Interrater and Intrarater Reliability and Discriminant Validity of a Pediatric Lower Extremity Physical Therapy Clearance Test

Matthew D. Ellington,* MD, Kelly Randich,† PT, DPT, Jessica Garfin,† PT, DPT, Nicole Archambault,† PT, DPT, Joanna Roocroft,† MA, Tracey P. Bastrom,† MA, Andrew T. Pennock,†‡ MD, and Eric W. Edmonds,†‡§ MD

Investigation performed at Rady Children's Hospital, San Diego, California, USA

Background: Few studies have validated when an athlete can safely return to sports, and even fewer have identified when he or she no longer requires physical therapy after surgery. Discontinuing physical therapy is often dictated by insurance restrictions, but most studies have suggested that the decision should be multifactorial, stemming from patient-derived subjective outcome questionnaires, clinical examination, and isokinetic and functional testing.

Purpose/Hypothesis: The purpose of this study was to establish discriminant validity and reliability of an objective physical therapy clearance (PTC) test in a clinical setting. The hypotheses were that the PTC test (1) will demonstrate different scores between normal and postoperative cohorts and (2) will have acceptable inter- and intraobserver reliability.

Study Design: Cohort study (diagnosis); Level of evidence, 3.

Methods: Four cohorts (27 total participants; age range, 12-18 years) underwent the PTC test: 9 adolescents 6 months after anterior cruciate ligament reconstruction, 4 adolescents 6 weeks after partial meniscectomy, 5 adolescents with nonstructural knee pain, and 9 control/healthy participants without any lower extremity complaint. The PTC test included a dynamic warm-up, objective measures (knee range of motion, thigh girth, and muscle motor tone), functional strength tests (heel raises, single-leg dips, hop tests, tuck jumps), and agility tests (shuffle and sprint T-test). Each testing session was videotaped and scored live by the physical therapist administering the test, and then scored via the video recording by an independent physical therapist and 2 orthopaedic surgeons.

Results: The PTC test was found to have discriminant validity between the control cohort and both cohorts with previous surgery. The single-leg dip, single-leg hop, and vertical tuck jump were the most discriminatory components. The PTC test had moderate to almost perfect intrarater reliability ($\kappa = 0.57$-1), but only fair to moderate interrater reliability among video graders ($\kappa = 0.29$-0.58) and slight to substantial reliability between video graders and the live PT rater ($\kappa = 0.19$-0.63).

Conclusion: The PTC test was found to have moderate inter- and intraobserver agreement, with the ability to discriminate between postoperative and control patients.

Keywords: physical therapy; return to sport; pediatric; adolescent; functional test

After lower extremity surgery in young athletes, it is often difficult to determine when it is safe for them to be discharged from physical therapy, let alone when they should return to unrestricted sport and athletic activity. Research has suggested that this decision is multifactorial, stemming from patient-derived subjective outcome questionnaires, their clinical examination regarding motion, stability and function, isokinetic testing, and functional testing.[1] Unfortunately, the extent of these recommendations ends at suggesting that functional testing should be done and not what functional tests give a validated or reproducible prediction of return-to-sport success. Yet, there is mounting evidence[8,10,13] to suggest that return to play before achieving biomechanical function places the adolescent at high risk for recurrent injury.

The risk of recurrent injury and altered neuromuscular control, as it affects postural stability of the hip and knee during a dynamic function, has been demonstrated in previous study[10] of anterior cruciate ligament (ACL) surgery. Moreover, there is evidence that an age-related association between neuromuscular training implementation and a reduction of ACL tearing incidence exists, but there are limited studies[4,6-9,12,13,15] that have been validated with reliability testing. Other functional tests to assess lower

The Orthopaedic Journal of Sports Medicine, 9(2), 2325967120983827
DOI: 10.1177/2325967120983827

extremity health have been explored in isolation, and these assessments may also be predictive of injuries.[5,11,16]

The physical therapists (PTs) at our institution (K.R., J.G., and N.A.), through an extensive literature review of laboratory-based studies, developed a physical therapy clearance (PTC) test that may take into account the known variables and potential predictors of lower extremity health (see the Appendix). As there can be both a psychological and a temporal variable associated with a patient's return to sports, this test was not utilized as a "return-to-sport" evaluation, but rather as an indicator that formal physical therapy was no longer of significant utility and that the patients could be released to sport-specific training. The PTC test utilizes a dynamic warm-up, objective measures (knee range of motion, thigh girth, muscle motor tone), functional strength tests (heel raises, dips, hop test, tuck jumps), and agility tests (T-test – both shuffle and sprint) to assess the patients. This test has already been implemented in the clinical practice of the therapists and the surgeons at our institution to help guide a safe return to full activity.

