Journal of Occupational Rehabilitation (2019) 29:194–204
https://doi.org/10.1007/s10926-018-9778-6

Work Functioning Among Firefighters: A Comparison Between Self-Reported Limitations and Functional Task Performance

Joy C. MacDermid[1,2] · Kenneth Tang[3] · Kathryn E. Sinden[4] · Robert D'Amico[5]

Published online: 25 May 2018

Abstract

Purpose Performance-based and disease indicators have been widely studied in firefighters; self-reported work role limitations have not. The aim of this study was to describe the distributions and correlations of a generic self-reported Work Limitations Questionnaire (WLQ-26) and firefighting-specific task performance-based tests. *Methods* Active firefighters from the City of Hamilton Fire Services ($n=293$) were recruited. Participants completed the WLQ-26 to quantify on-the-job difficulties over five work domains: *work scheduling* (4 items), *output demands* (7 items), *physical demands* (8 items), *mental demands* (4 items), and *social demands* (3 items). A subset of participants ($n=149$) were also assessed on hose drag and stair climb with a high-rise pack performance-based tests. Descriptive statistics and correlations were used to compare item/subscale performance; and to describe the inter-relationships between tests. *Results* The mean WLQ-26 item scores (/5) ranged from 4.1 to 4.4 (median = 5 for all items); most firefighters (54.5–80.5%) selected "difficult none of the time" response option on all items. A substantial ceiling effect was observed across all five WLQ-26 subscales as 44.0–55.6% were in the highest category. Subscale means ranged from 61.8 (social demands) to 78.7 (output demands and physical demands). Internal consistency exceeded 0.90 on all subscales. For the hose drag task, the mean time-to-completion was 48.0 s (SD = 14.5; range 20.4–95.0). For the stair climb task, the mean time-to-completion was 76.7 s (SD = 37.2; range 21.0–218.0). There were no significant correlations between self-report work limitations and performance of firefighting tasks. *Conclusions* The WLQ-26 measured five domains, but had ceiling effects in firefighters. Performance-based testing showed wider score range, lacked ceiling effects and did not correlate to the WLQ-26. A firefighter-specific, self-report role functioning scale may be needed to identify compromised work role capabilities in firefighters.

Keywords Firefighters · Work limitations · Task performance

Introduction

Firefighting is demanding and hazardous work [1–3]. Firefighters perform a wide range of tasks including fire suppression, rescue operations, patient transportation, and medical first aid; often under time pressure and in unpredictable hazardous environments. Firefighters must transition from a sedentary position to a highly pressured emergency context as part of the community need for rapid response. Fire suppressions may require awkward postures, heavy loads, heat exposures and contaminated air. As emergency responders, the ability to maintain a high level of work functioning can be paramount to the safety and success of firefighting operations. Firefighters are at risk for developing musculoskeletal [4–6] and mental health problems (e.g., post-traumatic stress disorders) [4, 7–10]. The majority of firefighters are exposed to some form of critical event such as serious injury

✉ Joy C. MacDermid
jmacderm@uwo.ca

1 Physical Therapy and Surgery, Western University, London, ON, Canada

2 Clinical Research Lab, Hand and Upper Limb Centre, St. Joseph's Health Centre, London, ON, Canada

3 Clinical Research Unit, CHEO Research Institute, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, ON, Canada

4 School of Kinesiology, Lakehead University, Thunder Bay, ON, Canada

5 Hamilton Professional Fire Fighters Association, Hamilton, ON, Canada

or death of a colleague or child victim, or eminent danger to themselves within a 2-month period [11]. About half would report a musculoskeletal complaint in either the back, neck or extremities [6].

Firefighters' high injury rates and high job demands suggest the potential for gaps to emerge between work role function and work demands. Gaps between work role function and demands are identified as work limitations. The ability to quantify work limitations is important for early identification and risk mitigation. Measures of work limitations might be used to screen for individuals who might need work accommodations or health services. These measures might also be used to evaluate interventions or policies intended to improve work role functioning. Performance-based tests have been used extensively in the fire service for training, return-to-work assessments and research. While self-report measures are subjective, they have the natural advantage of feasibility (e.g., lower cost and administration burden) and may enable assessments of a broader range of work demands, including problems related to the social or mental demands of work. That is, self-report questionnaires may be able to assess constructs beyond those that are amenable to objective testing. For example, while physical demands of work might be easily amenable to simulation testing, contextual decision-making and interpersonal demands would be less easily simulated or measurable.

At present, the comparability between different methods of assessing work limitations in active firefighters remains unclear. To this end, we conducted a study with the following aims: (1) to assess the score distribution and internal consistency of the Work Limitations Questionnaire (WLQ-26), (2) to examine the distribution of performance on two functional work tasks with high relevance to firefighters, and (3) to examine the relationship between performance on the WLQ-26 and two functional firefighting tasks.

Methods

Study Design

Recruitment

Firefighters from the City of Hamilton fire service were e-mailed invitations to participate in the FIRE-WELL project by the firefighter co-author (RD). To be included in the FIRE-WELL project, the study participant must have been an active firefighter from the City of Hamilton fire service, must understand English, and had provided informed consent for study participation. A total of $n = 298$ were recruited at study baseline. Five firefighters who originally signed up declined to answer the surveys, therefore, an initial sample of $n = 293$ met our inclusion criteria and were eligible for the current analyses. The study protocol was approved by the Research Ethics Board at McMaster University, and all firefighters signed informed consent.

