landing mechanics were achieved (eg, no pelvic drop or valgus at the knee). The number of times a line was not completely crossed was documented. Up to 3 strikes were given for faulty maneuvers, such as knee valgus, pelvic drop, or significant loss of balance. Because none of the subjects received 3 strikes, all scores were based on the number of lines crossed during the 60 seconds. If a subject were to receive 3 strikes, the test would be terminated and the time of the test termination recorded.

Hop Test Sequence (Tests 3 to 6)

The hop test sequence (**FIGURE 3**) was performed as described by Noyes et al.¹³ The tests included in this sequence were the single-leg hop, timed hop over a 6-m



the bottom of a 40 × 40-cm square, hops in, left, and so on clockwise around the square.⁵ Reprinted with permission from Caffrey et al.⁵ Copyright © 2009 *Journal of Orthopaedic & Sports Physical Therapy*[®].



FIGURE 5. Diagrammatic representation of the lower extremity functional test. For the lower extremity functional test, cones are placed at positions A, B, C, and D to form a diamond shape, with 30×10 -ft cross-sections. Subjects perform the following tasks in order: (1) forward run (A-C-A), (2) backward run (A-C-A), (3) side shuffles around the perimeter (A-D-C-B and back from A-B-C-D), (4) cariocas around the perimeter, (5) figure-of-eight runs around the perimeter (looping around cones C and A), (6) 45° cuts around the perimeter (cutting at B and D), (7) 90° cuts (A-D-B and back from A-B-D), (8) 90° crossover cuts over the inside foot (A-D-B and back from A-B-D), (9) forward run, and (10) backward run.

course, triple hop, and crossover hop, in which the subject hopped 3 times over 2 lines set 15 cm apart. Subjects first were allowed to practice on the dominant leg at approximately 50% and 75% effort, then performed 3 maximum-effort trials. The hop tests were next performed on the nondominant leg. The outcome for the timed hop was the time to cover 6 m, and the outcome for the remaining hop tests was maximum distance.



FIGURE 6. Level of importance versus clinical usage of lower extremity functional performance tests. The percent of sports medicine personnel (physicians, physical therapists, and athletic trainers) giving each test a high importance rating is compared to the frequency of their clinical usage of the test. Abbreviation: LEFT, lower extremity functional test.

TABLE 2	Test-Retest Reliability		
Test	ICC _{3,1} *	SEM	SRD
Lateral step-down, s	0.77 (0.67, 0.85)	22.4	62.1
Leap and catch, lines missed	0.71 (0.52, 0.84)	1.2	3.4
Single-leg hop, cm	0.91 (0.87, 0.94)	7.7	21.3
Timed hop, s	0.92 (0.88, 0.95)	0.1	0.3
Triple hop, cm	0.95 (0.92, 0.97)	25.1	69.7
Crossover hop, cm	0.94 (0.91, 0.96)	26.1	72.5
Square hop, lines crossed	0.83 (0.75, 0.89)	5.8	16.0
Abbreviations: ICC, intraclass of smallest real difference.	correlation coefficient; SEM, st	andard error of meası	ırement; SRD,

*Values in parentheses are 95% confidence interval.

For the single-leg hop, triple hop, and crossover hop, subjects were instructed to hold the landing position for 2 seconds. After 2 seconds, subjects could put the other foot down but were instructed not to move the landing foot until all 3 raters had measured the jump distance based on where the heel landed. Raters independently determined disqualification based on "not sticking the landing" (not holding the landing for 2 seconds) or on the hands coming off of the hips. The average distance of the 3 hop trials was used. Disqualified trials were excluded from the average.

Endurance Sequence (Tests 7 to 8)

Test 7 The square hop test (FIGURE 4) was performed as described by Caffrey et al.⁵ Subjects hopped clockwise for 30 seconds by jumping clockwise on 1 leg, in and out of a 40 \times 40-cm box drawn on the ground. During testing, the subject stated the number of each revolution aloud, and the rater silently marked the number of times the participant's foot hit a line. The subject stopped after 30 seconds, and the final score was number of lines crossed minus number of lines hit. Subjects were allowed to practice before starting the test and were not permitted to touch the contralateral foot to the ground during testing.