The purpose of this study was to establish the discriminant validity and reliability of the PTC test in a clinical setting. The hypotheses to be tested were that (1) the test will demonstrate significantly different scores between normal and postoperative/symptomatic patients, thus suggesting discriminant validity, and (2) the test will have acceptable intra- and interrater reliability.

METHODS

After receiving institutional review board approval, we performed a prospective study at a single pediatric/adolescent academic center between January 2016 and February 2017. Four different cohorts of patients (all <18 years of age) were targeted during the year of planned enrollment. The first cohort included patients who had undergone ACL reconstruction and were at least 6 months postoperative. The second cohort included patients who had undergone partial meniscal debridement only and who were at least 6 weeks postoperative. These 2 cohorts were undergoing the PTC test as part of their routine postoperative physical therapy program. The third cohort included patients who had knee pain with no identified structural abnormalities (adolescent knee pain) and who were undergoing a course of at least 6 weeks of physical therapy. These patients had been evaluated by a pediatric orthopaedic surgeon in our group (E.W.E.) before referral to physical therapy. The last cohort included patients with no knee injuries or subjective knee pain and was considered our control cohort, recruited from patient or employee families. These were both athletes and active nonathletes. As the latter 2 cohorts were undergoing the PTC test only for research purposes, they were provided a gift card as time compensation. There were no study exclusion criteria.

PTC Testing Sessions

For each participant, the PTC testing session was administered by a PT (live grader; alternating between J.G. and N.A.), who also videotaped each session for the reliability portion of the study. The PTC test consisted of objective measurements, functional strength tests, and agility tests (Table 1). Each component of the complete PTC test was graded on a scale from 0 to 5, for a total potential score of 75 points. However, the objective measurements of knee range of motion, thigh girth, manual muscle tone (for all 3 muscle groups: quadriceps, hamstrings, and gluteus medius), single-leg dip maximum angle, and all 3 hop tests were not recorded on video for assessment in this study and were therefore not accounted for in the total score. Therefore, as seen in Table 1, the total score possible was 30 points for the subjective elements recorded on the video.

In addition to the total raw score achieved for each participant, a normalized score (0%-100%) was calculated for analysis, based on the total points achieved by each participant divided by the total points possible for tests performed/observed for that participant. The rate of test failure was also recorded, where failure was defined as a score of 0 on any portion of the PTC or the inability to perform the task at hand. Individual components of the PTC were also evaluated, and a grade of 4 or 5 was considered passing for each component. A threshold of 66% mean percentage was established to determine whether the component was discriminatory.

A second PT (K.R.) along with a pediatric orthopaedic surgeon with subspecialty certification in orthopaedic sports medicine (M.D.E.) and a pediatric orthopaedic fellow (E.W.E.) then reviewed the videos to assess interrater reliability on the functional and agility portions of the PTC test. The orthopaedic fellow repeated the assessment by reviewing the videos 3 months after the first scoring to assess intrarater reliability. The live grader's score served as the values used for the discriminant validity portion of the study and as the anchor for statistical analysis in the reliability portion.

[§]Address correspondence to Eric W. Edmonds, MD, Rady Children's Hospital, 3020 Children's Way, MC 5062, San Diego, CA 92123, USA (email: ewedmonds@rchsd.org).

*Department of Surgery and Perioperative Care, Dell Medical School, University of Texas, Austin, Texas, USA.

[†]Pediatric Orthopedics & Scoliosis Center, Rady Children's Hospital, San Diego, California, USA.

[‡]Department of Orthopedic Surgery, University of California, San Diego, California, USA.

Final revision submitted August 11, 2020; accepted August 24, 2020.

One or more of the authors has declared the following potential conflict of interest or source of funding: M.D.E. has received educational support from Arthrex. A.T.P. has received educational support from Sportstek Medical, consulting fees from OrthoPediatrics, and nonconsulting fees from Smith & Nephew and has stock/stock options in Imagen. E.W.E. has received consulting fees from OrthoPediatrics and nonconsulting fees from Arthrex. AOSSM checks author disclosures against the Open Payments Database (OPD). AOSSM has not conducted an independent investigation on the OPD and disclaims any liability or responsibility relating thereto.

Ethical approval for this study was obtained from the University of California, San Diego (project No. 151251XL).