Data Collection

At baseline, all study participants were asked to complete a survey questionnaire by either paper or electronic format. This included the WLQ-26 and an array of socio-demographics, health-related, and job-related variables to help describe the participants. Key variables assessed include: age, sex, height, weight, years of service as a firefighter, professional ranking (five categories: recruit, firefighter, acting captain, captain, or chief), self-rated musculoskeletal health status (five categories: excellent to major problems) and a single rating of the extent to which bone and joint problems interfered with their work. In addition, a subset of participants ($n = 149$) were assessed using two performance-based tasks, described below.

Work Limitations Questionnaire (WLQ-26)

Work role functioning can be measured by self-report outcome measures such as the WLQ. The WLQ-26 is a close derivative of the original 25-item WLQ that was originally developed and tested in people with chronic health conditions [12]. Variants of the questionnaire have been reported in the literature. The WLQ-26 is similar to the original WLQ that was developed and tested in individuals with rheumatoid arthritis, chronic headache syndrome, or epilepsy [12]. It is grounded in the same conceptual framework and the items were drawn from the same item pool. The WLQ-26 differs in three respects from the WLQ-25: (1) it uses a 4-week recall period rather than a 2-week recall period; (2) it uses a single response set for all questions with "half the time" the middle category rather than "some of the time"; and (3) contains several additional items. No research has directly compared the WLQ-26 and WLQ, and the developers concede that because of different response values and the inclusion of additional items, there may be differences across the variants. However, 90% of the items overlap and four of the five response values are the same. Other variants of the scale have been used in the literature, including a 15-item version [13].

The WLQ-26 has been examined for content validity and validated in people with musculoskeletal disorders [14, 15]. The WLQ-26 was selected for a number of reasons: (1) it is applicable to a wide range of occupational sectors; (2) it assesses a wide range of on-the-job problems that are relevant to firefighters (five domains: time management, output, physical, mental, and social demands); (3) the WLQ-26 and close variants have shown sound psychometric properties in various clinical populations, including workers with

musculoskeletal [12, 16–19] and non-musculoskeletal health disorders [12, 20, 21]; (4) at 26 items, this tool has a reasonable balance of burden and comprehensiveness; and (5) the consistent response options and use of clear time categories are easier for participants than other versions where subscales responses vary. An expert and structured assessment of 21 measures of at-work productivity loss, ranging from single-item indicators to multidimensional scales, indicated that no superior at-work productivity measures was identified [22].

The WLQ-26 is designed to differentiate impairments over five work domains: *work scheduling* (WS, 4 items), which assesses difficulties with handling a job's time and scheduling demands; *output demands* (OD, 7 items), which assesses impact on productivity at work; *physical demands* (PD, 8 items), which examines difficulties when performing job tasks that involve bodily strength, movement, endurance, coordination, and flexibility; *mental demands* (MD, 4 items), which examines difficulties with cognitively-demanding tasks; and *social demands* (SD, 3 items), which assesses difficulties with on-the-job social interactions [12, 23]. For each WLQ-26 item, respondents indicate the proportion of time with difficulty handling parts of the job. All items are provided with the six response categories: "all of the time" (score = 1), "most of the time" (score = 2), "some of the time" (score = 3), "a little of the time" (score = 4), "none of the time" (score = 5), or "does not apply to my job" (score = missing, as per developers' recommendation). Skipped items are also scored as missing [24]. Subscale scores can be generated by taking the mean of subscale items, subtracting this value by one, and then multiplying it by 25 to get a score with a 0–100 range (100 = high work role functioning limitations, no work limitations).

Functional Work Tasks: Hose Drag and Stair Climb with a High-Rise Pack

We developed and administered two functional work tasks that simulated firefighters' job demands as part of the baseline assessment [25]: (1) hose drag, and (2) stair climbing while carrying a high-rise pack. In the hose drag task, the firefighters bent to floor level and lifted the nozzle (6.1 kg) of an uncharged firefighting hose (30 m). The firefighters were given standardized instructions to pull the uncharged fire hose to a pylon (18 m away from the starting line) as quickly as possible. At the pylon, the firefighters maneuvered the hose around the pylon at 90° and pulled the hose to an end marker (12 m from the pylon). The firefighters then repeated the task while returning the nozzle to the start/finish line. In the stair climb with a high-rise pack task, firefighters were asked to retrieve and lift a 22-kg high-rise hose pack from the floor to their shoulder. They then ascended and descended five flights of stairs while carrying the 22-kg high-rise hose pack over their shoulder. For both tasks, the time required to successfully complete the task was recorded by a research assistant.

Analysis

Descriptive Statistics

To describe the characteristics of the study participants, descriptive statistics (mean, %) were applied to socio-demographic and health-related variables.

Score Distribution for the WLQ-26

To examine the score distribution for the WLQ-26, we assessed the number of usable (non-missing) observations, and the proportion of observations for each of the six available response options for each of the 26 items. The proportion at the floor score (difficulty all the time, score = 0), ceiling score (difficulty none of the time, score = 5), or selecting the "does not apply to my job" option was reported. To summarize the data, the mean, standard deviation, median, and interquartile range were calculated at both item- and subscale-levels.

