VALIDATING SCREENING TESTS AND DEVELOPING PERFORMANCE STANDARDS FOR CONTEMPORARY FIREFIGHTERS

Hugh H.K. Fullagar, John A. Sampson, Herbert Groeller, Brendan Mott and Nigel A.S. Taylor

Affiliation:

Centre for Human and Applied Physiology,
School of Medicine, University of Wollongong,
Wollongong, NSW 2522, Australia.

Corresponding Author and Reprint Requests:

Nigel A.S. Taylor, *Ph.D.*
Centre for Human and Applied Physiology,
School of Medicine, University of Wollongong,
Wollongong, NSW 2522, Australia.

Telephone: 61-2-4221-3463
Facsimile: 61-2-4221-5945
Electronic mail: nigel_taylor@uow.edu.au

ACKNOWLEDGEMENTS

Hugh Fullagar was supported by an Australian Postgraduate Award (Department of Innovation, Industry, Science and Research, Australia) funded by Fire & Rescue New South Wales (Sydney, Australia).

CONFLICTS OF INTEREST

This research was principally funded by Fire & Rescue New South Wales (Sydney, Australia) with additional support provided by the Defence Science and Technology Organisation (Melbourne, Australia).

RUNNING HEAD: Screening tests and performance standards for firefighters

VALIDATING SCREENING TESTS AND DEVELOPING PERFORMANCE STANDARDS FOR CONTEMPORARY FIREFIGHTERS

Abstract

Objective: The purpose of this investigation was to validate physiological screening test items both individually and collectively as a circuit, and to develop legally defensible performance standards. ***Methods:*** Fourteen university students and one hundred and forty eight firefighters were studied whilst performing criterion fire-fighting screening test items. ***Results:*** Six items were chosen for the physical aptitude test. Mean time taken to complete the test circuit by operational firefighters was 13 min 31 s with mean physiological responses for both university student and contemporary firefighters differing by only 0.09 $L.min^{-1}$ and 3 $beats.min^{-1}$. A pass threshold of 15 min for items 1-4 and 1 min 53 s for items 5-6 was finally recommended. ***Conclusion:*** It was concluded that the proposed physical aptitude test is a valid representation of the full demands of fire fighting.

Practitioner's Summary

This communication describes the procedures used to validate a previously developed physical screening test for firefighters in Australia. Additionally, this article outlines the methods used to develop a series of performance standards for this screening test. Such an approach was necessary to ensure the defensible identification of capable recruits.

Keywords: fire fighting, fitness, work, employment, occupational physiology, exercise

INTRODUCTION

Since fire fighting is recognised as an extremely physically demanding occupation (Davis *et al.*, 1982; Gledhill and Jamnik, 1992a; Bilzon *et al.*, 2001a; Barr *et al.*, 2010), fire fighting organisations have a legal obligation to employ individuals who are capable of tolerating the demands of this profession (Constable and Palmer, 2000). Pre-employment physiological screening tests can serve to identify both capable and incapable (those exposed to an unacceptable risk of injury during fire fighting) individuals (Jamnik *et al.*, 2010a; 2010b). These tests aim to simultaneously increase the capability of the workforce whilst minimising the risk of injuries to both firefighters and members of the community. Indeed, capable employees are clearly associated with fewer injuries in fire fighting (Cady *et al.*, 1985) and other manual handling occupations (Chaffin, 1974), such as the Navy (Marcinik, 1986). Within this context, Australia's largest and most active urban fire and rescue service made a decision to establish a legally defensible physical aptitude test (PAT) for operational firefighters.

The legal and scientific steps concerning the provision of genuine, certifiable (*bona fide*) and legally defensible physiological employment standards have been identified and described (Gledhill and Jamnik, 1992a, 1992b; Gledhill *et al.*, 2001; Constable and Palmer, 2000; Taylor and Groeller, 2003; Payne and Harvey, 2010; Jamnik *et al.*, 2012; Tipton *et al.*, 2012). These steps involve the identification, quantification and characterisation of the most essential, physically demanding tasks, providing the framework for a detailed physical job demands analysis. In the series of manuscripts preceding this communication, we reported the identification of the fifteen critical physical tasks of fire fighting (Taylor *et al.*, 2014a). These tasks were then quantified and characterised while they were being performed by contemporary operational firefighters (Taylor *et al.*, 2014b). Most recently, we focussed on analysing these critical fire-fighting tasks, and the physical and physiological demands they placed upon firefighters (Groeller *et al.*, 2014). This most recent step involved a detailed task analysis, leading to the generation of a list of criterion tasks for the design of physiological screening tests for firefighters. From this distillation process, and by identifying the operational constraints imposed on subsequent screening by the urban fire and rescue service, a preliminary format for the PAT was proposed. These criterion tasks, recommended to be performed in a circuit, are summarised in Table 1 and illustrated in Figure 1.

However, to be legally defensible, these criterion tasks must provide valid measures of the key physiological attributes of capable firefighters (Constable and Palmer, 2000). Thus, a process permitting the evaluation of the validity of the criterion tasks outlined in Table 1 is warranted. Within this process, due consideration must be given to construct-, face-, content- and criterion-related validity of each test item (Taylor and Groeller, 2003). From a physiological employment standard perspective, construct validity relates to how well each test item measures the attributes (physiological or others) of the occupation. Face validity refers to the apparent similarities between the screening test and actual work-related performance. Although subjective, this validity is sometimes the most important for test implementation as it is readily accepted by workers and arbitrators (Gledhill *et al.*, 2001). Content validity is similar to face validity, but is slightly stronger, to the level which relates to the extent each test item approximates the occupational (fire fighting) tasks. Finally, criterion validity ensures test analyses can accurately compare outcome (generic experimental) measures with an established criterion reference (known occupational demand; Taylor and Groeller, 2003).

With this in mind, the purpose of the present investigation was to i) assess the validity of each test item within the developed physiological screening test for firefighters, and ii) establish recommended performance standards for the PAT. We anticipated this approach would assist in the identification of recruits who are capable of tolerating the physiological strain associated with fire fighting. It is envisaged these recruits will be well suited to undertake fire-fighting tasks in a safe and productive manner. Therefore, we hypothesise the full demands of fire fighting will be adequately represented in the provision of valid screening tests that have a predictive capacity for fire-fighting performance.

INSERT TABLE 1 ABOUT HERE

INSERT FIGURE 1 ABOUT HERE

METHODS

This research was conducted in accordance with the Declaration of Helsinki and received joint approval from the Human Research Ethics Committee (University of Wollongong) and the associated urban fire and rescue service. Following the results of our previous work (Taylor *et al.,* 2014a, Taylor *et al.*, 2014b, Groeller *et al.*, 2014), the present investigation focussed on steps 14-16 for the development of genuine, certifiable (*bona fide*) and legally defensible physiological employment standards (Taylor *et al.*, 2014a):

14) Evaluate validity and reliability of screening tests
15) Acknowledge and approve performance standard development
16) Develop performance standards

Subjects

Fourteen (eleven male, three female) physically active university students (mean 31.9 y [standard deviation (SD) 10.6]; 1.76 m [SD 0.10]; 73.8 kg [SD 11.1]) participated in the validation phase of this research (step 14 and 15). These individuals were screened using a Par-Q questionnaire (Chisholm *et al.*, 1978) to eliminate individuals with a history of cardiovascular, respiratory or thermoregulatory contraindications. One hundred and forty eight (males and females???) operational firefighters (mean 39.1 y [SD 9.1]; 1.78 m [SD 6.82]; 84.6 kg [SD 12.0) participated in the development of performance standards phase (step 16). This group of permanent and retained firefighters were made up of all employment ranks up to, and including Superintendent. It was envisaged crucial that the participating firefighters represented the associated workforce with regards to years of experience, firefighter rank, age, stature, mass, race and minority status. Following the same screening procedure as aforementioned, two firefighters were identified with cardiovascular risk factors, and they were not permitted to participate in the testing. All subjects provided written and verbal informed consent prior to participation in all experimental procedures, all of which were entirely voluntary. Whilst participating in this phase of the research, firefighters were relieved of duty for the duration of the testing days.

Experimental overview

Development of a test protocol and establish scientific accuracy of the protocol (Step14)

In this stage (step 14), we quantified and evaluated the physical and physiological demands of performing the proposed screening test items (simulations) that were aimed at replicating the movement patterns, the muscular demands and the metabolic costs of the criterion tasks identified within Table 2. The intention of these trials was to determine the criterion validity and the predictive capability of several proposed test items, and to establish suitable work rates for each task. Construct validity for all tasks had been previously established by Taylor *et al.* (2014a and b) and Groeller *et al.* (2014). Each of these task simulations within the present investigation were performed with participants wearing exercise clothing (gym shorts, t-shirt, running shoes), the thermal protective ensemble and personal protective equipment used by firefighters (7.7 kg), self-contained breathing apparatus (12 kg) and, on certain occasions (where stated throughout the methods section), a portable data collection device which measured oxygen consumption (respiratory gas analysis system: Metamax 3B, Cortex Biophysik, Leipzig, Germany, 1.82kg). This clothing, and the additional load carried by the participants, replicated the equipment worn and carried by firefighters within the urban fire and rescue service (total mass ~20 kg). The following description of the methods for each task item are presented in each of the four criterion classes as outlined previously (Groeller *et al.,* 2014).

