Development of a Modified Margaria-Kalamen Anaerobic Power Test for American Football Athletes

Ronald K. Hetzler, Rachele E. Vogelpohl, Christopher D. Stickley, Allison N. Kuramoto, Mel R. DeLaura, and Iris F. Kimura

Department of Kinesiology and Rehabilitation Science, Human Performance Laboratory, College of Education, University of Hawaii–Manoa. Honolulu, Hawaii

Abstract

Hetzler, RK, Vogelpohl, RE, Stickley, CD, Kuramoto, AN, DeLaura, MR, and Kimura, IF. Development of a modified Margaria-Kalamen anaerobic power test for American football athletes. *J Strength Cond Res* 24(4): 978–984, 2010—This study examined a modification of the Margaria-Kalamen test for football players. The football stair climb test (FST) protocol used in this study increased the vertical displacement (20 steps, 3.12 m) so that the mean best time for the test was 2.048 ± 0.267 seconds. Fifty-eight Division I-A football players volunteered to participate (mean ± *SD* age = 20.2 ± 1.8 yr, height = 184.1 ± 7.7 cm, weight = 102.5 ± 19.4 kg). Subjects performed 25 trials with 30 to 40 seconds of rest between trials. Test-retest reliability was determined using 34 subjects by way of intraclass correlation coefficients with a value of 0.73 for peak power and *SEM* of 105.4 W, indicating an acceptable level of reliability. Subjects were divided into 3 groups by position: linemen (Line), skill, and linebackers (LB). Alpha level was $p < 0.05$. Peak power was 1674.5 ± 300.8, 1712.6 ± 251.5, and 1388.6 ± 210.4 W for the LB, Line, and Skill groups, respectively. Groups were significantly different ($p < 0.0001$), with the LB and Line found to be more powerful than the Skill group. Peak power continued to increase throughout the 25 trials in the Skill and LB group but plateaued after approximately 17 trials in the Line group. It was concluded that the FST was a reliable test for measuring peak anaerobic power in collegiate football players, which, theoretically, should provide more accurate measures of peak power caused by increased vertical displacement and longer duration, resulting in a decreased influence of cheating strategies during test administration. To achieve maximal power in stair climbing tasks, coaches may need to incorporate a greater number of trials or a more intense warm-up than has been previously reported.

Key Words electronic timing, stair climbing, maximal power

Introduction

Football has been described as a highly explosive sport, consisting of sprinting, jumping, and colliding activities, requiring heavy reliance on anaerobic power for success (12). This type of power primarily uses the phosphagen energy system (adenosine triphosphate and creatine phosphate), allowing for quick movements lasting from a fraction of a second to approximately 5 seconds (12). Therefore, coaches and recruiters spend time, effort, and resources evaluating the anaerobic power of athletes, hoping to quantify an athlete's potential for success on the playing field (10). Numerous tests exist that may be used to measure anaerobic power including the Margaria-Kalamen test, 40-yard dash, Wingate anaerobic power test, vertical jump, broad jump, power clean, and Olympic snatch. Stair climbing has been reported to possess both face and concurrent validity for measuring explosive power (3), and the Margaria-Kalamen test has been shown to measure both horizontal and vertical power, potentially making it a more accurate predictor of football playing ability than other anaerobic power tests (12).

Margaria et al. (6), when developing the Margaria anaerobic power step test, found that the speed of progression up the steps increases from a standing start to reach a maximum constant value in about 1.5 to 2.0 seconds and then remains constant after 4 to 5 seconds. However, when using a 2-m run-up start, Margaria et al. (6) determined that only 0.5 to 1.0 seconds is required to measure peak speed of stair climbing. Subsequently, Kalamen (5) reported that a vertical distance of 1.05 m with a 6-m run-up resulted in greater power outputs; however, the resulting Margaria-Kalamen test still may be completed in under 0.5 seconds by football players (2,12), faster than the necessary interval suggested by Margaria et al. (6). A recently developed

Address correspondence to Ronald K. Hetzler, hetzler@hawaii.edu.