TABLE 1
Scoring Guide for the Physical Therapy Clearance Test[a]

Component	Highest Possible Score	Included on Video Assessment
Warm-up		
Dynamic warm-up (no scoring)	N/A	N/A
Jog 5 minutes[b]	5	5
Objective measurements		
Knee range of motion[c]	5	N/A
Girth[d]	5	N/A
MMT quadriceps: 0-5[e]	5	N/A
MMT hamstring: 0-5[e]	5	N/A
MMT gluteus medius: 0-5[e]	5	N/A
Functional strength		
Single-leg heel raise (reps)[f]	5	5
Single-leg dip: endurance (reps)[f]	5	5
Single-leg dip: maximum angle[f]	5	N/A
Single-leg hop test (single)[f]	5	N/A
Single-leg hop test (crossover)[f]	5	N/A
Single-leg hop test (triple jump)[f]	5	N/A
Vertical tuck jump[g]	5	5
Agility tests		
T-test: shuffle[h]	5	5
T-test: sprint[h]	5	5
Total score possible	**75**	**30**

[a]MMT, manual muscle test; N/A, not attainable.
[b]Graded as 5 = fluid; 4 = slight antalgic, Trendelenburg; 3 = significant limp; 0 = unable.
[c]Graded as 5 = full, symmetrical; 4 = less than 5° difference; 3 = 5°-10° difference; 2 = more than 10°- 15° difference; 1 = more than 15°-20° difference; 0 = more than 20° difference.
[d]Graded as 5 = within 0-1 cm difference; 4 = 1.1-2 cm difference; 3 = 2-3 cm difference; 2 = 3-4 cm difference; 1 = 4-5 cm difference; 0 = more than 5 cm difference.
[e]Grading based on traditional clinical assessment grades for the MMT[2]: 0 = no muscle function; 1 = muscle fiber contraction without motion; 2 = muscle fiber contraction; 3 = against gravity but no resistance; 4 = against some manual resistance; 5 = normal and against full resistance.
[f]Graded as percentage of opposite side: 5 = 90%-100%; 4 = 80%-90%; 3 = 70%-80%; 2 = 60%-70%; 1 = 50%-60%; 0 = less than 50%.
[g]Graded as 5 = no deviations; 4 = 1-2 deviations (no valgus); 3 = 3 deviations and/or slight valgus at landing; 2 = 4 deviations and/or moderate valgus at landing; 1 = more than 5 deviations and/or significant valgus at landing; 0 = unable to complete.
[h]Graded as 5 = good form, no hesitation; 4 = slight hesitation; 3 = mild genu valgum noted; 2 = moderate genu valgum noted, pain; 1 = hesitant with significant genu valgum; 0 = unable to complete.

No formal power analysis was performed for this study. The sample was one of convenience, with a target of 1 year of enrollment and a goal of 10 participants per group.

Statistical Methods

Discriminant Validity. Scores on the PTC were compared between our 4 diagnostic cohorts to evaluate for discriminant validity. The rate of test failure was compared using the chi-square test. The total raw and normalized scores were compared using the nonparametric Kruskal-Wallis test. The descriptive statistics reported for the total raw and normalized scores include means with standard deviations and medians with 25th and 75th percentiles. Alpha was set at $P < .05$ to declare significance. Effect sizes for the post hoc comparisons were converted to Hedges g, and forest plots of the effect size ±95% CI for each of the individual comparisons were created. This was done to evaluate the magnitude of the differences between groups, as a supplement to the P value, owing to the small sample sizes. The effect size convention for Hedges g is interpreted in the same manner as Cohen d (small effect = 0.2, medium effect = 0.5, large effect = 0.8). We further evaluated the ability of the PTC to discriminate between the control group and the other 3 groups using receiver operating characteristic (ROC) analysis. Area under the curve (AUC) and optimal total normalized score cut point for the control group, based on maximized sensitivity and specificity, are reported.

The overall percentage of passing grades was then evaluated within each component, and the mean percentage between cohorts within that component was considered the discriminant ability of the component.

Inter- and intrarater reliability. Because only the functional and agility tests were measured by the video rater, a total raw score and total normalized score were not calculated. Thus, absolute agreement on the individual grading scales (0-5) was evaluated utilizing the kappa (κ) statistic (0-0.2 = slight agreement, 0.21-0.4 = fair agreement, 0.41-0.6 = moderate agreement, 0.61-0.8 = substantial agreement, 0.81-1 = almost perfect agreement). As we considered this assessment to be somewhat critical (ie, a grade of 4 may not be much different than a grade of 5 but is considered lack of agreement in this type of analysis), we also evaluated the rate at which grading was within 1 point.

RESULTS

A total of 27 adolescents were enrolled in the study: 9 postoperative ACL patients (ACL group), 4 postoperative meniscus patients (meniscus group), 5 patients with knee pain (knee pain group), and 9 participants with healthy knees (control group). The ACL group completed the test at a mean of 26 weeks (range, 21-32 weeks) postoperatively, and the meniscus group completed the test at a mean of 7 weeks (range, 6-7 weeks) postoperatively.