INSERT TABLE 2 ABOUT HERE

Single sided carry tasks

Since a true replication of the some of the tasks were unfeasible due to time constraints (i.e. hazmat task) and resources (i.e. ventilation fan stair climb) for the PAT (Groeller *et al.,* 2014), tasks of a shorter duration yet equivalent physiological and biomechanical strain were established. An assessment of these tasks provided criterion validity. This included determining suitable walking speeds to assess the hazmat task (Table 2) whereby performance time was recorded. The same subjects completed a hazmat replication task on a separate day. This activity was a repetition of the above simulation, but with participants walking at three different absolute speeds. Oxygen consumption was measured continuously and trials were performed in a randomised order with 5-min rest between trials.

In addition, we investigated whether a loaded-box step provided a valid prediction of one's ability to perform the simulated ventilation fan stair climb (Table 2). Participants completing the simulated ventilation fan stair climb (loaded mass 17.5 kg) were instructed to perform it as quickly as possible, but at walking speed (one foot always in ground contact). Rest stops were permitted, as was the changing of the load between hands. Performance was assessed as the time to complete the task once. The same subjects performed three loaded (but each at three different speeds), 5-min box-stepping trials on a different day. Trials were performed in a randomised order with at least a 5-min rest between trials. Oxygen consumption was measured continuously.

Overhead pushing and holding tasks

Two criterion tasks were evaluated within this criterion class: motor-vehicle rescue and ladder under-run (10.5 m) and five different items were evaluated within this test category (Table 2). Three relative heights were chosen by for the motor vehicle rescue task; below the knees, at waist height and at eye level. In this simulation, participants first performed three trials, each lasting 6 min. Whilst the duration at each relative height remained constant (1 min), the duration of the static hold was varied (Table 2). In addition, each participant performed a fourth trial which involved a static hold in each position until failure; defined as the inability to hold the mass at the required height, or the loss of the safe holding posture. Task performance was recorded whilst oxygen consumption was measured continuously.

We then explored the possibility if performance on the ladder under-run (10.5 m) task could be predicted from performance on a generic, single-lift task (box lift and place, shoulder press). This contributed to the criterion-related validity of the assessments. All subjects were required to perform a single-person ladder raise from the ground to the vertical position immediately followed by lowering the ladder to the ground. A 30-s rest period was provided between the completion of the ladder raise and its subsequent lowering. Performance times for the separate raise and lower stages were recorded. The same subjects performed a maximal box lift and place task was used to replicate the movement pattern observed within the ladder raise. Subjects were required to lift a box (30.5^3cm) from the ground, and to place it on a platform 1.5 m above the ground. The box was returned to the ground by the researchers. Incorrect lifting technique (i.e. loss of a neutral spine, excessive lordosis, stooped lifting), sliding the box onto the shelf or the loss of box control each defined test failure. Maximal mass lifted for each subject was recorded (with mass added incrementally until maximal mass was achieved). Subjects also performed an endurance box lift on a separate day. This activity was a replication of the maximal lift, but the mass was set at 75% of the mass each person lifted in the maximal lift and the subject returned the box to the ground after each successful lift. The aim of this was to determine how many times each participant could successfully and safely lift and place this box. Finally, participants performed a maximal (front) shoulder press within a lifting machine (seated) on a different day, ensuring a controlled vertical movement. Failure was defined as an inability to move the loaded bar through the required range of motion and with full extension at the elbows.

Cardiorespiratory dragging tasks

The fire attack was included within the provisional screening test circuit by the subject matter experts due to its criticality to job performance. Thus, one task remained to be evaluated within this task: dragging a charged hose (38 mm: bushfire simulation). However, the duration of this task (>50 min) did not lend itself well to an efficient task within a test circuit. Since the oxygen consumption of sub-maximal work will stabilise after 3-5 min (Henry and DeMoor, 1956; Di Prampero *et al.*, 1970; Whipp and Wasserman, 1972), the hose-drag replication was performed over 6 min, and this was designed to replicate the average physiological demand of this criterion task (Taylor *et al.*, 2014b; 1.63 $L.min^{-1}$). The final aim of this phase was to determine which walking speed would be required to replicate the metabolic demand observed for the bushfire hose-drag simulation to provide criterion-related validity. Subjects performed this task at two different intensities in an alternating manner. The first stage of each work cycle was heavier, with the subject holding the branch of a 5.6-m section of 38-mm hose filled with sand (11 kg) and doubled over, giving a 2.8-m length of hose. The participant walked away 30 m from (pulling against stage) a resistance loader (Groeller *et al.*, 2014) that had been set to provide the equivalent resistive force. On reaching a marker, participants passed the hose branch to the researcher and then walked back to the resistance device (the unloaded or light work stage). This two-stage sequence was repeated until 6 min elapsed. Task performance was measured from the total distance covered. On a different day subjects repeated the above simulation but at two different absolute speeds. These speeds were the same for all subjects, with the first (slow speed) set at 30% of the average maximal walking speed observed in the first simulation, and the second (medium speed) set at 60% of the average maximal walking speed. Trials were performed in a randomised order with 5 min rest between trials. Oxygen consumption was measured continuously for all the simulations within this criterion task.

<u>Critical strength task</u>

The subject matter experts and research group confirmed the firefighter rescue (Table 1) needn't be simulated during step 14, as this activity was deemed to be an essential fire-fighting task, and it was anticipated that this task would be performed in the prototype PAT as a high-speed task that demanded a near-maximal effort. Therefore, this item was deemed of critical importance for job performance and fellow firefighter safety.

Acknowledge and approve performance standard development whilst running prototype screening test (Step 15)

Feedback and approval from researchers, various subject matter experts and personnel from the urban rescue service were used to modify the items of the prototype circuit test in this step. For instance, it was deemed gaining entry with a sledge hammer (Table 1) should be eliminated from the prototype circuit test as it could be performed with less effort if the firefighter chose to use alternate tools. Following these discussions, the same fourteen subjects who participated in step 14 completed the prototype PAT as a circuit (step 15; comprising of six chosen test items), providing the researchers and subject matter experts with a detailed description of all aspects of the circuit. These descriptions would assist in guiding test administration recommendations for the development of the final test circuit. Oxygen consumption and heart rate were monitored continuously.

Develop physical performance standards for the test protocol (Step 16)

Completing this step involved operational firefighters (both permanent and retained) being exposed to the prototype circuit test that would assist in recommendations for the urban fire and rescue service to determine performance standards, or 'pass thresholds', for the final PAT. The performance standards for the circuit were based upon the time taken for each firefighter to complete the prototype PAT. Firefighters were instructed to complete the test at operational work intensity. This recommendation was designed to ensure that performance standards that may arise from such trials do not suffer from bias introduced by participants working at unnecessarily high or low work rates. Running was not permitted as per occupational health and safety regulations within the urban rescue and fire service and firefighters were instructed that they could stop to rest at any time within or between tasks, and for any duration, except during the static-hold simulation. A sub-group (18) of operational firefighters within this original sample (148) also completed the PAT whilst oxygen consumption and heart rate were monitored.

Standardisation

Participants for all experiments were requested to refrain from strenuous exercise, and the consumption of alcohol and tobacco during the 12 h prior to testing. For the night prior to each trial, participants were instructed to drink 15 $mL.kg^{-1}$ of additional water before retiring, and to eat a high-carbohydrate and low-fat evening meal and breakfast.

Measurements

The following measures were recorded during Steps 14-16 where previously mentioned: Oxygen consumption was measured using a portable open-circuit, expired gas analysis and ventilation system (Metamax 3B, Cortex Biophysik, Leipzig, Germany). Data were recorded on a breath-by-breath basis and reported at 5-s intervals. Equipment was calibrated (3L syringe) at the start of each test day, with calibration verification performed throughout each day. Heart rates were monitored continuously from ventricular depolarisation (Polar Electro Sports Tester, Kempele, Finland). Performance time was recorded using a stopwatch (Hart Sports Timer 898, Brisbane, Australia). Subjective evaluation of face and content validity was determined immediately after completing the PAT, whereby each firefighter was asked to rate how well the test replicated the real-world demands of fire fighting using a visual-analogue scale (Figure 2).

INSERT FIGURE 2 ABOUT HERE

Design and analysis

Means, standard deviations (SD) and data ranges (minimal and maximal limits) task performance (time, distance) and physiological measurements (heart rate, oxygen consumption) are reported (Steps 14 and 15). In addition, two forms of correlation analysis (Pearson's product-moment and Spearman's rank-order correlation) were performed to investigate the predictive power of tasks within each class description (Table 1 and Table 2). Two additional correlation analyses were performed to investigate the predictive power of the simulated reel lift for the ladder-raise task. Mean, SD, data ranges and 95% confidence intervals (CI) were calculated for task performance (time) and physiological measurements (heart rate, oxygen consumption) for the development of performance standards (pass thresholds; Step 16). Where appropriate, significant differences are described using independent *t-tests* with *alpha* set at the 0.05 level for these comparisons.

RESULTS

Development of a test protocol and establish scientific accuracy of the protocol (Step 14)

The average performance time for the simulation was 589.9 s (SD 141.6: range: 436.0-973.0 s). The average walking speed over the entire simulation was 1.74 $m.s^{-1}$ (range: 1.05-2.35 $m.s^{-1}$). The average lap time for the maximal 5-min simulation was 41.4 s. Mean distance covered, heart rate and oxygen consumption were 387.0 m SD 109.2, 158 $beats.min^{-1}$ SD 19 and 1.74 $L.min^{-1}$ SD 0.4 respectively for the hazmat simulation. Heart rate and oxygen consumption for the slow (40 s) and medium walking speeds (68 s) were 130 $beats.min^{-1}$ SD 26, 1.37 $L.min^{-1}$ SD 0.29 and 134 SD 23, 1.47 $L.min^{-1}$ SD 0.25 respectively. These speeds remained greater than the average speed of the two subjects who failed to complete the 16-lap simulation, so these individuals were omitted from the last two trials. The 5-min replication displayed a good predictive relationship with the longer, 16-lap simulation (r^2=0.57; Figure 3). However, the loaded stair climb simulation poorly correlated with performance on the hazmat simulation (r^2=0.176). The three loaded box stepping rates used in this activity were 61 (fast: range: 45-106), 36 (23-54) and 18 $steps.min^{-1}$ (11-27) with derivations of these speeds presented in Figure 4. Mean heart rate and oxygen consumption for the slow, medium and fast rates were 115 beats.min^{-1} SD 23, 0.96 L.min^{-1} SD 0.1; 125 beats.min^{-1} SD 26, 1.21 L.min^{-1} SD 0.13; 138 beats.min^{-1} SD 23, 1.55 L.min^{-1} SD 0.38 respectively.

