



Determination of Perceived Levels of Physical Work Demand Thresholds, and Reliability and Responsiveness of the Modified-Spinal Function Sort Questionnaire in a Multidisciplinary Occupational Rehabilitation Setting

Cyrille Burrus^{1,2} · Philippe Vuistiner² · Bertrand Léger² · François Luthi^{1,2,3}

Accepted: 8 March 2021 / Published online: 24 March 2021

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Purpose To determine the levels of perceived work demand capacity corresponding to the Modified Spinal Function Sort (M-SFS) score and precise reliability validity and responsiveness.

Methods This prospective validation study included patients with chronic musculoskeletal impairments who underwent multidisciplinary occupational rehabilitation. After determining the percentiles of the work demand thresholds corresponding to the spinal function sort (SFS), the percentiles were transposed to the M-SFS. Reliability was assessed using the intraclass correlation coefficient and limits of agreement. Correlations with other questionnaires and a lifting task were measured to assess validity. Responsiveness was determined using anchor- and distribution-based approaches.

Results 288 patients were included. The following thresholds were obtained for the M-SFS: 0–43 points, minimal; 44–50, very light; 51–58, light; 59–64, light to medium; 65–70, medium; 71–76, heavy; and 77–80, very heavy. Reliability was confirmed. The correlation between the M-SFS and SFS scores was good at 0.89 (95% CI, 0.86–0.91) and moderate according to the PILE-test result of 0.60 (95% CI, 0.50–0.67). We could not calculate a valid anchor-based minimal clinically important difference. The standard error of measurement was 3.9 points, and the smallest detectable change was 10.8 points.

Conclusions On the basis of the comparison of the M-SFS and SFS scores, the M-SFS score can be interpreted in relation to the levels of work demand. This study confirms the good reliability and validity of the M-SFS questionnaire in assessing perceived physical capacity. Further studies are needed to determine its responsiveness.

Keywords Self-efficacy · Multidisciplinary occupational rehabilitation · Questionnaire

Introduction

Evaluation of musculoskeletal disorders is not only based on clinical examination or radiological examinations but also based on the patients' point of view. Patient-reported outcome measures may help provide insights into patients' perspectives [1]. Self-efficacy is defined as an individual's belief in his ability to achieve tasks [2] and may influence health outcomes from a biopsychosocial point of view [3]. In rehabilitation with vocational aspect, the pictorial questionnaire Spinal Function Sort (SFS) [4] is useful for appreciating patients' perceptions of the level of physical demand they can perform and to evaluate work-related self-efficacy beliefs [4].

The SFS consists of 50 items linked with work-related tasks that are illustrated and has good psychometric

✉ Cyrille Burrus
Cyrille.Burrus@crr-suva.ch

¹ Department for Musculoskeletal Rehabilitation, Clinique Romande de Réadaptation, Avenue Grand-Champsec 90, 1950 Sion, Switzerland

² Institute for Research in Rehabilitation, Clinique Romande de Réadaptation, Avenue Grand-Champsec 90, 1950 Sion, Switzerland

³ Department of Physical Medicine and Rehabilitation, Orthopaedic Hospital, Lausanne University Hospital, Rue du Bugnon 21, 1011 Lausanne, Switzerland

properties [4–6]. The score is categorized into different levels of physical work demands, as described by Matheson et al. [4]. The questionnaire was validated in a European rehabilitation setting by Oesch et al. [6] based on the Dictionary of Occupational Titles [7]. Thresholds were proposed by Matheson [8] and have then been adapted by the Swiss Association of Rehabilitation (SAR) [9] for their use during Functional capacity evaluations (FCE) with adjunction of a supplementary “light to medium” work demand category. The questionnaire is routinely administered to patients in multidisciplinary occupational rehabilitation to address their perceived level of physical performance. During FCEs, it is compared with the observed physical performance and assessed for its correspondence to a work demand category. In clinical practice, among patients with non-specific low back pain indicated for rehabilitation, the lowest scores (0–99 points) that corresponded to a minimal work demand were shown to have high predictive values for non-return to work at 3 and 12 months [6]. To reduce the number of items and update the illustrations that patients found sometimes out-of-date, the Modified SFS (M-SFS) was recently developed [10]. It has been proven to be a reliable and valid tool for workers with chronic musculoskeletal pain [11], but the correspondence of its score to work demand levels has not yet been evaluated.

The primary aim of this study was to determine the levels of physical work demand corresponding to the M-SFS score. The test–retest reliability validity and responsiveness of the questionnaire were the secondary outcomes.

Methods

Study Design

We conducted a cross-sectional study among patients who received multidisciplinary occupational rehabilitation after musculoskeletal trauma.

