



*Original Research*

---

## **Exploring Predictive Ability of Fitness Test Data Relative to Fire Academy Graduation in Trainees: Practical Applications for Physical Training**

ROBERT G. LOCKIE<sup>#1</sup>, ROBIN M. ORR<sup>#2</sup>, FERNANDO MONTES<sup>#3</sup>, TOMAS J. RUVALCABA<sup>†1</sup>, and J. JAY DAWES<sup>†4,5</sup>

<sup>1</sup>Center for Sport Performance, Department of Kinesiology, California State University, Fullerton, Fullerton, CA, USA; <sup>2</sup>Tactical Research Unit, Bond University, Robina, Qld, AUSTRALIA; <sup>3</sup>Los Angeles County Fire Department, Los Angeles, CA, USA. <sup>4</sup>School of Kinesiology, Applied Health and Recreation, Oklahoma State University, Stillwater, OK, USA; <sup>5</sup>Tactical Fitness and Nutrition Lab, Oklahoma State University, Stillwater, OK, USA

†Denotes graduate student author, #Denotes professional author

---

### ABSTRACT

*International Journal of Exercise Science* 15(4): 1274-1294, 2022. This study investigated the predictive abilities of fitness tests relative to academy graduation in firefighter trainees. Archival fitness test data from 305 trainees were analyzed, including: Illinois agility test (IAT); push-ups; pull-ups; leg tucks; multistage fitness test; 4.54-kg backwards overhead medicine ball throw (BOMB T); 10-repetition maximum deadlift; and a 91.44-m farmer's carry with 18-kg kettlebells. Within the department, trainees were allocated points for each test. Trainees were split into graduated (245 males, 16 females) or released (29 males, 15 females) groups. Independent samples t-tests and effect sizes calculated between-group fitness test differences (raw and scaled points). To provide a binary definition for the sensitivity/specificity analysis, trainees were defined as those scoring 60+ points for a test, and those scoring 0 points. For each test, the binary result (graduated/released) was plotted against the trainee's test performance (60+ Points/0 Points). Receiver operating curves were plotted for each fitness test, and the area under the curve (AUC) determined accuracy. Trainees who graduated performed more push-ups, pull-ups, and leg tucks than released trainees, and scored more points in all tests ( $p \leq 0.005$ ;  $d = 0.34-1.41$ ). Pull-ups, BOMB T, leg tuck, and the farmer's carry had high sensitivity (>80% true positive rate); the IAT had high specificity (83.3% for the true negative rate). Metronome push-ups, BOMB T points, and total points had fair accuracy for predicting academy graduation (AUC=0.709-0.754). While the data demonstrated that trainees who graduated tended to have better total-body muscular strength, endurance, and power, fitness tests may not be appropriate as a sole predictor for academy graduation.

**KEY WORDS:** Backwards overhead medicine ball throw, first responder, Illinois agility test, leg tuck, power, pull-ups, push-ups, muscular endurance, muscular strength, tactical

### INTRODUCTION

Firefighters must complete numerous physically demanding tasks when on the fireground. Some of these tasks include driving vehicles, operating hose lines, carrying equipment, climbing stairs, forcible entries, raising a ladder, crawling, searching, and dragging people or equipment (29, 33). Several fitness qualities contribute to firefighting job task performance, including maximal strength, anaerobic capacity, and aerobic fitness (29, 33). Accordingly, physical training is an integral part of fire training academies (30). In order to assist with training program design, fitness testing data could be used to identify the strengths and weaknesses of trainees. Although this is common practice in athletics (23), this approach is less common with tactical personnel (19). In addition to program design, identification of the fitness capacities of trainees is important, as these data could be used to identify trainees at risk of academy release (i.e., they do not fulfill the training requirements and are released from academy).

Previous research in law enforcement has shown that fitness can indicate whether recruits successfully graduate from an academy training program (9, 15-17). For example, the number of push-ups completed in 60 seconds (s) significantly ( $p \leq 0.01$ ) predicted academy graduation in 99 cadets (89 males, 10 females) from one law enforcement agency, with 13% explained variance (9). When considering just the male cadets, Dawes et al. (9) found that push-ups and the vertical jump significantly ( $p \leq 0.01$ ) predicted police academy graduation, with 15% explained variance. Lockie et al. (17) found that, in 311 law enforcement recruits (260 males, 51 females) from a different agency, age, number of shuttles completed in the 20-meter multistage fitness (MSFT), 2-kg medicine ball chest pass distance, and arm ergometer revolutions in 60 s predicted academy graduation ( $r = 0.411$ ;  $r^2 = 0.169$ ; adjusted  $r^2 = 0.158$ ). There is less research investigating relationships between fitness and graduation in firefighter populations. Butler et al. (5) did investigate whether certain tests predicted whether a firefighter trainee would be injured during academy. The authors found that a Functional Movement Screen (FMS) score below 14 (out of 21) differentiated between injured and non-injured trainees. Additionally, Butler et al. (5) found that deep squat and trunk stability push-up from the FMS ( $r = 0.330$ ), and the sit-and-reach ( $r = 0.218$ ), were significant predictors of injury. However, Butler et al. (5) did not find that the 2.4-km (aerobic fitness) run, or 2-minute push-up test (muscular endurance), predicted injury.

One method that could be used to investigate fitness test performance and graduation rates is via sensitivity and specificity analysis. In medical terms, sensitivity is the ability of a test to correctly identify individuals with a condition, while specificity is the ability of the test to correctly identify individuals without a condition (14). This statistical approach has been used in tactical research (34). Stevenson et al. (34) investigated firefighting physical ability when analyzing the relationship between certain surrogate fitness tests (seated shoulder press, rope pulldown, and 28-kg rope pulldown for repetitions) and job tasks (ladder lift, ladder lower, and ladder extension, respectively). Stevenson et al. (34) found that the best sensitivity and specificity achieved were 100% and 100% for a 35-kg seated shoulder press, 79% and 92% for a 60-kg rope pulldown, and 83% and 93% for 23 repetitions of the 28-kg rope pulldown, respectively. However, the use of sensitivity and specificity analysis relative to firefighter trainee fitness and academy graduation has yet to be analyzed. It is acknowledged that factors other

than fitness can contribute to a first responder trainee's release from academy, such as poor performance in physical skills tests (30, 33) or academic examinations (15). Nonetheless, given the relationship between fitness and firefighting job task performance (29, 33), it would be beneficial to investigate the sensitivity and specificity of certain fitness tests.

Receiver-operating characteristic (ROC) curves could also be used to provide information regarding the accuracy of certain fitness test in identifying specific characteristics about a trainee (i.e., likelihood of graduating from academy) (40). The area under the curve (AUC) can indicate the overall accuracy of the test (14), with a value approaching 1.0 indicating high sensitivity and specificity (14, 34). There have been some examples of the use of ROC curves in tactical research (16, 26). Orr et al. (26) used ROC curves to analyze injury, physical fitness failure, and attrition in army recruit training. Failure in the final physical fitness test battery by army trainees was predicted (AUC = 0.70) by age, height, body mass, push-ups, sit-ups, MSFT, and the type of training course completed (28 or 80 days). Relative to the current study, Lockie et al. (16) investigated whether fitness test performance predicted academy graduation in law enforcement recruits. The data indicated that performance in a 75-yard pursuit run (AUC = 0.708) and the MSFT (AUC = 0.727) had fair accuracy in predicting law enforcement academy graduation. It would be valuable for fire department training staff to identify whether a fitness test is specific, sensitive, and accurate, as it could be used to identify trainees at risk of academy release and motivate trainees to develop their weaknesses because they know they are at risk. This is important for fire departments, as a loss of trainees during academy can add unnecessary costs to a department (25). This information could be used to (a) educate incoming trainees as to specific fitness qualities they should develop for a certain training academy, and (b) motivate current trainees if they are deficient in a certain fitness test or quality.