Internal Consistency of the WLQ-26

Cronbach's alpha was assessed for each of the five WLQ-26 subscales. At minimum, an alpha of >0.7 was expected, but an excessively high alpha (i.e., >0.95) would indicate some item redundancy [26]. Item-subscale correlations were also evaluated and >0.4 was expected.

Score Distribution for the Functional Work Tasks

To examine the score distribution for the two functional work tasks, we assessed the number of usable (non-missing) observations, mean, standard deviation, median, and interquartile range.

Relationship Between WLQ-26 and Functional Work Task Performance

The level of correlation was assessed among the five WLQ-26 subscales and the two functional work tasks. Since all are interval-level measures where the magnitude of score difference is meaningful, Pearson correlations were performed to assess the linear relationship of each pair of measures. However, there were indications of non-normality in some of the data, and therefore, we also performed Spearman correlations to ascertain the relationship among these measures. The analyses in this study were conducted using Stata

version 13.1 (College Station, TX: StataCorp LP) and SPSS version 25 (Armonk, NY: IBM Corp.).

Differences in WLQ-26 and Functional Work Task Performance in Musculoskeletal Subgroups

Firefighters were divided into two subgroups based on their response to a question about the status of their muscles and joints (excellent no problems, very good, minor injury or age-related changes, versus moderate or major problems) and whether muscle and/or joint problems interfered with their ability to do their work (no problems, minor interference, moderate interference, major interference, prevent doing my job). Since no firefighters reported moderate or greater interference, this item was dichotomized into no problems versus minor interference. An independent *t* test was done to determine whether there were differences in any of the subscales of the WLQ-26 or the performance-based tests between the subgroups.

Results

Sample Characteristics

The mean age of the study participants ($n=293$) was 42.6 (SD = 9.7). A very high proportion was male (97.3%), which is typical for the fire service profession (Table 1). The participants have been firefighters for an average of 14.8 years (SD = 10.2, range 0–35). In terms of ranking, the majority were "firefighter" rank (74.1%) and "captain" was second most common (14.5%).

The participants reported a range of musculoskeletal health status based on self-rating using a global item. The distribution in response was as follows: excellent (10.0%), very good (28.5%), minor problems (46.4%), moderate problems (13.1%), and major problems (2.1%). In general, participants reported MSK pain at multiple joints: neck (20%), back (33%), upper limb (44%), and lower limb (45%). Out of 13 co-morbid conditions assessed, 38% had none, 35% had one, 18% had two, and 9% had three or more health conditions.

Table 1 Characteristics of study participants ($n=293$)

Characteristic	*n* available	Mean (SD) or %
Age	291	42.6 (9.7) Range 20–65
Sex	291	
Male		97.3%
BMI	292	
Underweight (< 18.5)		0.3%
Normal weight (18.5–24.9)		13.4%
Overweight (25–29.9)		57.5%
Obese (≥ 30)		28.8%
Total years of service as firefighter (includes years as a volunteer or professional)	293	14.8 (10.2) Range 0–35
Professional ranking	290	
Recruit		4.8%
Firefighter (1st class, etc.)		74.1%
Captain		14.5%
Acting captain		5.2%
Chief		1.4%
Musculoskeletal health status (self-rated)[a]	291	
Excellent (no problems)		10.0%
Very good		28.5%
Minor problems		46.4%
Moderate problems		13.1%
Major problems		2.1%
Number of co-morbidities (self-rated)[b]	293	
0		38.2%
1		35.2%
2		17.8%
3 or more		8.9%

[a]Musculoskeletal health status was assessed by the question: "what is the status of your joints and muscles?"

[b]From 13 possible conditions assessed

Score Distribution of the WLQ-26

All firefighters attempted the WLQ-26, and the number of completed items ranged from 17 (0.3%) to 26 (96.3%). At the item-level, the mean ranged from 4.1 to 4.4 (Table 2) and the median was 5 for all items. For 23 of 26 items, the interquartile range was 1; for the three other items, the interquartile range was 0. Notably, across all items, a large proportion of respondents (54.5–80.5%) selected the "difficult none of the time" option, which was the maximum score of 5. In contrast, the "does not apply to my job" option was selected relatively infrequently (1.0–6.5%). At the subscale level, the mean ranged from 61.8 (social demands) to 78.7 (output demands and physical demands) (Fig. 1). The proportion of respondents at the ceiling score (no work limitations) was high for all five WLQ-26 subscales (44.0–55.6%), well above the 15% often considered as the threshold for a significant "ceiling effect" [27].

Internal Consistency of the WLQ-26

Cronbach's alpha ranged between 0.93 and 0.98 for the five WLQ-26 subscales (Table 3). Item-subscale correlations ranged between 0.88 and 0.99 for the items (Table 2).