Participants

Any patient of working age (18–65 years) who was referred for rehabilitation at the Clinique Romande de Réadaptation in Sion Switzerland was eligible for participation in this study. The exclusion criteria were upper limb trauma, spinal cord or traumatic brain injuries, incapability to make judgments, or legal custody. We chose not to include patients with upper limb trauma, to whom the Hand Function Sort Questionnaire is usually administered [12, 13]. The patients were referred from all French-speaking cantons of Switzerland, including urban and industrial city centers or more rural regions. Most of our patients sent to our rehabilitation program are construction or industry workers referred

by general practitioners, surgeons, or insurance medical advisors, when they present with chronic musculoskeletal pain (≥ 3 months), functional impairments, and inability to return to work. The majority come in after work, leisure or traffic trauma. The rehabilitation program is based on a multidisciplinary approach with medical, psychological, social, and functional evaluations. The aim of our program is to take care of patients with a biopsychosocial approach in order to improve their functional status and chances to return to work. Following the evaluations patients have a rehabilitation program oriented on functional gains and if their functional limitations are not compatible with their previous work they are oriented to another work. Most of our patients receive financial compensation at the time of their rehabilitation if they are unable to work. We included patients consecutively, and the inclusion period was from June 2019 to March 2020. If a patient was treated twice during this period, we considered only the data from the first hospitalization. The patients gave informed consent. The study was conducted in accordance with the principles expressed in the Declaration of Helsinki 2008, and the protocol was approved by the local medical ethics committee (CCVEM 034/12).

Measurements

The patients' sociodemographic and characteristic data were collected at admission from the medical assessment. We recorded the following data: age, sex, body mass index (BMI), interval between injury and hospitalization in days, level of education (compulsory school vs. higher level), presence of an employment contract (yes vs. no), and trauma severity according to the Abbreviated Injury Scale (AIS) [14]. The AIS consists of a 7-point scoring system of injury severity (from 1 point for minor severity to 6 points for injury severity beyond therapeutic resources). We categorized the scores into three severity groups as follows: minor, moderate, and serious for scores ≥ 3 .

Modified Spinal Function Sort

The M-SFS [10] is a pictorial questionnaire that assesses patients' perceived ability to perform workability tasks. It can be used for patients with chronic musculoskeletal disorders [11]. It consists of 20 drawings with simple instructions and a 5-point Likert scale that measures the ability to perform each task as follows: unable (0 points), severely restricted (1 point), restricted (2 points), slightly restricted (3 points), and able (4 points). The score ranges from 0 to 80 points, with a higher score indicating a higher perceived physical capacity [10]. The M-SFS has a high internal consistency, with an ICC of 0.90 (95% confidence interval [CI], 0.84–0.94); good test–retest reliability; and a high

correlation (0.89) to the original SFS [11]. Various translations of this pictorial questionnaire have been used, including French, German, Portuguese, Italian, Albanian, or Serbian, on the basis of previous transcultural adaptations of the SFS [5, 12]. The M-SFS was administered to patients 2 days after admission to the clinic and on the last days before discharge.

Spinal Function Sort

The SFS [4] is a pictorial questionnaire from which the M-SFS was developed. It consists of 50 items with drawings linked to working tasks. For each task, the patient must rate his ability to perform it on a 5-point Likert scale (from “able” to “restricted” to “unable”). It has good reliability and validity [4, 6, 13]. The score ranges from 0 to 200 points, with a higher score indicating a perception of having the ability to perform heavier work demands. The levels of perceived physical work demand based on the Dictionary of Occupational Titles have been adapted, and the following thresholds are used in Switzerland: minimal, 0 to 99 points; very light (5 kg), 100 to 120 points; light (5–10 kg), 121 to 140 points; light to medium (10–15 kg), 141 to 160 points; medium perceived work demand (15–25 kg), 161 to 180 points; heavy perceived work demand (25–45 kg), 181 to 195 points; and very heavy (> 45 kg), 196 to 200 points.

Brief Pain Inventory

Pain intensity and interference were measured with the Brief Pain Inventory (BPI) [15, 16]. Of the 11 items of the questionnaire, 4 measure pain intensity and 7 measure the level of interference with functioning caused by pain. A higher score indicates higher pain intensity or interference.

Hospital Anxiety and Depression Scale

Anxiety and depressive symptoms were measured with the Hospital Anxiety and Depression Scale (HADS) [17, 18]. The HADS consists of 14 items addressing anxiety and depressive symptoms with 7 items each. A higher score indicates a higher level of anxiety or depressive symptoms.

Pain Catastrophizing Scale

Catastrophizing was measured using the Pain Catastrophizing Scale (PCS). The PCS is a 13-item questionnaire that addresses catastrophic thoughts related to pain. A high score suggests high catastrophic thoughts [19].

Tampa Scale for Kinesiophobia

Fear of movement or reinjury was measured using the Tampa Scale for Kinesiophobia (TSK), which is a 17-item questionnaire in which a high score suggests high kinesiophobia [20].