Therefore, the purpose of this research was to analyze the predictive capabilities of fitness tests relative to academy graduation in firefighter trainees. Given the need for a range of fitness capacities to complete firefighting tasks (29, 33), the testing battery administered by fire department training staff featured a range of fitness assessments, and both raw and scaled point scores (relative to an internal scoring chart) were provided to the researchers. This study involved the analysis of retrospective analysis of existing data, so the researchers did not have input into the fitness tests selected. It was hypothesized that trainees who graduated would display superior fitness relative to trainees who were released due to injury or performance across the assessments utilized in this study. It was further hypothesized that the fitness tests would be specific, sensitive, and accurate.

## **METHODS**

### *Participants*

Deidentified archival data from six academy classes from one U.S. based fire department were released with consent to the researchers for this retrospective study. The sample included data from 305 trainees (274 males, 31 females). Demographic information (age, height, and body mass) were not provided by the department to the researchers. This has also occurred in

previous first responder research (5, 20). Regardless, all trainees were over 18 years of age, and had completed a pre-placement medical evaluation (22). The inclusion criterion was available data for a fitness test. The exclusion criterion was clearly incorrect data through entry error. This was a convenience sample of de-identified data provided by the department, and the researchers had no control of the final sample size. As secondary data was utilized in this study, G\*Power software (v3.1.9.2, Universität Kiel, Germany) was used to confirm post hoc that the sample size of 305 (with groups of 261 and 44 participants) was sufficient for an independent t-test analysis such that data could be interpreted with a small effect level of 0.45 (13), and a power level of 0.87 when significance was set at 0.05 (10). Based on the archival nature of this analysis, the institutional ethics committee approved the use of pre-existing data (HSR-17-18-401). Even though this research involved a retrospective analysis, the study was still conducted in agreement with the ethical standards of the International Journal of Exercise Science (24).

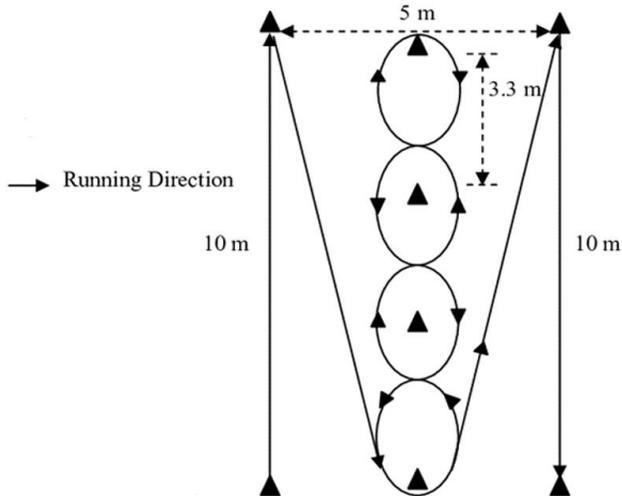
### *Protocol*

The six training cohorts started their academy during the year 2020 in southern California. All fitness tests were completed by trainees as part of their employment requirements and were conducted within one 90-minute physical training session for all academy classes. Trainees were expected to complete all tests unless there was some type of physical condition that prevented their participation (e.g., a pre-existing injury). The researchers were not informed the reasons why a recruit may not have had data for a particular test. All sessions were overseen by the same Certified Strength and Conditioning Specialist. As previously acknowledged, the researchers did not have input into the fitness tests that were used as the study used archival data. Nonetheless, the fitness test battery was standard practice within the fire department and designed to encompass a range of assessments for different fitness qualities. In line with this, the department training staff typically used the data to identify areas in need of improvement within the trainees (23). The tests were completed in the order presented hereafter, which was standard practice for the department. Testing was not been completed in the order relative to recommendations from the National Strength and Conditioning Association (23), which is a limitation. However, this was standard within the fire department and all data chronicled was used for record. As noted by Lockie et al. (20), fire academies have challenging situations relative to space, equipment, time, and staff, so the staff had to make the best of these conditions. Enough time was provided between test attempts to ensure adequate recovery, but exact times provided were dependent on the number of trainees within each class and the available staff. Testing occurred outdoors at the fire department's training facility in the early morning (~6:00am).

Internally, the fire department assigned scaled point scores relative to the performance in each of the fitness tests (60-100 points for each test) in addition to total points (out of 800), and the scoring system is shown in Table 1. The researchers were concerned with trainees who attained points for a test (i.e., they were awarded 60 or more points), and which trainees did not (i.e., they received a score of 0), in addition to the actual points allocated per test. The scoring system used was developed internally by the fire department and based on the system used by the U.S. Army in which the minimum requirement for a test was 60 points, while the maximum points for a test was 100 points (8). The researchers were not privy as to how the scale of points relative

to performance were developed. Nonetheless, this point scoring system was an accepted standard and used to guide the exercise programming within this fire department.

The IAT was used to measure change-of-direction speed (21). Previous research has shown that this test has good reliability (intra-class correlation coefficient [ICC] = 0.80-0.89) (35). The measurements and running direction for the IAT are shown in Figure 1. The IAT involved four markers being placed to indicate an area that was 10 m long and 5 m wide. In the middle of the grid, four markers were placed 3.3 m apart. Trainees began in the prone position behind the start point, outside the first cone. The tester gave a preparatory command of "Ready", before the start command of "Go". The trainee then jumped to their feet and ran the course as quickly as possible. Trainees were told not to cut over or contact the markers and were to follow the prescribed route throughout the trial. If a trainee failed to follow these directions, or they slipped during the trial, the trial was stopped and re-attempted. Time was recorded in s via stopwatch from initial movement until the trainee crossed the finish line. Depending on time constraints, 1-2 trials were completed by trainees, with the fastest trial analyzed in this study.



**Figure 1.** Distances and running direction for the Illinois agility test.

Maximal push-ups can provide a measure of upper-body muscular endurance and as a test have been found to have high trial-to-trial reliability (ICC = 0.95) (31). For this fire department, push-ups were performed in time with a metronome at a cadence of 80 beats per minute. An audio file for the metronome was played during the test, and standard procedures were utilized. On the "Get ready" command, trainees assumed the kneeling push-up position with the arms fully extended. On the "Get Set" command, trainees adopted the standard 'up' position (body taut and straight, hands positioned shoulder-width apart, fingers pointed forwards) (17). On the "Go" command and the initiation of the metronome, the trainee lowered themselves towards the ground by flexing the elbows until the upper arm was parallel to the ground. On the next metronome sound, the trainee immediately extended the elbows and returned to the up (or start) position. On the next metronome sound, the trainee returned to the bottom position, and so forth. The test was terminated when the trainee could no longer complete repetitions in time with the cadence. If the trainee maintained the cadence, but did not meet other technical standards (i.e., they did not fully extend the elbows, the upper arms were not parallel to the ground at the bottom, or there was a sag in the pelvis/trunk), the tester repeated the number of

the last correct repetition and told the trainee to make the required correction. The number of correct repetitions performed was the score used for analysis. The maximum number of repetitions set for the push-up test was 100 (Table 1), so trainees did not have to complete extra repetitions beyond this standard.

The pull-up test provided a measure of upper-body pulling strength (32), and this test has been found to have high reliability (ICC = 0.99) (7). On the "Get ready" command, trainees positioned themselves in a free-hang position on the pull-up bar. The hands were placed shoulder-width apart (or slightly wider) with a pronated grip. The thumbs were wrapped around the bar, and the elbows were extended. On the "Go" command, the trainee pulled their body upward with a vertical alignment until the chin was over the bar to complete one repetition. The trainee then descended back to the start position until their arms were fully extended. Trainees completed as many repetitions as they could with this technique until they could no longer raise their chin above the bar. If the trainee kicked their way up, the pull-up was not counted. The score used for analysis was the number of correct repetitions performed.