24(4)/978–984

Journal of Strength and Conditioning Research

modification of the Marageria-Kalamen test requires subjects to climb a vertical distance of 2.04 m (3). The mean time for completion of this test for college age students was 1.4 ± 0.2 seconds for males and 1.7 ± 0.3 seconds for females. Therefore, the time required to complete the 2.04 vertical distance in their study should have been adequate to elicit peak velocity according to the findings of Margaria et al. (6). However, the time required for trained football players to complete their test is not known, although it would likely be considerably faster. In addition, previous studies have shown there is an ability to "cheat" during the Margaria-Kalamen test by tripping the timing device with the lead leg before the center of gravity reaches the terminal step, which can inflate scores by 10–24% (7,15). It appears logical then that a longer test may better represent the maximal anaerobic power in football players while further decreasing the ability of subjects to inflate their scores through cheating strategies.

The number of trials necessary to elicit peak power for the Margaria-Kalamen test is unclear on the basis of previous research. Watson (15) suggested using at least 10 trials to calculate power in the Margaria-Kalamen test. Because Clemons and Harrison (3) found no differences between the second and third trials in their new stair climbing test, they used the average of these trials to represent explosive power. Other studies (1,2,12) involving football players have used fewer than the 10 trials suggested by Watson (15) to measure anaerobic power.

Therefore, the purpose of the present study was 3-fold: a) to assess the number of repetitions necessary to elicit maximal anaerobic power before fatigue using a modification of the Margaria-Kalamen test that requires trained football athletes approximately 2 seconds to complete. Modifications to the Margaria-Kalamen test were based on the theoretical premise that the increased vertical displacement would allow for more accurate assessment of peak power in trained athletes by increasing total test time while simultaneously decreasing the influence of cheating strategies. b) To examine test-retest reliability of the modified Margaria-Kalamen test. c) To assess the relationship between peak power, as determined by multiple test protocols, and player position on coaching depth charts.

Methods

Experimental Approach to the Problem

Twenty-five trials of a modified Maragaria-Kalamen stair climbing test were completed to evaluate the number of repetitions necessary to elicit maximal anaerobic power in each of 3 groups of American football players divided according to position played. Modifications to the Margaria-Kalamen test were designed to elicit maximal anaerobic power in trained football athletes by increasing the time required to complete the task while decreasing the possible influence of cheating strategies. Test-retest reliability of the modified stair climbing task was assessed in a subgroup of subjects ($n = 34$). Peak power as measured by the stair climbing task was viewed in combination with data derived from vertical jump and 36.6-m dash tests to determine significant contributors to coaches' depth chart ratings.

Subjects

Fifty-eight male NCAA Division I-A college football players volunteered to participate in this study, which was conducted during the conditioning period before spring practices. All subjects had been on the active team roster during the fall semester and participated in the strength and conditioning program for football athletes for at least the previous 6 months. Subjects were divided into 3 groups by position (group I = linebackers [LB]; group II = offensive and defensive lineman [Line]; and group III = quarterbacks, wide receivers, running backs, defensive backs, and kickers [Skill]). Not all team members participated, which may constitute a limitation of this study because it is unclear whether the subjects were a representative sample of the team as a whole. Stature was determined using a stadiometer (Country Technology, model 67032, Gay Mills, WI, USA), and weight was measured on a Cardinal Detecto Certifier Scale (model 442, Webb City, MO, USA). Descriptive data by group are presented in Table 1.

Before participation in this study, all subjects were screened for injuries or medical conditions by way of a health history questionnaire. Subjects were informed of the experimental risks and signed an informed consent document before participation in the study. All procedures were approved by the University Institutional Review Board, Committee on Human Studies, before data collection.

Procedures

Coaches Ratings. Position coaches were given a list of the subjects divided by position and asked to rank order the athletes on the basis of their depth charts. Each coach rated only the players under their direct supervision and ranked

Table 1. Subject descriptive data (mean ± *SD*).*

Group	*n*	Age (yr)	Height (cm)	Body mass (kg)
LB	10	19.7 ± 1.2	180.8 ± 7.7	101.2 ± 7.8
Line	19	20.0 ± 1.7	189.7 ± 5.8	123.4 ± 17.5
Skill	29	20.6 ± 2.2	189.7 ± 7.6	89.2 ± 2.3
Grand mean	58	20.2 ± 1.8	84.1 ± 7.7	102.5 ± 19.4

*LB = linebackers; Line = offensive and defensive lineman; Skill = quarterbacks, defensive backs, wide receivers, running backs. and kickers.

each player relative only to those players participating in the study.