PILE Test

The PILE test is a lifting test commonly used in rehabilitation with a vocational aspect to measure the lifting capacity of patients [21]. The test consists of incremental lifts of increasing loads in a crate from the floor to a 76-cm high desk. To pass each step, the patient must carry the load four times in 20 s. The load is increased by 2.5 or 5 kg at each step. The maximum load carried corresponds to the last step.

Data Collection and Bias

To minimize measurement bias, questionnaires and other clinical and demographic data were collected in the 2 days after admission and in the last 4 days preceding discharge. Records were collected with a digital pen, which permits data capture and direct transfer from paper to data files.

Data Analysis

Descriptive statistics are expressed as mean and standard deviation for continuous variables, or median and interquartile range (IQR) for data with a skewed distribution. To assess the hypotheses of normality, we performed a visual inspection of the data distribution. Categorical variables are expressed as counts and percentages.

Determination of Perceived Levels of Physical Work Demand Using the M-SFS

To determine the cutoff scores corresponding to the levels of perceived physical work ability, we used the SFS questionnaires. On the basis of the known thresholds for this questionnaire, we first determined the percentile of each level of perceived physical work demand. The M-SFS cutoff scores corresponding to the levels of perceived physical work ability were then determined by transposing the corresponding percentiles for each level of perceived work demand capacity to the M-SFS score. The consistency of the classifications of the levels of perceived physical work ability based on the two scores was measured with the intraclass correlation coefficient (ICC). Data from patients

admitted to our multidisciplinary occupational rehabilitation in 2017 and 2018 were used for the sample size calculation. In this 2017–2018 population ($n = 1065$), we noticed that only 1.5% of patients were classified in the “very heavy” work demand category. We thus chose to merge the last 2 categories which then represented 5.6% of this sample population of 2017–2018. This “Heavy-very heavy” category still represented the smallest group. In order to have at least 10 patients in this smallest group [22], a minimum of 179 patients was expected.

Reliability

50 patients answered a second M-SFS questionnaire 3 to 4 days after the M-SFS administered at admission, after oral confirmation they felt no change in their physical ability as expected after the short duration. Patients were kept blind from the results of the first test. Reliability was measured using the intraclass (ICC) correlation coefficient and the Bland–Altman limits of agreement [23]. The ICC was calculated on the basis of the absolute agreement with the two-way mixed-effects model [24, 25]. The ICC was considered excellent, good, moderate, and poor if r was > 0.90 , between 0.75 and 0.90, between 0.50 and 0.75, and < 0.50 , respectively [24]. For the Bland–Altman method, we measured the mean difference between the two M-SFS scores and 95% CI of the agreement defined as ± 1.96 standard deviation of the differences. The normality of the differences was verified.

The interrelatedness among the items of the M-SFS was measured using the Cronbach α [26]. A value > 0.70 was expected for good internal consistency [27].

Validity

The criterion validity of the M-SFS questionnaire was verified using Pearson correlation coefficients (r) between M-SFS and SFS [28, 29]. Consistency between the perceived thresholds obtained with the M-SFS and SFS questionnaire was checked by measuring the ICC between the two classifications. Construct validity was assessed with a priori hypotheses in order to verify if the M-SFS consistently measures the construct it is supposed to measure [29]. Higher correlations with the M-SFS were expected with perceived and observed physical abilities than with psychological dimensions. We also expected a lower correlation for age and BMI as older people and obese people were thought to have a lower perception of their physical abilities. Based on literature and on our clinical experience, we assumed the following hypotheses: (1) good to excellent correlation of the M-SFS with the SFS from which it was developed [11]; (2) a moderate-to-good correlation of the M-SFS with a lifting task ability (PILE test) [6, 11]; (3–4) weak-to-moderate correlation with BPI severity or interference scores [5, 11,

12]; (5–6) weak-to-moderate correlation with kinesiophobia (TSK) [6, 30, 31] and catastrophizing (PCS) scale [31]; (7–8) weak correlations with the HADS anxiety or depressive subscale score [5, 31]; (9) weak correlations with age [32] and [10] BMI [33]. More than 75% of these hypotheses were expected to be verified [29]. Correlations were considered excellent, good, moderate, weak, and absent if r was > 0.91 between 0.71 and 0.9, between 0.51 and 0.70, between 0.31 and 0.50, and < 0.30 [34, 35]. The 95% CIs for the correlation coefficients were calculated using Fischer’s transformation.

Floor and Ceiling Effect

The proportion of patients with the lowest or highest M-SFS scores was recorded to check for a floor or ceiling effect. Floor or ceiling effects were defined as $\geq 15\%$ of patients having minimal or maximal scores [36].