The predictive utility of the 5-min maximal box stepping simulation possessed a good predictive relationship with the criterion task (r^2=0.48; Figure 5). Whilst the strength demand was high for the static-hold simulation holding tasks, it there was clearly a low metabolic demand (Figure 6). For the ladder task the mean time was 42.9 s SD 20. Predicting ladder raise performance from a maximal box lift and place task (r^2=0.271) or a maximal shoulder press test (r^2=0.193) were both poor. The average distance covered in the hose drag trial was 596.6 m (SD 69.4) or an average walking speed of 1.66 m.s^{-1}. The cardiovascular strain and the metabolic cost of hose dragging performed at two walking speeds was: slow (0.88 m.s^{-1}; 133 beats.min^{-1}; SD 21; 1.54 L.min^{-1} SD 0.19) and medium (0.93 m.s^{-1}; 141 beats.min^{-1} SD 21; 1.75 L.min^{-1} SD 0.17). The oxygen cost of dragging a charged fire hose was found to be well predicted by determining the distance covered (r^2=0.612; Figure 7).

Based upon these results, it was recommended that both the hazmat simulation and loaded box stepping be considered as necessary (individual) items within the PAT, since the correlation between these tasks was poor. Comparatively, since hazmat work is typically interspersed with unloaded activity, and since the ventilation fan carry is a relatively brief task, then it was recommended that the 5-min duration used for these tasks within this evaluation be reduced when performed within the PAT. It was recommended an additional exploratory investigation of the ladder task be undertaken, as the predictive power of other activities were poor. Feedback from the subject matter experts indicated that due to the high face validity of the static-hold activity, it is recommended that it be included in the PAT. The cardiorespiratory demand observed during the hose-drag simulation was similar to that observed for the bushfire simulation performed by operational firefighters (Taylor *et al.*, 2014b). For this task to be used within the PAT, it would need to be performed as a repeated dragging task over a restricted course, as it was above. In addition, the metabolic threshold for recruit screening should not be set higher than the level deemed necessary to successfully perform this task under operational conditions. Accordingly, the recommended test work rate for this task was calculated to match the lower 95% CI observed in the second manuscript within these series of investigations (1.41 L.min^{-1}; Taylor *et al.*, 2014b).

INSERT TABLES AND FIGURES ABOUT HERE

Acknowledge and approve performance standard development whilst running prototype screening test (Step 15)

The prototype test circuit was performed in the following order to reflect operational sequences and intensities that may realistically be encountered during an incident: Unilateral load carriage (distance: 195 m, load: 26 kg): hazmat simulation > Test item two: Stepping with a unilateral load (36 steps [26 cm each], 17.5 kg): ventilation fan carriage simulation > Test item three: Static hold (load: 19.5 kg, duration: 3 min): motor-vehicle rescue simulation > Test item four: Charged hose drag (distance: 150 m): bushfire simulation > Test item five: Fire attack (height restriction: 1.25 m, distance: 30 m) > Test item six: Firefighter rescue (height restriction: 1.25 m, distance: 10 m). Mean total time for the prototype PAT was 14 min 10 s SD 3 min 12 s. Mean heart rate and oxygen consumption was 158 beats.min^{-1} SD 14 and 2.45 L.min^{-1} SD 0.36 respectively (Figure 8). Observations and feedback provided from this stage included the height restriction of 1.25 m during the simulated firefighter rescue forced a number of participants to adopt postures that prevented a rapid response to this emergency, and increased the risk of injury. Hence, it was recommended this height be re-considered.

INSERT TABLES AND FIGURES ABOUT HERE

Develop physical performance standards for the test protocol (Step 16)

On average, and across all operational firefighters, the time taken to complete the final PAT was 13 min 31 s (range: 9 min 53 s to 25 min 27 s). Retained firefighters performed significantly more slowly (15 min 10 s: range: 11 min 5 s to 20 min 29 s) than the permanent firefighters (13 min 10 s: range: 9 min 53 s to 25 min 27 s; $P<0.05$). The average times for each separate task and the circuit as a whole are illustrated in Table 3. In the group of 18 permanent firefighters from whom oxygen consumption and heart rate data were simultaneously recorded, the mean absolute oxygen consumption was 2.36 $L.min^{-1}$ (CI: 2.07-2.65 $L.min^{-1}$), while the mean heart rate was 161 $beats.min^{-1}$ (CI: 156-166 $beats.min^{-1}$). Mean subjective face validity ratings of the PAT were rated by the firefighters at 82 out of 100 (range: 15-100; SD 11). On the basis of this evaluation, it was concluded that the face validity of the PAT was very high, with two-thirds (66%) of all responses falling between 71-93.

Based upon the performance speeds and work rates that would elicit the threshold metabolic costs seen in firefighters performing the first four tasks effectively, and at an acceptable level of operational performance, it was determined that the first four tasks should be completed within 15 min (test item one [hazmat task]: 3 min; test item two [ventilation fan carriage]: 3 min; test item three [motor-vehicle rescue - static hold]: 3 min; test item four [charged hose drag]: 6 min). It is therefore recommended that this 15-min time be considered as a screening threshold (performance standard) for the first four items of the final PAT (Table 4). For the final two tasks (fire attack and firefighter rescue) it is recommended a combined time of these two tasks (the pass threshold) be set at 1 min 53 s (the lower 95% CI for these test items), with a 5% buffer (6 s) to be used. Such a combination also possessed high face-validity as these two tasks would be how such an approach would be performed in a real-life operational scenario. In this instance, the buffer would apply just to these two test items, and this would necessitate discrete timing for these tests. Thus, one could satisfy the two pass thresholds for items one-four and five-six. However, it is recommended that a failing score that falls outside the buffer zone for these last two criterion tasks should constitute an absolute test failure (Table 5). It is also recommended that the PAT be performed whilst carrying a load of 20 kg.

In addition to performing the final PAT, 108 firefighters within this sample also performed both a standard ladder raise and a simulated ladder raise using the resistance loader (Groeller *et al.*, 2014). This additional aim within this phase was due to the poor predictors of ladder-raise performance from step 14. Thus we sought to establish a satisfactory correlation between the ladder-raise task and a strength-based simulation task using the resistant loader. All firefighters successfully completed the ladder raise, with an average performance time of 11.3 s (range: 4.4 s to 31.9 s). Five firefighters (3.5% of the sample) failed to reach the maximum resistive load applied by the resistance loader during this task. These firefighters were also amongst the slowest to raise the ladder (average time 24 s). Although these numbers are small, a significant (P<0.05) correlation was observed, with 40% of the variance in the ladder raise time being explained by differences in the peak load raised using the resistance loader.

INSERT TABLES AND FIGURES ABOUT HERE

DISCUSSION

To the best of our knowledge, all the items within the final PAT in this investigation were deemed to be adequately represented in the provision of a valid screening test that have a predictive capacity for fire-fighting performance. Taken collectively, these tasks were considered the most critical and physically demanding for fire-fighting performance. The establishment of performance standards (pass thresholds) for the PAT were also obtained from the completion times of operational firefighters of all ages, experience levels, genders and races. Indeed, this approach is required to ensure a *bona fide* occupational requirement is legal defensible (Supreme Court of Canada, 1999), and as such will assist in the identification of recruits who are fully capable of tolerating the physiological strain associated with fire fighting.

Physical screening tests for job performance are utilised by many occupations, such as the Army (Rayson *et al.,* 2004) and police forces (Jamnik *et al.*, 2010a, 2010b, 2010c). Specifically, for decades firefighter screening tests have generated much interest for both researchers and fire-fighting organisations. Many investigators have developed firefighter screening tests to include grip strength, anthropometry, muscular strength and cardiorespiratory measurements (Lemon and Hermiston, 1977; Cady *et al.*, 1985; Ellam *et al.*, 1994; Henderson *et al.*, 2007). Comparatively, researchers have developed screening tests involving task-specific components, such as a ladder lift, stair climb and hose drag (Gledhill and Jamnik, 1992b; Rayson *et al.*, 2004). The current investigation proposed a task-specific test comprised of six criterion test items for firefighters which are believed to be valid representations of fire-fighting tasks critical to successful job performance. Throughout the present investigation and previous work there are both similarities and differences between job simulation tests; however it should be noted that due to both methodological and operational differences among these fire services and across investigations, the comparison of such results should be treated with caution.

