Occupational health care in high-demand jobs: the usefulness of job-specific workers' health surveillance for fire fighters

Plat, M.C.J.

Citation for published version (APA):
Chapter 4

Reproducibility and validity of the stair-climb test for fire fighters

MJ Plat, MHW Frings-Dresen, JK Sluiter
International Archives of Occupational and Environmental Health 2010; 83: 725-731
Abstract

Purpose A new job-specific test for fire fighters, the stair-climb test (FFstair-climb) was evaluated for reproducibility and validity for use in future workers’ health surveillance.

Methods The FFstair-climb was performed at three times by 20 male fire fighters (one and three weeks in between test and retest) to examine the reliability and agreement. An intraclass correlation coefficient (ICC) and standard error of measurement (SEM) were calculated. In addition, 45 fire fighters (2 female, 43 male) performed the test once while heart rate and testing time were monitored, to study the content validity. Construct validity was tested by correlating both self-rated and supervisor-rated work ability scores with testing time.

Results The reliability, reflected by the ICC, was high for the one week and the three week test-retest period, 0.82 and 0.91 respectively. The mean testing time for the FFstair-climb was about 65 s; the agreement showed SEMs of 5.4 s at the one week test-retest period and 3.8 s at the three week test-retest period. Content validity was good in 78% of the fire fighters, while the construct validity between testing times and work ability ratings was not sufficient (r < 0.30).

Conclusions The FFstair-climb can be used in the workers’ health surveillance of fire fighters as a reproducible instrument, with good content validity. Testing criterion validity in future research would be an appropriate next step.
Introduction
Fire fighters perform several physically and mentally demanding tasks. Physically, fire fighters are required to pull and drag hoses, as well as climbing ladders, as part of their job description.\textsuperscript{1-5} One of the physical demands of fire-fighting tasks is a high energetic peak load.\textsuperscript{6} As a result energetic peak load is a specific job demand. Specific job demands are exposures on a job that cannot be prevented by current state-of-the-art measures or may reveal safety risks at the workplace.\textsuperscript{7} The worker’s health and safety and, in some occupations like ambulance personnel and fire fighters, the health of the public may be in danger if the worker is not able to fulfill the specific demands of the job. Energetic peak load is a specific job demand. It is such a specific occupational health requirement during walking up and down stairs in a fast way, and is not preventable due to the nature of the job.\textsuperscript{5} Working under time constraints in order to increase rescue chances, i.e. having to respond quickly and doing heavy physical work in short periods of time is inherent to the job as fire fighter. Since energetic peak load is a specific occupational health requirement for fire fighters, it should be included as a test during a fire fighters’ workers’ health surveillance. When testing energetic peak load in a workers’ health surveillance, it should be evaluated in a job-specific scenario such as evidenced by testing the balance of male fire fighters,\textsuperscript{8} the use of the ambulance stair-climb test for Dutch ambulance workers,\textsuperscript{9} and a task-based fitness test in beach lifeguards.\textsuperscript{10}

For fire fighters, several job-specific tests have been previously studied and described. In Finland a submaximal job-related test drill for fire fighters containing five tasks was tested.\textsuperscript{11} Besides, Deakin and colleagues designed a job-specific circuit for fire fighters\textsuperscript{12} and this test was later adapted for evaluating Dutch fire fighters.\textsuperscript{13} The fire fighter circuit was shown to be reliable test at a three week period.\textsuperscript{14} The three week period was the time between the performances of the tests compared and found reliable. Although, these fire-fighting simulation tests mainly focus on evaluating physical exertion for longer time periods (approximately 15 minutes), fire-fighting is also exposed to energetic peak workloads, defined as an exercise or effort to be accomplished within a period of three minutes.\textsuperscript{15} However, studies on short duration protocols for fire-fighting are not yet reported.

