

Title

RELIABILITY AND VALIDITY OF THE WORK ROLE FUNCTIONING QUESTIONNAIRE (SPANISH VERSION).

Authors

Jose M Ramada^{1,2,3}, Consol Serra^{1,2,3}, Benjamin C Amick III^{4,5}, Femke I Abma⁶, Juan R Castaño⁷, Gemma Pidemunt⁸, Ute Bültmann⁶, George L Delclos^{1,3,4}.

Affiliations

- 1 CiSAL - Center for Research in Occupational Health, University Pompeu Fabra. PRBB Building. Doctor Aiguader, 88. 08003 - Barcelona, Spain.
- 2 Occupational Health Service. Parc de Salut MAR. Hospital del Mar. Passeig Marítim, 25-29. 08003 - Barcelona, Spain.
- 3 CIBER of Epidemiology and Public Health (CIBERESP).
- 4 Southwest Center for Occupational and Environmental Health, The University of Texas School of Public Health. 1200 Pressler Street. Houston, Texas 77030, USA.
- 5 Institute for work & Health. 80 University Avenue, Toronto, Ontario, Canada.
- 6 Department of Health Sciences, Work & Health, University Medical Center Groningen, University of Groningen. Antonius Deusinglaan 1, HPC FA10 9713 AV Groningen, The Netherlands.
- 7 Psychiatry Service. Parc de Salut MAR. Hospital del Mar. Passeig Marítim, 25-29. 08003 - Barcelona, Spain.
- 8 Orthopedic Surgery and Traumatology Service. Parc de Salut MAR. Hospital del Mar. Passeig Marítim, 25-29. 08003 - Barcelona, Spain.

Correspondence

José M^a Ramada

CISAL - Centre for Research in Occupational Health. University Pompeu Fabra
PRBB Building. Dr. Aiguader, 88. 08003 - Barcelona. Spain

Email address: jramada@parcdesalutmar.cat

Telephone number: +34 933160875

Fax number: +34 933160410

ABSTRACT

Purpose: Recently, the cross-cultural adaptation of the Work Role Functioning Questionnaire to Spanish was carried out, achieving satisfactory psychometric properties. Now we examined the reliability and validity of the adapted Spanish version (WRFQ-SpV) in a general working population with and without (physical and mental) health issues to evaluate its measurement properties.

Methods: A cross-sectional study was conducted among active workers. For reliability, we calculated Cronbach alphas to assess 'internal consistency', and the standard error of measurement (SEM) to evaluate 'measurement error'. We assessed the 'structural validity' through confirmatory factor analyses and 'construct validity' by means of hypotheses testing. The consensus-based standard for the selection of health status measurement instruments (COSMIN) taxonomy were used in the design of the study.

Results: A total of 455 workers completed the questionnaire. It showed excellent internal consistency ($\alpha=0.98$). The SEM for the overall scale was 7.10. The original five factor structure reflected fair dimensionality of the construct (Chi-square, 1445.8; 314 degrees of freedom; RMSEA=0.08; CFI > 0.95 and WRMR > 0.90). For construct validity, all hypotheses were confirmed differentiating groups with different jobs, health conditions and ages. Moderate to strong correlations were found between WRFQ-SpV and a related construct (work ability).

Conclusions: Our study provides evidence of the reliability and validity of the WRFQ-SpV to measure health-related work functioning in day-to-day practice and research in occupational health care and the rehabilitation of disabled workers. It should be useful to monitor improvements in work functioning after implementing rehabilitation and/or accommodation programs. Longitudinal studies are needed to assess the responsiveness of the questionnaire.

Key terms: validity; reliability; work-functioning instrument; measurement instrument; psychometric properties; self-report.

INTRODUCTION

Increasing life expectancy in developed countries and delayed retirement age are increasing the overall age of the workforce. Aging workers are more likely to have chronic health issues and a certain degree of disability, but most are able to maintain job competence with some workplace adjustments and/or rehabilitation programs [1-4]. Also, there is evidence showing that work has positive health effects when conditions are reasonably acceptable; therefore, promoting an active working life is recommendable [5,6].