Whilst previous investigators have found significant correlations between stair-climbing ability and maximal consumption as measured by horizontal locomotive tasks (Cateneo *et al.*, 2010; Koegelenberg *et al.*, 2008), the ability of the loaded step task to predict the flat hazmat simulation was poor in the present investigation. This indicates that these tasks were sufficiently different in their physiological demands (although both require upper-body muscular work), negating the possibility that one task may be used as a substitution for the other. This was also the case with predictions of ladder-raise performance from a maximal box lift and place task or a maximal shoulder press test, which were both poor. These results are most likely due to the high skill component of the ladder task. Thus, we sought to establish a satisfactory correlation between the ladder-raise task and a strength-based simulation task using the resistant loader, with a significant correlation found between a standard ladder raise and a simulated raise on the resistant loader. This supports previous work, suggesting strong correlations exist between manual-handling task performance and strength measures, especially with progressions in resistance training (Sharp *et al.*, 1993a; Kraemer *et al.*, 2001). Hence, it was concluded that a simulated ladder raise be included as part of the final PAT.

Although slight operational differences exist among fire services, the physiological demand of the motor-vehicle rescue task (static-hold simulation) are well established (Gledhill and Jamnik, 1992a; Groeller *et al.*, 2014). Interestingly, the metabolic demand in this investigation was low compared to the physiological responses in earlier stages of this research (Taylor *et al.*, 2014b), limiting the construct validity of this task for inclusion in the final PAT. However, it is clear the muscular endurance and strength (Groeller *et al.*, 2014) required to perform this task at different heights (range: close to the ground to above the shoulder) places a considerable strain on firefighters working musculature (Gledhill and Jamnik, 1992a). This supports previous work on the mechanical loading of the body (Ayoub *et al.*, 1979; Warwick *et al.*, 1980; Mital, 1984; Ljunberg *et al.*, 1989). For instance, the maximal load that can be successfully lifted is significantly lower when performing maximal lifts to a height above the shoulder (Ayoub *et al.*, 1979; Mital, 1984). Since none of the other criterion tasks possess these unique strength and muscular endurance requirements (Groeller *et al.*, 2014) and given the high face validity of the task, the static-hold simulation task (testing at three different heights) was deemed a necessary inclusion in the final PAT.

The cardiovascular strain of hose dragging performed was high; similar to values in the second article within this series of communications (Taylor *et al.*, 2014b). Hence, this task possessed high construct validity. This is consistent with previous work highlighting the high physiological strain involved when performing hose drag tasks (Gledhill and Jamnik, 1992a; Smith *et al.*, 1996; Williford *et al.*, 1999; Bilzon *et al.*, 2001b; Smith *et al.*, 2001). Moreover, Canadian, British and North American fire-fighting organisations recommend simulated hose tasks be utilised in firefighter fitness assessments (Gledhill and Jamnik, 1992b; Rayson *et al.*, 2004; Michaelides *et al.*, 2008). In the present investigation, the oxygen cost of dragging a charged fire hose was also found to be well predicted by determining the distance covered, showing good content validity. However, technique clearly influenced test performance during the hose-drag tasks, especially during step 15. Since the PAT is designed as a recruit screening aid, then test performance should not be significantly influenced by variations in performance technique, as this would introduce an unacceptable bias into the screening process (Constable and Palmer, 2000). It is therefore recommended that all participants should be given sufficient opportunity to attempt these tasks using various dragging techniques, so that each individual will have an opportunity to select and (if necessary) learn the most appropriate technique for himself/herself. Indeed, such familiarisation periods can improve performance by approximately 5% in strength-based tests (Mendez-Villanueva *et al.*, 2007; Ploutz-Snyder and Giamis, 2001).

The inclusion of such critical, lifesaving tasks in proposed physiological screening tests is necessary to maintain the welfare of employees and members of the wider community. Indeed, the inclusion of lifesaving tasks in screening tests is common in other emergency occupations, such as life guarding (Reilly *et al.*, 2006a, 2006b). This is crucial, as legally defensible (valid) screening procedures must also reflect the physical demands of the most important trade tasks (Gumieniak *et al.*, 2011). Although short duration, the fire attack and firefighter rescue entail significant metabolic and upper and lower body muscular demands (Table 3; Taylor *et al.*, 2014b; Groeller *et al.,* 2014). In particular, feedback from the firefighter focus groups in the identification of the physical tasks for firefighters indicated that the firefighter rescue was the single most important (critical) task (Taylor *et al.*, 2014a). Indeed, the inability to perform these tasks would place their own, and fellow firefighters', safety at an unacceptable level of risk. Hence, the fire attack and firefighter rescue were valid inclusions in the final PAT.

Although screening tests are comprised of numerous intellectual, physiological and psychological components which serve to identify those most capable of performing the job (Constable amd Palmer, 2000), these components are often ignored in order to serve purposes outside this *bona fide* occupational standard development (i.e. political motivation). As such, many of these tests derive standards from an inadequate participant group (Barr and Flannery, 2006), or using characteristics of the potential recruits rather than what constitutes adequate job performance (Jamnik *et al,* 2010b). With this approach leaves open the possibility of legal challenges (Constable and Palmer, 2000, Doherty *et al*., 2007). In light of this, the PAT in the present investigation was based on the identification the most essential, physically demanding tasks (Taylor *et a.,* 2014a), the quantification and characterisation of these tasks (Taylor *et al.*, 2014b) and the generation of a list of criterion tasks for the design of physiological screening tests for firefighters (Groeller *et al.*, 2014). All of these steps involved the derivation of the desired performance outcomes from an adequate representation of the current employment population based on age, gender, race and employment rank. This approach is similar to previous approaches for correctional officers (Jamnik *et al,* 2010a, 2010b, 2010c), with an aim to identify job simulation tests which closely resembled the physical demands endured during on-the-job performance.

Indeed, it is crucial an implemented recruit screening procedure is a valid and reliable representation of the physical demands faced during actual job performance. In the present investigation, the subjective face validity of the final PAT was high (82/100 SD 11). This was an important finding, with this type of validity sometimes the most important for test implementation as it is readily accepted by workers and arbitrators (Gledhill *et al.,* 2001). Feedback from the subject matter experts also supported the final PAT as indicative of the demands likely to be faced during on-the-job performance, providing good content validity for this circuit test. In the group of eighteen permanent firefighters from whom oxygen consumption and heart rate data were simultaneously recorded (Step 16), these values were very close to the responses of the University participants (Step 15), differing by only 0.09 $L.min^{-1}$ and 3 $beats.min^{-1}$. These results provide strong evidence that the physiological impact of the PAT was not skill dependent, and this is important information, since the test is initially aimed at recruit screening. While there is no doubt skill is included in almost all physical tasks, the degree to which the skill dictates the completion of the task must be carefully managed (Equal Employment Opportunity, 1978; Constable and Palmer, 2000). For instance, it is known physiological changes can occur with the acquisition of motor skills (Kleim, *et al.*, 1996; Nudo *et al.*, 1996; Vandenberg *et al.*, 2002). Given skill is a critical part of firefighter training (e.g. operating heavy machinery) and can be taught proceeding recruitment, physiological employment screening tests should primarily target key physiological attributes so employees are measured on physical performance (Constable and Palmer, 2000).

The average time for completing the final PAT was 13 min 31 s. Within the interpretation of this mean time, consideration must be given to the use of either an absolute screening threshold, or a threshold with a zone of uncertainty. It is recommended that the latter be adopted, since there will always exist a range of scores around such screening thresholds for which one can never have absolute certainty that these individuals have either passed or failed the PAT. It is therefore recommended that a 5% buffer be used for this purpose. The justification for this 5% buffer is two-fold. Firstly, the recommended acceptable work rates for some tasks were set to simulate values at the lower end of the 95% probability range, as this corresponded with values observed in operational firefighters who could still successfully complete the task, but at the lower end of the range of these physiological responses. Thus, by definition, the recommended performance standard could be passed by 95% of the operational firefighters tested. Secondly, since it is the widely accepted convention in science to set statistical thresholds for physiological research at the 5% level, then this value was also considered appropriate in this context.

CONCLUSION

The final PAT is a physiological screening test comprising of six criterion test items which we conclude to be a valid representation of the full demands of fire fighting as performed within Australia. These tasks were chosen based on the performance, physiological responses and feedback from operational firefighters whilst completing a rage of previously established physically demanding tasks. The performance standards for the final PAT were derived from the performance times of a sample of contemporary firefighters (whom were representative of the range of ages, genders, race and employment rank currently present in the urban fire and rescue service) who completed the PAT at operational pace. From a collective perspective, the PAT possesses high criterion-, content-, construct- and face-related validity. Hence, we believe this physiological employment standard meets the requirements to qualify as genuine, certifiable (*bona fide*) and legally defensible. We anticipate this research, and the series of studies that preceded it, will assist in the identification of recruits who are capable of tolerating the physiological strain associated with fire fighting.

REFERENCES

Ayoub, M.M., Dryden, R., McDaniel, J., Knipfer, R., and Dixon, D. (1979). Predicting lifting capacity. *American Industrial Hygiene Association Journal*. 40:1075-1084.

Barr, D., Gregson, W., and Reilly, T. (2010). The thermal ergonomics of firefighting reviewed. *Applied Ergonomics.* 41:161-172.

Barr and Flannery v. Treasury Board of the Department of National Defence. (2006). *Public Service Labour Relations Board of Canada*. Citation: PSLRB 85.

Bilzon, J.L.J., Scarpello, E.G., Smith, C.V., Ravenhill, N.A., and Rayson, M.P. (2001a). Characterization of the metabolic demands of simulated shipboard Royal Navy fire-fighting tasks. *Ergonomics*. 44(8):766-780.

Bilzon, J.L.J., Allsopp, A.J., and Tipton, M.J. (2001b). Assessment of physical fitness for occupations encompassing load-carriage tasks. *Journal of Occupational Medicine.* 51(5):357-361.

Cady, L.D., Thomas, P.C., and Karwasky, R.J. (1985). Program for increasing health and physical fitness of fire fighters. *Journal of Occupational Medicine*. 27:110-114.

