

Original Research

The Reliability and Validity of the Modified Butterfly Agility Test

Richard Cahanin^{1a}, Troy Burley¹, Andy Waldhelm² ¹ Physical Therapy, University of South Alabama, ² Physical Therapy, South College Keywords: Sport, performance, screening, agility, test https://doi.org/10.26603/001c.133920

International Journal of Sports Physical Therapy

Vol. 20, Issue 5, 2025

Background

No gold standard criteria exist for return to sport decision-making. The Butterfly Agility Test (BAT) has been shown to be a valid and reliable test for assessing various aspects of athletic performance; however, the space required to conduct it may be prohibitive for clinical environments and it has not been validated among an athletic population.

Purpose

To determine the reliability of the modified BAT (mBAT) and the number practice trials required to optimize its reliability. A secondary purpose was to assess the preliminary validity of the mBAT associated with the Y-Balance Test (YBT) and its relationship with player position among competitive soccer athletes.

Study Design

Multi-phase reliability and validity study

Methods

Phases I and II involved 25 healthy adults (72% female, mean age = 23.4 years) and 45 competitive soccer athletes (55% female, mean age = 18.2 years), respectively. In Phase I, subjects performed 10 consecutive trials of the mBAT to assess test-retest and inter-rater reliability of the mBAT. In Phase II, subjects completed two consecutive trials of the mBAT and YBT during the pre-season period. Intraclass correlation coefficients were used to assess the reliability of the mBAT and repeated measures analysis of variance (ANOVA) was used to assess differences between trials in Phase I. In Phase II, Pearson's product-moment correlation coefficient was used to assess the relationship between mBAT and YBT performance and an ANOVA was used to assess differences in mBAT and YBT performance between player positions.

Results

In Phase I, the mBAT demonstrated excellent reliability (ICC = 0.92-1.00). In Phase II, female collegiate athletes performed better than male high-school athletes on the mBAT (p = 0.01). No significant relationships were found between the mBAT or YBT with player position (p > 0.26).

Conclusion

The mBAT is a reliable physical performance test and three practice trials are recommended prior to recorded trials. The validity of the mBAT for distinguishing athletic performance remains inconclusive, although it appears the mBAT measures a distinct aspect of physical performance compared to the YBT. Although this data is

Corresponding author:
 Richard Cahanin, PT, DPT, PhD
 University of South Alabama, Department of Physical Therapy,
 5721 USA Dr. North
 Mobile, AL., USA 36688
 rcahanin@southalabama.edu

preliminary, the mBAT shows promise as a useful tool for clinicians, trainers, and coaches to assess physical performance among competitive soccer athletes. Larger prospective studies are warranted prior to utilizing the mBAT for clinical decision-making.

Level of Evidence

3b

INTRODUCTION

Currently, there is no "gold standard" testing protocol for return to sport (RTS) decision-making. It is, however, recommended to include measures to identify physical deficits as part of a comprehensive evaluation.^{1,2} Physical performance tests (PPTs) can identify deficits in components of athletic performance, such as speed, agility, balance, and power, which may help inform RTS decision-making.^{3,4}

The Star Excursion Balance Test and the lower quarter Y-Balance Test (YBT), measures of dynamic standing balance and neuromuscular control of the lower extremities, have been linked to increased risk of athletic injury, particularly among soccer, basketball, and football players.⁵⁻⁷ Although relationships with athletic injury have been observed, current evidence suggests that age, gender, and sport/activity should be considered when interpreting test results.⁷ For example, Plisky et. al.⁵ found reach distance in the anterior direction relative to leg-length greater than four centimeters was significant (OR ~ 2.5, p < 0.05) for predicting lower extremity injury among high-school basketball players. In contrast, Read et. al.⁸ found no significant relationship between absolute reach distance difference of four centimeters in the anterior direction and athletic injury among elite male soccer players. Additionally, differences in the quality of injury prediction among athletes categorized according to bone growth maturation were revealed, indicating greater relative reach distances among less skeletally mature athletes compared to more skeletally mature ones, although absolute reach distances were greatest among those who were more skeletally mature.

However, no high-quality evidence exists supporting the use of PPTs for assessing risk of secondary athletic injury, particularly among female athletes.⁹ Healthcare providers should exercise caution interpreting the results of current PPTs for RTS decision-making.^{10,11} The reasons for the shortcomings of RTS protocols are varied. One reason is that some RTS criteria previously studied have not been strict or comprehensive enough.¹² Another reason is that they may not effectively measure the physiologic demands an athlete will be exposed to while participating in sports.^{13,14} Gokeler et. al.¹⁴ purported that current PPTs for ACL injury screening are "closed skills" and do not reflect actual physical demands on the field. Regardless, strict RTS criteria fail to explain relatively high initial athletic injury rates, highlighting the need to identify athletes at risk of injury prior to participation in sports.^{5,6}

Before implementing a PPT into routine practice, its reliability should be verified. Also, to account for the effects of motor learning, multiple practice trials of a PPT are recommended prior to recording trials. In this manner, subsequent performance can be more confidently attributed to changes in physical performance and not a learning effect. The optimal number of practice trials for every PPT is variable and test dependent¹⁵⁻¹⁷; therefore, a PPT should undergo testing to confirm how many practice trials are required to achieve performance stabilization.