Test 8 The LEFT (FIGURE 5) is composed of 8 multidirectional drills performed in each direction continuously in a 16step sequence within a diamond-shaped course.4 The test order consisted of a forward run, backward run, side shuffle right/left, carioca right/left, figure-ofeight run right/left, 45° cuts right/left, 90° cuts right/left, 90° crossover cuts right/left, forward run, and backward run. Subjects were familiarized with the LEFT 1 week prior to and immediately before the timed trial. The score for the LEFT was the time to complete the full sequence, as measured with a stopwatch. The LEFT was performed only to obtain normative values, not for test reliability.

Statistical Analysis

SPSS for Windows Version 21.0 (SPSS Inc, Chicago, IL) was used for all statistical analyses. Reliability for the first 7 tests was assessed using the intraclass correlation coefficient (ICC) and 95% confidence intervals. An ICC model 3,1 was used to assess test-retest reliability, and an ICC model 2,1 was used to assess interrater reliability. In both cases, the single-measures ICC statistic was used to investigate whether the scores for 2 consecutive test sessions were similar and whether the judgment of each rater was the same as the others.

The ICC values range from 0 to 1. Values ranging from 0.90 to 1.00 were considered excellent reliability, 0.80 to 0.90 high reliability, and 0.60 to 0.80 moderate reliability.10,18,21 The standard error of measurement (SEM) and smallest real difference (SRD) were calculated from ICCs using the formulas SD $\times \sqrt{1}$ – ICC and $1.96 \times \sqrt{2} \times \text{SEM}$, respectively. For the 6 single-leg tests, we also analyzed limb symmetry scores by sex by calculating means and 95% confidence intervals of the averaged scores from the 3 different raters. Mean \pm SD values also were calculated for each test for both sexes and limbs to provide normative data.

RESEARCH REPORT

RESULTS

Face Validity Survey

RESPONSES FROM THE FACE VALIDITY survey are shown in **FIGURE 6**. The range of percent of respondents who rated tasks as being of high importance was 58% to 71%, whereas the range of percent of respondents who rated tasks as being frequently used was 26% to 45%.

Reliability

A summary of the results of the FLEE tests (excluding the LEFT) is provided in **TABLES 2** and **3**. The ICC values for test-retest reliability ranged from 0.71 for the leap and catch test to 0.95 for the triple hop test. The interrater reliability ICC values ranged from 0.83 for the lateral step-down test to 1.00 for the single-leg, triple, and crossover hop tests.

Normative Values

Mean \pm SD values by sex and limb for all tests are provided in **TABLE 4**. Average limb symmetry indices ranged from 98.6% for the triple hop and timed hop to 114.4% for the lateral step-down (**TABLE 5**).

DISCUSSION

Face Validity Survey

HE RATINGS OF HIGH IMPORTANCE of the tests contained within the FLEE were found to be 1.5 to 2 times greater than the frequency of use. This discrepancy indicates that the FLEE tests are considered important but underutilized by sports medicine personnel. Part of the reason that these tests are underutilized may be that the reliability and normative values of the tests have not been reported. Therefore, providing reliability and normative data for healthy athletes may increase the frequency of use of the FLEE tests to better match their level of perceived importance in assessing lower extremity function.

Reliability

The FLEE tests evaluated had test-retest ICC values of 0.71 to 0.95 and interrater

TABLE 3	INTERRATER RELIABILITY		
Test	ICC _{2,1} *	SEM	SRD
Lateral step-down, s	0.83 (0.76, 0.88)	20.7	57.3
Leap and catch, lines missed	0.85 (0.76, 0.91)	0.8	2.1
Single-leg hop, cm	1.00 (1.00, 1.00)	1.7	4.8
Timed hop, s	0.93 (0.89, 0.95)	0.1	0.3
Triple hop, cm	1.00 (0.99, 1.00)	5.6	15.5
Crossover hop, cm	1.00 (1.00, 1.00)	4.7	13.0
Square hop, lines crossed	0.98 (0.98, 0.99)	1.9	5.1
Abbreviations: ICC. intraclass correlation coefficient: SEM. standard error of measurement: SRD.			

smallest real difference.