Cateneo, D.C., Kobayasi, S., de Carvalho, L.R., Paccanaro, R.F., and Cateneo, A.J.M. (2010). Accuracy of six minute walk test, stair test and spirometry using maximal oxygen uptake as gold standard. *Acta Cirurgica Brasileira*. 25:194-200.

Chaffin, D.B. (1974). Human strength capability and low-back pain. *Journal of Occupational Medicine*. 16:248-254.

Chisholm, D.M., Collis, M.L., Kulak, L.L., Davenport, W., Gruber, N., & Stewart, G. (1978). *PAR-Q Validation Report: The evaluation of a self-administered pre-exercise screening questionnaire for adults*. Vancouver, BC: Ministry of Health.

Constable, S.H., and Palmer, B. (2000). *The process of physical fitness standards development*. Human Systems Information Analysis Center Program Office, Wright Patterson Air Force Base, OH, USA.

Davis, P.O., Dotson, C.O., and Maria, D.L.S. (1982). Relationship between simulated fire fighting tasks and physical performance measures. *Medicine and Science in Sport and Exercise.* 14:65-71.

Di Prampero, P.E., Davies, C.T., Cerretelli, P., and Margaria, R. (1970). An analysis of O_2 debt contracted in submaximal exercise. *Journal of Applied Physiology*. 29:547-551.

Docherty, D., Goulet, L., Gaul, K., McFadyen, P., and Petersen, S. (2007). *Phase III Report: Development and Validation of a Physical Fitness Test and Maintenance Standards for Canadian Forces Diving Personnel*. A report prepared on behalf of Canadian Forces Personnel Support Agency. Pp. 1-186.

Ellam, L.D., Fieldman, G.B., Fordham, M., Goldsmith, R., and Barham, P. (1994). The perception of physical fitness as a guide to its evaluation in firemen. *Ergonomics*. 37(5):942-952.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor and Department of Justice. (1978). *Uniform guidelines on employee selection procedures.* Federal Register, 43(166), 38295–38309. 29CFR1607. United States Government. Pp. 199-224. Available from U.S. Government Printing Office: http://frwebgate.access.gpo.gov/cgi-bin/get-cfr.cgi?TITLE=29&PART=1607&SECTION=1&YEAR=2000&TYPE=PDF [Accessed May 26th, 2011].

Gledhill, N., Bonneau, J., and Salmon, A. (2001). *Bona fide occupational requirements*. Proceedings of the consensus forum on establishing *bona fide* requirements for physically demanding occupations. York University, Toronto, Canada. September 13th-16th, 2000.

Gledhill, N., and Jamnik, V.K. (1992a). Characterisation of the physical demands of fire fighting. *Canadian Journal of Sport and Science*. 17(3):207-213.

Gledhill, N., and Jamnik, V.K. (1992b). Development and validation of a fitness screening protocol for firefighter applicants. *Canadian Journal of Sport and Science*. 17(3):199-206.

Groeller, H., Fullagar H.H.K., Sampson, J.A., and Taylor, N.A.S., (2014c). Screening tests for Australian firefighters. *Journal of Occupational and Environmental Medicine*. Submitted 2014.

Gumieniak, R., Jamnik, V.K., and Gledhill, N. (2011). Physical Fitness Bona Fide Occupational Requirements for Safety– Related Physically Demanding Occupations; Test Development Considerations. *Health and Fitness Journal of Canada*. 4(2):47-52.

Henderson, N.D., Berry, M.W., and Matic, T. (2007). Field measures of strength and fitness predict firefighter performance on physically demanding tasks. *Personnel Psychology*. 60:431-473.

Henry, F.M., and DeMoor, J.C. (1956). Lactic and alactic oxygen consumption in moderate exercise of graded intensity. *Journal of Applied Physiology*. 8:608-614.

Jamnik, V.K., Thomas, S.G., Shaw, J.A., and Gledhill, N. (2010a). Identification and characterization of the critical physically demanding tasks encountered by correctional officers. *Applied Physiology, Nutrition and Metabolism*. 35:45–58.

Jamnik, V.K., Thomas, S.G., Burr, J.F., and Gledhill, N. (2010b).Construction, validation, and derivation of performance standards for a fitness test for correctional officer applicants. *Applied Physiology, Nutrition and Metabolism*. 35:59-70.

Jamnik, V.K., Thomas, S.G., and Gledhill, N. (2010c). Applying the Meiorin Decision requirements to the fitness test for correctional officer applicants; examining adverse impact and accommodation. *Applied Physiology, Nutrition and Metabolism*. 35:71–81.

Jamnik, V.K., Gumienak, R., and Gledhill, N. (2012). Physiological employment standards II: Developing and implementing physical employment standards for safety-related occupations. *European Journal of Applied Physiology*. *In Press*.

Kleim, J.A., Lussnig E., Schwarz, E.R., Comery, TA, and Greenough, W.T. (1996). Synaptogenesis and Fos expression in the motor cortex of the adult rat after motor skill learning. *Journal of Neuroscience*. 16: 4529–4535.

Koegelenberg C.F.N., Diacon A.H., Irani S., Bolliger C.T. (2008). Stair climbing in the functional assessment of lung resection candidates. *Respiration*.75:374-379.

Kraemer, W.J., Mazzetti, S.C., Nindl, B.C., Gotshalk, L.A., Volek, J.S., Bush, J.A., Marx, J.O., Dohi, K., Gomez, A.L., Miles, M., Fleck, S.J., Newton, R.U., and Hakkinen. (2001). Effect of resistance training on women's strength/power and occupational performances. *Medicine and Science in Sports and Exercise.* 33:1011-1025.

Lemon, P.W., and Hermiston, R.T. (1977). The human energy cost of fire fighting. *Journal of Occupational and Environmental Medicine.* 19:558-562.

Ljungberg, A.N., Kilbom, Å., and Hagg, G.M. (1989). Occupational lifting by nursing aides and warehouse workers. *Ergonomics*. 32:59-78.

Marcinik, E.J. (1986). Sprain and strain injuries in the Navy: the possible role of physical fitness in their prevention. *Aviation, Space, and Environmental Medicine*. 57:800-804.

Mendez-Villanueva, A., Bishop, D., and Hamer, P. (2007). Reproducibility of a 6-s maximal cycling sprint test. *Journal of Science and Medicine in Sport*. 10:323-326.

Michaelides, M.A., Parpa, K.M., Thompson, J., and Brown, B. (2008). Predicting performance on a firefighter's ability test from fitness parameters. *Research Quarterly for Exercise and Sport.* 79(4):468-475.

Mital, A. (1984). Maximum weights of lift acceptable to male and female industrial workers for extended work shifts. *Ergonomics.* 27:1115-1126.

Nudo, R.J., and Milliken, G.W. (1996). Reorganization of movement representations in primary motor cortex following focal ischemic infarcts in adult squirrel monkeys. *Journal of Neurophysiology.* 75: 2144–2149.

Payne, W., and Harvey, J. (2010). A framework for the design and development of physical employment tests and standards. *Ergonomics*. 53:858-871.

Ploutz-Snyder, L.L., and Giamis, E. (2001). Orientation and familiarization to 1RM strength testing in old and young women. *Journal of Strength and Conditioning Research*. 15:519-523.

Rayson, M, P. (2004). Operational physiological capabilities of firefighters: literature review and research recommendations. *Optimal Performance Ltd. on behalf of the Office of the Deputy Prime Minister.*

Reilly, T., Wooler., and Tipton, M. (2006a). Occupational fitness standards for beach lifeguards. Phase 1: the physiological demands of beach lifeguarding. *Occupational Medicine.* 56:6–11.

Reilly, T., Iggleden, C., Gennser, M. and Tipton, M. (2006b). Occupational fitness standards for beach lifeguards. Phase 2: the development of an easily administered fitness test.

Occupational Medicine. 56:12-17.

Sharp, M.A., Harman., E.A., Boutillier, B.E., Bovee, M.W., and Kraemer, W.J. (1993a). Progressive resistance training pogram for improving manual materials handling performance. *Work*. 3:62-68.

Smith, D.L., Petruzello, S.J., Kramer., J.M., and Misner, J.E. (1996). Physiological, psycholphysical and psychological responses of firefighters to fire fighting training drills. *Aviation, Space and Environmental Medicine*. 67:1063-1068.

Smith, D. L., Manning, T.S., and Petruzzello, S.J. (2001). Effect of strenuous live-fire drills on cardiovascular and psychological responses of recruit firefighters. *Ergonomics*. 44(3):244-254.

Supreme Court of Canada. (1999). British Columbia (Public Service Employee Relations Commission v. BCGSEU). 3 S.C.R. 3. (Meiroin decision).

Taylor, N.A.S., Fullagar, H.H.K., Sampson, J.A., and Groeller, H. (2014a). The essential, physically demanding tasks of fire fighting in Australia. *Journal of Occupational and Environmental Medicine*. Submitted 2014.

Taylor, N.A.S., Fullagar, H.H.K., Sampson, J.A., and Groeller, H. (2014b). Physiological demands of the essential fire-fighting tasks performed by Australian firefighters. *Journal of Occupational and Environmental Medicine*. Submitted 2014.

Taylor, N.A.S., and Groeller, H. (2003). Work-based assessments of physically-demanding jobs: a methodological overview. *Journal of Physiological Anthropology*. 22:73-81.

Tipton, M.J., Milligan, G.S., and Reilly, T.J. (2012). Physiological employment standards I: Occupational fitness standards: objectively subjective? *European Journal of Applied Physiology*. *In Press*.