Additionally, PPTs should be validated within the population they are used. Soccer is a physically demanding sport requiring a combination of speed, cardiopulmonary endurance, muscular strength, muscular endurance, muscular power, and agility. Depending on the player position, these athletes may be required to run continuously, often upwards of 10 km during a single match, with variable intensities.¹⁸ In addition, players may be required to perform explosive movements, such as jump for headers, accelerate or decelerate quickly, and make quick directional changes. Also, players may need to be able to coordinate their movements while controlling a ball and react quickly to other players and the ball, which is particularly important for goalkeepers. Clearly, soccer requires a comprehensive set of physical abilities, but many clinical tests to assess physical performance only assess a single aspect of it, such as muscular power or agility. The Butterfly Agility Test (BAT) was developed to overcome this limitation and provide clinicians a more comprehensive assessment of physical performance among athletes participating in high-intensity sports involving running, jumping, and cutting, such as soccer.

The BAT has demonstrated good reliability (ICC = 0.89, 95% CI = 0.23-0.97), and moderate-to-strong relationships with multiple aspects of physical performance among young adults, including agility, muscular power, and speed (r = 0.50-0.77).¹⁹ The BAT includes tasks commonly performed in high-intensity sports and has potential utility as a pre-participation screening test or as a component to a comprehensive evaluation for RTS decision-making during the end stage of rehabilitation from an injury. However, the BAT has not been validated to identify the risk of athletic injury or for RTS assessment. Additionally, the BAT requires ~ 6 m² of floor/ground space and a box 42 cm in height, which may not be conducive for some clinical practices with space limitations. Modifications to the original BAT may enable more clinicians to adopt the test in their clinical practice. Therefore, the purpose of this study was to determine the reliability of the modified BAT (mBAT) and the number practice trials required to optimize its reliability. A secondary purpose was to assess the preliminary validity of the mBAT associated with the Y-Balance Test (YBT) and its relationship with player position among competitive soccer athletes.

The research hypotheses were that 1) the mBAT would demonstrate excellent reliability and that its reliability would improve after multiple trials 2) the mBAT would demonstrate a statistically significant relationship with the results of the YBT, and 3) "defenders" would perform significantly better on the mBAT than athletes playing other positions among competitive soccer athletes.

MATERIALS AND METHODS

DESIGN

A multi-phase reliability and validity study was designed. The study protocol was approved by the University's research ethics committee prior to subject recruitment.

PARTICIPANTS

In Phase I, twenty-nine healthy adult participants were recruited from January to March 2022. In Phase II, fortytwo competitive soccer athletes were recruited from July 2023-January 2024. A power analysis according to G-Power (version 3.1.9.4) indicated a minimal total sample of 14 subjects was needed to conduct a 2 x 10 repeated measures analysis of variance (ANOVA), 52 subjects for an independent t-test, 108 subjects for a chi-squared test, and 128 for a multivariate ANOVA with five groups.

Participants in either phase were excluded if they reported current pregnancy, low weekly exercise frequency and intensity, recent lower body pain while running, jumping, cutting, or shuffling, or were not cleared for unrestricted participation in sport activities. Additionally, subjects were excluded if they answered "yes" to any COVID-19 screening questions or any of the first seven questions of the Physical Activity Readiness Questionnaire (PAR-Q). Finally, if participants had a BMI > 30, they were excluded. All participants provided written consent prior to induction into the study.

INSTRUMENTATION

Each participant's height and weight were measured using a physician's scale (Detecto SP7044, Webb City, MO, USA) in Phase I. In Phase II, height and weight were self-reported by athletes. The YBT Kit (Functional Movement Systems, Inc. Chatham, VA) was used to measure dynamic singlelimb standing balance in Phase II. In both phases, the mBAT was measured with a hand-controlled stopwatch (Robic 1000W; Waco, Tx, USA).

MATERIALS AND CONDITIONS

During Phase I, testing was performed inside a temperature-controlled room on a wooden platform floor; room temperature and humidity was 68° F and 30%, respectively. All participants wore athletic footwear and apparel during testing. Also, a 46 x 57 x 30 cm hand-made wooden box with a vulcanized rubber top surface was utilized for mBAT testing. During Phase II, testing was performed outside on artificial turf or track; ambient temperature and humidity ranged 64°-92° F and 21-79%, respectively. All athletes wore standard soccer cleats and team jerseys during mBAT testing, although cleats were removed for the YBT; also, shin guards were removed for all testing. Finally, a 32 x 32 x 31 cm metal-framed platform (Rage Fitness, CF-PB012/S, Gibson Athletic, Denver, CO) with a vulcanized rubber top surface was used for mBAT testing.