The leg tuck measures grip, arm, shoulder, and trunk muscle strength (20), and as a test has been found to exhibit high reliability (ICC = 0.998-0.999) (27). On the "Get ready" command, the trainee moved to a free-hang position, with their hands positioned with an alternated grip on the bar. The trainee's body was extended and faced the length of the bar. On the command "Go", the trainee lifted their legs up so that their elbow flexed to an approximate 90° angle while simultaneously bringing their knees to contact their elbows. This constituted the leg tuck. The knees had to contact the elbows for a successful repetition to be counted. The trainee then returned to the free-hang position and repeated this movement as many times as possible. The trainee's body was to be extended in the free-hang position between each repetition, and they could not rest the legs on the bar or swing past the starting position when they lowered themselves from the tuck. The analyzed score was the number of correct repetitions performed.

Estimated  $\text{VO}_{2\text{max}}$  was derived from the MSFT (28), which was conducted according to established practices (18). The MSFT has demonstrated high reliability (ICC = 0.96) (1). The trainees ran back and forth between two lines spaced 20 m apart, which were indicated by markers. The speed of running for the MSFT was standardized by pre-recorded auditory cues (i.e., beeps) played from an audio file that could be heard by trainees. The test was terminated when the trainee was unable to reach the lines twice in a row in accordance with the auditory cues. This test was scored according to the final stage the trainee was able to achieve. From this, estimated  $\text{VO}_{2\text{max}}$  (measured in milliliters per kg body mass per minute;  $\text{ml} \cdot \text{kg}^{-1} \cdot \text{min}^{-1}$ ) for each trainee was derived from the table documented by Ramsbottom et al. (28). This conversion has been used previously in first responder research (18). Ramsbottom et al. (28) stated that the correlation between  $\text{VO}_{2\text{max}}$  and the shuttle level attained in the MSFT in adults was 0.92.

**Table 1.** The scoring system for the fitness tests (Illinois agility test [IAT], metronome push-ups, pull-ups, 4.54-kg backwards overhead medicine ball throw [BOMBT], leg tuck, estimated maximal aerobic capacity [VO<sub>2max</sub>] from the 20-m multistage fitness test, 10-repetition maximum [10RM] RM deadlift), and the farmer’s carry used within this fire department.

Points	IAT (s)	Push-ups (repetitions)	Pull-ups (repetitions)	BOMBT (m)	Leg Tucks (repetitions)	Estimated VO <sub>2max</sub> (ml · kg <sup>-1</sup> · min <sup>-1</sup> )	10RM Deadlift (kg)	Farmer’s Carry (s)
60	17.5	60	6	4.59	6	45.2	79.38	33.9
61		61		5.28		45.6	81.65	
62		62	7	5.59		45.9	83.91	
63	17.4	63		5.89	7	46.2	86.18	
64		64	8	6.20		46.5	88.45	
65		65		6.48		46.8	90.72	
66	17.3	66	9	6.99	8	47.1	92.99	31.9
67		67		7.49		47.7	95.25	
68		68		7.98		48.0	97.52	
69	17.2	69	10	8.28	9	48.3	99.79	
70		70		8.48		48.6	102.06	
71		71		8.59		48.9	104.33	
72	17.1	72	11	8.79	10	49.2	106.59	29.9
73		73		8.89		49.5	108.86	
74		74		9.09	11	49.9	111.13	
75	17.0	75	12	9.19		50.2	113.40	
76		76		9.40	12	50.5	115.67	
77		77	13	9.50		50.6	117.93	
78	16.9	78		9.68		50.8	120.20	27.9
79		79		9.78	13	51.1	122.47	
80		80	14	9.98		51.4	124.74	
81	16.8	81		10.08		51.7	127.01	
82		82	15	10.29	14	52.0	129.27	
83		83		10.39		52.3	131.54	
84	16.7	84		10.59		52.6	133.81	25.9
85		85	16	10.69	15	52.9	136.08	
86		86		10.90		53.2	138.35	
87	16.6	87		11.00		53.5	140.61	
88		88	17	11.18	16	53.8	142.88	
89		89		11.28		54.0	145.15	
90	16.5	90		11.48		54.3	147.42	23.9
91		91	18	11.58	17	54.6	149.69	
92		92		11.79		54.9	151.95	
93	16.4	93		11.89		55.1	154.22	
94		94	19	12.09	18	55.4	156.49	
95		95		12.29		55.7	158.76	
96	16.3	96		12.50	19	56.0	161.03	21.9
97		97	20	12.78		56.3	163.29	
98		98		12.98	20	56.6	165.56	
99	16.2	99		13.18		56.8	167.83	
100	16.1	100	21	13.49	21	57.1	170.10	19.9

The BOMBT with a 4.54-kg (10 lbs) medicine ball was used to assess combined upper- and lower-body power and coordination and this test has high reliability (ICC = 0.996) (36). The trainee stood with their back to the throwing area. Their feet were positioned shoulder-width apart and their heels were placed on the start line. The medicine ball was held in front of the body, with the arms extended at shoulder height. In one continuous movement, the trainee flexed the hips, knees, and trunk to lower the ball below the waist. The trainee then extended their legs and thrust the hips forwards forcefully, while flexing the shoulders and elevating the ball above shoulder height as they threw it back over their head as far as possible. Following the throw, the trainee's feet could leave the ground. However, their body could not go past the start line. The horizontal distance was measured via a tape measure in feet from the start line to the point where the ball first contacted the ground. The score was converted to meters (m) for this study.

The 10-repetition maximum (10RM) deadlift provided a measure of lower-body strength. The literature has indicated that bilateral strength exercises have high reliability (ICC = 0.92) (3). The deadlift movement was performed as described by Lockie et al. (20). However, trainees self-selected their stance foot position and grip on the bar, although they were to use a conventional grip with their hands positioned on the bar outside of their legs. Trainees performed warm-up sets (up to 10 repetitions) as necessary with different loads (52 kg [115 lbs], 74 kg [175 lbs], 84 kg [185 lbs], 102 kg [225 lbs]). After the warm-up sets, the weight was progressively increased, and trainees completed 10 repetitions, which were counted by a staff member. A successful repetition occurred when the trainee was standing up straight via knee and hip extension, with their shoulders retracted and positioned behind the vertical orientation of the bar. A pause of approximately 2 s between repetitions was allowed at the top of the deadlift. The trainee then lowered the weight with control back to the ground. No rest was allowed between repetitions while the weight was on the ground. The test was stopped if the trainee did not attain the correct upright position, they exceeded the 2-s time limit at the top or bottom of a repetition, they dropped the weight, or they failed to keep the bar ascending during a repetition. Approximately 2-3 minutes were provided between attempts. The load for the last successful 10RM attempt was recorded in lbs, before being converted to kg for this analysis. Trainees achieved maximum points for the deadlift if they had a 10RM of 170.10 kg (375 pounds; Table 1). Accordingly, trainees could choose not to attempt a 10RM above this load if they successfully achieved this load.

Lockie et al. (20) stated that the fire department training staff included a farmers carry with kettlebells in their test battery as firefighters need to execute loaded carries as part of their job. Reliability data were not available for the farmer's carry, but this test was established in the battery used by this fire department (20). Loaded carries assess force production during a functional movement, and qualities such as grip strength are also stressed in this activity (39). The farmers carry was designed to assess a trainee's ability to carry 36 kg (80 lbs) kettlebells 91.44 m (100 yards). To complete the test, trainees held an 18-kg (40-lb) kettlebell in each hand and travelled four times up-and-back over a 22.86-m (25-yard) distance. The trainees were instructed to complete the four shuttles with the kettlebells as quickly as possible by walking, jogging, or running (depending on their capacity). To begin the farmer's carry, the trainee stood

at the start line with the kettlebells positioned on the ground on each side of the trainee. On the command "Go", the trainee squatted down and lifted the two kettlebells and proceeded to traverse the 4 x 22.86-m course as quickly as possible. If a trainee dropped a kettlebell at any point, they could pick the kettlebell up and continue. Time was recorded in s via stopwatch from the initiation of movement until the trainee finished the course. As previously stated, the scoring system used for the farmers carry was not provided to the researchers, so it is not included in Table 1. Nonetheless, the scores assigned to each trainee based on their farmers carry time was provided to the researchers and thus was included in the study analysis.