Quality work functioning tools are required to obtain valid measurements to evaluate the impact of health on work functioning and to monitor the extent to which workers improve their ability to meet job demands after a rehabilitation or accommodation program. This will enable healthcare professionals, human resources managers, employers and other stakeholders to support an active and healthy labor force. Moreover, valid outcome measures are needed to assess how workers function at work over the course of their job careers and the existing continuum between working successfully at one extreme and disability and work-absence at the other [7].

There are a number of tools to measure constructs related with self-perceived work functioning, including the Functional Status Index [8], the Work Productivity and Activity Impairment Questionnaire [9], the Health and Labor Questionnaire [10], the Endicott Work Productivity Scale [11], the Work Ability Index [12], the Role-based Performance Scale [13], the Stanford Presenteeism Scale [14], the Work Instability Scale [15], and the Work Activity Limitations Scale [16].

Since 'being present at work without being able to meet job demands' (presenteeism) [17] is not the same as 'performing work demands successfully', a series of work-role specific functioning questionnaires were developed in the 2000's. Among those, there are different versions of the Work Limitations Questionnaire [18] and the Work Role Functioning Questionnaire (WRFQ) [19].

The WRFQ measures perceived difficulties to perform the job due to health problems. This questionnaire is a generic instrument conceptually developed to represent a wide range of health conditions and work demands. Furthermore, it is freely available in the literature for professionals and researchers. Recently, it has been successfully translated, adapted and validated to be used in different contexts (e.g. Canadian French [20], Brazilian Portuguese [21], Dutch [7,22] and Spanish spoken in Spain [23]). These versions have shown good psychometric properties in different populations.

Before using an adapted instrument it is important to assess its measurement properties [24]. Recent reviews have shown that health-related work outcome measures and health-related work functioning instruments need better validation studies to make them more meaningful for researchers, practitioners and patients [25,26]. The cross-cultural adaptation of the WRFQ to Spanish was recently carried out, and the questionnaire showed good test-retest reliability (intraclass correlation coefficients, ICCs between 0.77 and 0.93 for all subscales) [23], but further assessment of the validity and reliability of the questionnaire in a larger sample was recommended.

Therefore, the objective of this study was to examine the reliability and validity of the Spanish version of the WRFQ (WRFQ-SpV) in a general working population of Barcelona (Spain), with and without (physical and mental) health issues.

METHODS

Procedures and sample characteristics

After carrying out the cross-cultural adaptation of the WRFQ to Spanish spoken in Spain [23], it was necessary to assess its reliability and validity in a larger sample so that it could be used in both occupational health and rehabilitation settings; hence a cross-sectional study was conducted among active workers of a general working population of Barcelona (Spain). The consensus-based standards for the selection of health status measurement instruments (COSMIN) taxonomy was used in the study design [27-29].

Participants were recruited at a large public hospital in Barcelona, among patients, persons accompanying patients, hospital workers and other workers that were carrying out different duties at the hospital (ambulance drivers, bar tenders, kitchen and cleaning staff). Patients were recruited through the outpatient services of psychiatry, physical medicine and rehabilitation, orthopedic surgery and traumatology. The inclusion criteria were: 1) active workers of both sexes, working at least 10 hours per week in the past four weeks, 2) age 18 years and older, and 3) able to read and understand Spanish (the language of the questionnaire). Participants were excluded if they had plans to stop working within the following six months.

The study protocol and the informed consent process was reviewed and approved by the Clinical Research Ethical Committee of the Parc de Salut Mar (Barcelona). All participants received information about the study purpose and signed the informed consent to participate in it.

Measures

The WRFQ-SpV is a self-administered questionnaire containing 27 items grouped into 5 subscales reflecting different work demands: work scheduling, output, physical, mental and social demands [23]. The recall period is four weeks and each subscale is measured by the percentage of time in a working day the employee has difficulty performing those demands. Response options vary on a five-point scale: 0=all of the time (100%), 1=most of the time, 2=half of the time (50%), 3=some of the time, 4=none of the time (0%) and 5=does not apply to my job. For each subscale and for the overall scale, item scores were summed, divided by the number of items included in the subscale (or the overall scale), and then multiplied by 25 to obtain the scores, ranging from 0% (difficulty all the time) to 100% (no difficulty at any time). **Higher values indicate better work functioning (less disability at work).** The scores for "does not apply to my job" were transformed to missing values. Scales and/or subscales containing more than 20% missing values were set to missing.