VandenBerg, P.M., Hogg, T.M., Kleim, J.A., and Whishaw, I.Q. (2002). Long-Evans rats have a larger cortical topographic representation of movement than Fischer-344 rats: a microstimulation study of motor cortex in naive and skilled reaching-trained rats. *Brain Research Bulletin*. 59:197–203.

Warwick, D., Novak, G., Schultz, A., and Berkson, M. (1980). Maximum voluntary strengths of male adults in some lifting, pushing and pulling activities. *Ergonomics*. 23:49-54.

Whipp, B.J., and Wasserman, K. (1972). Oxygen uptake kinetics for various intensities of constant-load work. *Journal of Applied Physiology*. 33:351-356.

Williford, H.N., Duey, W.J., Olsonc, M.S., Howard, R., and Wang, N. (1999). Relationship between fire fighting suppression tasks and physical fitness. *Ergonomics*. 42(9):1179-1186.

TABLES

Table 1: Criterion task movement classifications (Groeller *et al*., 2014).

Criterion task class	Class description	Criterion tasks
1	Single-sided carrying tasks	Hazmat task Stair climb with ventilation fan
2	Overhead push and hold tasks	Motor-vehicle rescue
		Ladder raise (10.5 m)
		Using a sledge axe to gain entry
3	Cardiorespiratory dragging tasks	Fire attack
		Dragging charged hose (38 mm)
4	Critical strength task	Firefighter down (rescue)

Table 2: A description of all tasks simulated during Step 14.

Task	Description
Hazmat task	A 26-kg, liquid-filled plastic container with a mass equivalent to 50% of the heaviest object moved in a two-man hazmat simulation, was carried unilaterally over a 32-m synthetic grass track. The activity was performed as a series of "out and back" walks, with each walk being comprised of a 32-m walk out, followed by a 180^{o} turn at a fixed marker, and a 32-m return trip. This constituted one lap (64 m), with participants alternating between unloaded and loaded laps for a total of 16 laps.
Hazmat incident replication	This activity was a repetition of the above simulation, but with participants walking at three different absolute speeds, and for a 5-min duration at each speed. Two of these speeds were the same for all participants, with the first (slow speed) set at 30% of the average maximal walking speed observed in the above simulation, and the second (medium speed) set at 60% of the average maximal walking speed.
Stair climb with ventilation fan (loaded stair climb)	Participants carried a 17.5-kg mass (liquid-filled plastic container: 50% of the mass of the ventilation fan) in one hand, and completed the task walking backward. The simulation commenced with a 7.3-m walk on level ground to the base of the stairs, and continued with the ascent of 64 steps (4 storeys).
Loaded box stepping (three stepping rates)	Participants performed three loaded, 5-min box-stepping trials (17.5-kg mass). Stepping involved a continuous step cycle onto, and down from a stable box (26 cm), with the stepping cadence set via an electrical metronome. Three stepping cadences were used: one represented each individual's maximal stepping rate observed during loaded stair climb and the other two stepping rates (30% and 60% of the average maximal stepping rate observed in the loaded stair climb) were set at the same absolute cadences for all participants.
Motor vehicle rescue (Hydraulic shears static-hold simulation)	A series of intermittent holds and rests, using a replica hydraulic tool (19.5 kg). The tool was held in three positions (1 min) relative to the stature of each participant: below knee height, between knee and hip heights, and between the shoulder and eye level. The position sequence was randomised across subjects and performed twice. The duration of the static hold was varied: (a) 20-s hold and 40-s rest, (b) 30-s hold and 30-s rest, and (c) 40-s hold and 20-s rest. The resting phase involved placing the load onto a bench at the same height, and resting briefly.
Ladder-raise	A single-person ladder raise (10.5 m long, 49.5 kg) from the ground to the vertical position (resting against an external wall), immediately followed by lowering the ladder to the ground.
Box lift and place - maximal load	The task began with a gender-specific mass: 10 kg (females) and 20 kg (males). An additional mass, selected by the participant within a 2.5-10 kg range, was added following each successful lift. A 2-min rest was provided between successive lifts, and this cycle continued until the participant could no longer safely lift and place the box.
Box lift and place - endurance	Box mass was set at 75% of the mass each person lifted in the above simulation. The subject returned the box to the ground after each successful lift. How many times each participant could successfully and safely lift and place this box was determined.

Table 2: *continued*

1RM maximum shoulder press	Following a warm-up with the bar plus 10 kg, the maximal mass a participant was able to correctly lift once, and through the desired range of motion.
Dragging charged hose (38 mm)	The participant held the branch of a 5.6-m section of 38-mm hose filled with sand (11 kg) and doubled over, giving a 2.8-m length of hose and then walked away 30m from (pulling against) a resistance loader that had been set to provide the resistive force. On reaching a marker, participants passed the hose branch to the researcher and then walked back to the resistance device: the unloaded or light work stage. On reaching the loader, the researcher handed the hose branch back to the subject, who then performed the next hose-drag simulation. This two-stage sequence was repeated until 6 min elapsed.
Hose-drag replication (two walking speeds)	This activity was a repetition of the above simulation, but with participants walking at two different absolute speeds, and for a 6-min duration at each speed. These speeds were the same for all subjects, with the first (slow speed) set at 30% of the average maximal walking speed observed in the above simulation, and the second (medium speed) set at 60% of the average maximal walking speed.

Table 3: A summary of time, respiratory and cardiovascular strain of each part of the Physical Aptitude Test (PAT) completed by operational firefighters (Step 16)

Task	**Mean time (range)**	**Mean oxygen consumption (95% probability range; $L.min^{-1}$) ***	**Mean heart rate (95% probability range; $beats.min^{-1}$) ***
Hazmat	2 min 14 s (1 min 12 s - 3 min 55 s)	2.25 (1.77-2.73)	148 (142-154)
Loaded stepping	1 min 20 s (0 min 18 s to 3 min 36 s)	2.73(2.43-3.03)	162 (156-168)
Static hold	Fixed duration (3 min)	1.50 (1.21-1.69)	154 (148-160)
Cardiorespiratory hose drag task	3 min 57 s (1 min 36 s to 8 min 3 s)	2.90 (2.52-3.28)	169 (164-174)
Fire attack	49 s (44 s to 54 s)	2.69 (2.26-3.12)	175 (169-181).
Firefighter rescue	15 s (6 s to 35 s)	2.48 (2.05-2.91)	173 (168-179)
Overall physical aptitude test	13 min 31 s (9 min 53 s to 25 min 27 s)	2.36 (2.07-2.65)	161 (156-166)

Notes: * = A sample of 18 firefighters within the overall sample of 148 were used to collect the physiological measurements (two right columns).

Table 4: The recommended thresholds for the Physical Aptitude Test (PAT).

Task	The recommended thresholds for the Physical Aptitude Test
1. Hazmat	3 min*
2. Loaded stepping	3 min*
3. Static hold	3 min*
4. Cardiorespiratory hose drag task	5 min*
Total time for items 1-4	15 min (screening threshold)#
Total time for items 5-6 (Fire attack and firefighter rescue)	1 min 53 sec (screening threshold)#

Notes: * = These times define purely a recommendation in order to finish items 1-4 in the 15 min period. The single 15 min performance standard is the overall pass threshold for items 1-4. # = It is further recommended that individuals completing the test in a time frame that falls within the 5% buffer zones for either of the test-item groups (i.e. items 1-4 or 5-6) should be identified as candidates for whom retesting might be appropriate.

Table 5: The recommended failure criteria for the Physical Aptitude Test (PAT). Failure on any one of these criteria should represent test failure, independently of the performance on any of the other test components

Four way failure criteria for the Physical Aptitude Test for firefighters
Failure to complete the three positional holds for the required duration: such individuals could be allowed to continue the test, but should still be required to pass this component as part of the complete test sequence
Failure to satisfy the time threshold for test items one-four, and with a score outside the buffer zone
Failure to satisfy the time threshold for test items five-six, and with a score outside the buffer zone
Failure to satisfy the time threshold for test items five-six, and with a score outside the buffer zone

FIGURE LEGENDS

Figure 1: Proposed physical aptitude test (circuit) for assessing criterion firefighter tasks. It is recommended this test series be performed in the following order: single-sided (unilateral) carriage task, holding task, hose-

Figure 2: The visual-analogue scale used for firefighter evaluations of the face validity of the Physical Aptitude Test. This was with participants instructed to mark a vertical line through a 100-mm horizontal line, the left end of which indicated "very poor" whilst the opposite end was marked "very good". This scale provided subjective ratings that could then be assigned a score from 1 ("very poor") to 100 ("very good"), by measuring the distance (mm) from the left-hand end to the point at which the vertical line was marked.

Figure 3: The correlation between the time to complete the 16-lap hazmat incident simulation at maximal walking speed, and the distance covered in a maximal, 5-min simulation of the same task. Individual scores are presented with linear modelling and 95% CI (N=12).

Figure 4:. Predicting the walking speed required to elicit an oxygen cost of 1.47 $L.min^{-1}$ during a 5-min hazmat incident simulation. Individual scores are presented (N=12) for the three stepping rates (fast = purple, medium [60%] = yellow, slow [30%] = green), with linear modelling (solid line) and 95% CI (dashed lines) applied to these data.

Figure 5: The correlation between performance time on the loaded stair climb task and the oxygen cost of a box-stepping activity (5 min) performed at an equivalent speed. Individual scores are presented with linear modelling and 95% CI (N=14).

Figure 6: Oxygen consumption and heart rate data during a simulated static hold of the hydraulic shears at each of three relative heights: below the knees (low), between the knees and shoulders (middle) and at eye level (high). Data are means with standards errors of the means (N=14).