TEST ADMINISTRATION

Prior to testing, all participants received verbal instructions with a demonstration of testing procedures and were guided through a five-to-ten-minute dynamic warm-up. Each physical performance test was administered as follows.

MODIFIED BUTTERFLY AGILITY TEST (MBAT)

To setup the test, distances between the front and center of the box/platform and each cone were measured as measured using a fiberglass tape measure (Empire 100' Long Reel, Mukwonago, WI, USA). Next, the angle between each cone and the midline was measured with a universal goniometer. Once ready to begin a trial, each subject began the test while standing on top of the box/platform. Then, each subject was instructed to complete the steps of the test as quickly as possible without errors (Figure 1). The assessor(s) used either a black or white flag, approximately 18 x 15 cm in dimension, held near the center of their chest during step 3, to indicate the next cone subjects were to move toward during step 4. The assessor(s) used handtimers (Robic 1000W; Waco, Tx, USA) to measure the time required for subjects to complete each trial.

Y-BALANCE TEST (YBT)

The YBT protocol described by Plisky et. al.²⁰ was utilized. Without shoes, each subject stood in single-limb stance on the YBT platform with their longest toe touching the starting demarcation line. Next, subjects were instructed to slide the reach indicator with their contralateral lower extremity as far as possible along the measurement rod. If a subject lost balance during the test, lost contact with the reach indicator, or demonstrated inappropriate use of the reach indicator, the trial was repeated. Subjects performed maximal reaches in the anterior, posterolateral, and posteromedial directions associated with the device.

DATA COLLECTION

During Phase I, three independent assessors recorded trials. The principal investigator, a physical therapist (PT), recorded all trials for each participant and two student physical therapists (SPTs) alternated recording trials. This procedure allowed for assessment of concurrent inter-rater reliability between each SPT and the PT and intra-rater reliability across trials for the PT. Also, testing errors were assessed visually and a survey about the testing experience was administered by a third SPT. During Phase II, testing was administered and recorded by two PTs and five SPTs.

During Phase I, participants performed the mBAT ten consecutive times with three-minutes of rest between each trial. During Phase II, the mBAT and the YBT were per-



Figure 1. mBAT Schematic and detailed step-by-step procedures

Step 1	The participant performs drop jump from the top of box to the floor/ground directly in front of the box The assessor starts the timer upon initial foot-to-ground/floor contact	
Step 2	Participant runs forward around the 1 st cone	
Step 3	Participant runs forward back towards the box, positioning themselves to the rear of the box and turns to face the 1 st cone/assessor. Meanwhile the assessor holds a flag (either black or white) up in front of chest to indicate the next direction the participant should move towards	
Step 4	Assuming the assessor raises a black flag, the participant laterally shuffles towards their right side (i.e., 2 nd cone in schematic)	
Step 5	Upon transition around the 2 nd cone, the participant back-pedals towards the 3 rd cone	
Step 6	Upon transition around the 3 rd cone, the participant runs forward towards the rear of the box	
Step 7	Upon transition around the rear of the box, the participant laterally shuffles towards the $4^{ m th}$ cone	
Step 8	Upon transition around the $4^{ m th}$ cone, the participant back-pedals towards the $5^{ m th}$ cone	
Step 9	Upon transition around the $5^{ m th}$ cone, the participant runs forward towards the rear of the box	
Step 10	Once positioned behind the box, the participant performs a bilateral vertical jump landing on top of the box with both feet striking the top surface of the box simultaneously	
Step 11	Once the feet land on top of the box, the participant immediately transitions to another drop-jump to the floor/ground, landing directly in front of the box	
Step 12	The participant performs bilateral broad jumps forward towards the $1^{ m st}$ cone	
Step 13	Upon transition around the 1 st cone, the participant performs bilateral broad jumps forward back towards the rear of the box	
Step 14	Finally, once positioned behind the box, the participant performs a bilateral vertical jump landing on top of the box with both feet striking the top surface of the box simultaneously. The assessor stops the timer upon foot contact to the top of the box	

formed, in random order, twice with one-minute of rest between each trial; the mean of two recorded trials of each test was used for statistical analysis. Subjects completed a post-test survey regarding areas of difficulty during testing.

STATISTICAL ANALYSIS

Statistical analysis was performed by the principal investigator. Data were analyzed using Statistical Package for Social Sciences (SPSS) version 27 (IBM Corp., Armonk, N.Y., USA). Alpha level of significance was set at 0.05 and power at 0.80 for all analyses. Outliers were identified using the