Responsiveness

The Minimal Clinically Important Difference (MCID) was determined by using the receiver-operating characteristic (ROC) method that compares “improved” and “non-improved” patients between the time of admission and the end of hospitalization. To determine which patients were improved, we used the Patient Global Impression of Change (PGIC) [37] as an anchor. The PGIC questionnaire is answered at the end of rehabilitation by the patients to evaluate on a 7-point Likert scale whether their health status has changed during their stay. Each level corresponds to the following perceived level of change: 1, worse than ever; 2, much worse; 3, slightly worsened; 4, unchanged; 5, slightly improved; 6, much improved; and 7, completely improved. The score was transformed into a binary variable, with scores of 6 and 7 considered as “improved” as opposed to the other scores considered “non-improved.” The correlation between the M-SFS score changes and the PGIC was used to measure the “credibility” of the anchor. A correlation > 0.3 was expected [38]. The optimal cutoff score on the ROC curve was determined using Youden’s index [39].

In complement to the anchor-based approach, a distribution-based method was also used. The standard error of measurement (SEM) and the smallest detectable change (SDC) were measured. The SEM represents the measurement error of an instrument and can be defined as the square root of the error variance [40, 41]. The SDC represents a measure of the variation of a score, not a measurement error [40, 41]. We used the following formula: $SEM = \sigma_{\text{pooled}} \times \sqrt{1 - ICC}$ and $SDC = 1.96 \times \sqrt{2} \times SEM$.

All the analyses were performed using Stata 16 (Stata Corp, College Station, TX, USA). The significance level was set at a probability of < 0.05 .

Table 1 Summary statistics (mean value \pm standard deviation (SD) for continuous variables with normal distribution; median value (interquartile range) for continuous variable with skewed distribution; absolute number (relative number) for binary variables)

Variable	N	Possible values	Mean \pm SD	Min.–max. values
Age	288	Years	43.9 \pm 11.8	18.0–80.2
BMI	285	Kg/m ²	28.5 \pm 5.5	16.0–48.2
BPI severity subscale	288	0–10	4.7 \pm 2.1	0–9.7
BPI interference subscale	288	0–10	5.1 \pm 2.2	0–10
TSK	277	17–68	45.6 \pm 9.2	22–66
PCS	278	0–52	24.6 \pm 13.0	0–52
HADS-A	277	0–21	9.6 \pm 4.6	0–20
HADS-D	277	0–21	7.5 \pm 4.4	0–19
Interval injury—rehabilitation	258	Days	431 (250–681)	39–5976
Gender	288	Female	57 (19.8%)	
		Male	231 (80.2%)	
Injuries location	288	Spine	63 (21.9%)	
		Lower limb	218 (75.7%)	
		Multiple	7 (2.4%)	
AIS (injury severity)	282	Minor	85 (30.1%)	
		Moderate	130 (46.1%)	
		Severe	67 (23.8%)	
Native language	288	French	128 (44.4%)	
		Other	160 (55.6%)	
Education level	288	Low	162 (56.3%)	
		High	126 (43.7%)	
Employment contract at admission	282	Yes	139 (49.3%)	
		No	143 (50.7%)	
M-SFS	288	0–80	45.3 \pm 19.8	2–78
SFS	288	0–200	107.5 \pm 54.8	2–200

N available data for each variable; possible values = range for continuous variables and categories for dichotomized variables. *BMI* body mass index (kg/m²); *AIS* Abbreviated Injury Scale; *BPI* Brief Pain Inventory; *M-SFS* Modified; Spinal Function Sort; *SFS* Spinal Function Sort

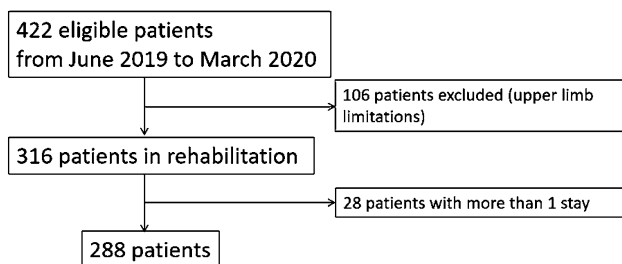


Fig. 1 Flow-chart

Results

The patients' characteristics are detailed in Table 1. From the 422 patients eligible for the study, 288 patients were included from July 2019 to March 2020 (Fig. 1). The patients were predominantly middle-aged men (mean age, 43.9 years; 80.2% of men), more than half of whom

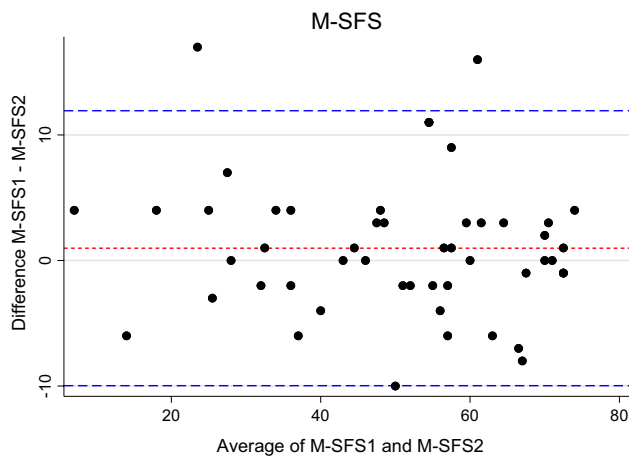
were not native French speakers (55.6%) or had a low level of education (56.3%). Most of the patients included were admitted to our center > 1 year after trauma (median duration, 431 days; IQR, 250–681 days), had lower limb impairments (75.7%), and most injuries were classified as minor to moderate (76.2%) according to the Abbreviated Injury Score (AIS).