Table 2 WRF-26 item-level properties

Item	n^a	%						Mean (SD)	Median (IQR)	Item-subscale correlations
		1 difficult all the time	2 difficult most of the time	3 difficult half of the time	4 difficult some of the time	5 difficult none of the time	6 does not apply to my job			
Work scheduling (WS)										
Get to work on time	292	13.7	0.3	0.3	3.1	80.5	2.1	4.4 (1.4)	5 (0)	0.92
Stick to a routine or schedule without having to rearrange your work tasks	291	5.8	8.9	2.1	13.1	63.6	6.5	4.3 (1.3)	5 (1)	0.91
Work without taking frequent rests or breaks to avoid discomfort	292	4.8	6.5	3.8	19.2	63.4	2.4	4.3 (1.1)	5 (1)	0.88
Work the required number of hours	291	16.2	1.7	1.4	2.8	77.0	1.0	4.2 (1.5)	5 (0)	0.94
Output demands (OD)										
Handle very demanding or stressful work situations	292	11.0	4.1	3.4	13.7	66.1	1.7	4.2 (1.4)	5 (1)	0.91
Do your work without becoming tense or frustrated	292	6.5	9.9	3.4	24.0	54.5	1.7	4.1 (1.3)	5 (1)	0.89
Do your work carefully	292	11.6	8.6	0.7	8.2	69.2	1.7	4.2 (1.5)	5 (1)	0.95
Satisfy those people who judge your work	290	6.6	12.8	2.4	14.1	60.3	3.8	4.1 (1.3)	5 (1)	0.94
Feel a sense of accomplishment	293	7.9	12.3	3.1	17.4	58.4	1.0	4.1 (1.3)	5 (1)	0.93
Finish work on time	293	9.6	9.9	2.4	7.2	65.2	5.8	4.2 (1.4)	5 (1)	0.96
Handle the workload	292	12.3	8.9	0.7	8.2	68.5	1.4	4.1 (1.5)	5 (1)	0.96
Physical demands (PD)										
Lift, carry or move objects at work weighing 10 pounds or less	293	13.0	3.8	1.7	3.4	76.1	2.1	4.3 (1.4)	5 (0)	0.94
Lift, carry or move objects at work weighing 10 pounds or more	293	12.6	3.4	2.1	11.3	68.6	2.1	4.2 (1.4)	5 (1)	0.96
Walk more than one block or climb up or down one flight of stairs while working	293	12.3	2.1	1.0	10.9	72.0	1.7	4.3 (1.4)	5 (1)	0.95
Sit, stand, or stay in one position for longer than 15 min while working	293	8.2	3.8	4.4	18.1	63.8	1.7	4.3 (1.2)	5 (1)	0.89
Bend, twist, or reach while working	293	9.6	5.8	3.4	19.1	61.1	1.0	4.2 (1.3)	5 (1)	0.96
Use hand operated tools or equipment	293	8.2	4.8	3.4	13.0	68.3	2.4	4.3 (1.3)	5 (1)	0.96
Use your upper body to operate tools or equipment	292	6.9	7.9	2.7	20.6	59.9	2.1	4.2 (1.2)	5 (1)	0.95
Use your lower body to operate tools or equipment	291	4.5	8.9	3.1	20.3	61.2	2.1	4.3 (1.2)	5 (1)	0.93
Mental demands (MD)										
Keep your mind on your work	292	9.3	10.6	1.4	16.1	61.3	1.4	4.1 (1.4)	5 (1)	0.98
Keep track of more than one task or project at the same time	291	6.5	10.0	3.4	15.8	62.5	1.7	4.2 (1.3)	5 (1)	0.97
Concentrate on your work	291	11.3	7.6	2.1	15.8	62.2	1.0	4.1 (1.4)	5 (1)	0.99
Remember things having to do with your work	292	7.5	10.6	2.4	19.9	58.6	1.0	4.1 (1.3)	5 (1)	0.97

Table 2 (continued)

Item	n[a]	%							Mean (SD)	Median (IQR)	Item-subscale correlations
		1 difficult all the time	2 difficult most of the time	3 difficult half of the time	4 difficult some of the time	5 difficult none of the time	6 does not apply to my job				
Social demands (SD)											
Talk with people in person, in meetings, or on the phone	292	6.2	6.5	3.1	11.6	68.2	4.5		4.4 (1.2)	5 (1)	0.95
Control irritability or anger toward people when working	292	6.5	10.3	1.7	15.8	64.0	1.7		4.2 (1.3)	5 (1)	0.97
Help other people get work done	292	9.9	6.9	1.7	7.9	72.3	1.4		4.3 (1.4)	5 (1)	0.98

[a]n includes observations providing "does not apply to my job" response option which are treated as missing value and therefore do not contribute to calculations of other statistics (mean, median, IQR, item-subscale correlations)

Functional Work Task Performance

For the hose drag task, the mean time-to-completion was 48.0 s (SD = 14.5). The fastest time observed was 20.4 s and the slowest time was 95.0 (Fig. 2). For the stair climb with high-rise pack task, the mean time-to-completion was 76.7 s (SD = 37.2). The fastest time observed was 21.0 s and the slowest was 218.0 s.

Relationship Between WLQ-26 and Functional Work Task Performance

Among the five WLQ-26 subscales, Pearson r ranged between 0.86 and 0.93, and Spearman r were consistently lower but followed a similar pattern, ranging between 0.71 and 0.83. Between the two functional work tasks, the Pearson r was moderate, but $p < 0.05$, 0.47 and the Spearman r was 0.56. Comparing the WLQ-26 subscales and the two functional work tasks, only non-significant Pearson r (−0.09 to 0.11) and Spearman r (−0.16 to 0.05) were found.

Subgroup Testing

There were small differences—mostly in the direction expected—between the subgroups. However, none of these differences were statistically significant (Table 4).