Table 1.	Subject	characteristics
----------	---------	-----------------

	Phase I (n = 25) Mean ± s.d.	Phase II (n = 42) Mean ± s.d.	Test statistic (p-value)
Gender (% female)	72	55	χ2 = 1.96 (0.16)
Height (m)	1.71 ± 0.10	1.70 ± 0.07	t = 0.61 (0.55)
Weight (kg)	67.44 ± 13.09	63.88 ± 10.25	t = 1.16 (0.25)
Body mass index (kg/m ²)	22.80 ± 2.75	22.07 ± 1.36	t = 1.06 (0.29)
Age (years)	23.40 ± 1.80	18.21 ± 2.59	t = 9.69 (0.00*)
Exercise Intensity (% > 80% / 30 – 80% / < 30% of max effort/session)	40/52/8	100/0/0	χ2 = 32.47 (0.00*)
Sports Participation (% yes)	60	100	χ2 = 29.11 (0.00*)
Exercise Duration (% > 180 / 60 - 180 min./week)	68/32	100/0	χ2 = 12.37 (0.00*)
Mean mBAT (sec.)	17.18 ± 1.91	13.68 ± 0.91	t = 9.33 (0.00*)

s.d.=standard deviation, m= meter, kg= kilogram, max = maximal, min = minute, sec. = second

* indicates significance at p < 0.05

outlier labeling technique. Participants with outlying datum in three or more trials were excluded from final data analysis. For Phases I and II, data from 25 and 42 participants were included in the final analysis, respectively.

Descriptive statistics, chi-squared tests of independence, and independent t-tests were used to analyze subject characteristics in both phases (Table 1). In Phase I, a repeated measures analysis of variance (ANOVA) with posthoc paired t-tests were used to analyze differences between trials among subjects. Pearson's product-moment correlation coefficients were calculated to determine the relationships between each set of trials; correlations were considered strong if greater than 0.75, moderate if between 0.50 and 0.75, and fair if less than 0.50. Inter- and intra-rater reliabilities were analyzed using intraclass correlation coefficients (ICC_{3, 1}); reliability was considered to be "excellent" for measures yielding a coefficient of 0.90 or higher, "good" for values between 0.75 to 0.89, "moderate" for values between 0.50 to 0.74, and poor if less than 0.50. The levels of acceptable coefficients for correlations and reliability were based upon previously recommended guidelines.¹⁸ In phase II, an ANOVA with post-hoc Scheffe comparisons was used to analyze mBAT and YBT performance differences between player positions. Additionally, post-hoc independent t-tests were used to analyze mBAT and YBT performance differences between two player positions. A comparison of mean mBAT performance between Phases I and II was performed with an independent t-test.

RESULTS

Overall, two of three research hypotheses were rejected and one was unable to be rejected. Among the rejected, the mBAT did not demonstrate a statistically significant relationship with the YBT and defenders did not perform better than other player positions. However, both the intra- and inter-rater reliability of the mBAT were found to be excellent.



Figure 2. Phase I mBAT performance across trials

Mean mBAT times (left y-axis) \pm s.d. measured concurrently by a PT (green circles) and a SPT (red squares) across ten trials (x-axis) during phase I of the study are displayed. The sum of errors (black triangles, right y-axis) committed by all subjects across trials in phase I of the study are displayed. A trend towards reduced mBAT times and a lower number of errors committed during each trial was observed across trials, with the lowest times and errors observed during trial ten. mBAT = modified Butterfly Agility Test, s.d. = standard deviation, sec. = second, PT = physical therapist, SPT = student physical therapist

Subject characteristics differed between Phases I and II (Table 1). Although there were no differences in height, weight, or gender between subjects in Phases I and II, subjects in Phase II participated in competitive sport, exercised with greater intensity and for longer duration, and performed better on the mBAT compared to subjects in Phase I. Fifty-five percent of athletes in Phase II were females competing in NCAA Division I, whereas 45% were males competing in Alabama 6A high-school soccer. Regarding player positions, most athletes, 40% (n = 17), reported playing "forward," 17% (n = 7) were "wide defenders," 14% (n = 6) were "wide forwards," 12% (n = 5) were goal keepers,10% (n = 4) were "defenders;" and 7% (n = 3) did not report their position. Lastly, 31% (n = 13) reported having sustained a lower extremity injury within the prior 12-months.

In Phase I, significant differences in mBAT times were found between trials (F [1,24] = 3,017-3,070, p < 0.00). The mean mBAT times were 19.30 ± 2.50 seconds across trials 1-3, 16.98 ± 1.85 seconds across trials 4-6, and 16.33 ± 1.64 seconds across trials 7-10 (Figure 2).

	Inter-rater (SPT) ICC _(3,1) (95% CI)/SEM (sec.)/MDD ₉₅ (sec.)	Intra-rater (PT) ICC _(3,1) (95% CI)/SEM (sec.)/MDD ₉₅ (sec.)
Trials 1 & 2	0.66(0.22-0.85)/1.68/4.65	0.67(0.25 - 0.85)/1.67/4.62
Trials 2 & 3	0.81(0.57 - 0.92)/0.85/2.34	0.81(0.57 - 0.92)/0.86/2.38
Trials 3 & 4	0.91(0.80 - 0.96/0.53/1.46	0.92(0.81 - 0.96)/0.50/1.39
Trials 4 & 5	0.93(0.84 - 0.97)/0.51/1.41	0.92(0.82 - 0.97)/0.54/1.51
Trials 5 & 6	0.87(0.70 - 0.94)/0.67/1.87	0.86(0.68 - 0.94)/0.69/1.92
Trials 6 & 7	0.92(0.83 - 0.97)/0.45/1.25	0.92(0.82 - 0.97)/0.44/1.23
Trials 7 & 8	0.92(0.82 - 0.96)/0.44/1.22	0.91(0.80 - 0.96)/0.46/1.27
Trials 8 & 9	0.94(0.86 - 0.97)/0.40/1.12	0.94(0.86 - 0.97)/0.40/1.11
Trials 9 & 10	0.88(0.73 - 0.95)/0.54/1.51	0.88(0.73 - 0.95)/0.53/1.48