M-SFS Cutoff Score Levels

The M-SFS cutoff score levels corresponding to the percentiles are presented in Table 2. A score of 43 points corresponds to the upper limit of the minimal perceived work demand capacity; between 44 and 50 points, a very light perceived work demand capacity; between 51 and 58 points, a light perceived work demand capacity; between 59 and 64 points, a light-to-medium perceived work demand capacity; between 65 and 70 points, a medium perceived work demand capacity; and ≥ 71 points, a heavy perceived work demand capacity.

Table 2 Proposed correspondence of work demand category thresholds for the Modified Spinal Function Sort (M-SFS) questionnaire

Work demand category (maximal load)	SFS	Corresponding percentiles	M-SFS	95% CI
Minimal	0–99	44.1	0–43	40–47
Very light (5 kg)	100–120	11.5	44–50	47–53
Light (5–10 kg)	121–140	11.5	51–58	53.7–61
Light to medium (10–15 kg)	141–160	11.1	59–64	62–67
Medium (15–25 kg)	161–180	12.1	65–70	69–72
Heavy (25–45 kg)	181–195	6.9	71–76	73–76.5
Very heavy (> 45 kg)	> 195	2.8	77–80	

**Fig. 2** Bland–Altman plot of the M-SFS score. The central small-dashed line represents the mean difference between test and retest questionnaires, and the upper and lower dashed lines represent 95% limits of agreement (mean difference \pm 1.96 SD of the differences between scores)

The ICC measured for the consistency between the two classifications of the levels of perceived physical work ability was 0.86 (95% CI, 0.83–0.89).

Reliability

50 patients participated in the measurement of reliability. The ICC for the total score was 0.95 (95% CI: 0.91–0.97). The mean difference between the test and retest questionnaires was 0.98 points. The 95% limits of agreements were between -9.70 and 11.66 points, without evidence of bias. The Bland–Altman plot is presented in Fig. 2.

Internal Consistency

The Cronbach α of the M-SFS was 0.94.

Validity

Correlations with the SFS and other questionnaires and the PILE test are presented in Table 3. A high consistency was observed between the classifications based on the SFS and those based on the M-SFS with an ICC of 0.86 (95% CI, 0.83–0.89).

Floor and Ceiling Effects

We found no floor or ceiling effect, as extreme scores were observed for < 1% of patients.

Responsiveness

Of the 288 patients included in the study, 279 PGIC ratings were available because 9 patients did not answer the M-SFS at discharge.

Anchor-Based Approach

Of the patients, 35.1% were classified as “improved.” The area under the ROC curve was 0.65 (95% CI, 0.58–0.73). On the basis of the optimal specificity/sensitivity, the MCID range was calculated to be 2 points for a maximal Youden index of 0.26. The correlation of the M-SFS score change and PGIC was 0.26, which does not allow precise determination of the MCID on the basis of the PGIC anchor [38].

Distribution-Based Approach

We calculated an SEM of 3.9 points and an SDC of 10.8 points.

Discussion

Knowledge of the M-SFS score thresholds adds clinical utility to the score, as it can be interpreted in relation to the level of perceived physical work demand capacity. As it is used for

Table 3 Correlations with 95% confidence intervals between the M-SFS or the SFS scores and questionnaires and PILE-test ($p < 0.001$)

	SFS [#]	PILE-test [#]	BPI severity [#]	BPI interference [#]	TSK [#]	PCS [#]	HADS-A [#]	HADS-D [#]	Age	BMI
M-SFS	0.89 [0.86; 0.91]	0.61* [0.52; 0.70]	-0.48 [-0.57; -0.39]	-0.47 [-0.56; -0.38]	-0.37 [-0.46; -0.26]	-0.49 [-0.57; -0.40]	-0.41 [-0.50; -0.31]	-0.42 [-0.51; -0.32]	-0.23 [-0.33; -0.11]	-0.15 [-0.26; -0.03]
SFS	1	0.61 [0.52; 0.70]	-0.47 [-0.55; -0.37]	-0.42 [-0.51; -0.32]	-0.32 [-0.42; -0.21]	-0.43 [-0.52; -0.33]	-0.35 [-0.45; -0.24]	-0.37 [-0.47; -0.26]	-0.22 [-0.33; -0.11]	-0.10 [-0.22; 0.01]