Discriminant Validity. The meniscus group descriptively had the highest rate of failure (75%), followed by the ACL group (56%), knee pain group (40%), and control group (22%; $P = .28$). There was a significant difference observed in mean total raw and normalized scores based on patient cohort ($P = .039$ for both) (Table 2). Post hoc comparison revealed that the difference between the meniscus and

TABLE 2
Differences in Failure Rate, Total Raw Score, and Total Normalized Score According to Study Group[a]

	ACL Group	Meniscus Group	Knee Pain Group	Control Group	*P*
Failure rate, %	56	75	40	22	.28
Raw score					**.039**
Mean ± SD	65 ± 4	55 ± 11	67 ± 4	63 ± 3	
Median (25th-75th percentile)	65 (60-69)	58 (44-64)	64 (60-66)	67 (65-71)	
Normalized score, %					**.039**
Mean ± SD	86 ± 6	74 ± 15	89 ± 5	84 ± 5	
Median (25th-75th percentile)	87 (80-92)	77 (58-85)	85 (79-88)	89 (86-95)	

[a]Bolded *P* values indicate statistically significant difference between groups ($P < .05$). ACL, anterior cruciate ligament.

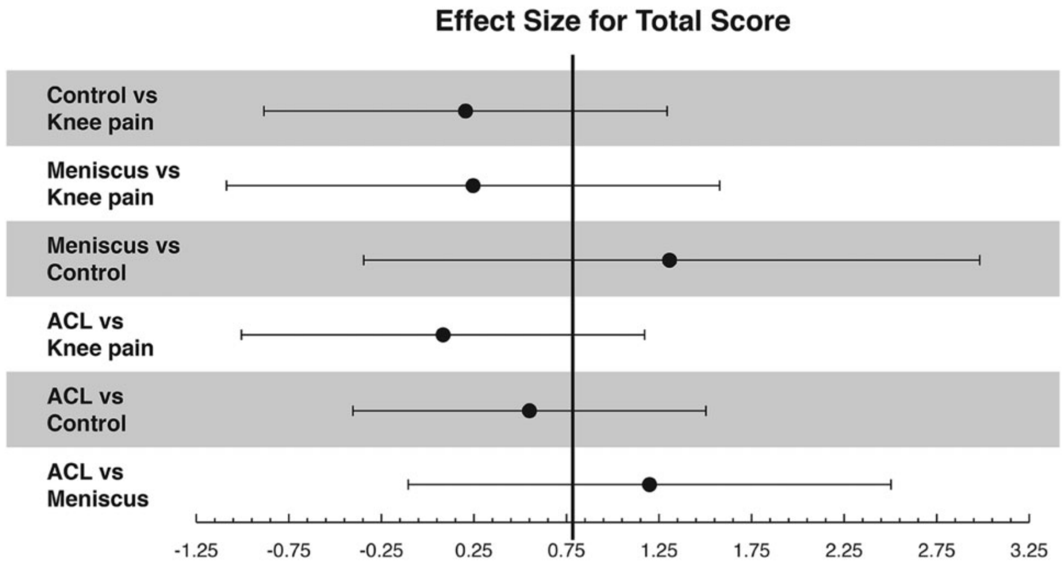

Figure 1. Forest plot demonstrating effect size (±95% CI) for total raw scores based on post hoc comparison of patient groups. ACL, anterior cruciate ligament.

control groups reached statistical significance for raw score ($P = .012$) and normalized score ($P = .011$). The difference between the ACL and meniscus groups was not significant for total raw score ($P = .068$) or total normalized score ($P = .059$). All other post hoc comparisons were nonsignificant. This was also observed in the forest plots of the effect sizes for both the total raw score (Figure 1) and the total normalized score (Figure 2). The comparisons of meniscus versus ACL and meniscus versus control groups had effect sizes above the large threshold (0.8), suggesting a relevant, observable difference in scores for these groups. The ability of the PTC test to discriminate between controls and knee pain patients regarding total normalized score, as evaluated by ROC, resulted in an AUC of 0.77 ($P = .004$). The total normalized score with optimal sensitivity and specificity was identified as 87.5% (specificity = 78%, sensitivity = 67%).

Evaluation of failures directly related to a single component of the test, rather than the overall score, demonstrated that the single-leg dip, single-leg hop, and the vertical tuck jump were the items that would have triggered at least 1 of the evaluators to not move on with the remainder of PTC for 11 of the patients (Figure 3). When assessing how many of the patients scored either a 4 or 5 on the 5-point scale (a passing score), it was found that the same tests (single-leg dip, single-leg hop, and vertical tuck jump) achieved lower scores, with less than two-thirds of the adolescents achieving a passing score. Evaluation of the single-leg dip found that endurance (passing rate of only 58.6%) was more discriminatory than the maximum angle achieved during the test (passing rate of 76.4%). The single-leg hop had a passing rate of only 57.5%, and the vertical tuck jump had a passing rate of 49.5%. Components that likely had no, or limited, discriminant validity (passing rates of ≥95%) included the manual muscle testing (for all 3: hamstrings, gluteus, and quadriceps), T-test shuffle, and T-test sprint. Girth had a passing rate of 93.8%.