Development of Football Stair Climb Test Protocol. A pilot study was conducted using former football athletes and athletes from other sports ($n = 6$). Variations of a stair climbing test over a 3.12-m staircase were evaluated for subsequent use with the experimental subject population. The vertical distance was chosen because it was the longest uninterrupted staircase available, and it was reasoned that a greater distance would allow for greater levels of fatigue over multiple trials in less-conditioned athletes. This distance was also desirable because it was found to require greater than 1.5 seconds for completion, even in trained college athletes. Because of the high number of errors associated with the 3-step increment when testing pilot subjects, a 2-step protocol was adopted. In addition, varying numbers of trials (10, 15, 20, and 25 repetitions) of the stair climbing task were evaluated during pilot testing with approximately 30 seconds of rest between trials. On the basis of the pilot subjects' responses, the number of trials was set at 25.

Modified Margaria-Kalamen Test. The following modifications were made to the Margaria-Kalamen test. Before beginning the test, the subjects were instructed to stretch and warm-up on their own. The failure to control the warm-up period constitutes a limitation to the present study. All subjects were experienced with climbing stairs as part of their off-season conditioning program. Subjects were instructed to run up an ordinary flight of stairs, 2 steps at a time, as fast as they could, with a 6-m run-up. The flight of stairs was located outdoors, and the timed distance consisted of 20 steps (3.12 m vertical distance). The starting position and the even-numbered stairs were marked with colored chalk to ensure consistency between subjects in completing the stair climbing task. An electronic timing system (Lafayette Instruments, Lafayette, IN, USA, model 54035A) was used to time the trials. Switch pads were used to start and stop the timing device with the start pad secured to the 2nd step and a stop pad to the 20th step with double-sided outdoor carpet tape. The time for each trial was recorded to the nearest thousandth of a second and subsequently used to calculate power. Power was calculated as Power = body mass (kg) $\cdot$ 9.81 m$\cdot$s^{-2} $\cdot$ vertical distance$\cdot$time^{-1} (3). Subjects were tested in random groups of 3. The subjects participated in the test in a rotational order, separated by 15 seconds, which allotted approximately 30 seconds of rest between trials. A total of 25 trials was completed for each subject. To evaluate reliability, 34 subjects repeated this test within a week of the first testing session, with a minimum of 48 hours between tests. In addition, subject errors were tracked and recorded as touched with hand or missed timing pad, tripped, or omitted trials because of fatigue. Trials in which errors occurred were eliminated from the analysis.

Vertical Jump. Vertical jump data were collected using the Vertec Vertical Jump Tester (Sports Imports, Inc, Columbus, OH, USA). Subjects were given 5 trials to determine maximal vertical displacement. The corrected Harman (4) formula was used to convert vertical jump values to power in watts.

36.6-m Dash Test. The 36.6-m (40-yd) dash data were collected outdoors on a calm day on a MONDO rubberized track surface using a SPEEDTRAP II Wireless Timing System (Brower Timing Systems, Salt Lake City, UT, USA), which recorded time to the nearest 0.01 second. The start of each trial was self initiated with the release of a start pad initiating the electronic timer and ended when subjects ran through the infrared beam. The fastest time of 5 attempts was recorded.

Statistical Analyses

Descriptive statistics and correlations were generated using SPSS v12.0 (Chicago, IL, USA). Intraclass correlation coefficients (ICC) along with *SEM*, Standard Error of Prediction (SEP), and 95% confidence limits were used to establish reliability (16). Differences between position groups were evaluated using a one-way analysis of variance (ANOVA). Where differences were found, the Duncan multiple range test was used to determine which group values were significantly different. A repeated-measures ANOVA was performed to test for differences in peak power across time for the 25 trials. A one-way ANOVA for repeated measures was used to assess differences in peak power calculated from the first 10 trials and peak power from all 25 trials. The general linear model was used to account for missing data points caused by subject errors during trial completion. Changes in velocity over the 25 trials were examined by way of second order polynomial trendlines. A regression analysis (maximum R^2 improvement procedure) was performed to predict depth chart ratings by the position coaches using results from the football stair climb test (FST), peak vertical jump, and best 36.6-m dash time as the independent variables. The alpha level was set at $p < 0.05$.