Figure 7: Oxygen consumption and heart rate data during a hose-drag simulation performed at maximal walking speed. Data are means with standard errors of the means (N=14).

Figure 8: Raw oxygen consumption data of subjects performing the Physical Aptitude Test (N=14).

Figure 9: Predicting the oxygen cost of dragging a charged fire hose from distance covered. Individual scores (N=13) are presented (0.88 $m.s^{-1}$ = green, 0.93 $m.s^{-1}$ = yellow, maximal speeds = purple) with linear modelling (solid lines) and 95% CI (dashed lines).

FIGURE 1

.

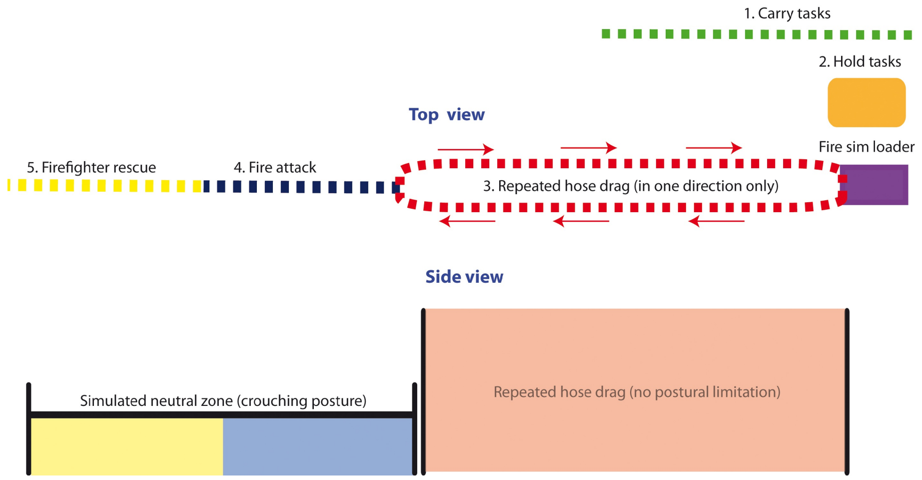

1. Carry tasks
2. Hold tasks
Top view
Fire sim loader
5. Firefighter rescue
4. Fire attack
3. Repeated hose drag (in one direction only)
Side view
Simulated neutral zone (crouching posture)
Repeated hose drag (no postural limitation)

FIGURE 2

Visual analogue scale:

How well do you think this Physical Aptitude Test replicated the demands of fire fighting?

Please place a vertical line at some point along the horizontal line shown below, and between the two ends points ("very poor" and "very good").

Very poor   Very good

FIGURE 3

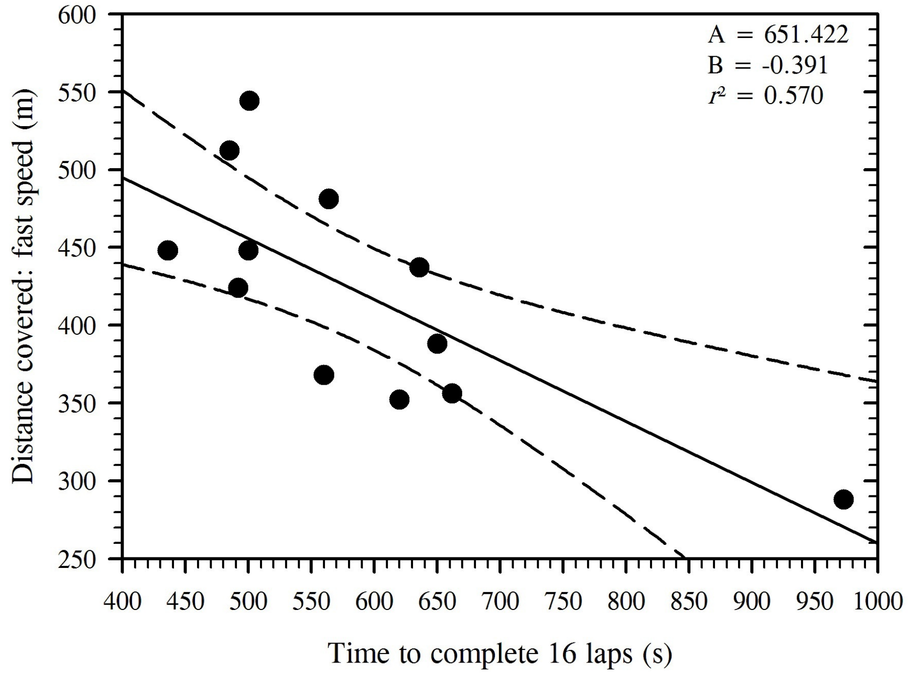

A = 651.422
B = -0.391
r^2 = 0.570
Distance covered: fast speed (m)
600
550
500
450
400
350
300
250
400
450
500
550
600
650
700
750
800
850
900
950
1000
Time to complete 16 laps (s)

FIGURE 4

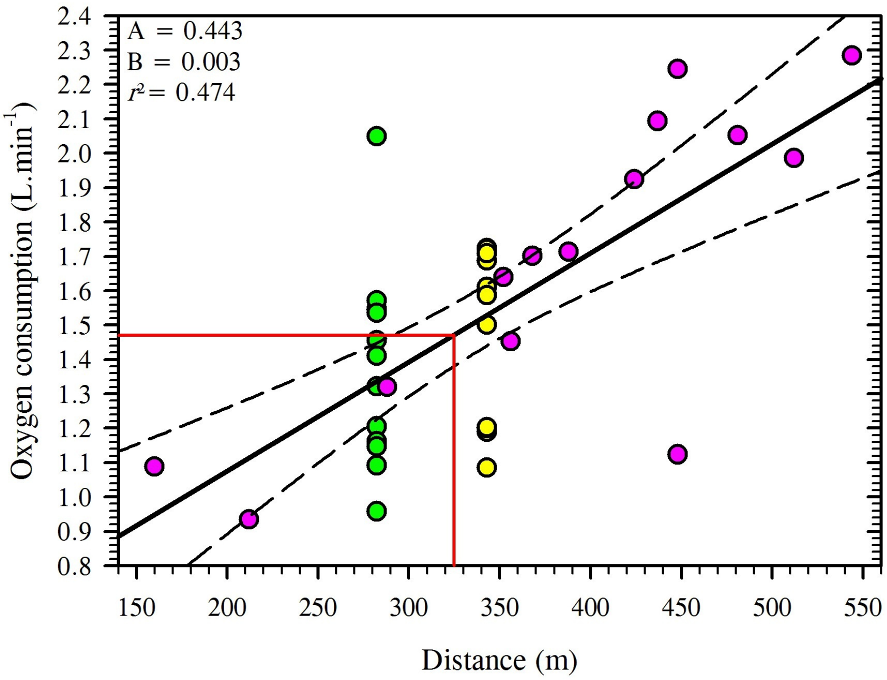
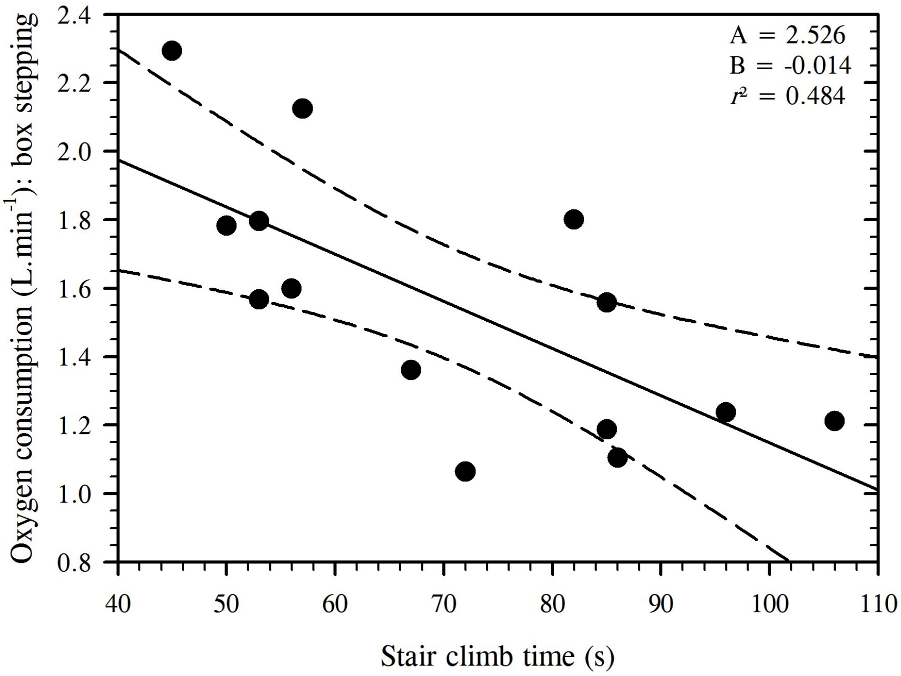
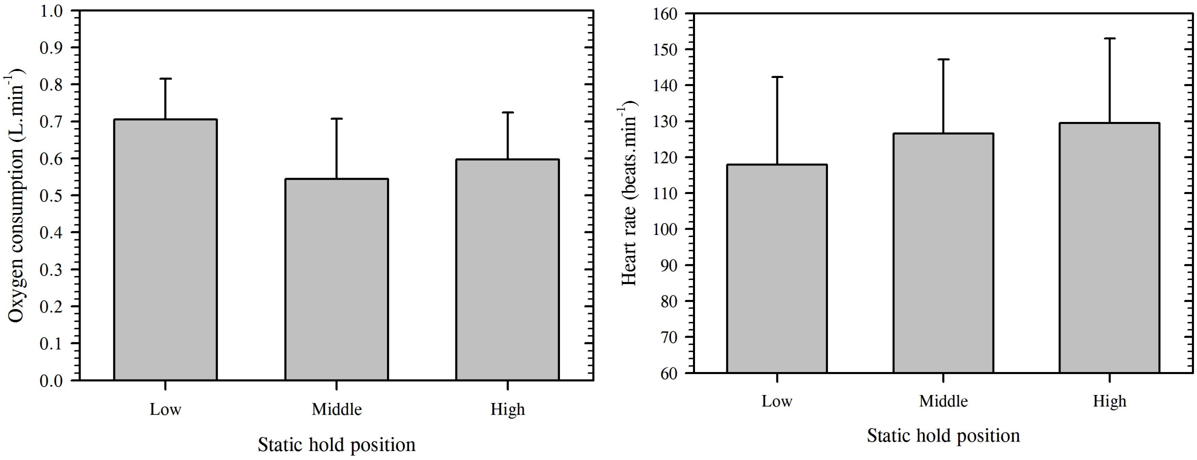