Table 2. Test-retest reliability of the mBAT across trials and assessors

SPT = student physical therapist, PT = physical therapist, mBAT = modified Butterfly Agility Test, ICC(3,1) = intraclass correlation coefficient, CI = confidence interval, SEM = standard error of the measurement, MDD95 = minimal detectable difference, sec. = seconds

Mean mBAT times (left y-axis) \pm s.d. measured concurrently by a PT (green circles) and a SPT (red squares) across ten trials (x-axis) during phase I of the study are displayed. The sum of errors (black triangles, right y-axis) committed by all subjects across trials in phase I of the study are displayed. A trend towards reduced mBAT times and a lower number of errors committed during each trial was observed across trials, with the lowest times and errors observed during trial ten. mBAT = modified Butterfly Agility Test, s.d. = standard deviation, sec. = second, PT = physical therapist, SPT = student physical therapist

Pairwise comparisons of mBAT times measured by either the PT or SPTs revealed no significant differences between trials 1 and 2 (p = 1.00), between trials 1 and 3 (p > 0.05), or between trials 2 and 3 (p > 0.34); however, significant differences were found between trial 1 and trials 4-10 (p < 0.00), trial 2 and trials 4-10 (p < 0.00), and trial 3 and trials 4-10 (p < 0.04); differences between trials 4-10 were variable (p= 0.01 – 1.00). Also, moderate correlations were found between each of the first two pairs of mBAT trials (r = 0.55 -0.70; p < 0.00) and strong correlations were found in the remaining seven pairs (r = 0.77 - 0.89; p < 0.00), indicating data stabilization after the third trial and agreement with the first research hypothesis, which anticipated improved mBAT performance across multiple trials.

Similarly, agreement with the first research hypothesis anticipating excellent mBAT reliability was observed. The assessors' (i.e., PT and SPTs) concurrent inter-rater reliability was excellent during each mBAT trial. Also, test-retest inter-rater and intra-rater reliability among the SPTs and the PT, respectively, was moderate between trials 1 and 2, was good between trials 2 and 3, was excellent between trials 3 and 4, and was excellent between trials 4 and 5. Test-retest reliability among both SPTs and the PT between the remaining trials (i.e., 5-10) varied from good to excellent (Table 2 and Figure 3).

Inter-rater and intra-rater test-retest reliability point estimates surpass 0.90 threshold (black dotted horizontal line), indicating excellent reliability, between trials 3 and 4. Also, the 95% confidence intervals (black t-bars) become narrower between trials 3 and 4, indicating improved mea-



Figure 3. Test-retest reliability of the mBAT across trials and assessors

Inter-rater and intra-rater test-retest reliability point estimates surpass 0.90 threshold (black dotted horizontal line), indicating excellent reliability, between trials 3 and 4. Also, the 95% confidence intervals (black t-bars) become narrower between trials 3 and 4, indicating improved measurement precision relative to trials 1 and 2 and trials 2 and 3. ICC(3,1) = intraclass correlation coefficient, SPT = student physical therapist, mBAT = modified Butterfly Agility Test

surement precision relative to trials 1 and 2 and trials 2 and 3. ICC(3,1) = intraclass correlation coefficient, SPT = student physical therapist, PT = physical therapist, mBAT = modified Butterfly Agility Test

According to the post-test survey in Phase I, participants reported the most difficulty with coordinating broad jumps (80%) and lateral shuffling (12%). Also, most participants (64%) reported they gained confidence in their performance after the third trial. Objectively, a total of 236 errors were committed among all participants, of which 110 (47%) were committed within the first three trials. Most errors were committed during the first trial (22%). A strong correlation was found between mBAT and the total number of errors committed during each trial (r = 0.96; p < 0.00), indicating more errors were associated with worse mBAT performance (Figure 2).