### Statistical Analysis

Statistical analyses were processed using the Statistics Package for Social Sciences (Version 28; IBM Corporation, New York, USA). Normality of the fitness test data was checked by visual analysis of the Q-Q plots. Descriptive statistics (mean  $\pm$  standard deviation [SD]) were calculated for each variable. The sample was split into those trainees who graduated, and those who were released for any reason (i.e., injury, skills test performance failures) during academy. Sexes were combined within these groups, as all trainees needed to attain the same standards to graduate academy, regardless of sex. Combining the sexes within an analysis approach has been conducted in previous tactical research (4). Any differences between the graduated and released groups for the raw and scaled point scores allocated per test (and total overall points) were investigated by independent samples t-tests, with significance set at  $p < 0.05$ . Levene's test for equality of variances were checked to determine whether equal variances were to be assumed or not. Effect sizes ( $d$ ) were also calculated for the between-group comparisons, where the difference between the means was divided by the pooled SD (6). A  $d$  less than 0.2 was considered a trivial effect; 0.2 to 0.6 a small effect; 0.6 to 1.2 a moderate effect; 1.2 to 2.0 a large effect; 2.0 to 4.0 a very large effect; and 4.0 and above an extremely large effect (13).

The specificity, sensitivity, and accuracy analyses were included in this study due to their potential predictive capabilities relative to fitness testing and firefighter trainee academy graduation (38, 40), and the fact these procedures have been adopted in other tactical research (16, 26, 34). The specificity and sensitivity analysis were adapted from Stevenson et al. (34). For each fitness test, the binary result (graduated/released) was plotted against the trainee's performance in the corresponding fitness test (60+ Points or 0 Points). For fitness each test, sensitivity (true positive rate) and specificity (false positive rate) were calculated relative to whether the trainee attained points or not according to the department's scoring standards. Sensitivity, which was the ability of the fitness test to correctly identify those who graduated academy, was calculated using the following formula:  $Sensitivity = True\ Positives \div (True\ Positives + False\ Negatives)$ . Specificity, which was the ability of the fitness test to correctly identify those who were released from academy, was calculated using the following formula:  $Specificity = True\ Negatives \div (False\ Positives + True\ Negatives)$ . Figure 2, which was adapted from Trevethan (38), displays the basis for deriving sensitivity and specificity specific to this study. Additionally, ROC curves were then plotted for each fitness test (raw and scaled point scores), with sensitivity on the y-axis and  $1 - specificity$  on the x-axis (34). The positive condition was whether a trainee graduated. AUC, which represented the overall accuracy of the test (14), was computed in

addition to the 95% confidence intervals (CI). An AUC value that approached 1.0 indicated a high sensitivity and specificity (14, 34). Zero discrimination equated to an AUC of 0.5 (14). A minimum AUC of 0.70 was expected for a model to have minimally acceptable predictive accuracy (37).

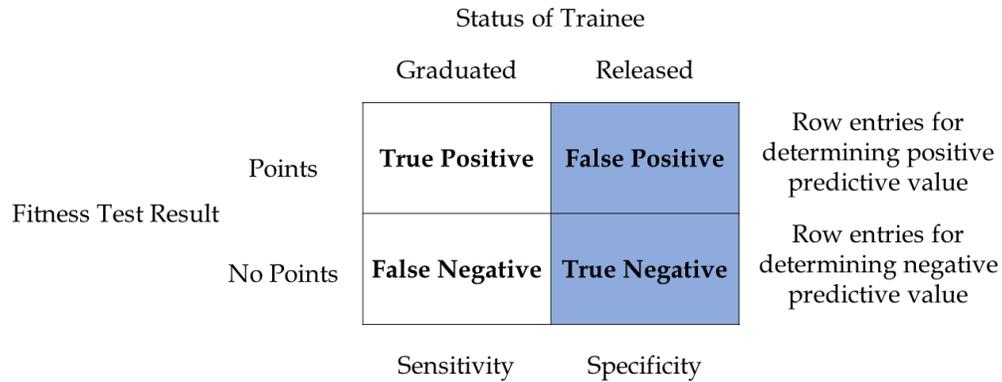


Figure 2. Diagram demonstrating the basis for deriving sensitivity and specificity specific to this study.

## RESULTS

Across the six classes, 261 trainees graduated academy (245 males, 16 females; 86% of the sample), while 44 (29 males, 15 females; 14% of the sample) trainees were released at various time points. Table 2 shows the descriptive data for each group. Raw data for all tests was not included for all trainees; however, all available data were included. The n range is presented in Table 2 for each group. Equal variances were assumed for all variables except push-ups and the leg tuck. The graduated trainees completed significantly more push-up (moderate effect), pull-up (small effect), and leg tuck (moderate effect) repetitions. There were no other significant between-group differences in the fitness tests.

The descriptive data for the points attained for each test is shown in Table 3. All trainees had points assigned for each test. Thus, if they were not able to perform a certain test and did not have a raw score, they were assigned a score of 0 for that test. Equal variances were not assumed for all variables except push-ups and total points. The graduated recruits scored significantly more points in all tests. Moderate effects were shown for the between-group differences in estimated  $VO_{2max}$  and total points, and large effects for the BOMBT and 10RM deadlift.

The sensitivity and specificity data for each test, relative to whether a trainee attained points in each test, is shown in Table 4. The table cells relating to sensitivity are highlighted green, while specificity cells are highlighted in orange. The 10RM deadlift was not included in this analysis as all trainees achieved at least 60 points in this fitness test (i.e., no trainees scored 0). For the IAT, sensitivity was low (76.1% false negative rate), although specificity was high (83.3% for the true negative rate). Several tests – pull-ups, BOMBT, leg tuck, and farmer’s carry – had high sensitivity (>80% true positive rate) but low specificity (<30%; high false positive rate).

Metronome push-ups and estimated  $VO_{2max}$  were neither sensitive (54.8-56.7%) or specific (47.7-78.6%).

**Table 2.** Descriptive data (mean  $\pm$  SD) and between-class comparisons for Graduated and Released trainees from one fire department for the Illinois agility test (IAT), metronome push-ups, pull-ups, 4.54-kg backwards overhead medicine ball toss (BOMBT), leg tuck, estimated maximal aerobic capacity from the 20-m multistage fitness test (estimated  $VO_{2max}$ ), 10-repetition maximum (10RM) deadlift, and the farmer's carry.

	Graduated (n = 255-261)	Released (n = 42-44)	<i>p</i>	<i>d</i>
IAT (s)	18.43 $\pm$ 1.46	18.51 $\pm$ 1.18	0.754	0.06
Push-ups (repetitions)	64.68 $\pm$ 22.67	44.50 $\pm$ 17.44*	<0.001	1.00
Pull-ups (repetitions)	12.12 $\pm$ 6.39	9.20 $\pm$ 5.88*	0.005	0.48
BOMBT (m)	9.52 $\pm$ 1.66	9.54 $\pm$ 1.98	0.949	0.01
Leg Tuck (no.)	12.46 $\pm$ 5.88	8.88 $\pm$ 4.27*	<0.001	0.70
Estimated $VO_{2max}$ (ml $kg^{-1}$ $min^{-1}$ )	46.20 $\pm$ 5.88	44.78 $\pm$ 5.89	0.139	0.24
10RM Deadlift (kg)	143.72 $\pm$ 15.20	142.41 $\pm$ 15.09	0.599	0.09
Farmer's Carry (s)	28.77 $\pm$ 4.13	29.69 $\pm$ 4.21	0.183	0.22

\* Significantly ( $p < 0.05$ ) different from the graduated group.

**Table 3.** Descriptive data (mean  $\pm$  SD) and between-class comparisons for Graduated and Released trainees from one fire department for the points assigned for the Illinois agility test (IAT), metronome push-ups, pull-ups, 4.54-kg backwards overhead medicine ball toss (BOMBT), leg tuck, estimated maximal aerobic capacity from the 20-m multistage fitness test (estimated  $VO_{2max}$ ), 10-repetition maximum (10RM) deadlift, and the farmer's carry. Individual tests had a maximum score of 100. Total points had a maximum score of 800.