The variables of interest are listed from left to right according to the decreasing importance of the expected correlation with the score. A negative sign of the correlation coefficient should be interpreted as an inverse relationship of the association

Interpretation of correlation coefficient: $r > 0.91$: excellent; $0.71 < r \leq 0.9$: good; $0.51 < r \leq 0.70$: moderate; $0.31 < r \leq 0.50$: weak; no correlation if $r < 0.30$ [34, 35]

M-SFS Modified Spinal Function Sort; SFS Spinal Function Sort; BPI Body mass index; BPI Brief Pain Inventory; HADS-A Hospital Anxiety and Depression Scale-Anxiety subscale; HADS-D Hospital Anxiety and Depression Scale-Depression subscale; TSK Tampa Scale for Kinesiophobia. PCS Pain Catastrophizing Scale

*Spearman correlation

#Hypotheses confirmed

the SFS with scores < 100 points, M-SFS scores < 48 points should be interpreted as a minimal perceived work demand capacity [8]. After the determination of the M-SFS score thresholds, M-SFS scores of 58, 64, and 70 points should be interpreted as the perceived capacity to carry maximal loads of 10, 15, and 25 kg, respectively.

The correspondence of the M-SFS score with work demand categories adds value to the questionnaire in clinical practice. It first provides an additional element for assessing the consistent nature of the complaints, particularly if the score is minimal. When it is used with observed measures such as the PILEtest, it can be used to assess the accordance of the patient's perceived physical ability with the maximal lifting abilities. This can identify potential barriers to rehabilitation when patients underestimate their abilities. On the other hand, when patients overestimate their capacities, it can help tailor rehabilitation measures by increasing attention to the adjustment of load levels during fitness training. Its predictive value for non-return to work, as was found for the SFS [6], would need further analysis.

The reliability we found for the M-SFS in French was excellent, with an ICC of 0.95, in accordance with that found in the German version by Trippolini et al. [11]. For the 50 patients who participated in the reliability study, the mean difference between the test and retest questionnaires was low and the variances proportionally related to the score were in accordance with the development study [11].

The internal consistency was high (Cronbach α , 0.93), comparable with that found during the validation of the German version [11].

Eight of ten hypotheses were accepted and the validity of the M-SFS questionnaire was confirmed. The criterion validity of the M-SFS was confirmed, as we found a high correlation between the M-SFS and SFS scores. This confirms the correlation of 0.89 published in the validation study [11]. We found a high consistency in the levels of perceived physical work capacity obtained with the two questionnaires. Concerning construct validity, a medium correlation of 0.61 was found with the PILE test. This confirms our hypothesis and represents higher correlations than the correlations presented by Trippolini et al. [11] who employed three lifting tasks during a FCE ($r = 0.43$ – 0.59). The higher correlations we found may be explained by the difference in the instructions between the PILE and FCE lifting tests. In the former test, patients are asked to give their best possible performance, whereas during FCE, the tests are stopped by the assessor when a maximal effort is achieved. The PILE test performances may thus correspond more to patients' perceived abilities, whereas the FCE performances are more influenced by the examiner's instructions. For the other variables, the correlations we find with the M-SFS were weak or absent, confirming the construct of the questionnaire mainly addressing patients' perceptions to achieve work-related

tasks. Weak correlations were found with the BPI severity and interference scale scores. We found no other study that assessed the correlation between the M-SFS and BPI scores. Our results are in accordance with the reported correlations with the VAS score for pain of -0.33 by Borloz et al. in a validation study of the SFS [5] and the correlation of -0.25 in the validation study of the Hand Function Sort (HFS) [12]. Weak correlations with HADS scores were also found and correspond to the correlation published by Borloz in the SFS validation study [5] or in another study on the association of psychological variables with perceived functional impairments in chronic low back pain patients [31]. The weak correlations found with the TSK and PCS are of the same magnitude, in accordance with other studies that showed associations between perceived physical function and catastrophic thoughts or kinesiophobia [6, 30, 31, 42, 43]. No correlation was found with age. Our population is mainly composed of men, 80% of whom were between 28.2 and 58.9 years old. It is possible that the small range of ages in our collective did not make it possible to find a correlation between age and perception to achieve work-related tasks. No correlation was found with BMI which could be explained by the fact that our population of workers coming from the construction or industrial sector may perceive their physical abilities to be higher than another more sedentary population.

This study has some limitations. First, our patient sample was mainly composed of men who sought for late rehabilitation after trauma and may limit the generalization of our results. Second, our choice to determine thresholds for the M-SFS score from the SFS score may be controversial, as it was based on the SFS questionnaire from which it was developed. Both questionnaires are different as the M-SFS contains items on the ability to maintain a posture for example. The high correlation between the two scores in our population and in another population [11] indicates they are very close however. Another method to determine thresholds based on observed measures may be another option. In our patient population, another method based on observed measures would not have been realistic as it is frequently observed discrepancies between perceived functional abilities and observed measurements [44]. Other methods of measurements based on FCEs may find other results in the future.