In Phase II, the mean mBAT and relative composite YBT were 13.68 \pm 0.91 sec. and 0.98-0.99 \pm 0.07%, respectively, however no significant correlations were observed between the mBAT and the YBT (r = 0.21-0.22, *p* > 0.17), indicating that these tests may not assess the same aspect of physical

performance. Therefore, the second research hypothesis was rejected, which anticipated a statistically significant relationship between these PPTs. Similarly, no significant relationships were found between player positions and either the mBAT (F[4] = 0.81, p = 0.53, $\eta^2 = .09$, $1-\beta = 0.23$) or YBT (F(8) =0.54, p = 0.82, $\eta^2 = .06$, $1-\beta = 0.23$), which led to the rejection of the third research hypothesis anticipating better performance among defenders compared to other position players. However, moderate relationships were found for age (r = -0.48, p < 0.00) and height (r = 0.32, p = 0.04), indicating relatively older and shorter athletes performed better on the mBAT. Finally, significant differences in mBAT performance between athletes in Phases I and II were found (t(38) = 9.76, *p* < 0.00, d = 3.27), which indicated competitive soccer athletes performed better on the mBAT than gender and body composition matched recreationally active adults.

DISCUSSION

The results of the current study are consistent with a previous study assessing the reliability of the BAT,¹⁹ which provides additional evidence that the mBAT is a reliable test among young recreationally active adults. The current study implemented modifications to the BAT, which included reducing box/platform dimensions and changes to the cone orientation. The box/platform height was reduced from 42 cm to 30 cm in order to reduce the cost of implementation and to improve participant safety. Also, a 30 cm box/platform height is consistent with what is used in other PPTs, such as the drop jump^{21,22} and Landing Error Scoring System^{23,24} tests, which reduces the need for additional equipment to conduct all of these tests. Additionally, the distance between the box/platform and the first cone was reduced from 3 m to 1.83 m, a reduction of 1.17 m. These changes may enable clinicians to implement the mBAT in smaller spaces where the original BAT would be space prohibitive.

In Phase I, the results indicated that three mBAT practice trials should be performed before recording trials in order to optimize test reliability. Not only was this indicated objectively, but subjectively as well. An analysis of errors committed across trials and participants indicated a disproportionate number of errors occurring during the first three trials (i.e., 47%). These errors were likely influenced by the cognitive process of motor learning.²⁵ In Phase I of the current study, only 20% of participants reported confidence in their performance within the first three trials, further supporting the need for at least three practice trials. Since no detrimental effect was observed in performance after the third trial, the effect of fatigue during latter trials was negligible, which indicates the three-minute rest period was excessive for physiologic recovery between trials. In fact, the mean mBAT times continued to improve after trial four, indicating continued performance improvement with successive trials. Overall, the mean mBAT measured by both sets of assessors were worst during trial one and best during trial ten, with a difference in mean time of 4.7 sec. between the two trials. However, this difference should

be interpreted with caution considering overlapping error margins across all trials (Figure 2). Additionally, the results demonstrate that the mBAT does not require extensive training, knowledge, or time to conduct, which makes it conducive for clinical practice among clinicians with various experience levels, including SPTs.

In Phase II, female collegiate competitive soccer athletes performed better on the mBAT than high-school competitive males, providing preliminary evidence the mBAT may be able to distinguish between levels of athletic ability. Although there was no significant difference in mBAT between five different player positions, the analysis was underpowered (1- β = 0.23).

LIMITATIONS

The small sample size was a limitation of the current study, particularly in Phase II, which may have caused a Type II error. Previous studies assessing test reliability of the BAT, YBT, or Star Excursion Balance Test have used a sample size less than twenty-five, as used in the current study.^{19,20,26, 27} Therefore, the authors believe the sample size was sufficient in Phase I to address the research question. However, in Phase II, a univariate ANOVA analyzing differences in mBAT performance between five player positions demonstrated a poor effect size with inadequate power ($\eta^2 = 0.09$, 1- $\beta = 0.23$), which favors the probability of a Type II error having occurred.

Another limitation was the differing environments where testing was conducted between Phases/groups. Subjects performed the mBAT indoors during Phase I, whereas they performed it outdoors in Phase II. Also, the mBAT was performed on a wooden platform in Phase I and either on artificial turf (i.e., girls) or a rubber track (i.e., boys) in Phase II. Additionally, within Phase II, data for the boys were collected in January, whereas they were collected in late July for the girls, indicating a 28° ambient temperature and a 58% humidity difference between testing environments for the two groups.

Diverging methodologies between Phases I and II was another limitation. First, the box used for the mBAT had different dimensions between Phases I and II. Although both boxes were approximately the same height, the box used in Phase I was 14 cm longer and 25 cm wider than the box used in Phase II. A commercially produced box was implemented in Phase II in order to allow multiple tests to be performed simultaneously. Additionally, the authors believed that the rubberized top and base stabilizers provided with the commercially produced box would enhance participant safety and facilitate accessibility of materials used to conduct testing. Therefore, differences in mBAT performance between Phases and groups during Phase II should be interpreted with caution. Second, rest periods between trials differed between Phases I and II. In Phase I, three minutes of rest was provided between trials, whereas only one minute of rest was provided in Phase II. Although different, the authors learned from Phase I that three minutes of rest between trials was excessive. In order to improve data collection efficiency during Phase II, the rest time was reduced to approximately one minute which appeared sufficient as no participant was forced to initiate a trial prior to self-reporting readiness to begin.