Results

The mean peak times (s) when performing the FST were significantly different between groups ($p = 0.0008$); the post hoc test revealed that the LB and Skill groups were significantly faster than the Line. The mean peak velocities (m$\cdot$sec^{-1}) were also significantly different between groups ($p = 0.0034$); the post hoc test revealed that the LB and Skill groups had significantly greater velocities than the Line. The peak power (Watts) was also significantly different between groups ($p < 0.0001$); the post hoc test revealed that the LB and Line were significantly more powerful than the Skill group for peak power. The peak power data were also analyzed by group and over trials by creating a time variable. The repeated-measures ANOVA revealed significant differences over time ($p < 0.001$), indicating that the subjects' responses changed over the 25 trials. Results of the modified stair climb test are presented in Table 2.

Polynomial trend lines (second order) indicated that the velocity continued to increase over the 25 trials for the Skill

Table 2. Fastest time, peak velocity, and peak power from football stair climb test by group (mean ± *SD*).

Group	*n*	Fastest time (s)	Peak velocity ($m \cdot s^{-1}$)	Peak power (W)	
				From 25 trials	From first 10 trials
LB	10	1.885 ± 0.243*	1.69 ± 0.29*	1674.5 ± 300.8†	1508.5 ± 168.4
Line	19	2.221 ± 0.284	1.43 ± 0.20	1712.6 ± 251.5†	1613.5 ± 228.4‡
Skill	29	1.991 ± 0.204*	1.58 ± 0.17*	1388.6 ± 210.4	1312.7 ± 193.1‡
Grand mean	58	2.048 ± 0.267	1.55 ± 0.22	1544.0 ± 284.2	1445.0 ± 241.7‡

*Value significantly different than LINE group ($p < 0.05$).
†Value significantly different than SKILL group ($p < 0.05$).
‡Value significantly different from peak power calculated from 25 trials ($p < 0.001$).
LB = linebackers; Line = offensive and defensive lineman; Skill = quarterbacks, defensive backs, wide receivers, running backs and kickers.

and LB groups but plateaued at approximately trial number 17 and subsequently decreased for the Line group. Goodness of fit for linear and second order polynomial trend lines were compared using R^2. The second order polynomial produced a higher R^2 than the linear models for velocity (second order $R^2 = 0.90$, 0.51, and 0.52 for the Skill, Line, and LB groups, respectively). Figure 1 presents the average velocity by group over the 25 trials.

Test–retest results for peak power was analyzed using $ICC_{(2,1)}$ to determine reliability of the test ($n = 34$) (16). The ICC was 0.73 for peak power. There is no established ICC value for evaluating reliability (16). However, when viewed with the *SEM*, calculated in this case to be 105.37 W (6.8% of the mean peak power), an ICC coefficient of 0.70 and above appears an acceptable indication of reliability. Furthermore, when applying the SEP (calculated in the present study as 139.82 W) as described by Weir (16), the 95% confidence interval was ±274.1 W. Therefore, the test was judged to be a reasonably reliable measure of anaerobic power for football athletes.

Peak power data were correlated with the subjects' vertical jump ($n = 51$) and the 36.6-m dash ($n = 39$). The mean values for the 2 tests were vertical jump = 6039.1 ± 627.1 W and 36.6-m dash = 5.2 ± 0.37 s. Group values are reported in