This is the first study to examine the responsiveness of the M-SFS score. As the correlation between the M-SFS score change and PGIC was low, we could not precisely determine an anchor-based MCID [38]. The PGIC measure we used at the end of our program may have been too general and explains this low correlation. Other research studies with more specific questions such as changes in physical capacities may find stronger associations [45] and help determine a

valid anchor-based MCID. On the basis of the measurement distribution method, we found an SEM of 3.9 points and an SDC of 10.8 points and the Limits of agreement were between -9.7 and 11.7 points. As we could not determine an anchor-based MCID, we propose that a difference of 11 points be considered as not due to measurement error.

Conclusions

Our study proposes thresholds corresponding to the perceived levels of physical work demand for the M-SFS score. By extending the scope of the psychometric qualities of the M-SFS, the validity and clinical utility of the M-SFS were increased. Other analyses are needed to determine the responsiveness of the questionnaire.

Acknowledgements We thank the nursing team who collected the questionnaires and the research assistants at the Institute for Research in Rehabilitation at Clinique Romande de Réadaptation, Sion, for their support in the data acquisition.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval All the participants gave their informed consent to the access and use of their clinical data for the present study. The protocol was approved by the ethics committee of the local medical association (Commission Cantonale Valaisanne d’Ethique Médicale, CCVEM 034/12). The study was performed in accordance with the ethical standards listed in the 2008 Declaration of Helsinki.

References

1. Black N. Patient reported outcome measures could help transform healthcare. *BMJ (Clinical research ed)*. 2013;346:f167.
2. Bandura A. Self-efficacy: the exercise of control. New York: W H Freeman/Times Books/ Henry Holt & Co; 1997.
3. Holden G. The relationship of self-efficacy appraisals to subsequent health related outcomes: a meta-analysis. *Soc Work Health Care*. 1991;16(1):53–93.
4. Matheson LN, Matheson ML, Grant J. Development of a measure of perceived functional ability. *J Occup Rehabil*. 1993;3(1):15–30.
5. Borloz S, Trippolini MA, Ballabeni P, Luthi F, Deriaz O. Cross-cultural adaptation, reliability, internal consistency and validation of the Spinal Function Sort (SFS) for French- and German-speaking patients with back complaints. *J Occup Rehabil*. 2012;22(3):387–393.
6. Oesch PR, Hilfiker R, Kool JP, Bachmann S, Hagen KB. Perceived functional ability assessed with the spinal function sort: is it valid for European rehabilitation settings in patients with non-specific non-acute low back pain? *Eur Spine J Off Publ Eur Spine Soc Eur Spinal Deform Soc Eur Sect Cerv Spine Res Soc*. 2010;19(9):1527–1533.