Fire fighters are often required to climb stairs and perform all their activities while carrying full equipment. As became clear from workplace observations in fire fighters in the Netherlands, walking stairs is done for a mean time of 1
minute and 15 seconds, while lifting/carrying is done for nearly 9 minutes during a 24-h shift. Furthermore, from a simulated fire-fighting exercise test, it became clear that climbing was one of the most demanding tasks. Therefore, Sluiter and Frings-Dresen decided to test the energetic peak load while climbing stairs as a part of the job-specific test. A stair-climb test with good reliability (ICC 0.80) was previously developed in Singapore. Sluiter and Frings-Dresen adapted the stair-climb test from Singapore to a fire-fighting-specific test, also known as the fire-fighting stair-climb test (FFstair-climb). FFstair-climb requires the fire fighter to climb stairs with turn out gear, self-contained breathing apparatus (together 21 kg) and in addition to lifting extra fire-fighting-related materials (20 kg). To use the FFstair-climb in practice during the Dutch workers' health surveillance for fire fighters, the reproducibility and validity needed to be tested in the Dutch fire-fighting population first. Therefore, the purpose of the present study was to determine the reproducibility and validity of the FFstair-climb. We hypothesized that validity would not become high, as a gold standard is lacking for testing energetic peak load in fire fighters.

**Methods**

**Participants**

In an a-priori power analysis, it turned out that 23 subjects for the reliability part of the study were needed (confidence level 0.95 and an ICC between 0.9 and 0.8). For the validity part we used as many as possible fire fighters from one department, from which supervisors could judge work ability. A sample of 23 fire fighters from three regional fire departments in the Netherlands were invited to execute the FFstair-climb three times to assess the reproducibility of this test. Those 23 fire fighters were a selection of a sample of 264 fire fighters, stratified on age and sex. Those fire fighters were those who performed the first test performance in the first week of a testing period of about two weeks in the different departments and who wanted to perform the test three times. However, only 21 fire fighters performed the test three times. Participants for this part were all male, with a mean age of 35 years old (SD 9; range 21-52), a mean body weight of 86 kg (SD 10; 74-112) and a mean height of 182 cm (SD 5; 170-189). Three participants were volunteer fire fighters and the other 17 were professional. All participants executed operational tasks.
To address validity, all 45 professional fire fighters (43 men and 2 women) from one of the three departments performed the FFstair-climb one time. The 45 participants of the validity part of the study were all professional fire fighters. The 45 participants had a mean age of 38 years old (SD 9; range 24-54), a mean body weight of 87 kg (10 SD; range 67-112), and a mean height of 182 cm (SD 6; range 172-198). All participants executed operational tasks. Seven of the fire fighters who performed the reproducibility part of the study are included as well in the validity part of the study.

The study was performed in accordance with the Declaration of Helsinki and the ethics committee of the Academic Medical Center approved this study. All participants signed written informed consent before participating in this study.

**FFstair-climb**

The Dutch law requires that a building with a floor higher than 20 m should have a fire fighter elevator (Bouwbesluit 2003/ Construction order 2003). Therefore, real-life fire-fighting tasks include climbing stairs up to 20 m. Consequently, the FFstair-climb rises about 20 m and contains between 108, 110 and 117 steps of 17-19 cm step height (differs per fire department). During the FFstair-climb, fire fighters climbed the stairs with their own turn out gear (including pants, blouse, protective clothing, helmet, boots and gloves) and the self-contained breathing apparatus is worn and connected (weighing 21 kg together). Fire fighters also carry fire-fighting-related materials, such as two rolled fire hoses (one in each hand), on their way up (weighing 20 kg in total). So fire fighters took added to their own body weight a total of 41 kg with them, including the turn out gear and fire-fighting-related material. A sports instructor, specialized in training fire fighters, gave the standardized test instructions before the start. Participants climbed the stairs as quickly as possible, without running, at a constant pace, without stopping and without holding the side-railings. Participants took only one step at a time and had to touch each step with one foot. Testing time (in seconds) was recorded by the sports instructor, who climbed along with the participant throughout the test, starting at the moment the fire fighter started to climb the first step and finishing at the moment the fire fighter touched the last step with both feet. Heart rate was measured using a heart rate monitor (Polar RS800, Finland) strapped around the chest and recorded at the moment the participant completed the test.
Reproducibility and validity

Reproducibility, consisting of reliability and agreement,\textsuperscript{18} was tested with a within-subjects design. Fire fighters were tested three times, at baseline, after one week (to examine one week test-retest reproducibility) and three weeks after having performed the test at one week (to examine three week test-retest reproducibility).