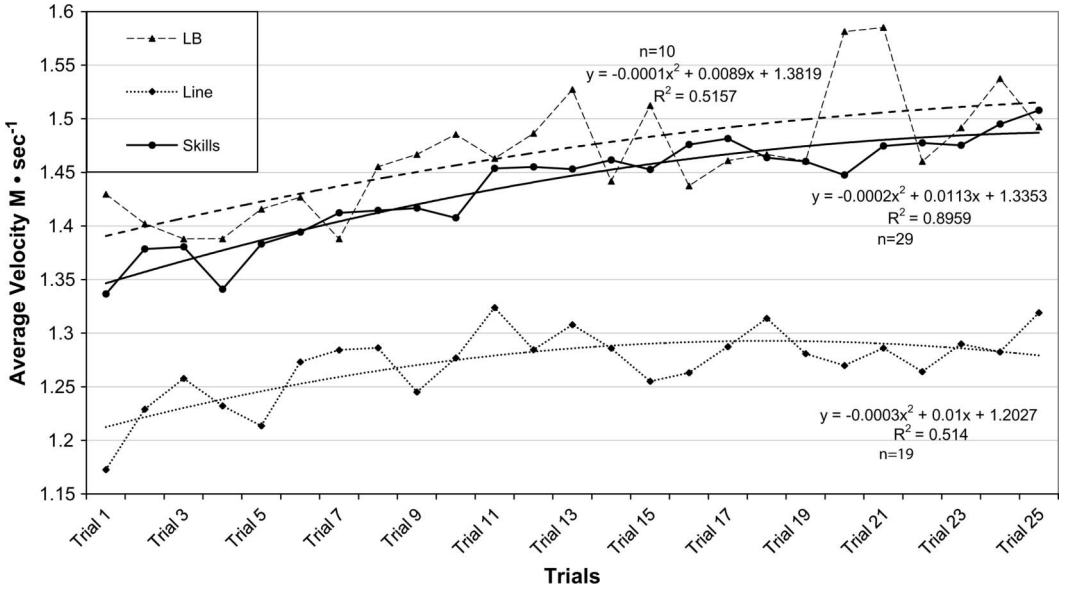

Figure 1. Average velocity by group over 25 trials with second order polynomial trend lines, including equations for trend lines and R^2 values where y = velocity and x = trial number.

Table 3. Vertical jump and 36.6-m dash data by group (mean ± *SD*).

Group	Vertical jump (cm), $n = 51$	Vertical jump (w), $n = 51$	36.6-m (s), $n = 39$
LB	66.2 ± 4.9	6058.4 ± 475.3	5.10 ± 0.15
Line	57.6 ± 12.1	6382.4 ± 773.8	5.48 ± 0.34
Skill	68.7 ± 7.9	5788.0 ± 443.7	4.95 ± 0.25
Grand means	63.7 ± 10.4	6039.1 ± 627.1	5.20 ± 0.37

LB = linebackers; Line = offensive and defensive lineman; Skill = quarterbacks, defensive backs, wide receivers, running backs, and kickers.

Table 4. Errors types occurring by group over 25 trials (1,450 total trials).

Group	*n*	Trips	MP/TP	Rest	Total
LB	10	2	10	3	15
Line	19	8	12	24	44
Skill	29	37	42	0	79
Grand total	58	47	64	27	138

LB = linebackers; Line = offensive and defensive lineman; Skill = quarterbacks, defensive backs, wide receivers, running backs. and kickers; Trips = tripped while performing trial; MP/TP = missed pad or touched pad with hand; Rest = required rest and was not able to complete trail.

Table 3. The correlations with the FST were $r = 0.41$ ($p = 0.0027$) and $r = 0.45$ ($p = 0.0041$) for vertical jump and 36.6-m dash, respectively. Thus, the correlations between the FST and other tests of power were similar and statistically significant. However, the strength of the correlations was only moderate.

Results of the regression analysis to determine which tests could be used to accurately predict depth chart rankings resulted in a relatively low R^2 of 0.42. The most parsimonious model included the 36.6-m dash and the peak power variable from the FST. The best predictor was 36.6-m dash time, which accounted for 35% of the variance. The addition of the FST peak power variable accounted for an additional 7% of the variance of coaches' ratings.

Subject errors were tracked and recorded as touched with hand or missed timing pad, tripped, and omitted trials because of fatigue. Data for errors are presented in Table 4.

Discussion

The most important findings of this study were a) 25 trials of a modification of the Margaria-Kalamen test (the FST), designed to be of sufficient duration to capture peak velocities of collegiate football players, were not adequate to elicit a plateau in peak power in the LB and Skill groups, as indicated by continued increases in velocity in these groups, 2) the FST was found to possess acceptable test-retest reliability, and 3) the peak power from the FST was a significant predictor of player position on coaches' depth charts.

The initial stair climb study by Margaria et al. (6) determined that speed would increase to reach a maximum constant value in approximately 1.5 to 2.0 seconds, then remain constant for 4 to 5 seconds after the initial acceleration period. The energy released during this time period (1.5–5 s after the initial acceleration) was found to be an expression of maximal anaerobic power. Therefore, although widely used to evaluate power, it appears the Margaria-Kalamen test may not be the optimal means for assessing maximal power because values are derived from a short distance of 1.05 m, which may only require approximately 0.4 to 0.5 seconds for a trained athlete to complete (2,12). In addition, Watson (15) suggested that it is possible to "cheat" the timing device in such a short test and inflate scores by up to 10%. This was confirmed by Mayhew et al. (7), who reported that when their subjects were trained to cheat the test, the difference was approximately 0.1 seconds, resulting in an approximate 24% increase in power. Because of the increased time it takes to complete the FST, a decrease of 0.1 seconds would only result in approximately a 5% increase in calculated power. Therefore, the length of the FST would make such cheating practices less influential on the results of the test.