Discussion

This study indicates that active duty firefighters scored very well on WLQ-26 which indicates low work limitations, even though a majority reported at least minor MSK problems, and a majority had at least one comorbid health condition. Conversely, variability in task performance was more evident on the performance-based tests. This is also evident in their performance-based tests where there were small score differences between those with limitations and those without. Firefighters who reported having moderate to severe muscle and joint problems took approximately 10 s longer to complete the stair climb task than did firefighters not reporting those problems. However, none of the subgroup comparisons were statistically different. We performed these analyses for hypothesis generation since we were not adequately powered for this analysis with respect to performance-based tests.

Given that it was designed to have wide applicability and assess a broad range of work demands, the WLQ-26 was initially thought to be appropriate for use in this population. However, several issues are apparent in our data. First, there was evidence of a strong ceiling effect across all five WLQ-26 subscales. Evidence of ceiling effects can indicate an inability to discriminate between individuals. Our baseline

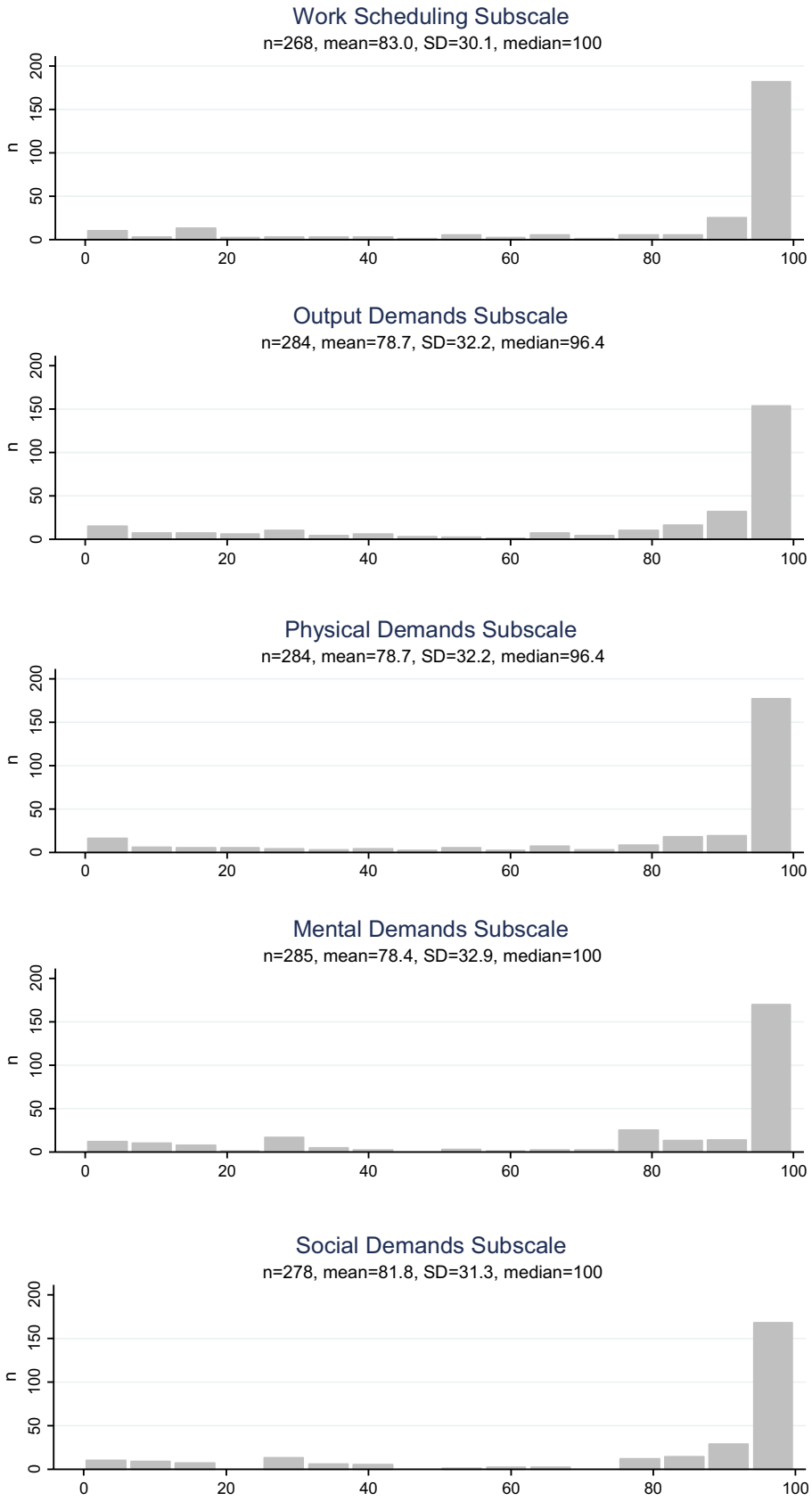

Fig. 1 Score distribution of the WRF-26 subscales

Table 3 Correlation between WRF-26 subscales and functional work task performance (Pearson *r*: lower-left side of off-diagonal, Spearman *r*: upper-right side of off-diagonal)