The content validity of the FFstair-climb, testing energetic peak load in fire fighters, was evaluated in this study by measuring the heart rate of each fire fighter at the moment as previously described, and by recording the testing time. The attained heart rate at the end of the FFstair-climb was calculated as a percentage of the participants’ estimated theoretically maximum heart rate (220 minus age (in years)).\textsuperscript{19} The heart rate should reach at least 85% of their estimated theoretical maximum heart rate (HRmax), to be valid as a condition for maximal exercise.\textsuperscript{20} To be considered as an energetic peak workload, the test should be completed within three minutes.

Convergent validity was studied, as part of construct validity, by relating the outcome (testing time in seconds) of the FFstair-climb with a composite reference standard of related variables.\textsuperscript{21} Due to the fact that there is no gold standard for testing the energetic peak load in a job-specific way, we searched for an alternative. Rutjes et al. proposed the use of several related variables to test the validity if a gold standard is not available.\textsuperscript{21} In the present study, related variables were found in the Work Ability Index,\textsuperscript{22} therefore following three variables were collected:

a) The fire fighter filled in two items of the Work Ability Index adjusted for fire-fighting work. Fire fighters scored their own current work ability for executing fire-fighting tasks in relation to the physical demands. Responses were given on a 5-point scale varying from very poor (1) to very good (5).

b) Fire fighters scored their own current general work ability on an 11-point scale, in which ‘0’ was completely unable to work and ‘10’ was work ability at its best to execute fire-fighting tasks (the first question of the Work Ability Index). These data were collected before the first FFstair-climb.

c) The judgment of the direct supervisor on the current work ability of the fire fighter was requested, similar to the second variable described above, with a score between 0 and 10. Two direct fire fighter supervisors from one fire department co-judged the 45 fire fighters. Together, they gave one mark for each participant. The judgments of the direct supervisors were collected after the first test. The supervisors did not see the performance or the results of the
test, but judged the work ability from the fire fighter as they know it from real-life fire-fighting activities.

For the composite reference standard, both self-rated (b) and supervisor-rated work ability (c) were summed. The sum for this composite reference test ranged between 0 and 20.

**Statistical analysis**

All analysis were performed with SPSS 16.0. Means, standard deviations and ranges were calculated for the testing time. First, the test results were visually checked for outliers and thereafter tested for changing scores higher than 2 SD. Reliability and agreement were calculated by using the testing time (in seconds).

To determine the reliability, the intraclass correlation coefficient (ICC) was calculated\(^1\) by using the ICC model 2.1 A, a random two-way analysis, according to Shrout and Fleiss.\(^2\) The single measure ICC between baseline and after the one week measurement was calculated, and the measurement between the one week and the three weeks afterwards as well. An ICC with a value of less than 0.70 was considered as low reliability and greater than or equal to 0.70 as high reliability between measured variables.\(^3\) The standard error of measurement (SEM), which reflects the degree of agreement, was calculated by means of components of variance using the following equation: \(\text{SEM} = \sqrt{\sigma^2 \text{testing moments} + \sigma^2 \text{error}}\).\(^1\)

For the end heart rates of all participants, the mean, standard deviation and ranges were calculated. Relative frequencies (%) were calculated: how many participants' reached the criterion as having at least an end heart rate of 85% of their HRmax to prove content validity.