Future studies should include large prospective designs (i.e., a minimum sample size of 130) with a standardized protocol to determine whether or not the mBAT can differentiate levels of performance among athletes and to determine the predictive ability of the mBAT to identify individuals at risk of initial or secondary athletic injuries. Also, normative mBAT values and cutoff scores for both pre-participation sports screening and return to sport decision-making should be established among various sports and athletic skill levels. If more measurement precision is desired, an electronic timing gate system could be integrated into the test. Similarly, an electronic light system could be integrated to reduce the burden on the assessor during the reactive component of the test. Finally, additional measurements, other than time, could be added to the test, such as a qualitative or quantitative assessment of joint kinematics, kinetics, muscle activity, or other physiologic tests, which may be useful for comparison of physiologic demands of testing with those required of players during competitive play.

CONCLUSION

The mBAT is reliable, quick to administer, inexpensive to deploy, and can be implemented by clinicians with various experience levels, including SPTs. It requires a 30 cm box, a stopwatch, five small cones, two different colored flags, and $\sim 6 \text{ m}^2$ floor area. A dynamic warm-up with specific instruction focusing on broad jump technique should be administered and three practice trials should be performed before averaging the results of multiple recorded trials. When

using a hand-timer for measurement, the precision of the measurement is estimated to be about 0.5 seconds. A change in time of two seconds is suggested as the threshold to indicate a significant performance difference. The mBAT can be deployed indoors or outdoors with modifications and may have potential to distinguish physical abilities among athletes, which could be useful information for coaches and trainers for player position assignment and training regimen prescription. Although data associated with the current study is considered preliminary, the mBAT shows promising utility and may be considered for inclusion within a pre-sport participation testing battery or for return to sport decision-making among competitive soccer athletes. However, large prospective validation studies are warranted prior to using the mBAT for clinical decision making.

CONFLICTS OF INTEREST

The authors report no conflicts of interest.

ACKNOWLEDGEMENT

We would like to thank Cynthia McKinnis, Rachael Smith, Caroline Manemann, Nathaniel Roberts, Madeleine Gill, Cade Harris, Andrew Allen, and Alexandra Hall for their contribution to this work

Submitted: November 26, 2024 CDT. Accepted: March 16, 2025 CDT. Published: May 01, 2025 CDT. © The Author(s)

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. Cheney S, Chiaia TA, de Mille P, Boyle C, Ling D. Readiness to return to sport after ACL reconstruction: A combination of physical and psychological factors. *Sports Med Arthrosc Rev.* 2020;28(2):66-70. doi:10.1097/JSA.00000000000263

2. Paterno MV, Huang B, Thomas S, Hewett TE, Schmitt LC. Clinical factors that predict a second ACL injury after ACL reconstruction and return to sport: preliminary development of a clinical decision algorithm. *Orthop J Sports Med*. 2017;5(12):2325967117745279. doi:10.1177/ 2325967117745279

3. Grindem H, Snyder-Mackler L, Moksnes H, Engebretsen L, Risberg MA. Simple decision rules can reduce reinjury risk by 84% after ACL reconstruction: the Delaware-Oslo ACL cohort study. *Br J Sports Med*. 2016;50(13):804-808. doi:<u>10.1136/</u> <u>bjsports-2016-096031</u>

4. Toole AR, Ithurburn MP, Rauh MJ, Hewett TE, Paterno MV, Schmitt LC. Young athletes cleared for sports participation after anterior cruciate ligament reconstruction: how many actually meet recommended return-to-sport criterion cutoffs? *J Orthop Sports Phys Ther.* 2017;47(11):825-833. doi:10.2519/jospt.2017.7227

5. Plisky PJ, Rauh MJ, Kaminski TW, Underwood FB. Star Excursion Balance Test as a predictor of lower extremity injury in high school basketball players. *J Orthop Sports Phys Ther*. 2006;36(12):911-919. doi:10.2519/jospt.2006.2244

6. Gribble PA, Terada M, Beard MQ, et al. Prediction of lateral ankle sprains in football players based on clinical tests and body mass index. *Am J Sports Med.* 2016;44(2):460-467. doi:10.1177/0363546515614585

7. Plisky P, Schwartkopf-Phifer K, Huebner B, Garner MB, Bullock G. Systematic review and meta-analysis of the Y-Balance Test Lower Quarter: reliability, discriminant validity, and predictive validity. *Int J Sports Phys Ther.* 2021;16(5):1190-1209. doi:10.26603/001c.27634

8. Read PJ, Oliver JL, Myer GD, Farooq A, De Ste Croix M, Lloyd RS. Utility of the anterior reach Y-BALANCE Test as an injury risk screening tool in elite male youth soccer players. *Phys Ther Sport.* 2020;45:103-110. doi:10.1016/j.ptsp.2020.06.002

9. Fältström A, Hägglund M, Hedevik H, Kvist J. Poor validity of functional performance tests to predict knee Injury in female soccer players with or without anterior cruciate ligament reconstruction. *Am J Sports Med.* 2021;49(6):1441-1450. doi:10.1177/03635465211002541