	Work scheduling demands	Output demands	Physical demands	Mental demands	Social demands	Functional hose task	Functional stair task
Work scheduling demands	α = 0.93 *n* = 268	0.75* *n* = 264	0.73* *n* = 263	0.71* *n* = 263	0.71* *n* = 258	−0.04 *n* = 138	0.02 *n* = 136
Output demands	0.90* *n* = 264	α = 0.98 *n* = 284	0.76* *n* = 278	0.81* *n* = 279	0.79* *n* = 274	−0.07 *n* = 144	0.00 *n* = 142
Physical demands	0.86* *n* = 263	0.90* *n* = 278	α = 0.98 *n* = 284	0.74* *n* = 280	0.71* *n* = 273	−0.16 *n* = 144	−0.06 *n* = 142
Mental demands	0.86* *n* = 263	0.93* *n* = 279	0.90* *n* = 280	α = 0.98 *n* = 285	0.83* *n* = 275	−0.04 *n* = 145	0.05 *n* = 143
Social demands	0.86* *n* = 258	0.91* *n* = 274	0.90* *n* = 273	0.93* *n* = 275	α = 0.96 *n* = 278	−0.02 *n* = 140	0.04 *n* = 138
Functional hose task	−0.04 *n* = 138	−0.00 *n* = 144	−0.09 *n* = 144	0.03 *n* = 145	−0.01 *n* = 140	α = n/a *n* = 148	0.56* *n* = 146
Functional stair task	0.08 *n* = 136	0.08 *n* = 142	0.03 *n* = 142	0.10 *n* = 143	0.11 *n* = 138	0.47* *n* = 146	α = n/a *n* = 146

α Cronbach's alpha, *n/a* not applicable

*$p < 0.05$

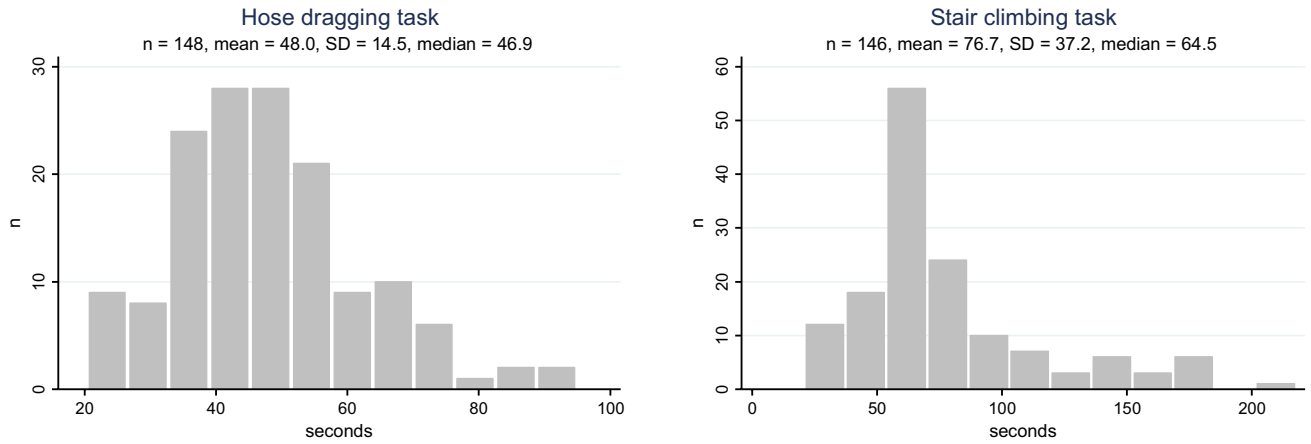

Fig. 2 Functional work tasks performance among firefighters

data and the fact that firefighters were on full active duty suggest that the firefighter participants were in relatively good health and this may account for the ceiling effects. To the best of our knowledge, a ceiling effect has not been often found in previous psychometric testing of the WLQ-26. Thus, while this tool may be appropriate for a wide range of occupations, the unique nature of firefighting tasks demands may make it an outlier from a measurement perspective. We further believe that this ceiling effect (i.e., highly skewed data) is largely responsible for the excessively high Cronbach's alpha, item-subscale correlations, and inter-subscale correlations observed. Thus, the high internal consistency further supports why this tool does not perform optimally in firefighters, but does not lead to any generalized concern about the WLQ-26 from a psychometric perspective. There is a plethora of evidence that support the proper convergence and divergence among WLQ-26 subscales in the literature [15, 17, 28]. However, we did note a small concern suggesting respondents may not always be closely attentive to the instructions provide. We noticed a very small proportion scoring exceedingly low on all five WLQ subscales (indicating very low work functioning), which we thought was somewhat suspect given that these were active firefighters that are currently working.

A much broader range of performance was demonstrated on the two functional work tasks. As a result, very low correlations were found between these two methods of assessment. This suggests the WLQ-26 may not be able

Table 4 Differences in test scores in MSK subgroups

	Bone and joint status		Bone and joint-related work interference	
	Excellent/very good or minor problems ($n=244$)	Moderate or severe problems ($n=44$)	None ($n=139$)	Minor interference ($n=130$)
Work Limitations Questionnaire (% score)				
Time	82.1	77.0	82.4	80.1
Mental	79.0	76.2	80.2	77.2
Interpersonal	81.9	79.4	83.4	79.6
Physical	82.7	76.0	85.1	78.5
	Excellent/very good or minor problems ($n=131$)	Moderate or severe problems ($n=15$)	None ($n=75$)	Minor interference ($n=63$)
Firefighting performance tests (s)				
Hose drag	47.6	51.7	47.7	47.5
Stair climb	75.7	84.6	76.5	77.5