For the work ability outcome, median and interquartile scores were calculated. The convergent validity was determined with a Spearman's rank correlation coefficient between the outcome of the test (testing time) and the work ability judgment. The correlation coefficients were calculated between the FFstair-climb testing time and the self-rated judgment on the work ability with respect to the physical demands of fire-fighting, the self-rated judgment on the work ability of fire fighters, the supervisor-rated judgment on the work ability of the participant, and the self-rated and supervisor-rated work ability of the fire fighter added together. Convergent validity correlation was considered to be low if correlation coefficient was less than 0.30, moderate if correlation coefficient was between 0.30 and 0.60 and good if correlation coefficient was greater than or equal to 0.60.\(^5\)
Results
One of the 21 participants of the reliability part of the study was determined as an outlier, because of inconsistent performance compared to all other participants: the change in this participant was > 2 SD compared to the mean change comparing the first and second performance. Therefore this outlier was excluded for the reproducibility part of the study and so, the results of 20 participants were used.

Reliability and agreement
As shown in Table 1, the mean testing time at the first performance of the FFstair-climb was 67 s. The mean testing time at the second and third performances was 63 s.

The reliability of the test-retest period of one week and three weeks were both high, with values of the ICC of 0.82 and 0.91, respectively (second and third column of Table 1, respectively).

<table>
<thead>
<tr>
<th>Testing periods compared</th>
<th>N</th>
<th>Variance persons</th>
<th>Variance testing periods</th>
<th>Variance error</th>
<th>SEM (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One week</td>
<td>20</td>
<td>128.7</td>
<td>4.6</td>
<td>24.5</td>
<td>5.4</td>
</tr>
<tr>
<td>Three weeks</td>
<td>20</td>
<td>143.9</td>
<td>0.0</td>
<td>14.1</td>
<td>3.8</td>
</tr>
</tbody>
</table>

The standard error of measurement (SEM) is shown in Table 2, reflecting the agreement between the testing moments within the subjects. The SEM was 5.4 s at the one week test-retest period (comparing the first with the second performance) and 3.8 s at the three week test-retest period (comparing the second and the third performance). These data indicate that if a fire fighter performs the FFstair-climb several times and the change in testing time is 3.8 seconds or smaller it can be considered as a measurement error.
Content validity
The mean value of the heart rate of the fire fighters immediately after performance of the FFstair-climb compared to the estimated theoretically maximum heart rate was 92% (SD 8; range 72-105). Out of the total of 45 participants, 35 fire fighters (78%) reached the criterion of the 85% HRmax, and were thus classified as being at energetic peak load. All participants performed the FFstair-climb within a testing time of 3 minutes (mean testing time 73 s; SD 12; range 44-93).

Convergent validity
The median value for self-rated work ability for 45 participants with respect to the physical demands of fire-fighting was 4 (interquartile score of 1.5). The median score for the self-rated general work ability was 8.0 (interquartile score of 1.8), while the median score for the work ability rated by the supervisors was 7.0 (interquartile score of 1.0). The median score for the summed score for self-rated and supervisor-rated work ability was 15.0 (interquartile score of 2.5).

Spearman’s rank correlation coefficient between testing time and work ability ratings were low, ranging from -0.18 to -0.28, and were not significant (Table 3). It should be noted that lower work ability rating was associated with longer testing time.

Table 3 Correlations (N=45) between FFstair-climb testing time and different scores on work ability

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>FFstair-climb testing time Spearman’s rho</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work ability self-rated with respect to physical demands</td>
<td>45</td>
<td>-0.28</td>
<td>0.07</td>
</tr>
<tr>
<td>Work ability self-rated</td>
<td>45</td>
<td>-0.18</td>
<td>0.24</td>
</tr>
<tr>
<td>Work ability supervisor-rated</td>
<td>45</td>
<td>-0.20</td>
<td>0.19</td>
</tr>
<tr>
<td>Work ability self-rated plus supervisors</td>
<td>45</td>
<td>-0.21</td>
<td>0.17</td>
</tr>
</tbody>
</table>

Discussion
This study evaluated the reproducibility and validity of the fire-fighting stair-climb test (FFstair-climb). First, in terms of reproducibility, reliability was high with ICC values of 0.82 after one week and 0.91 after three week test-retest periods. Second, the SEM values which reflect agreement were for the above mentioned testing intervals 5 and 4 seconds with a mean testing time of about 65 s. The content validity was good, as over 75% of the participants reached the criterion
of an energetic peak load (reaching 85% of the estimated maximum heart rate at the end of the test) and all participants performed the test within 3 minutes. The convergent validity between testing time and work ability rating was low, with correlations < 0.30. Overall, the present study supports using the FFstair-climb as a reproducible test in the workers’ health surveillance for fire fighters for testing the energetic peak load.