Inter- and Intrarater Reliability. The PTC test had moderate to almost perfect intrarater reliability ($\kappa = 0.57$-1), but only fair to moderate interrater reliability among video graders ($\kappa = 0.29$-0.58) and slight to substantial reliability between video graders and the live PT rater ($\kappa = 0.19$-0.63). Intrarater reliability was in the substantial range for all measures of the functional/agility component

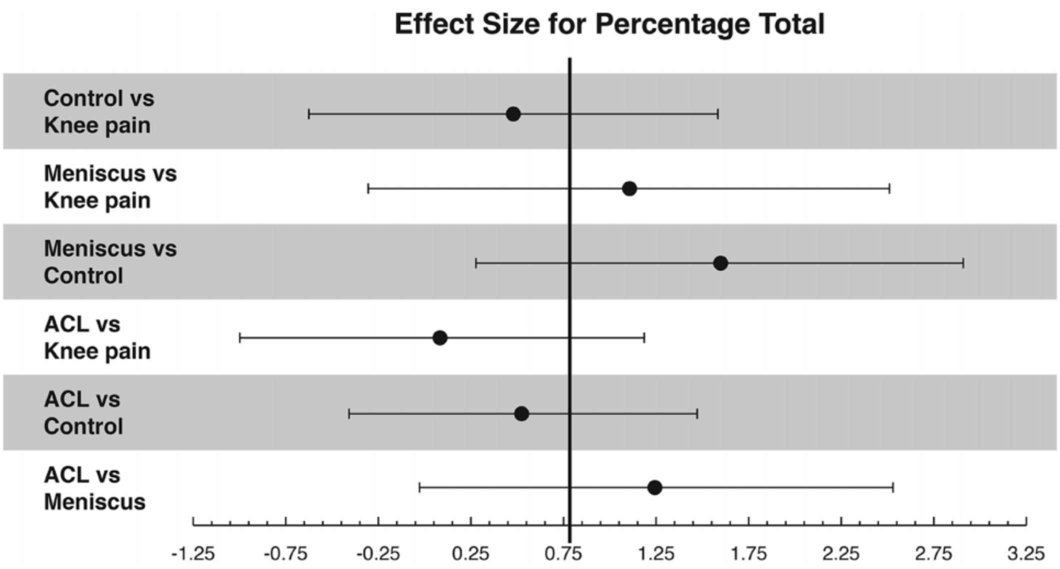

Figure 2. Forest plot demonstrating effect size (±95% CI) for total normalized scores based on post hoc comparison of patient groups. An effect size of ≥0.8 was considered a large magnitude of difference. ACL, anterior cruciate ligament.