Based on previous studies examining stair climbing in football athletes, the number of repetitions necessary to elicit maximal anaerobic power is unclear. For example, Beckenholdt and Mayhew (2) used 3 trials, Seiler et al. (12) used 5 trials, and Arnold et al. (1) used 8 trials, all less than the minimum 10 trials recommended by Watson et al. (15) when calculating power in the Margaria-Kalamen test. Stuart et al. (13) proposed a fitness test for football athletes consisting of 10 repetitions of timed 40-yard dashes (36.6 m). Each repetition was separated by 25 seconds of rest, to simulate the conditions of a normal football game. It was noted that running speed for all athletes steadily declined with each additional 40-yard dash. They concluded that better conditioned athletes would be able to maintain faster 40-yard dash times over the 10 trials when compared with less-conditioned athletes. In the present study, it was reasoned that the athletes would fatigue and times would plateau or get slower over the 25 trials.

An important finding of the present study was that 25 trials were not enough to elicit fatigue; in fact, the times continued to improve for the LB and Skill groups, indicating a need to increase the amount of repetitions to evaluate true peak power. When peak power was calculated from the first 10 trials of the FST, based on the suggestions of Watson (15), values were significantly lower than the maximum peak power obtained from all 25 trials (Table 2). This would indicate that studies using less than 10 trials may have failed to elicit peak power during stair climbing.

The continued increases in velocity may be caused by a learning effect. When subjects were trained to perform a stair climbing task, performance was shown to increase with learning, without a subsequent increase in explosive power (11). Although the Margaria-Kalamen test involves some degree of skill, the football players in the present study commonly performed numerous repetitions of stair climbing as part of their off-season conditioning program and as a result were highly trained in stair climbing. A more plausible explanation would be that the increases in velocity were caused by the effect of warm-up. Recent studies have suggested that intense warm-up is needed to elicit maximal running speed or jumping height (9,13,14). Therefore, given an adequately intense warm-up, 25 trials or less may be sufficient to elicit true peak power.

Beckenholdt and Mayhew (2) conducted the Margaria-Kalamen test on 50 collegiate athletes with a mean body mass of 79.5 kg. They reported absolute peak power values of 1535.7 ± 256.7 W. This is in good agreement with the results of the present study, in which the grand mean for peak power was 1544.0 ± 284.2 W. However, data from the FST test may not be directly comparable with Margaria-Kalamen test results that have been reported in the literature because power output for this test was lower when compared with other studies of anaerobic power involving the Margaria-Kalamen test using football athletes. Arnold et al. (1) reported a mean peak power of 2100.3 W, whereas Seiler et al. (12) reported a mean peak power of 2256 W. Differences between peak power results in these studies and the present study may be caused by the increased vertical displacement of test used in the present study. However, it has previously been shown that velocities remain constant over this time period (6). Therefore, differences in power in the present study were most likely a result of the fact that 2-step increments were used instead of the 3-step increments commonly used in the Margaria-Kalamen test. This required the subjects to take 4.5 steps to cover the same distance covered in 3 steps in the traditional test. Examination of the error pattern (Table 4) suggests that the 2-step increment may not have been appropriate for the Skill position players, many of whom commented that a 3-step test would have been better for them. However, the LB and Line groups were comfortable with the 2-step increment. A third explanation would be that the subject populations where more powerful in the other 2 studies. This is supported by the fact that the subjects in the Seiler et al. (12) study had greater vertical displacement in the vertical jump test than subjects in the present study (70.9 ± 8.6 cm vs. 66.2 ± 10.4 cm, respectively). In addition, Arnold et al. (1) reported 36.6-m dash times that were faster than those for subjects in the present study (4.9 ± .2 s vs. 5.2 ± 0.4 s, respectively).

Seiler et al. (12) also found strong "inter" correlations among the various power tests administered: Wingate anaerobic power test, Margaria-Kalamen test (with and without a run-up), and 5-, and 35-yard dash, indicating that all tests share a common component, which was assumed to be lower-body anaerobic power. They reported that the common variance between tests showed great range, revealing no one test could be used to quantify anaerobic power. The present study revealed only moderate correlations between the FST and the other administered power tests, vertical jump and 36.6-m dash.