The reproducibility in the present study can be separated into two concepts. First, the high reliability found in this study argues in favor of using the FFstair-climb for discriminative purposes because the high reliability means that the test can distinguish participants’ despite the measurement error. Second, the agreement found in this study, as reflected by SEM of 5 and 4 seconds on a mean testing time of about 65 seconds, reflects the extent to which differences after repeated measurements in the same person can be expected theoretically. The agreement should be studied when an instrument is used for evaluative purposes. Therefore, when a difference within one participant is smaller than 5 seconds it can be considered a measurement error. For the FFstair-climb, the results imply that the test can be used for discriminative and evaluative purposes in the workers’ health surveillance for fire fighters.

Subjects from the reproducibility part of the test were applicants from a larger stratified sample, who performed the test in the first week of the test period and who wanted to perform the test three times. It may be possible that those fire fighters were more willing to participate, because of good physical habits, nevertheless it is unknown, because this was not registered. However, testing times of the fire fighters’ first performance is compared to their own second and third performance (within-subjects comparison), so in that view it did not influence the results. It may have influenced the variation in testing time, when those fire fighters were all better in physical performance as the whole cohort of fire fighters. Therefore, the variation in the testing time of the 20 participants was compared to that of 264 fire fighters in the Netherlands. The first performance is compared, because the larger cohort performed the test once. The testing times for the FFstair-climb in the population of 264 fire fighters was 40 - 242 seconds, while testing time ranged in the 20 fire fighters from this study from 44 - 84 seconds. So there was less variation in the 20 fire fighters as was demonstrated in the larger cohort. Therefore ICC might even rise somewhat in the whole cohort, as it is a ratio between measurement error and variation of outcomes. The 45 fire
fighters in the validity part of the present study, were all professional fire fighters from one fire department. So validity results can probably be generalized towards the professional fire fighters in the Netherlands.

To estimate the maximum heart rate, the following formula is often used: 220-age in the literature. However, there are proponents\textsuperscript{15,19} and opponents\textsuperscript{26,27} to the use of this formula. The formula provides an easy and inexpensive insight into the estimated theoretically HRmax of an individual. The estimated HRmax was used to control whether the end heart rate of the FFstair-climb reached at least 85% of the HRmax, as it should do if it is measuring an energetic peak load. Åstrand et al. described a range of about ± 10 beats·min\textsuperscript{-1} per individual around the HRmax obtained with the formula, 220-age.\textsuperscript{15} As used in this study and applied to a group of 45 participants, it can be expected that the range is averaged out in the group: some participants can attain a somewhat higher heart rate while others will attain a somewhat lower heart rate. As 35 participants out of 45 reached at least 85% of the HRmax at the end of the test, we believe that the method used demonstrates good content validity for the FFstair-climb.

The 10 fire fighters who did not reach 85% of the HRmax at the end of the FFstair-climb were post hoc compared to the group who did on (i) use of medication, (ii) testing time, (iii) work ability scores and (iv) age. It was found that (i) One fire fighter took medication for high blood pressure, which is known to suppress one's heart rate. (ii) No differences in the range of scores for testing time were found. (iii) The differences found in the work ability scores were just that these fire fighters did not score the lowest occurring values for the work ability with regard to their physical demands and the supervisor-rated work ability. (iv) A study from Tanaka et al. showed that the formula used (220-age) gives an overestimation of the HRmax of persons younger than 40 years old.\textsuperscript{28} However, no age effect was found between the two groups. Therefore, it can be concluded from the post hoc analysis that nine out of the ten fire fighters can not be considered as a special group within the 45 participants of the validity study based on these variables. A potential reason for these 9 participants not reaching a heart rate of 85% of the HRmax might be that they did not perform the testing protocol at their maximum level, because they did not know what kind of effort to expect when performing the FFstair-climb for the first time. Criteria to judge the test should be developed in the future. If heart rate percentage will be one of the criterion, it should be kept in mind that individuals being fitter do perhaps not reach their
HRmax due to the nature of the test, as running is not allowed. Therefore a more extensive criterion is proposed for the future based on the criterion 1) > 85% of the theoretically HRmax OR 2) performing the test within a fast time.