A = 0.443
B = 0.003
r² = 0.474
Oxygen consumption (L.min-1)
Distance (m)
2.4
2.3
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
1.3
1.2
1.1
1.0
0.9
0.8
150
200
250
300
350
400
450
500
550

FIGURE 5



A = 2.526
B = -0.014
r² = 0.484
Oxygen consumption (L.min-1): box stepping
Stair climb time (s)
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
40
50
60
70
80
90
100
110

FIGURE 6



Oxygen consumption (L.min-1)
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
Low
Middle
High
Static hold position
Heart rate (beats.min-1)
160
150
140
130
120
110
100
90
80
70
60
Low
Middle
High
Static hold position

FIGURE 7

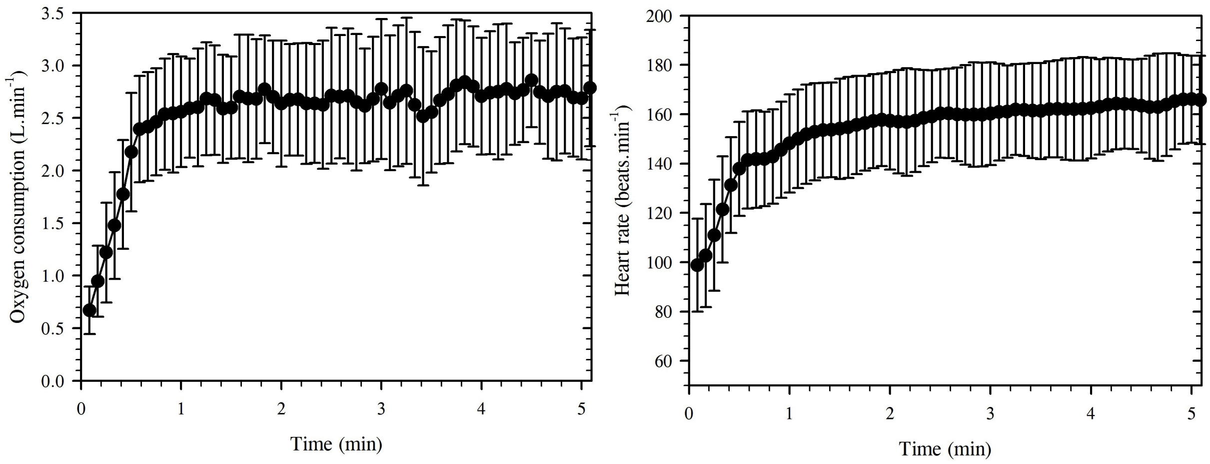
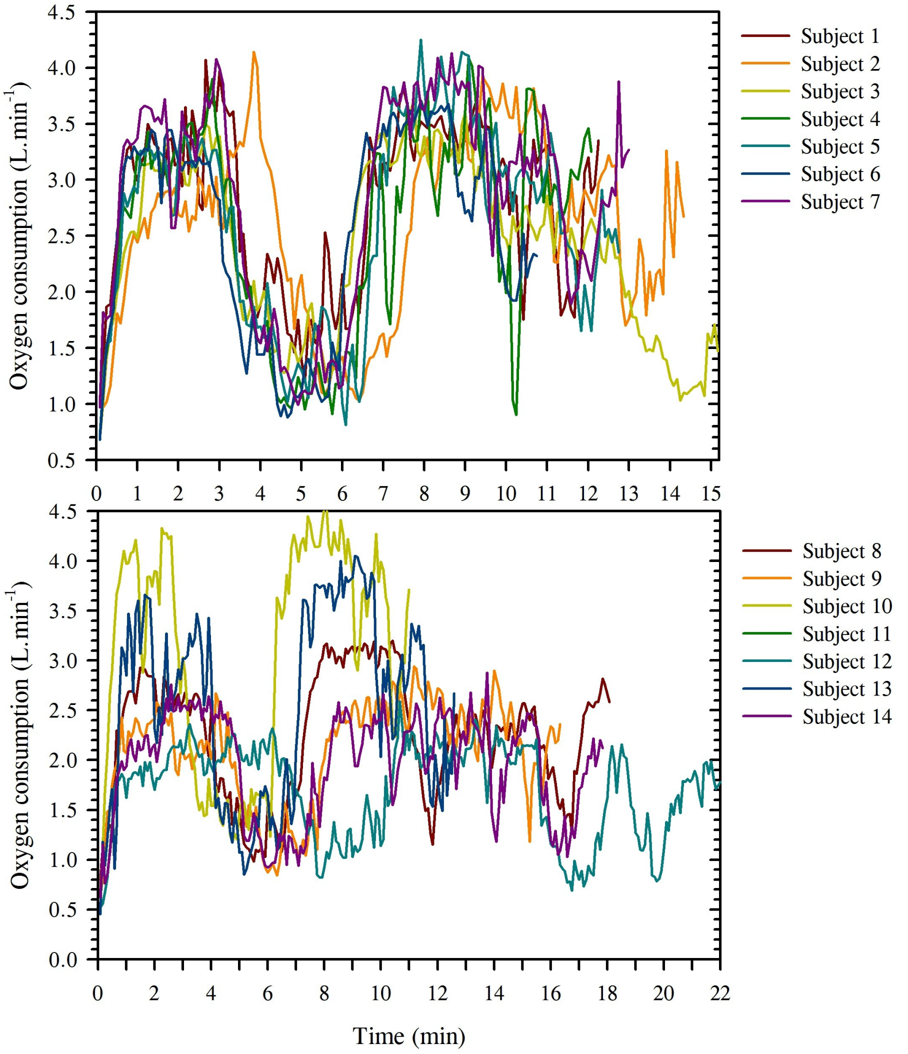
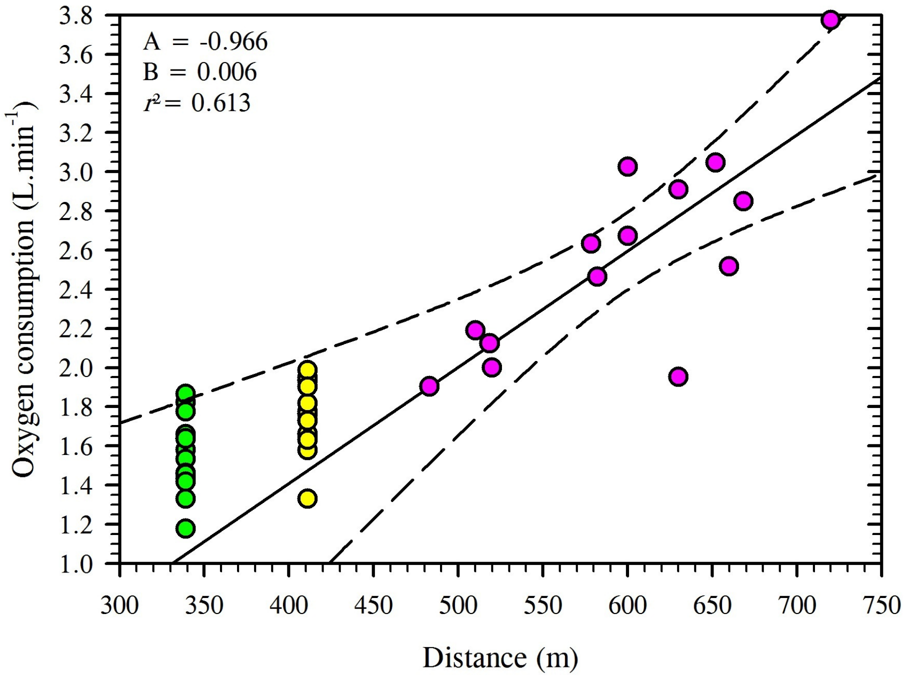

Oxygen consumption (L.min-1)
Time (min)
Heart rate (beats.min-1)
Time (min)

FIGURE 8


Oxygen consumption (L.min-1)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Subject 1
Subject 2
Subject 3
Subject 4
Subject 5
Subject 6
Subject 7
Oxygen consumption (L.min-1)
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
0 2 4 6 8 10 12 14 16 18 20 22
Time (min)
Subject 8
Subject 9
Subject 10
Subject 11
Subject 12
Subject 13
Subject 14

.

FIGURE 9


A = -0.966
B = 0.006
r^2 = 0.613
Oxygen consumption (L.min^{-1})
Distance (m)
3.8
3.6
3.4
3.2
3.0
2.8
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
300
350
400
450
500
550
600
650
700
750