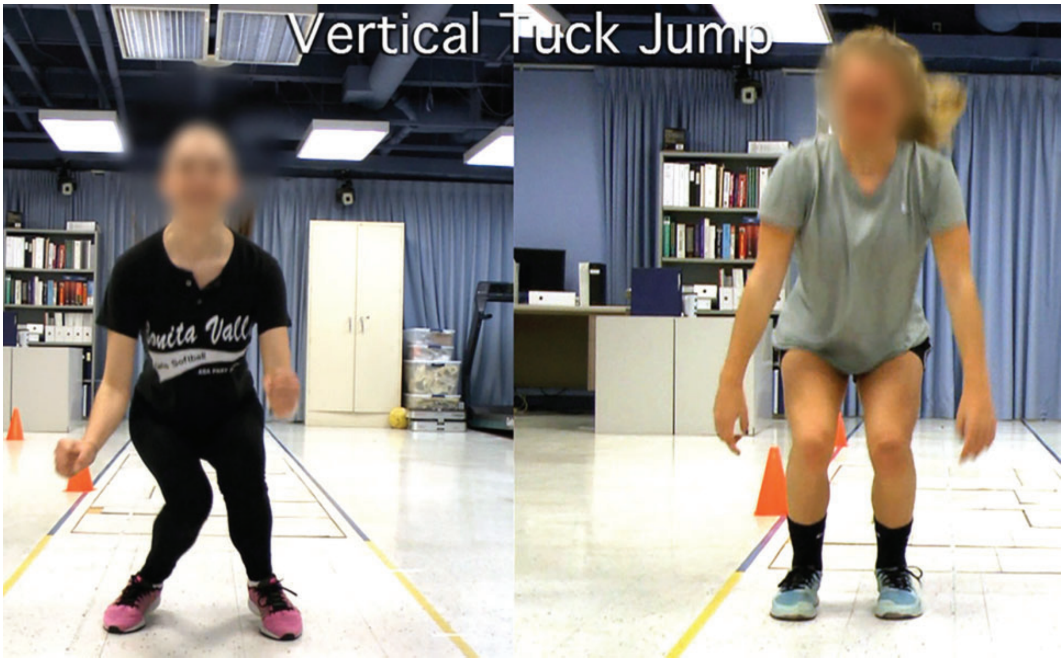

Figure 3. Comparative photographs demonstrating 2 different outcomes for the vertical tuck jump: the child on the left failing the test and the one on the right demonstrating a passing test. Images courtesy of SD PedsOrtho.

of the PTC test (κ = 0.63-0.85). Interrater reliability demonstrated a wide range of reliability from slight to moderate for the measures of the functional/agility component of the PTC test (κ = 0.02-0.57). The agreement between the 2 orthopaedic surgeons ranged from fair to moderate; however, the agreement among all other raters ranged from slight to moderate. The rate at which the video graders were within 1 point of the live PT grade is demonstrated for each test in Table 3. For the jog, single-leg heel rise, and agility sprint test, the 3 graders were within 1 point of the live grader ≥90% of the time. For the agility shuffle test, 1 grader had <90% agreement within 1 point, and for the vertical tuck jump, 2 graders had <90% agreement within 1 point. The only test with all 3 graders demonstrating <90% agreement within 1 point was the single-leg dip endurance test.

TABLE 3
Percentage of Grades Within 1 Point of Each Other for Video Graders as Compared With the Live PT Grade[a]

	Percentage of Grades Within 1 Point		
	Attending Surgeon vs Live PT Grade	Orthopaedic Fellow vs Live PT Grade	Video PT vs Live PT Grade
Jog	96	96	92
Single-leg heel rise	92	100	92
Single-leg dip: endurance	77	69	64
Vertical tuck jump	64	80	90
Agility test: shuffle	96	92	79
Agility test: sprint	100	100	100

[a]PT, physical therapist.

DISCUSSION

In this study, we developed a PTC test that incorporated objective measures, functional strength tests, and agility tests with moderate to almost perfect interobserver, but slight to moderate intraobserver agreement to determine when a patient had successfully cleared physical therapy. We found that the single-leg dip, single-leg hop, and vertical tuck jump were the most discriminatory components. To date, there has not been consensus in the literature on which test or tests can adequately assess when a patient can safely be cleared from physical therapy, let alone when he or she can safely return to sports after knee injury. The hop test and agility T-test have been the only tests independently found to be both validated and reliable.[9,12] A recent meta-analysis[14] found that current "return-to-sport" tests have equivocal findings in relation to validity and the ability to reduce the risk of graft rupture and contralateral tear in post-ACL reconstruction patients. Another recent study[3] showed that landing mechanics can play a role in return-to-sports testing, reporting that healthy women exhibited reduced peak knee flexion in single-leg jump cutting after exercise compared with those who had both passed and failed a "return-to-sports" test after ACL reconstruction. As a whole, our PTC had discriminant validity, with a significant difference in the mean total and total percentage scores based on the different cohorts tested, with an indicated cutoff for a "passing" score. Therefore, the PTC provides an objective assessment of progress with rehabilitation and allows providers to determine when best to release the patient from formal physical therapy.

The risk of injury and the risk of recurrent injury after ACL reconstruction have been well studied. Altered neuromuscular control of the hip and knee during a dynamic landing, as well as defects in postural stability, can be predictive of a second ACL injury after an athlete is released to return to sport.[10] Yet, to date, the only available functional tests with validation and reliability testing are the hop test and agility T-test.[9,12] Recently, the single-leg vertical jump test has been shown to correlate with subjective International Knee Documentation Committee scores, Tegner activity scale, ACL-Return to Sport After Injury scale, isokinetic extensor muscle strength, and other functional tests, potentially making it a convenient test for return to sport.[6] Also, the combination of multiple tests, including single-limb functional tests, has been shown to be more beneficial than standard hop and isokinetic strength testing for return to sports.[4,15] Our study suggests that a simple single-leg dip test (as well as a single-leg hop and vertical tuck jump) could have discriminatory validity between passing and failing the overall PTC test, but further evaluation into each of these would be required to detail their ability to discriminate between someone who should be passing (a control) and those who have not fully recovered from surgery.