The regression analysis revealed that the 36.6-m dash accounted for the greatest amount of variance (35%) in the coaches' rankings. The addition of the FST peak power variable significantly accounted for an additional 7% of the variance of coaches ratings. Seiler et al. (12) state that because of the low common variance between 5-yard dash and 35-yard dash times, the 40-yard (36.6-m) dash is a poor indicator of initial acceleration and therefore not specific to the demands of most positions in football. Sawyer et al. (10) reported that the vertical jump was a strong predictor of football playing ability when compared with coaches' rankings. However, vertical jump was not included in the most parsimonious model predicting depth chart rank in the present study. Thus, in the present study, the FST provided a better indication of an athlete's power than vertical jump.

Practical Applications

To achieve maximal power during stair climbing tests, the practitioner should be aware that a greater number of trials or a more intense warm-up may be needed than has been previously reported. The FST was found to be a reasonably reliable test for measuring peak anaerobic power in collegiate football players. The FTS should provide a more accurate measure of peak power than the traditional Margaria-Kalamen stair climb test for American football athletes because of increased vertical displacement and longer duration, resulting in a decreased influence of cheating strategies. In addition, the FTS, which is a measure of power, and the results of a combined 36.6-m dash may give the practitioner insight to help improve a player's position on depth chart ratings.

References

1. Arnold, JA, Brown, B, Micheli, RP, and Coker, TP. Anatomical and physiologic characteristics to predict football ability. Report of study methods and correlations, University of Arkansas, 1976. *Am J Sports Med* 8: 119–22, 1980.
2. Beckenholdt, SE and Mayhew, JL. Specificity among anaerobic power tests in male athletes. *J Sports Med Phys Fitness* 23: 326–332, 1983.

3. Clemons, J and Harrison, M. Validity and reliability of a new stair sprinting test of explosive power. *J Strength Cond Res* 22: 1578–1583, 2008.

4. Harman, EA, Posenstein, MT, Frykman, PN, Rosenstein, RM, and Kraemer, WJ. Estimation of human power output from vertical jump. *J Appl Sport Sci Res* 5: 116–120, 1991.

5. Kalamen, JL. *Measurement of Maximum Muscular Power in Man.* Columbus, OH: Ohio State University, 1968.

6. Margaria, R, Aghemo, P, and Rovelli, E. Measurement of muscular power (anaerobic) in man. *J Appl Physiol* 21: 1662–1664, 1966.

7. Mayhew, JL, Schwegler, TM, and Piper, FC. Relationship of acceleration momentum to anaerobic power measurements. *J Sports Med Phys Fitness* 26: 209–213, 1986.

8. Powers, SK and Howley, ET. *Exercise Physiology* (3rd ed). Theory and Application to Fitness and Performance. New York, NY: Power and Howley, McGraw Hill, 2003.

9. Saez Saez de Villarreal, E, Gonzalez-Badillo, JJ, and Izquierdo, M. Optimal warm-up stimuli of muscle activation to enhance short and long-term acute jumping performance. *Eur J Appl Physiol* 100: 393–401, 2007.

10. Sawyer, DT, Ostarello, JZ, Suess, EA, and Dempsey, M. Relationship between football playing ability and selected performance measures. *J Strength Cond Res* 16: 611-616, 2002.

11. Schwegler, TM, Mayhew, JL, and Piper, FC. Effects of acceleration momentum on anaerobic power measurements. *Carnegie Res Papers* 1: 23–26, 1985.

12. Seiler, S, Taylor, M, Diana, R, Layes, J, Newton, P, and Brown, B. Assessing anaerobic power in collegiate football players. *J Strength Cond Res* 4: 9–15, 1990.

13. Stewart, M , Adams, R, Alonso, A, Van Koesveld, B, and Campbell, S. Warm-up or stretch as preparation for sprint performance? *J Sci Med Sport* 10: 403–410, 2007.

14. Thompsen, AG, Kackley, T, Palumbo, MA, and Faigenbaum, AD. Acute effects of different warm-up protocols with and without a weighted vest on jumping performance in athletic women. *J Strength Cond Res* 21: 52–56, 2007.

15. Watson, A. *Physical Fitness and Athletic Performance: A Guide For Students, Athletes and Coaches.* London: Longman, 1983. pp. 147–148.

16. Weir, JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res* 19: 231–240, 2005.