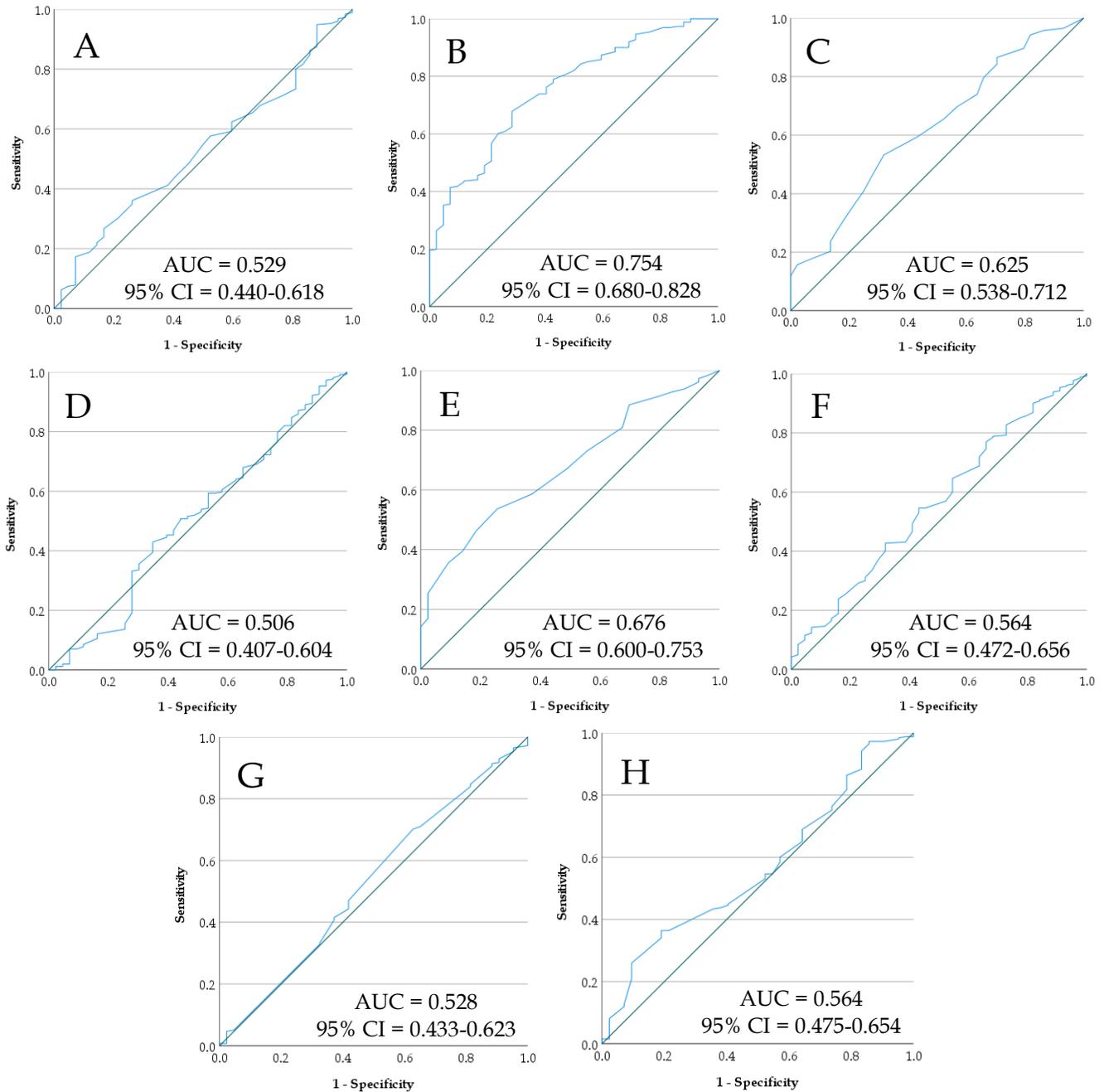
	Graduated (n = 261)	Released (n = 44)	<i>p</i>	<i>d</i>
IAT	19.70 $\pm$ 34.42	1.57 $\pm$ 10.40*	<0.001	0.56
Push-ups	44.08 $\pm$ 41.07	30.09 $\pm$ 39.57*	0.018	0.34
Pull-ups	66.89 $\pm$ 30.99	49.93 $\pm$ 37.37*	0.003	0.53
BOMBT	78.63 $\pm$ 8.84	59.86 $\pm$ 28.08*	<0.001	1.41
Leg Tuck	71.91 $\pm$ 25.15	59.07 $\pm$ 36.42*	0.014	0.48
Estimated $VO_{2max}$	45.77 $\pm$ 38.05	21.95 $\pm$ 32.72*	<0.001	0.64
10RM Deadlift	96.97 $\pm$ 5.75	77.75 $\pm$ 35.09*	<0.001	1.35
Farmer's Carry	64.71 $\pm$ 27.65	50.59 $\pm$ 33.95*	0.006	0.49
Total Points	488.59 $\pm$ 132.71	349.45 $\pm$ 146.38*	<0.001	1.03

\* Significantly ( $p < 0.05$ ) different from the graduated group.

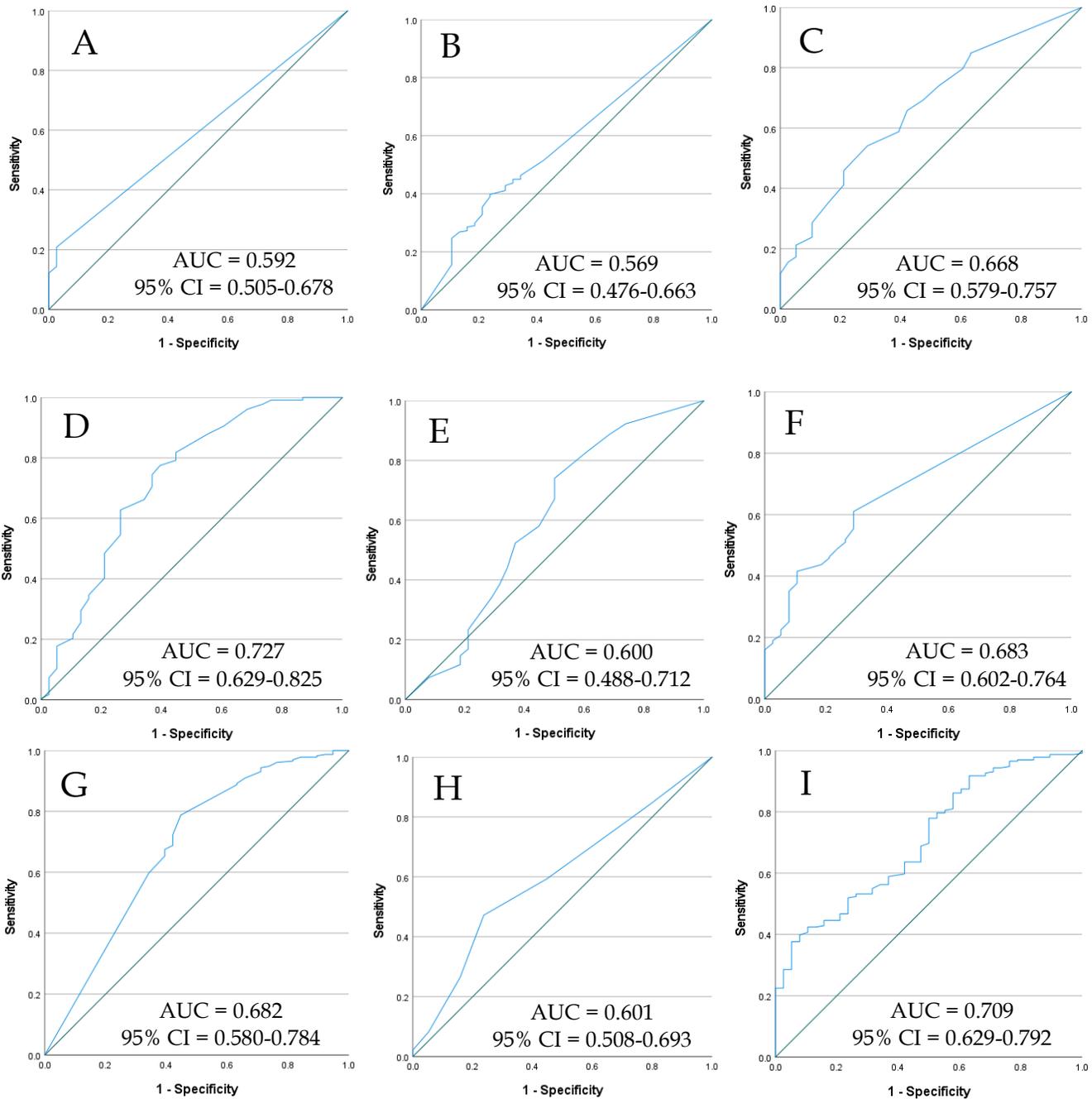
ROC curves are shown in Figure 3 for the raw fitness test data, and in Figure 4 for the assigned scores for each test. For each graph, the diagonal reference line was included. With regards to the raw fitness test data, the metronome push-up test had fair accuracy for predicting academy graduation for a trainee (AUC = 0.754). All other fitness tests had relatively poor accuracy for predicting academy graduation in firefighter trainees (AUC  $\leq$  0.676). Regarding the scaled points score allocated for each test, the BOMBT (AUC = 0.727) and overall score (AUC = 0.709) had fair accuracy for predicting academy graduation. All other point scores had relatively poor accuracy (AUC  $\leq$  0.683).