When evaluating other factors that contribute to lower extremity health, such as the functional "step-down" test, core strength, and proprioception training, none have gone so far as to offer prediction for safe return to sport or risk for recurrent injury, despite multiple studies[5,11,16] extensively researching their role in post-ACL surgery function, patellar instability, and generalized knee pain, particularly in women. Moreover, it is interesting to note that when using the older standard-of-care protocol of 6 months physical therapy for ACL recovery and 6 weeks physical therapy for a partial meniscectomy, the vast majority of those adolescents failed our PTC test (either by the total raw score, total normalized score, or with a single component). This has led our institution to push physical therapy to last longer than the previous insurance industry standards, hoping to restore normative biomechanical function in our patients before clearance. It is further interesting to note that 22% of the participants with no symptoms, recruited as part of the control cohort, also failed the test. This may be an indicator of injury risk within this population, in general.

With regard to discriminant validity, the current study found a significant difference in the mean total raw and normalized scores based on the different groups tested. A cutoff score of 87.5% on the PTC was identified as the normalized score with the most optimal sensitivity and specificity to give an objective measurement of "safe to release from physical therapy." This normalized score may be better than utilizing a time-based physical therapy duration since a majority of patients failed our test at the conclusion of their time-based rehabilitation regimen. Because of these results, our institution has begun utilizing objective criteria (either the PTC test itself or elements of the test, based on PT preference) to release or request further visits for an individual patient. Further study into outcomes for specific pathologies at our institution are needed to demonstrate the individual utility of the PTC test for the postoperative adolescent patient.

The intrarater reliability for the PTC test demonstrated substantial agreement on all the functional/agility components, and therefore we believe that it is a reliable test. This finding, however, only suggests that the same PT treating the patient in the postoperative course should perform the PTC test on their patient, as they will know when they are most likely to pass. However, the interrater reliability was lower, thus the PTC test may not be as reliable in a facility that utilizes multiple therapists at different times for the same patient. However, it is worth noting that for a majority of the time and each component of the PTC test, the graders were within 1 point/test of each other so this could just represent measurement error.

There are several limitations to this study. First, we had a small sample size, with 27 adolescents. Second, this study was only intended to determine the discriminant validity and the reliability of measurements for our PTC test and not the predictability of future injury -– which is of course the main objective of these patient assessments. But, without first determining the reliability and validity of the test, it would be unreasonable to assess the predictability of the test. Further, 3 of the evaluators only had the video recordings to evaluate and could not shift their vantage point to evaluate the patient for grading purposes. This does not appear to be a significant issue, however, as the agreement between the video graders was not significantly different than between them and the live grader with respect to scoring each patient within 1 point of each other. We did have slight to moderate interobserver agreement, which brings into question the reproducibility of our test.

Our PTC test can hopefully serve as a starting point for future research. We hope that this combination of previously described functional and agility assessments can further elucidate when a patient has reached his or her maximum benefit from physical therapy and can safely be discharged from active rehabilitation. We do not presume that adolescents should immediately return to competitive athletics, rather that passing this test implies that they may train within their sport-specific activities. Further study is being conducted that will help determine if the PTC test can aid in the prevention of ACL graft rupture or contralateral tear. In the meantime, the results of this study (particularly the discriminant validity) have shifted the care of adolescents within our facility, extending physical therapy until objective measures contained within the PTC test are met rather than relying on arbitrary protocols based on insurance time or number of sessions.