10. Losciale JM, Zdeb RM, Ledbetter L, Reiman MP, Sell TC. The association between passing return-tosport criteria and second anterior cruciate ligament injury risk: a systematic review with meta-analysis. *J Orthop Sports Phys Ther*. 2019;49(2):43-54. doi:10.2519/jospt.2019.8190

11. Harrison JJ, Yorgey MK, Csiernik AJ, Vogler JH, Games KE. Clinician-friendly physical performance tests for the knee. *J Athl Train*. 2017;52(11):1068-1069. doi:<u>10.4085/</u>1062-6050-52.11.19

12. Kaplan Y, Witvrouw E. When is it safe to return to sport after ACL reconstruction? reviewing the criteria. *Sports Health*. 2019;11(4):301-305. doi:<u>10.1177/1941738119846502</u>

13. Mann JB, Ivey PA, Mayhew JL, Schumacher RM, Brechue WF. Relationship between agility tests and short sprints: reliability and smallest worthwhile difference in National Collegiate Athletic Association Division-I football players. *J Strength Cond Res*. 2016;30(4):893-900. doi:<u>10.1519/</u> JSC.000000000001329

14. Gokeler A, Benjaminse A, Della Villa F, Tosarelli F, Verhagen E, Baumeister J. Anterior cruciate ligament injury mechanisms through a neurocognition lens: implications for injury screening. *BMJ Open Sport Exerc Med*. 2021;7(2):e001091. doi:<u>10.1136/bmjsem-2021-001091</u>

15. Bolgla LA, Keskula DR. Reliability of lower extremity functional performance tests. *J Orthop Sports Phys Ther*. 1997;26(3):138-142. doi:<u>10.2519/</u> jospt.1997.26.3.138

16. Mok KM, Bahr R, Krosshaug T. Reliability of lower limb biomechanics in two sport-specific sidestep cutting tasks. *Sports Biomech*. 2018;17(2):157-167. doi:<u>10.1080/14763141.2016.1260766</u>

17. Linek P, Sikora D, Wolny T, Saulicz E. Reliability and number of trials of Y Balance Test in adolescent athletes. *Musculoskelet Sci Pract.* 2017;31:72-75. doi:<u>10.1016/j.msksp.2017.03.011</u> 18. Chmura P, Andrzejewski M, Konefał M, Mroczek D, Rokita A, Chmura J. Analysis of motor activities of professional soccer players during the 2014 World Cup in Brazil. *J Hum Kinet*. 2017;56:187-195. doi:10.1515/hukin-2017-0036

19. Cahanin R, Esleck B, Hardy J, Bass B, Rogers T, Looney K. Development of the Butterfly Agility Test. *Phys Ther Sport*. 2021;52:38-44. doi:<u>10.1016/</u> j.ptsp.2021.08.004

20. Plisky PJ, Gorman PP, Butler RJ, Kiesel KB, Underwood FB, Elkins B. The reliability of an instrumented device for measuring components of the Star Excursion Balance Test. *N Am J Sports Phys Ther.* 2009;4(2):92-99.

21. Mackala K, Rauter S, Simenko J, et al. The effect of height on drop jumps in relation to somatic parameters and landing kinetics. *Int J Environ Res Public Health*. 2020;17(16). doi:<u>10.3390/</u>ijerph17165886

22. Kingston B, Murray A, Norte GE, Glaviano NR. Validity and reliability of 2-dimensional trunk, hip, and knee frontal plane kinematics during single-leg squat, drop jump, and single-leg hop in females with patellofemoral pain. *Phys Ther Sport*. 2020;45:181-187. doi:10.1016/j.ptsp.2020.07.006

23. Hanzlíková I, Hébert-Losier K. Clinical implications of landing distance on Landing Error Scoring System scores. *J Athl Train*.
2021;56(6):572-577. doi:10.4085/1062-6050-068-20

24. Padua DA, Marshall SW, Boling MC, Thigpen CA, Garrett WE, Beutler AI. The Landing Error Scoring System (LESS) ss a valid and reliable clinical assessment tool of jump-landing biomechanics: the JUMP-ACL study. *Am J Sports Med*. 2009;37(10):1996-2002. doi:<u>10.1177/</u> <u>0363546509343200</u>

25. Fitts P, M P. *Human Performance*. Brooks/Cole Publishing Company; 1967.

26. Calatayud J, Borreani S, Colado JC, Martin F, Flandez J. Test-retest reliability of the Star Excursion Balance Test in primary school children. *Phys Sportsmed*. 2014;42(4):120-124. doi:10.3810/ psm.2014.11.2098

27. Kinzey SJ, Armstrong CW. The reliability of the Star-Excursion Test in assessing dynamic balance. *J Orthop Sports Phys Ther*. 1998;27(5):356-360. doi:<u>10.2519/jospt.1998.27.5.356</u>