There were no significant differences amongst any of the comparisons. No moderate or severe interference was reported so the variable was dichotomized between none and minor. Higher scores indicate poor performance on the performance-based test and better self-reported work role functioning as measured on the Work Limitations Questionnaire

to detect emerging work limitations in firefighters, which is most likely because they are required to perform very high demand tasks that are not adequately represented on a generic WLQ. The lack of correlation between the self-report measures in the performance-based measures in our data was driven by ceiling effects. Ceiling effects would preclude accurate assessment of the real relationships between self-reported work limitations and performance-based measures. It is noteworthy that there were moderate correlations between the two different performance-based tests. Since the nature and duration of these tests were quite different, they might tap into different physical domains. Ideally, when constructing a battery of tests to assess firefighters, complementary measures that provide a comprehensive assessment would be identified. The moderate correlations between the two performance-based tests suggest that while related, they are sufficiently different that the time invested in additional testing would be worthwhile since additional information would be achieved.

A second noteworthy observation was that the "does not apply to my job" response option was not often chosen across the WLQ-26 items (range 1.0–5.8% of respondents). In addition, the instrument was generally well completed by respondents with very few missed items overall. This is notable because we believe this is a positive indication of WLQ-26's item relevance for this population, which offers some support for its content validity. This suggests that the domains may be relevant, but the items are not designed to capture the specific areas of firefighting that represent areas at most risk of compromise. That said, this should not be regarded as a complete endorsement of WLQ-26's content validity in this population, since firefighters are very task oriented and have an attitude of getting the job done, regardless of obstacles. This could lead them to complete forms even when the items are not salient. Future qualitative research involving firefighters could be useful to further inform whether its content is salient. This highlights a measurement principle that high rates of "missingness" might lead one to question the content validity of a measure; high rates of completeness do not exclude these concerns. Given our findings, development of a self-report firefighter-specific WLQ is that is conceptually linked to the domains of the WLQ is justified as it would retain some comparability across occupations, while also identifying emerging work limitations and risk in firefighters.

It is clear from our study that WLQ-26 scores were not related to performance scores. Although measurement issues were a pitfall in the comparison, it is also important to recognize that these are distinct constructs and measurement approaches. First, there is a difference in specificity in the demand in question. The WLQ-26 assesses more general constructs, whereas the functional work tasks are highly specific. Even within the physical demands subscale, the items assess a much broader range of potential demands compared to the functional work tasks. Second, there is a difference in terms of the "point of reference". Whereas the functional work tasks measure "absolute" performance with no reference point, we believe the WLQ-26 is in a way evaluating "relative" performance referenced against one's personal best (i.e., how much

limitation compared to when one was completely healthy). To further illustrate this difference, it is perhaps helpful to envision a worker with absolutely no health impairments/limitations but still performs these tasks poorly compared to one's peers because of lack of natural physical ability (e.g., weak or slow). For these above reasons, this comparison is not ideal as a test of construct validity. These factors were considered when hypothesizing that, no more than moderate relationships might be expected.

Our study provides unique information on the relationship between standardized performance-based work tasks and self-reported work limitations in a cohort of firefighters from a single fire service. However, limitations in our work should also be considered. The validity of the simulated firefighting tasks to actual performance on the fire ground has not been determined. We did not conduct qualitative investigations to identify why firefighters responded as they did to the self-report items. Measurement limitations in the generic reporting of work limitations precluded accurate assessment of their relationship to performance-based tests. Subgroup differences in the performance-based tests were relatively underpowered since fewer people participated in these tests and these comparisons were performed for hypothesis generation only. Physical performance was assessed as a function of speed, not safety or quality of task performance. While speed is only one aspect of physical performance, it is an important one in firefighting. Further, speed is widely used as an indicator of performance capability. Finally, given the different versions of the WLQ that have been reported in the literature we cannot be certain that our conclusions apply across all versions. However, we are very confident that our conclusions would be stable across different versions of the questionnaire since there is a very substantial overlap of items between different versions; and the nature of the difficulty issue would apply across all versions.

Conclusions

The WLQ-26 demonstrates ceiling effects, and did not relate strongly with functional task performance in active duty firefighters. However, the items were considered relevant and may form the framework for a more specific WLQ-26 that targets the unusually high task demands of firefighters.

Acknowledgements This work was supported by a Canadian Institutes of Health Research Grant (FRN 114112). Joy MacDermid was supported by a Canadian Institutes of Health Research Chair in Gender, Work and Health and the Dr. James Roth Chair in Musculoskeletal Measurement and Knowledge Translation. Kenneth Tang was supported by a Canadian Institutes of Health Research Banting Postdoctoral Fellowship. We thank the Hamilton Professional Fire Fighters Association and the Hamilton Fire Department for their support of this research. We also thank Margaret Lomotan for assistance with data collection and manuscript preparation.