*Values in parentheses are 95% confidence interval.

reliability ICC values of 0.83 to 1.00, indicating that each demonstrated acceptable repeatability for sports injury research.10 The timed lateral step-down and lateral leap and catch tests had moderate (ICC<0.80) test-retest and high (ICC = 0.80-0.90) interrater reliability, whereas the other tests had high to excellent testretest reliability (ICC>0.80) and excellent interrater reliability (ICC>0.90). The relatively lower reliability for the control sequence tests is likely due to greater subjectivity in scoring. Because raters gave strikes based on qualitative decisions, discrepancies in scoring could occur between raters or between multiple tests conducted by a single rater. Although these qualitative tests were more subjective, they were deemed necessary for a comprehensive test battery to indicate faulty movement patterns that may result in reinjury.

Previous studies have assessed testretest reliability for the hop test sequence but not interrater reliability. Test-retest reliability for single-leg, triple, and crossover hops in healthy individuals has been reported to range from 0.92 to 0.97, with timed hop scores as low as 0.66.^{3,12,17} Ross et al¹⁷ reported timed hop test-retest reliability of 0.92 in Air Force cadets, which is consistent with the findings of the present study. Lower ICC values (0.66) have been reported in untrained subjects compared to cadets or collegiate athletes, who had higher ICC values (0.92).^{3,12,17} Testretest data based on SEM values have been reported as being 4.51 to 7.93 cm for the single-leg hop, 0.06 to 0.13 seconds for the timed hop, 11.17 to 23.18 cm for the triple hop, and 15.95 to 21.16 cm for the crossover hop.^{3,12,17} The SEM values in our study were slightly higher for triple and crossover hops (25.1 and 26.1 cm, respectively), which might have been due to fatigue resulting from the multiple tasks evaluated as part of the FLEE. Additionally, Gustavsson et al9 reported testretest reliability for the square hop (ICC = 0.85), which was similar to our ICC value of 0.83. Our study did not measure testretest reliability for the LEFT, but Tabor et al²⁰ reported reliability to be excellent (ICCs ranging from 0.95 to 0.97, with an SEM of 1.7 to 1.9 seconds).

Our SEM and SRD values for the FLEE provide information for judging an athlete's score in the context of regular variation in performance or true difference. For example, the SEM for single-leg hop test-retest reliability indicates that a 7.7-cm difference in an athlete's performance is within 1 SD of the athlete's previous score. Additionally, the SRD indicates that 21.3 cm is the smallest difference in a score required to signify a real difference in performance.

Normative Values

Normative data from the sample of the present study were stratified by sex to determine normal FLEE ranges for healthy male and female athletes. Based on these

TABLE 4

NORMATIVE VALUES OF THE FUNCTIONAL LOWER EXTREMITY EVALUATION Test by Sex and Dominant Limb*

	Men		Women	
Test	Dominant Leg (19 Right, 3 Left)	Nondominant Leg (3 Right, 19 Left)	Dominant Leg (14 Right, 4 Left)	Nondominant Leg (4 Right, 14 Left)
Lateral step-down, s	114 ± 50	126 ± 48	131 ± 39	122 ± 50
Leap and catch, lines missed	$2\pm2^{\dagger}$		$1\pm2^{\dagger}$	
Single-leg hop, cm	184.5 ± 14.6	183.9 ± 16.5	144.1 ± 16.0	141.7 ± 16.3
Timed hop, s	1.9 ± 0.2	1.9 ± 0.2	2.3 ± 0.4	2.4 ± 0.3
Triple hop, cm	546.1 ± 58.1	542.4 ± 62.9	410.2 ± 45.2	397.9 ± 45.6
Crossover hop, cm	493.2 ± 64.9	493.2 ± 71.5	368.0 ± 49.0	367.3 ± 53.0
Square hop, lines crossed	73 ± 13	72 ± 11	64 ± 15	62 ± 15
LEFT, s	$109.4\pm9.7^{\dagger}$		$117.2 \pm 8.3^{+}$	

Abbreviation: LEFT, lower extremity functional test.

*Values are mean \pm SD.

⁺Bilateral task; data apply to both legs.

TABLE 5

LIMB SYMMETRY FOR THE FUNCTIONAL LOWER EXTREMITY EVALUATION TEST

Test	Value*	
Lateral step-down	114.4 (102.3, 126.4)	
Single-leg hop	99.9 (97.9, 102.0)	
Timed hop	98.6 (97.0, 100.3)	
Triple hop	98.6 (96.8, 100.4)	
Crossover hop	100.2 (98.0, 102.3)	
Square hop	102.2 (93.7, 110.7)	
*Values are mean (95% confidence interval) %.		

data, functional performance of the recovering athlete could be compared to that of healthy athletes to see how they should perform once full function has been regained. Clinically, a rehabilitating athlete's FLEE scores could be compared to these normative scores; however, there is great variation in what would be considered normal for a specific athlete. Limb symmetry data provide additional information that may be clinically important. For example, during the triple hop the nondominant limb performed at 98.6% of the dominant limb. Though this slight variation between limbs may be considered normal, a significantly larger side-to-side difference would not.