All participants were invited to complete the WRFQ-SpV on paper, providing self-reported information on age, gender, level of education (primary, secondary, higher), job type (manual, non-manual, mixed), working hours and primary health condition (none, musculoskeletal, mental, others).

Three single items of the work ability index (WAI) [12] were included in the survey for a convenience subsample of participants, who voluntarily accepted to answer to these items. The first was the overall item 'current work ability compared with the life-time best', with a possible score of 0=completely unable to work to 10=work ability at its best. Recent studies showed that this overall single item highly correlates with the overall WAI score [30] and also showed the convergent validity and the similarity in results between the overall WAI scores and the scores of the overall single item of the WAI in large samples of participants [31]. Also, there is an increasing number of studies using the overall single item of the WAI to assess 'work ability' in different populations [7,30,32,33]. The other two items measure work ability in relation to physical and mental job demands, with a possible score of 1=very poor to 5=very good, and are questions already validated in the original version of the questionnaire [12].

Reliability is defined as the degree to which the measurement is free from measurement error and the extent to which scores for participants who have not changed are the same for repeated measurement under several conditions [27,35]. Since the test-retest reliability of the WRFQ was already assessed in our previous manuscript [23], in the current manuscript reliability was assessed calculating the standard error of measurement (SEM) and evaluating the internal consistency.

Validity of a questionnaire is the degree to which an instrument truly measures the construct it purposes to measure. Content validity was already assessed in our previous manuscript [23]. Criterion validity cannot be assessed for the WRFQ since 'Work Functioning' is a construct that has not a gold standard. Construct validity refers to whether the instrument provides the expected scores, based on existing knowledge about the construct [35]. There is an international consensus of experts [27-29] recommending to assess construct validity by means of 'hypotheses testing' and evaluating the 'structural validity'. Structural validity was assessed by means of a Confirmatory Factor Analysis (CFA); and four groups of hypotheses were formulated and tested.

Hypothesis 1, addressing health issues: 1a) Participants without health issues report higher scores on the overall scale of the WRFQ than those with health issues; 1b) Participants with physical health issues report the lowest score on the subscale of physical demands; 1c) Participants with mental health issues report the lowest score on the subscale of mental demands.

Hypothesis 2, addressing job types: Participants with physical health issues and manual job report a lower score on the WRFQ subscale of physical demands than those with physical health issues and non-manual or mixed jobs.

Hypothesis 3, addressing correlation between WRFQ scores and scores of a related construct (work ability): 3a) There are moderate to strong correlations between the score of the overall work ability item of the WAI (that measures a related construct) and the overall score of the WRFQ; 3b) There are moderate to strong correlations between the scores of the mental and physical demands items of the WAI and those of the subscales of physical and mental demands of the WRFQ.

Hypothesis 4, addressing age: Consistently with other studies finding that both, chronological and functional age, are associated with a decrease in work ability and/or work outcomes [42-46], there is a trend on the overall scores of the WRFQ showing worse work functioning with increasing age.

Statistical analysis

WRFQ-SpV mean scores, standard deviations (SD), median scores and ranges were calculated. Floor and ceiling effects were also explored. These effects occur when more than 15% of the participants' responses to a certain question cluster at the top or the bottom of the scale [34]. Since the original version of the WRFQ was developed for a working population with health problems [19], and our population contains a percentage of participants declaring no health issues, we carried out a sensitivity analysis of floor and ceiling effects, restricting the sample to only those participants reporting health problems to explore if there were differences in the presence of these effects due to the characteristic of the sample.

Participant scores were presented by job type (manual, non-manual, mixed), reported health issues (none, physical, mental) and groups of age (18-35 years, 36-45 years, 46-55 years, 56-65 years), assessing the statistical significance of the differences by means of the Kruskal Wallis H test (to compare median scores) and analysis of variance (ANOVA) to compare mean scores. Post-hoc paired analyses (comparing median or mean scores for each of the two groups) were performed to determine which group or groups were responsible for significant differences. When comparing median scores between two groups, Mann-Whitney test for two independent samples were used, and when comparing mean scores between two groups, t-Tests were used.