REFERENCES

1. Barber-Westin SD, Noyes FR. Factors used to determine return to unrestricted sports activities after anterior cruciate ligament reconstruction. *Arthroscopy*. 2011;27(12):1697-1705.
2. Bohannon RW. Measuring knee extensor muscle strength. *Am J Phys Med Rehabil*. 2001;80(1):13-18.
3. Chang EW, Johnson S, Pollard C, Hoffman M, Norcross M. Landing biomechanics in anterior cruciate ligament reconstructed females who pass or fail a functional test battery. *Knee*. 2018;25(6):1074-1082.
4. Ebert JR, Edwards P, Currie J, et al. Comparison of the “back in action” test battery to standard hop tests and isokinetic knee dynamometry in patients following anterior cruciate ligament reconstruction. *Int J Sports Phys Ther*. 2018;13(3):389-400.
5. Ekstrom RA, Donatelli RA, Carp KC. Electromyographic analysis of core trunk, hip, and thigh muscles during 9 rehabilitation exercises. *J Orthop Sports Phys Ther*. 2007;37(12):754-762.
6. Lee DW, Yang SJ, Cho SI, Lee JH, Kim JG. Single-leg vertical jump test as a functional test after anterior cruciate ligament reconstruction. *Knee*. 2018;25(6):1016-1026.
7. Logerstedt D, Di Stasi S, Grindem H, et al. Self-reported knee function can identify athletes who fail return-to-activity criteria up to 1 year after anterior cruciate ligament reconstruction: a Delaware-Oslo ACL cohort study. *J Orthop Sports Phys Ther*. 2014;44(12):914-923.
8. Myer GD, Sugimoto D, Thomas S, Hewett TE. The influence of age on the effectiveness of neuromuscular training to reduce anterior cruciate ligament injury in female athletes: a meta-analysis. *Am J Sports Med*. 2013;41(1):203-215.
9. 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(5):513-518.
10. Paterno MV, Schmitt LC, Ford KR, et al. Biomechanical measures during landing and postural stability predict second anterior cruciate ligament injury after anterior cruciate ligament reconstruction and return to sport. *Am J Sports Med*. 2010;38(10):1968-1978.
11. Rabin A, Kozol Z. Measures of range of motion and strength among healthy women with differing quality of lower extremity movement during the lateral step-down test. *J Orthop Sports Phys Ther*. 2010; 40(12):792-800.
12. 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(3):337-349.
13. Schmitt LC, Paterno MV, Hewett TE. The impact of quadriceps femoris strength asymmetry on functional performance at return to sport following anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther*. 2012;42(9):750-759.
14. Webster KE, Hewett TE. What is the evidence for and validity of return-to-sport testing after anterior cruciate ligament reconstruction surgery? A systematic review and meta-analysis. *Sports Med*. 2019; 49(6):917-929.
15. Wren TAL, Mueske NM, Brophy CH, et al. Hop distance symmetry does not indicate normal landing biomechanics in adolescent athletes with recent anterior cruciate ligament reconstruction. *J Orthop Sports Phys Ther*. 2018;48(8):622-629.
16. Zazulak BT, Hewett TE, Reeves NP, Goldberg B, Cholewicki J. Deficits in neuromuscular control of the trunk predict knee injury risk: a prospective biomechanical-epidemiologic study. *Am J Sports Med*. 2007;35(7):1123-1130.

APPENDIX

PHYSICAL THERAPY CLEARANCE TEST

Lower Extremity: Physical Therapy Clearance Test						
Patient: ____________________ Date of Surgery: _____________ Weeks Post-Op: _____						
Surgeon: ____________________ Therapist: ____________________						
Thank you so much for completing this form.						
Objective Measurements:						
Knee flexion range of motion:						
Right: ______ Left: ______ Difference ____	**5**	**4**	**3**	**2**	**1**	**0**
Girth (10 cm above superior patellar pole):						
Right: ______ Left: ______ Difference ____	**5**	**4**	**3**	**2**	**1**	**0**
Seated quadriceps MMT:						
Right: ______ Left: ______	**5**	**4**	**3**	**2**	**1**	**0**
Seated hamstring MMT:						
Right: ______ Left: ______	**5**	**4**	**3**	**2**	**1**	**0**
Side-lying gluteus medius MMT:						
Right: ______ Left: ______	**5**	**4**	**3**	**2**	**1**	**0**
Functional Strength Tests:						
Single heel raise (max. reps to 30):						
Divide side with lower reps by opposite side for % difference						
Right: ______ Left: ______ Score: _____%	**5**	**4**	**3**	**2**	**1**	**0**
Single leg dip max. angle:						
Divide side with lower reps by opposite side for % difference						
Right: ______ Left: ______ Score: _____%	**5**	**4**	**3**	**2**	**1**	**0**
Single leg dip endurance (max. reps to 30):						
Divide side with lower reps by opposite side for % difference						
Right: ______ Left: ______ Score: _____%	**5**	**4**	**3**	**2**		**0**
If applicable → secondary grading scale:_____						
Single leg – Single hop test (cm)						
Divide side with lower distance by opposite side for % difference						
Average right: ______ Average left: ______ Score: _____%	**5**	**4**	**3**	**2**	**1**	**0**
Single leg – Crossover hop test (cm)						
Divide side with lower distance by opposite side for % difference						
Average right: ______ Average left: ______ Score: _____%	**5**	**4**	**3**	**2**	**1**	**0**
Single leg – Triple Jump hop test (cm)						
Divide side with lower distance by opposite side for % difference						
Average right: ______ Average left: ______ Score: _____%	**5**	**4**	**3**	**2**	**1**	**0**
Vertical Tuck Jump (10 consecutive reps):						
Number of deviations (listed below): _____	**5**	**4**	**3**	**2**	**1**	**0**
Lower extremity valgus at landing						
Thighs do not reach parallel at peak						
Thigh height not equal						
Feet not landing at the same time						
Excessive trunk flexion						
Agility Tests:						
Shuffle T-test	**5**		**3**			**0**
Sprint/Cutting T-test	**5**		**3**			**0**
	SCORE: _____/out of 70					

MMT, manual muscle test.