7. United States Department of Labor. United States Employment Service, and the North Carolina Occupational Analysis Field Center. Dictionary of Occupational Titles (DOT): Revised Fourth Edition, 1991.
8. Matheson LN. History, design characteristics, and uses of the pictorial activity and task sorts. *J Occup Rehabil.* 2004;14(3):175–195.
9. Swiss association of rehabilitation 6210 Sursee. Available from: <https://www.sar-reha.ch/fr/contact.html>.
10. Janssen S, Trippolini MA, Hilfiker R, Oesch P. Development of a modified version of the spinal function sort (M-SFS): a mixed method approach. *J Occup Rehabil.* 2016;26(3):253–263.
11. Trippolini MA, Janssen S, Hilfiker R, Oesch P. Measurement properties of the modified spinal function sort (M-SFS): is it reliable and valid in workers with chronic musculoskeletal pain? *J Occup Rehabil.* 2018;28(2):322–331.
12. Konzelmann M, Burrus C, Hilfiker R, Rivier G, Deriaz O, Luthi F. Cross-cultural adaptation, reliability, internal consistency and validation of the Hand Function Sort (HFS(c)) for French speaking patients with upper limb complaints. *J Occup Rehabil.* 2015;25(1):18–24.
13. Matheson LN, Kaskutas VK, Mada D. Development and construct validation of the Hand Function Sort. *J Occup Rehabil.* 2001;11(2):75–86.
14. Baker SP, O'Neill B, Haddon W Jr, Long WB. The injury severity score: a method for describing patients with multiple injuries and evaluating emergency care. *J Trauma.* 1974;14(3):187–196.
15. Cleeland CS, Ryan KM. Pain assessment: global use of the Brief Pain Inventory. *Ann Acad Med Singapore.* 1994;23(2):129–138.
16. Tan G, Jensen MP, Thornby JI, Shanti BF. Validation of the Brief Pain Inventory for chronic nonmalignant pain. *J Pain Off J Am Pain Soc.* 2004;5(2):133–137.
17. Razavi D, Delvaux N, Farvacques C, Robaye E. Validation de la version française du HADS dans une population de patients cancéreux hospitalisés. [Validation of the French version of the Hospital Anxiety and Depression Scale (HADS) in a population of hospitalized cancer patients]. *Rev Psychol Appl.* 1989;39(4):295–307.
18. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. *Acta Psychiatr Scand.* 1983;67(6):361–370.
19. Sullivan MJL, Bishop SR, Pivik J. The pain catastrophizing scale: development and validation. *Psychol Assess.* 1995;7(4):524–532.
20. Kori SH, Miller RP, Todd DD. Kinesiophobia: a new view of chronic pain behavior. *Pain Manag.* 1990;3:35–43.
21. Mayer TG, Barnes D, Kishino ND, Nichols G, Gatchel RJ, Mayer H, et al. Progressive isoinertial lifting evaluation. I. A standardized protocol and normative database. *Spine.* 1988;13(9):993–997.
22. Wright EM, Royston P. Calculating reference intervals for laboratory measurements. *Stat Methods Med Res.* 1999;8(2):93–112.
23. Bland JM, Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet.* 1986;327(8476):307–310.
24. Koo TK, Li MY. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med.* 2016;15(2):155–163.
25. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1996;1(1):30–46.
26. Cronbach LJ. Coefficient alpha and the internal structure of tests. *Psychometrika.* 1951;16(3):297–334.
27. Taber KS. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ.* 2018;48(6):1273–1296.
28. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: an international Delphi study. *Qual Life Res Int J Qual Life Asp Treat Care Rehabil.* 2010;19(4):539–549.
29. Terwee CB, Bot SD, de Boer MR, van der Windt DA, 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.
30. Lundberg M, Frennered K, Hägg O, Styf J. The impact of fear-avoidance model variables on disability in patients with specific or nonspecific chronic low back pain. *Spine.* 2011;36(19):1547–1553.
31. Thomas EN, Pers YM, Mercier G, Cambiere JP, Frasson N, Ster F, et al. The importance of fear, beliefs, catastrophizing and kinesiophobia in chronic low back pain rehabilitation. *Ann Phys Rehabil Med.* 2010;53(1):3–14.
32. Turcotte M, Schellenberg G. A portrait of seniors in Canada. Ottawa: Minister of Industry; 2007.
33. Giuli C, Papa R, Bevilacqua R, Felici E, Gagliardi C, Marcellini F, et al. Correlates of perceived health related quality of life in obese, overweight and normal weight older adults: an observational study. *BMC Public Health.* 2014;14(1):35.
34. Evans JD. Straightforward statistics for the behavioral sciences. Pacific Grove: Brooks/Cole Pub. Co.; 1996.
35. Fermanian J. Validation of assessment scales in physical medicine and rehabilitation: how are psychometric properties determined? *Ann Readapt Med Phys.* 2005;48(6):281–287.
36. McHorney CA, Tarlov AR. Individual-patient monitoring in clinical practice: are available health status surveys adequate? *Qual Life Res Int J Qual Life Asp Treat Care Rehabil.* 1995;4(4):293–307.
37. Dworkin RH, Turk DC, Farrar JT, Haythornthwaite JA, Jensen MP, Katz NP, et al. Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain.* 2005;113(1):9–19.
38. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol.* 2008;61(2):102–109.
39. Fluss R, Faraggi D, Reiser B. Estimation of the Youden Index and its associated cutoff point. *Biom J.* 2005;47(4):458–472.
40. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol.* 2006;59(10):1033–1039.
41. van Kampen DA, Willems WJ, van Beers LW, Castelein RM, Scholtes VA, Terwee CB. Determination and comparison of the smallest detectable change (SDC) and the minimal important change (MIC) of four-shoulder patient-reported outcome measures (PROMs). *J Orthop Surg Res.* 2013;8:40.
42. Tichonova A, Rimdeikienė I, Petruševičienė D, Lendraitienė E. The relationship between pain catastrophizing, kinesiophobia and subjective knee function during rehabilitation following anterior cruciate ligament reconstruction and meniscectomy: a pilot study. *Medicina (Kaunas).* 2016;52(4):229–37.
43. Wolfensberger A, Vuistiner P, Konzelmann M, Plomb-Holmes C, Leger B, Luthi F. Clinician and patient-reported outcomes are associated with psychological factors in patients with chronic shoulder pain. *Clin Orthop Relat Res.* 2016;474(9):2030–2039.
44. Tuscher J, Burrus C, Vuistiner P, Leger B, Rivier G, Luthi F. Predictive value of the fear-avoidance model on functional capacity evaluation. *J Occup Rehabil.* 2018;21:513–522.
45. Scott W, McCracken LM. Patients' impression of change following treatment for chronic pain: global, specific, a single dimension, or many? *J Pain Off J Am Pain Soc.* 2015;16(6):518–526.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.