Internal consistency was assessed calculating Cronbach alpha coefficients considering appropriate values ≥ 0.70 [34]. The standard error of measurement (SEM) was calculated for a stable subgroup of participants (n=40) that completed the questionnaire twice in similar conditions, within an interval that varied from 7 to 15 days [35]. This subgroup of participants was composed of the first 40 participants of the study who completed the first round and accepted to complete the questionnaire a second time within this interval.

A CFA was conducted testing whether data collected in this general working population (N=455) had an adequate fit in the predetermined five factor model structure defined by the authors of the original questionnaire [19]. A four factor model structure was also tested because the Work Limitations Questionnaire [18], designed from the same pool of items to measure on-the-job impact of chronic health problems, has a structure with four factors (one of them named mental-interpersonal) and earlier studies [20,21,23] recommended caution when interpreting the internal consistency of the social demands subscale. Thus, we hypothesized it might be necessary to collapse the subscales of mental and social demands into a single factor of psychosocial demands with seven items.

Following recommendations in the literature regarding CFA, we did not use the standard maximum likelihood theory (applicable to continuous variables). Instead, we used the robust categorical least squares (applicable to categorical variables), based on the fact that the observed variables are measured on a Likert scale and the variables are approximately symmetrical [36-38].

Rhemtulla [36] suggests that when there is a minimum of five categorical variables in the response options, which is the case of the WRFQ, the CFA could also be assessed applying “the method of the standard theory of maximum likelihood” treating these variables as if they were continuous (but we would be at the limit of acceptance of this method). To verify the possible existence of differences depending on the method, calculations were performed applying both methods.

Chi-squared tests for goodness of fit, the root mean square error of approximation (RMSEA), the comparative fit index (CFI) and the weighed root mean residual (WRMR) were used to evaluate the models. Reference values for $RMSEA \leq 0.05$ indicating close fit, between 0.06 and 0.08, fair fit and between 0.09 and 0.1, mediocre fit. Reference values for $CFI \geq 0.95$ and $WRMR > 0.90$ for acceptance [39].

Correlations were evaluated for item-subscale, item-total, among subscales and subscale-total, using Pearson’s correlation coefficient (r), considering $r \geq 0.40$ as evidence of moderate or strong correlations [40,41].

For hypotheses testing, significance of the differences among groups were evaluated using the non-parametric Kruskal Wallis H test when comparing differences among median scores and analysis of the variance (ANOVA) when differences among mean scores were compared. Correlations between constructs were assessed using Pearson’s correlation coefficient (r) interpreting: $r < 0.4 =$ ‘weak’; $0.4 \leq r \leq 0.7 =$ ‘moderate and $r > 0.7 =$ ‘strong’ [41].

All analyses were performed with SPSS (Version 15.0. Chicago, IL; 2006) and Mplus (Version 7. Los Angeles, CA; 2012).

RESULTS

Sample characteristics. Four hundred fifty-five participants completed the WRFQ-SpV and were included in the analyses. All were active employees working an average of 39 hours per week ($SD=8.5$), mean age of 42 years ($SD=11$) and with different levels of education, job types and health issues (table 1).

Compared with the general Spanish working population, women and participants with higher educational level were overrepresented [47]. A subgroup of 181 participants also completed the WAI items [Supplementary materials (1)].

Table 2 shows the mean, SD and median scores for each WRFQ-SpV subscale and the overall scale. Mental and social demands subscales scored the highest mean and median, and the output demands subscale scored the lowest.

Floor effects were not found for any subscale, but ceiling effects were found for the subscales of work scheduling (20%), mental (29%) and social demands (31%), exceeding the 15% criterion [34]. A sensitivity analysis was carried out, restricting the sample to only those participants reporting health problems (n=299; 66% of the sample), and ceiling effects also appeared for the same subscales.