**Table 4.** Sensitivity and specificity data for the Illinois agility test, metronome push-ups, pull-ups, 4.54-kg backwards overhead medicine ball toss, leg tuck, estimated maximal aerobic capacity from the 20-m multistage fitness test, and the farmer’s carry.

		Graduated	Released
<i>Illinois Agility Test</i>			
60+ Points	Count	61	7
	Percentage within 60+ Points Group	23.9%	16.7%
0 Points	Count	194	35
	Percentage within 0 Points Group	76.1%	83.3%
<i>Metronome Push-ups</i>			
60+ Points	Count	143	9
	Percentage within Graduated Group	54.8%	21.4%
No Points	Count	118	33
	Percentage within Released Group	45.2%	78.6%
<i>Pull-ups</i>			
60+ Points	Count	220	31
	Percentage within Graduated Group	84.3%	70.5%
No Points	Count	41	13
	Percentage within Released Group	15.7%	29.5%
<i>Backwards Overhead Medicine Ball Throw</i>			
60+ Points	Count	220	31
	Percentage within Graduated Group	84.3%	70.5%
No Points	Count	41	13
	Percentage within Released Group	15.7%	29.5%
<i>Leg Tuck</i>			
60+ Points	Count	238	34
	Percentage within Graduated Group	91.2%	79.1%
No Points	Count	23	9
	Percentage within Released Group	8.8%	20.9%
<i>Estimated Maximal Aerobic Capacity</i>			
60+ Points	Count	148	23
	Percentage within Graduated Group	56.9%	52.3%
No Points	Count	112	21
	Percentage within Released Group	43.1%	47.7%
<i>Farmer’s Carry</i>			
60+ Points	Count	221	33
	Percentage within Graduated Group	85.7%	78.6%
No Points	Count	37	9
	Percentage within Released Group	14.3%	21.4%



**Figure 3.** Receiver operating curves and area under the curve (AUC) values (with 95% confidence intervals [CI]) for the Illinois agility test (A), metronome push-ups (B), pull-ups (C), 4.54-kg backwards overhead medicine ball toss (D), leg tuck (E), estimated maximal aerobic capacity from the 20-m multistage fitness test (F), 10-repetition maximum deadlift (G), and the farmer's carry (H).



**Figure 4.** Receiver operating curves and area under the curve (AUC) values (with 95% confidence intervals [CI]) for the scaled points scored in the Illinois agility test (A), metronome push-ups (B), pull-ups (C), 4.54-kg backwards overhead medicine ball toss (D), leg tuck (E), estimated maximal aerobic capacity from the 20-m multistage fitness test (F), 10-repetition maximum deadlift (G), farmer's carry (H), and total points (I).

## DISCUSSION

The purpose of this study was to analyze the predictive ability of fitness tests in relation to academy graduation in firefighter trainees. It should be clearly acknowledged that this fire department does not use fitness test data for diagnostic purposes per se; rather, the fitness tests

data informs programming within physical training (23). Nonetheless, this research explored the ability of the analyzed fitness tests to predict academy graduation. This is an important concept, and numerous studies have indicated that law enforcement recruits who are unsuccessful with academy training tend to display poorer physical fitness (9, 15-17). These studies were supported to an extent by the current research on firefighter trainees. Those trainees who graduated performed 45% more push-ups, 32% more pull-ups, and 40% more leg tucks than trainees who were released. These tests provided measures of upper-body and abdominal muscular endurance and strength (19, 20, 32). Notably, differences in fitness characteristics were also demonstrated when considering the scaled point scores.

In the military, fitness test performance is often converted into a numerical score, which when combined with scores from other fitness tests, provides an overall rating for an individual (8). The fire department in this study used a similar process in their fitness testing. Accordingly, the lowest score that could be attained on a test was 0, while points ranging from 60-100 were allocated depending on performance on a test. The trainees who graduated scored significantly more points in all tests compared to the trainees who were released. The largest effects were seen for the BOMBT (total-body power), 10RM deadlift (maximal strength), and total points (the combination of all fitness tests). This is reflective of the need for firefighters to have an appropriate level of fitness in multiple fitness characteristics to successfully perform their job tasks (29, 33). In addition to the between-group analysis, the scoring system provided a natural cut-off for sensitivity/specificity analyses (i.e., the trainee either attained points or did not for a fitness test). However, there has not been analysis of fitness test data in this manner, and whether it could provide more specific information for training prescription.

As stated, the specificity and sensitivity were included in this study due to their potential predictive capabilities relative to fitness testing and firefighter trainee academy graduation (38). To provide some support to this approach, Stevenson et al. (34) was able to demonstrate that found that 35-kg seated shoulder press, 60-kg rope pulldown, and 23 repetitions of the 28-kg rope pulldown were specific and sensitive relative to predicting performance in the ladder lift, ladder lower, and ladder extension, respectively (specificity = 79-100%, sensitivity = 92-100%). However, the use of sensitivity and specificity analysis relative to firefighter trainee fitness and academy graduation has yet to be analyzed. In the context of this study, sensitivity was the ability of a fitness test to identify whether a recruit graduated from academy (i.e., they successfully completed academy training). Specificity was the ability to identify whether a trainee did not graduate from academy (i.e., they did not successfully complete academy training). The impact of fitness test data relative to academy graduation is different from the standards required in medical research when defining false positive or false negative rates. However, this type of analysis could be used to identify at-risk trainees and motivate these trainees to target specific fitness qualities. This would be especially true if departments release trainees due to fitness test performance, which can happen in law enforcement agencies (15). As stated, the training staff from the fire department in this study assigned points for performance in each fitness test as a means of providing feedback for the trainees. These scores were used within this initial exploratory analysis for investigating sensitivity and specificity. The

hypothesis from this study was shown to be partially correct. Four tests – pull-ups, BOMBT, leg tuck, and farmer’s carry – appeared to have some level of sensitivity. Between 84.3-91.2% of trainees who scored 60+ points in these tests were correctly identified as graduating from academy. However, none of these tests were specific, with high false positive rates (i.e., trainees who scored points in these tests were incorrectly identified as being released). Nonetheless, this information could be used to provide evidence as to the importance of the different fitness qualities measured by these tests in successful trainees. This includes upper-body pulling strength (32), total-body power (36), upper-body and trunk muscular strength and endurance (20), and grip strength and specific force production in a loaded carry (39), respectively.