For the hop test sequence, the mean values of our male subjects were generally better than those of recreational athletes and worse than those of Air Force cadets, as previously reported.12,17 In contrast, women scored the same or better than recreational athletes and better than nonspecified subjects for all tests.^{3,12} With respect to the LEFT, means and ranges for men were similar to those of male collegiate varsity lacrosse players, as reported by Tabor et al.²⁰ Furthermore, our subjects had similar LEFT scores to those of Division III collegiate athletes, as reported by Brumitt et al.4

Limitations

A limitation of the present study was that only 1 of the 3 raters gave instructions to the subjects, although the 3 raters independently assessed subject scores. Therefore, the real-life situation of each rater administering and scoring the test was not exactly met. A second limitation of our study was that normative data can vary by sport, and there were not enough participants per sport to provide sportspecific data. As such, our reported values may not apply individually to the sports represented.

CONCLUSION

EST-RETEST AND INTERRATER RELIability values of the tests that comprise the FLEE were found to be acceptable. The next step in establishing the FLEE as a standardized functional performance test is to evaluate how injured athletes perform on the FLEE by measuring interrater reliability and FLEE scores for injured athletes at different stages of rehabilitation. Test-retest reliability cannot be tested easily in an injured population, as improvements in function can occur quickly owing to the rehabilitation process. Follow-up validity studies will examine concurrent and predictive validity, as well as changes in reinjury rates of athletes after using the FLEE. The normative data provided in this study permit comparison of a rehabilitating athlete's functional ability to the expected performance values of healthy, uninjured athletes. However, preinjury and postinjury comparisons would also be informative, as these values would be athlete specific. This study is the first step in establishing the FLEE as a standardized functional performance test that could help reduce lower extremity reinjury rates in athletes. •

KEY POINTS

FINDINGS: Test-retest reliability was moderate to excellent (ICC $_{3,1}$ = 0.71-0.95) and interrater reliability was high to excellent $(ICC_{2,1} = 0.83-1.00)$ for the tests that comprise the FLEE.

IMPLICATIONS: This study demonstrated the reliability of the FLEE's comprehensive testing protocol in healthy athletes and provided normative values to

RESEARCH REPORT]

guide the rehabilitation goals of injured athletes.

CAUTION: Reliability was reported in healthy athletes. As such, these results cannot be generalizable to injured persons.

ACKNOWLEDGEMENTS: We would like to thank all of the athletes who participated as subjects in this study and the athletic trainers who helped with recruitment. Most importantly, we would like to thank Tammy Moreno and Nina Holley for dedicating so many hours to rating the FLEE for all subjects. Finally, we would like to thank Dr Anne Friedlander for her feedback on this research project and Dr Alex Sox-Harris for his aid in statistical analysis and graphical depictions of the data.

REFERENCES

- Arnason A, Sigurdsson SB, Gudmundsson A, Holme I, Engebretsen L, Bahr R. Risk factors for injuries in football. *Am J Sports Med.* 2004;32:55-16S. http://dx.doi. org/10.1177/0363546503258912
- Bahr R, Bahr IA. Incidence of acute volleyball injuries: a prospective cohort study of injury mechanisms and risk factors. Scand J Med Sci Sports. 1997;7:166-171.
- Bolgla LA, Keskula DR. Reliability of lower extremity functional performance tests. J Orthop Sports Phys Ther. 1997;26:138-142. http:// dx.doi.org/10.2519/jospt.1997.26.3.138
- Brumitt J, Heiderscheit BC, Manske RC, Niemuth PE, Rauh MJ. Lower extremity functional tests and risk of injury in Division III collegiate athletes. Int J Sports Phys Ther. 2013;8:216-227.