References

1. Walton SM, Conrad KM, Furner SE, Samo DG. Cause, type, and workers' compensation costs of injury to fire fighters. Am J Ind Med. 2003;43(4):454–458.
2. Reichard AA, Jackson LL. Occupational injuries among emergency responders. Am J Ind Med. 2010;53(1):1–11.
3. Leigh JP, Markowitz SB, Fahs M, Shin C, Landrigan PJ. Occupational injury and illness in the United States: estimates of costs, morbidity, and mortality. Arch Intern Med. 1997;157(14):1557–1568.
4. Sluiter JK, Frings-Dresen MHW. What do we know about ageing at work? Evidence-based fitness for duty and health in fire fighters. Ergonomics. 2007;50(11):1897–1913.
5. Bos J, Mol E, Visser B, Frings-Dresen M. Risk of health complaints and disabilities among Dutch firefighters. Int Arch Occup Environ Health. 2004;77(6):373–382.
6. Negm A, MacDermid J, Sinden K, D'Amico R, Lomotan M, MacIntyre NJ. Prevalence and distribution of musculoskeletal disorders in firefighters are influenced by age and length of service. J Mil Veteran Fam Heal. 2017;3(2):33–41.
7. Plat MCJ, Frings-Dresen MHW, Sluiter JK. Which subgroups of fire fighters are more prone to work-related diminished health requirements? Int Arch Occup Environ Health. 2012;85(7):775–782.
8. Fullerton CS, Ursano RJ, Wang L. Acute stress disorder, posttraumatic stress disorder, and depression in disaster or rescue workers. Am J Psychiatry. 2004;161(8):1370–1376.
9. Carey MG, Al-Zaiti MSA, Dean GE, Loralee Sessanna DSF. Sleep problems, depression, substance use, social bonding, and quality of life in professional firefighters. J Occup Environ Med. 2011;53(8):928–933.
10. Carleton RN, Afifi TO, Turner S, Taillieu T, Duranceau S, LeBouthillier DM, et al. Mental disorder symptoms among public safety personnel in Canada. Can J Psychiatry. 2018;63(1):54–64.
11. MacDermid JC, Nazari G, Rashid C, Sinden K, Carleton N, Cramm H. Two-month point prevalence of exposure to critical incidents in firefighters in a single fire service (unpublished).
12. Lerner D, Amick BC 3rd, Rogers WH, Malspeis S, Bungay KCD. The work limitations questionnaire. Med Care. 2001;39(1):72–85.
13. Amick BC, Habeck RV, Ossmann J, Fossel AH, Keller R, Katz JN. Predictors of successful work role functioning after carpal tunnel release surgery. J Occup Environ Med. 2004;46(5):490–500.
14. Arumugam V, Macdermid JC, Grewal R. Content analysis of work limitation, stanford presenteeism, and work instability questionnaires using international classification of functioning, disability, and health and item perspective framework. Rehabil Res Pract. 2013;2013:614825.
15. Roy J-S, MacDermid JC, Amick BC, Shannon HS, McMurtry R, Roth JH, et al. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. Phys Ther. 2011;91(2):254–266.
16. Wolfe F, Michaud K, Choi HK, Williams R. Household income and earnings losses among 6396 persons with rheumatoid arthritis. J Rheumatol. 2005;32(10):1875–1883.
17. Walker N, Michaud K, Wolfe F. Work limitations among working persons with rheumatoid arthritis: results, reliability, and validity of the work limitations questionnaire in 836 patients. J Rheumatol. 2005;32(6):1006–1012.

18. Beaton DE, Tang K, Gignac MAM, Lacaille D, Badley EM, Anis AH, et al. Reliability, validity, and responsiveness of five at-work productivity measures in patients with rheumatoid arthritis or osteoarthritis. Arthritis Care Res. 2010;62(1):28–37.
19. Lerner D, Reed JI, Massarotti E, Wester LM, Burke TA, Knottnerus JA, et al. The work limitations questionnaire's validity and reliability among patients with osteoarthritis. J Clin Epidemiol. 2002;55(2):197–208.
20. Sanderson K, Tilse E, Nicholson J, Oldenburg B, Graves N. Which presenteeism measures are more sensitive to depression and anxiety? J Affect Disord. 2007;101(1–3):65–74.
21. Erickson SR, Guthrie S, VanEtten-Lee M, Himle J, Hoffman J, Santos SF, et al. Severity of anxiety and work-related outcomes of patients with anxiety disorders. Depress Anxiety. 2009;26(12):1165–1171.
22. Beaton D, Bombardier C, Escorpizo R, Zhang W, Lacaille D, Boonen A, et al. Measuring worker productivity: frameworks and measures. J Rheumatol. 2009;36(9):2100–2109.
23. Amick BC, Lerner D, Rogers WH, Rooney T, Katz JN. A review of health-related work outcome measures and their uses, and recommended measures. Spine. 2000;25(24):3152–3160.
24. Lerner D, Rogers WH, Chang H. Scoring the work limitations questionnaire (WLQ) and the WLQ index for estimating work productivity loss (Technical Report). 2003.
25. Sinden KE. Identifying determinants of firefighter work health and task performance: implications for injury management. McMaster University; 2014. http://hdl.handle.net/11375/15948.
26. Tavakol M, Dennick R. Making sense of Cronbach's alpha. Int J Med Educ. 2011;2:53–5.
27. Terwee CB, Bot SDM, de Boer MR, van der Windt DAWM., Knol DL, Dekker J, et al. Quality criteria were proposed for measurement properties of health status questionnaires. J Clin Epidemiol. 2007;60(1):34–42.
28. Beaton DE, Kennedy CA. Beyond return to work: testing a measure of at-work disability in workers with musculoskeletal pain. Qual Life Res. 2005;14(8):1869–79.