The only test that appeared to have some degree of specificity was the IAT, in that 83.3% of trainees who did not graduate were correctly identified (i.e., 83.3% of trainees who scored 0 points for the IAT were released). The IAT provided a measure of change-of-direction speed (21, 35), and qualities such as lower-body strength, power, and running speed contribute to faster performance in this test (21). It is plausible that graduates who were released may have been lacking in these qualities. However, what should be acknowledged is that the IAT was not sensitive, with a high false negative rate (76.1% of trainees who scored 0 points graduated from academy). Additionally, certain fitness tests were neither specific nor sensitive relative to graduation as determined by the analysis in this study. This included the metronome push-ups and estimated  $VO_{2max}$ , in addition to the deadlift (which was not analyzed as all trainees scored 60+ points). There could be several reasons for this. Trainees can be released from academy for a variety reasons, including injuries and poor performance in academics and skill-based training scenarios (30, 33). Fitness is just one component of what is required to be a successful firefighter. Perhaps most notably, firefighter trainees will generally arrive at academy with some level of physical fitness. In order to be accepted to an academy, trainees must successfully complete the Candidate Physical Ability Test, or CPAT. The CPAT is a nationally recognized test that simulates job-specific tasks (stair climb, hose drag, equipment carry, ladder raise and extension, forcible entry, search, rescue drag, and ceiling breach and pull) (11). The tasks are completed in succession, and the CPAT must be finished in 10:20 min:s (11). As anaerobic and aerobic fitness are needed to successfully perform the CPAT (33), this could have limited the differences in fitness between graduated and released trainees. This could have also influenced the ROC and AUC results.

ROC curves are often used to analyze the effectiveness of diagnostic tests (12, 40). However, the approach of this study was to determine whether performance in certain fitness tests could be used to profile successful trainees. In what was somewhat counter to the studies’ hypothesis, there were few scores (raw or scaled point scores) that could predict academy graduation with a level of accuracy. For the raw fitness scores, the metronome push-up test had fair accuracy for predicting academy graduation for a trainee, with an AUC of 0.754. Relative to the scaled point scores, the BOMBT (AUC = 0.727) and overall score (AUC = 0.709) had fair accuracy for predicting academy graduation. Stevenson et al. (34) found that certain fitness tests (one-repetition maximum shoulder press, one-repetition maximum rope pulldown, repeated rope pulldown with 28 kg) had high predictive abilities for firefighting job tasks (ladder lift, lower,

and extension tasks; AUC = 0.82-1.00). However, these were discrete firefighting job tasks (34), as opposed to a broader concept of academy graduation. The AUC values found in this study were comparable to those shown by Orr et al. (26) in army recruits, and Lockie et al. (16) in law enforcement recruits, relative to academy attrition and graduation. Similar to Lockie et al. (16), the data from this study suggest that the analyzed fitness tests may have limitations in predicting academy graduation in firefighter trainees. As noted, this could be related to the fact that fitness is but one component that could affect trainee's academy success (30, 33). Additionally, as they all firefighter trainees must successfully complete the CPAT before being accepted to an academy (11), a 'floor' (or 'basement') effect, may also have influenced results (2).

Nonetheless, when combined with the sensitivity and specificity analysis, the results from this study provide some indication that successful firefighter trainees tend to display better fitness characteristics. Although the fitness tests may not necessarily be used for diagnostic purposes relative to graduation (and the department from this study did not use fitness test data for this reason), they could still be used to indicate trainees who tend to be successful through the academy process. Fitness test data could be used to provide feedback to trainees, especially those that may be deficient in certain fitness characteristics early in the training academy. This strategy could be very important for fire department command staff, especially considering the high costs associated with the loss of personnel (25). Some novel tests were spotlighted specific to firefighter graduation in this study, including the pull-ups, BOMBT, leg tuck, and farmer's carry (sensitivity), IAT (specificity), and metronome push-ups, BOMBT points, and total points (accuracy). If fire academy training staff were to identify deficiencies in any of these tests, they could provide targeted training to the individual requiring improvement. Improving performance in these fitness tests could ultimately contribute to a more successful academy training experience. The data from the current study reinforce the need for a trainee to be an 'all-rounder', with a requirement for anaerobic and aerobic capabilities (29, 33). In particular this recommendation is supported by the AUC results for the total points score, which incorporated scores for all the fitness tests and suggested fair predictive accuracy. While the points system and how it was derived requires further analysis, it is notable that a scoring system that incorporated multiple measures of fitness could provide an indication of academy graduation in firefighter trainees.

There are study limitations that should be recognized. Even though this has occurred in other tactical studies (5, 20), trainee demographic information (age, height, and body mass) were not provided to the researchers. This may limit comparisons across other firefighter research. This study involved a retrospective analysis of data, so the researchers did not have input into the selection of fitness tests. Additionally, tests were not completed in the order recommended by the National Strength and Conditioning Association (23). Firefighter training academies experience many challenges (e.g., space, equipment, time, and staffing), so staff often need to make the best of less-than-ideal conditions (20). Nonetheless, this could have affected the data that was collected. Investigating whether poor performance on several tests cumulatively impacts the predictive ability of fitness tests (i.e., the combination of low scores on the BOMBT

and 10RM deadlift) could be a focus of future studies. The points system used by this fire department was relatively arbitrary, but it was modelled off procedures adopted by the U.S. Army (8). The U.S. Army uses points derived from fitness tests to determine a soldier's physical readiness to be deployed for combat (8). In law enforcement, scores in fitness tests can be used to award high-performing recruits (19). Accordingly, there is context for analyzing the points allocated for fitness test performance in tactical personnel. In this study, a split of trainees who scored points and those that did not was used as an initial group split to analyze sensitivity and specificity in this exploratory research. Future studies could analyze different point scales to indicate whether they provide greater sensitivity or specificity as it pertains to academy graduation in firefighter trainees. Likewise, the specific reason for trainee separation may better inform the sensitivity, specificity, and accuracy of fitness tests (e.g., separated due to injury). The focus on this study was on measures of general fitness. Although general measures of fitness (i.e., anaerobic and aerobic fitness, muscular strength) relate to firefighting job tasks (29, 33), future research could analyze these job tasks and their sensitivity, specificity, and accuracy relative to academy graduation. In addition to this, academic scores and cognitive measures were not measured in this study and could contribute to the ability of a trainee to graduate from a fire training academy. The predictive capabilities of academic scores and cognitive measures relative to academy graduation should be examined.

In conclusion, the results from this study suggested that certain fitness tests could provide useful indicators for trainees who have a greater probability of graduation from a fire academy. Pull-ups, BOMBT, leg tuck, and the farmers carry all appeared to be sensitive, with between 84.3-91.2% of trainees who scored 60+ points in these tests being correctly identified as academy graduates. However, there were also a high number of false positives from these tests (i.e., trainees who scored 60+ points incorrectly identified as being released). The IAT was specific (83.3% of trainees who scored 0 points for the IAT were correctly identified as being released), but not sensitive (76.1% of trainees who scored 0 points still graduated from academy). The metronome push-up test, BOMBT points, and total points had fair accuracy for predicting academy graduation for a trainee (AUC = 0.709-0.754). Fitness tests should ideally not be used for diagnostic purposes relative to academy graduation. This is influenced by reasons other than fitness that a trainee may be released from academy, and a degree of homogeneity in trainees as they must all complete the CPAT prior to academy. Nevertheless, the data from this study collectively demonstrated that trainees who graduated from academy tended to have better total-body muscular strength, endurance, and power. Fitness testing is a tool for fire academy training staff that could be used to build a profile of successful trainees. This information could also be used to provide feedback for trainees who may be deficient in some of these fitness qualities, and to staff so they could provide targeted training for an individual.

## REFERENCES

1. Aandstad A, Holme I, Berntsen S, Anderssen SA. Validity and reliability of the 20 meter shuttle run test in military personnel. *Mil Med* 176(5): 513-518, 2011.