- Caffrey E, Docherty CL, Schrader J, Klossner J. The ability of 4 single-limb hopping tests to detect functional performance deficits in individuals with functional ankle instability. J Orthop Sports Phys Ther. 2009;39:799-806. http:// dx.doi.org/10.2519/jospt.2009.3042
- Creighton DW, Shrier I, Shultz R, Meeuwisse WH, Matheson GO. Return-to-play in sport: a decision-based model. *Clin J Sport Med*. 2010;20:379-385. http://dx.doi.org/10.1097/ JSM.0b013e3181f3c0fe
- 7. Ericsson YB, Dahlberg LE, Roos EM. Effects of functional exercise training on performance and muscle strength after meniscectomy: a randomized trial. Scand J Med Sci Sports. 2009;19:156-165. http://dx.doi. org/10.1111/j.1600-0838.2008.00794.x
- Fuller CW, Bahr R, Dick RW, Meeuwisse WH. A framework for recording recurrences, reinjuries, and exacerbations in injury surveillance. *Clin J Sport Med*. 2007;17:197-200. http://dx.doi. org/10.1097/JSM.0b013e3180471b89
- 9. Gustavsson A, Neeter C, Thomeé P, et al. A test battery for evaluating hop performance in patients with an ACL injury and patients who have undergone ACL reconstruction. *Knee Surg Sports Traumatol Arthrosc.* 2006;14:778-788. http://dx.doi.org/10.1007/s00167-006-0045-6
- Hayen A, Dennis RJ, Finch CF. Determining the intra- and inter-observer reliability of screening tools used in sports injury research. J Sci Med Sport. 2007;10:201-210. http://dx.doi. org/10.1016/j.jsams.2006.09.002
- Matheson GO, Shultz R, Bido J, Mitten MJ, Meeuwisse WH, Shrier I. Return-to-play decisions: are they the team physician's responsibility? *Clin J Sport Med*. 2011;21:25-30. http:// dx.doi.org/10.1097/JSM.0b013e3182095f92
- Munro AG, Herrington LC. Between-session reliability of four hop tests and the agility T-test. J Strength Cond Res. 2011;25:1470-1477. http:// dx.doi.org/10.1519/JSC.0b013e3181d83335
- **13.** Noyes FR, Barber SD, Mangine RE. Abnormal lower limb symmetry determined by function

hop tests after anterior cruciate ligament rupture. *Am J Sports Med.* 1991;19:513-518.

- 14. Östenberg A, Roos E, Ekdahl C, Roos H. Isokinetic knee extensor strength and functional performance in healthy female soccer players. *Scand J Med Sci Sports*. 1998;8:257-264. http:// dx.doi.org/10.1111/j.1600-0838.1998.tb00480.x
- 15. Reid A, Birmingham TB, Stratford PW, Alcock GK, Giffin JR. Hop testing provides a reliable and valid outcome measure during rehabilitation after anterior cruciate ligament reconstruction. *Phys Ther.* 2007;87:337-349. http://dx.doi. org/10.2522/ptj.20060143
- **16.** Reiman MP, Manske RC. *Functional Testing in Human Performance*. Champaign, IL: Human Kinetics; 2009.
- 17. Ross MD, Langford B, Whelan PJ. Test-retest reliability of 4 single-leg horizontal hop tests. J Strength Cond Res. 2002;16:617-622.
- Shultz R, Mooney K, Anderson S, et al. Functional movement screen: inter-rater and subject reliability [poster]. Br J Sports Med. 2011;45:374. http://dx.doi.org/10.1136/ bjsm.2011.084038.180
- **19.** Smith RW, Reischl SF. Treatment of ankle sprains in young athletes. *Am J Sports Med.* 1986;14:465-471.
- Tabor MA, Davies GJ, Kernozek TW, Negrete RJ, Hudson V. A multicenter study of the test-retest reliability of the Lower Extremity Functional Test. J Sport Rehabil. 2002;11:190-201.
- **21.** Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19:231-240. http://dx.doi.org/10.1519/15184.1
- **22.** Westin M. A new judgement scale to evaluate performance of the square hop test. 2nd European Conference of Sports Rehabilitation; September 19-20, 2008; Stockholm, Sweden.



DOWNLOAD PowerPoint Slides of JOSPT Figures

JOSPT offers **PowerPoint slides of figures** to accompany all full-text articles with figures on *JOSPT*'s website (**www.jospt.org**). These slides are generated automatically by the site, and can be downloaded and saved. They include the article title, authors, and full citation. *JOSPT* offers full-text format for all articles published from January 2010 to date.