It is not ideal that a gold standard is missing for measuring job-specific energetic peak load, within fire fighters. Consequently, we made a composite reference standard as suggested by Rutjes et al.\textsuperscript{21} In spite of that it is stated as best in absence of a gold standard, the values found for the convergent validity between the testing time and the work ability ratings, ranging from -0.28 to -0.18, can be considered as low correlations. One of the reasons for not finding a moderate or high correlation can be the small ranges of values found in the work ability rating. The self-rated scores on work ability had a lowest value of six (which was given once) and a highest value of 10 (also given once) with an accumulation of scores at 7, 8 and 9. As a result there was low variance between the scores. Inherent to the calculations of correlation coefficients, analyses with low variances have a higher chance of revealing low correlations.\textsuperscript{29} On the other hand, when studying convergent validity correlations that are high are equally disliked as low correlations. The idea of convergent validity is that not the whole construct is covered by the other construct, but it is partly the same and therefore correlations should not be high. Another reason for finding low correlations in the present study can be that the work ability questions were too general to cover the same construct as the actual FFstair-climb. As can be seen in the results of the convergent validity (Table 3), the most specific question addressing the physical demands of fire-fighting gave the highest correlation with testing time in this study.

Since the FFstair-climb was developed to use in the workers' health surveillance for fire fighters, its predictive value with respect to future work ability is one of the aspects we recommend for future study.

For practical reasons the tests executed in our study were performed at three different locations and on stairs of three regional fire departments throughout the Netherlands, for the reproducibility part of the study. The different locations had small differences in step height. Spanjaard et al. showed that varying step heights required different joint moments.\textsuperscript{30} The differences in step height were minimal in our study, a maximum of 2 centimeter per step in our study, while 8.5 and 12.5 centimeter were added in the study from Spanjaard et al., and therefore differences in testing outcomes were not expected.\textsuperscript{30} In addition to the step height,
strategies in climbing have been tested as well. Aziz and Teh demonstrated a difference in heart rate whether the participant climbs one step at a time or covers two steps in a single stride. In our study, all participants were required to touch all the steps, as emphasized in the instructions. However, the aforementioned aspects in the discussion underline standardization of the test circumstances and instructions when using the FFstair-climb in future workers’ health surveillance for fire fighters.

The FFstair-climb is adapted from the stair-climb test of Teh and Aziz. As the test has been changed in amount of steps, lifting extra weight and used in a different population there is no relevance to compare the test-retest results.

Based on the results of the present study, we recommend the use of the FFstair-climb. This test should be used to measure energetic peak load in fire fighters along with other instruments in the workers’ health surveillance.

**Conclusion**

The results of the present study demonstrate good reliability of the FFstair-climb at one week and three week test-retest periods. Agreement results show the possibility of using the FFstair-climb as an evaluative instrument for fire fighters. The FFstair-climb can also be considered as a valid instrument for testing job-specific energetic peak load with good content validity, but low convergent validity. Overall, it can be concluded that the FFstair-climb can be used in the workers’ health surveillance for fire fighters. Testing of criterion validity would be an appropriate next step.
References

7. Sluiter JK, Frings-Dresen MHW. Pre-employment and Periodical Workers’ Health Surveillance in the fire-fighting sector. [In Dutch: Aanstellingskeuring en Periodiek Preventief Medisch Onderzoek (PPMO) voor de Brandweersector.] Amsterdam, Coronel Institute for Occupational and Environmental Health, Academic Medical Center, reportno. 06-03, 2006.
Reproducibility and validity fire-fighting stair-climb test