2. Adams KA, Lawrence EK. *Research Methods, Statistics, and Applications*. 2nd ed. Thousand Oaks, CA: SAGE Publishing; 2018.
3. Blazeovich AJ, Gill ND. Reliability of unfamiliar, multijoint, uni- and bilateral strength tests: effects of load and laterality. *J Strength Cond Res* 20(1): 226-230, 2006.
4. Bloodgood AM, Dawes JJ, Orr RM, Stierli M, Cesario KA, Moreno MR, Dulla JM, Lockie RG. Effects of sex and age on physical testing performance for law enforcement agency candidates: Implications for academy training. *J Strength Cond Res* 35(9): 2629-2635, 2021.
5. Butler RJ, Contreras M, Burton LC, Plisky PJ, Goode A, Kiesel K. Modifiable risk factors predict injuries in firefighters during training academies. *Work* 46(1): 11-17, 2013.
6. Cohen J. *Statistical Power Analysis for the Behavioral Sciences* 2nd ed. Hillsdale, New Jersey: Lawrence Earlbaum Associates; 1988.
7. Coyne JO, Tran T, Secomb JL, Lundgren L, Farley OR, Newton R, Sheppard JM. Reliability of pull up and dip maximal strength tests. *J Aust Strength Cond* 23(4): 21-27, 2015.
8. Davis JD, Orr R, Knapik JJ, Harris D. Functional Movement Screen (FMS™) scores and demographics of US Army pre-ranger candidates. *Mil Med* 185(5-6): 788-794, 2020.
9. Dawes JJ, Lockie RG, Orr RM, Kornhauser C, Holmes RJ. Initial fitness testing scores as a predictor of police academy graduation. *J Aust Strength Cond* 27(4): 30-37, 2019.
10. Faul F, Erdfelder E, Lang AG, Buchner A. G\*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Methods* 39(2): 175-191, 2007.
11. Firefighter Candidate Testing Center. Candidate Physical Ability Test. Available from: <https://www.fctconline.org/departments/about-cpat/>. Accessed March 3, 2021.
12. Hanley JA, McNeil BJ. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1): 29-36, 1982.
13. Hopkins WG. How to interpret changes in an athletic performance test. *Sportscience* 8: 1-7, 2004.
14. Lalkhen AG, McCluskey AJ. Clinical tests: Sensitivity and specificity. *CEACCP* 8: 221-223, 2008.
15. Lockie RG, Balfany K, Bloodgood AM, Moreno MR, Cesario KA, Dulla JM, Dawes JJ, Orr RM. The influence of physical fitness on reasons for academy separation in law enforcement recruits. *Int J Environ Res Public Health* 16(3): 372, 2019.
16. Lockie RG, Dawes JJ, Dulla JM, Orr RM. Extending research on law enforcement academy graduation and fitness: A research note on receiver operating characteristic curves. *J Strength Cond Res* 36(7): 2018-2022, 2022.
17. Lockie RG, Dawes JJ, Dulla JM, Orr RM, Hernandez E. Physical fitness, sex considerations, and academy graduation for law enforcement recruits. *J Strength Cond Res* 34(12): 3356-3363, 2020.
18. Lockie RG, Dawes JJ, Moreno MR, Cesario KA, Balfany K, Stierli M, Dulla JM, Orr RM. Relationship between the 20-m multistage fitness test and 2.4-km run in law enforcement recruits. *J Strength Cond Res* 35(10): 2756-2761, 2021.

19. Lockie RG, Dawes JJ, Orr RM, Dulla JM. Recruit fitness standards from a large law enforcement agency: Between-class comparisons, percentile rankings, and implications for physical training. *J Strength Cond Res* 34(4): 934-941, 2020.
20. Lockie RG, Orr RM, Montes F, Ruvalcaba TJ, Dawes JJ. Differences in fitness between firefighter trainee academy classes and normative percentile rankings. *Sustainability* 14(11): 6548, 2022.
21. Lockie RG, Post BK, Dawes JJ. Physical qualities pertaining to shorter and longer change-of-direction speed test performance in men and women. *Sports* 7(2): 45, 2019.
22. Los Angeles County Fire Department. Your Path to Becoming a Firefighter. Available from: <http://fire.lacounty.gov/wp-content/uploads/2019/08/fire-fighter-trainee.pdf>. Accessed February 12, 2021.
23. McGuigan MR. Principles of Test Selection and Administration. In: G G Haff and N T Triplett editors. *Essentials of Strength Training and Conditioning*. Champaign, IL: Human Kinetics; 2015, pp. 249-258.
24. Navalta JW, Stone WJ, Lyons S. Ethical issues relating to scientific discovery in exercise science. *Int J Exerc Sci* 12(1): 1-8, 2019.
25. Orr R, Simas V, Canetti E, Schram B. A profile of injuries sustained by firefighters: A critical review. *Int J Environ Res Public Health* 16(20): 3931, 2019.
26. Orr RM, Cohen BS, Allison SC, Bulathsinhala L, Zambraski EJ, Jaffrey M. Models to predict injury, physical fitness failure and attrition in recruit training: A retrospective cohort study. *Mil Med Res* 7(1): 26, 2020.
27. Piddubny A, Velez FAM, Velez MAM, Zhanna T, Sergey P, Alexander A, Maxim Y, Artem G. Reliability of alternative tests to assess the strength and endurance of the pectoral girdle (shoulder girdle) muscles of military personnel. *Ciencia Latina Revista Científica Multidisciplinar* 6(2): 1758-1773, 2022.
28. Ramsbottom R, Brewer J, Williams C. A progressive shuttle run test to estimate maximal oxygen uptake. *Br J Sports Med* 22(4): 141-144, 1988.
29. Rhea MR, Alvar BA, Gray R. Physical fitness and job performance of firefighters. *J Strength Cond Res* 18(2): 348-352, 2004.
30. Rio Hondo Fire Academy. Policy & Procedures Manual 2015-2017. Available from: <https://uploads.documents.cimpress.io/v1/uploads/588ee5e6-33e0-463a-9cdb-4e6201d5de12~110/original?tenant=vbu-digital>. Accessed January 13, 2022.
31. Ryman Augustsson S, Bersås E, Magnusson Thomas E, Sahlberg M, Augustsson J, Svantesson U. Gender differences and reliability of selected physical performance tests in young women and men. *Adv Physiother* 11(2): 64-70, 2009.
32. Sanchez-Moreno M, Pareja-Blanco F, Diaz-Cueli D, González-Badillo JJ. Determinant factors of pull-up performance in trained athletes. *J Sports Med Phys Fitness* 56(7-8): 825-833, 2016.
33. Sheaff AK, Bennett A, Hanson ED, Kim YS, Hsu J, Shim JK, Edwards ST, Hurley BF. Physiological determinants of the Candidate Physical Ability Test in firefighters. *J Strength Cond Res* 24(11): 3112-3122, 2010.
34. Stevenson RD, Siddall AG, Turner PF, Bilzon JL. Physical employment standards for UK firefighters: Minimum muscular strength and endurance requirements. *J Occup Environ Med* 59(1): 74-79, 2017.

35. Stewart PF, Turner AN, Miller SC. Reliability, factorial validity, and interrelationships of five commonly used change of direction speed tests. *Scand J Med Sci Sports* 24(3): 500-506, 2014.
36. Stockbrugger BA, Haennel RG. Validity and reliability of a medicine ball explosive power test. *J Strength Cond Res* 15(4): 431-438, 2001.
37. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 240(4857): 1285-1293, 1988.
38. Trevethan R. Sensitivity, specificity, and predictive values: Foundations, pliabilities, and pitfalls in research and practice. *Front Public Health* 5: 307, 2017.
39. Winwood PW, Cronin JB, Brown SR, Keogh JWL. A biomechanical analysis of the farmers walk, and comparison with the deadlift and unloaded walk. *Int J Sports Sci Coach* 9(5): 1127-1143, 2014.
40. Zou KH, O'Malley AJ, Mauri L. Receiver-operating characteristic analysis for evaluating diagnostic tests and predictive models. *Circulation* 115(5): 654-657, 2007.