Reliability assessment: The SEMs were 7.1 for the overall score, 8.5 for work scheduling, 8.9 for output, 8.6 for physical, 10.6 for mental and 13.3 for social demands [Supplementary materials (2)]. Cronbach alpha coefficients were 0.98 for the overall scale and above 0.81 for all subscales (table 2).

Structural validity assessment: Fit was fair for the five factor model applying method of the robust categorical least squares for categorical variables (Chi-square, 1285.8; 314 degrees of freedom, $p < 0.001$) and mediocre for the four factor model (Chi-square, 1353.5; 318 degrees of freedom, $p < 0.001$). The resulting root mean square error of approximation (RMSEA) were 0.08 (CI90%= 0.07-0.08) and 0.09 (CI90%= 0.08-0.09) for the five and four factor models respectively. The comparative fit index (CFI) was 0.97 for both factor structures, and the weighed root mean residual (WRMR) were 1.5 and 1.6 for the five and four factor models respectively.

The results of the CFA applying the method of the standard theory of maximum likelihood for continuous variables also showed fair fit for the five factor model (Chi-square, 1445.8; 314 degrees of freedom, $p < 0.001$) and mediocre for the four factor model (Chi-square, 1546.2; 318 degrees of freedom, $p < 0.001$). RMSEA were 0.08 (CI90%= 0.08-0.09) and 0.09 (CI90%= 0.08 ; 0.09) for the five and four factor models respectively.

All Pearson's correlations (item-subscale, item-total, among subscales and subscale-total) were ≥ 0.40 (moderate to strong) and considered appropriate [41] (table 3).

Construct validity assessment by hypotheses testing: Participant median scores by job type (manual, non-manual, mixed) and existing health condition (none, physical, mental) are shown in table 4.

Table 4 and the post-hoc paired analyses, presented in Supplementary materials (3), showed that participants without health issues had statistically significant higher overall scores than those with mental health issues and those with physical health issues, confirming hypothesis 1a.

Participants with physical health issues showed the lowest score on the subscale of physical demands (median score=64), confirming hypothesis 1b (table 4).

Respondents with mental health issues obtained the lowest score on the subscale of mental demands (median score=63), confirming hypothesis 1c (table 4).

Confirming hypotheses 1a, 1b, 1c and 2 showed that the instrument has ability to differentiate between workers with and without physical or mental health problems.

Participants with physical health issues and manual job reported a statistically significant lower score on the WRFQ subscale of physical demands (median score=54; $p=0.021$) compared to those with physical health issues and non-manual (median score=67; $p=0.021$) or mixed jobs (median score=71; $p=0.021$), confirming hypothesis 2 [Supplementary material (4)].

A strong correlation was found between the overall WAI item score and the overall score of the WRFQ (0.71), and moderate correlations were found between the physical and mental demands items of the WAI and the WRFQ subscales of physical and mental demands (0.62 and 0.67 respectively). These correlations confirm hypothesis 3a and 3b (table 3).

Table 5 shows the mean scores and SD of the WRFQ-SpV by age groups for each subscale and the overall scale, revealing a trend of worse work functioning with increasing age in the overall scores. The

group of oldest participants (56-65 years) showed higher mean scores on work scheduling and mental demands than workers aged 46-55 years, and the main decrease in the mean scores for the oldest workers (56-65 years) appeared to be on physical demands, although these differences were not significant.

Post-hoc paired analyses, exposed in Supplementary material (5), showed that the differences in the median scores for the overall scale between the group of 18-35 years and the group of 36-45 were not significant ($p=0.267$). They appeared to be significant for the youngest group (18-35 years) compared with the group of 46-55 years ($p=0.032$) and the group of 56-65 years ($p=0.005$). Decreases in physical and social demands appeared to be significant in these paired analyses, except for the comparison of physical demands between the groups of 18-35 years and 36-45 years. Other post-hoc paired analyses evaluating differences in the median scores between groups of 36-45 years, 46-55 years and 56-65 years did not show statistically significant differences, except for the decrease in physical demands, concluding that hypothesis 4 was (partially) confirmed.

DISCUSSION

The WRFQ-SpV is a brief and easily interpretable questionnaire to measure health-related work functioning, that is freely available in different languages [7,20,21,23]. Results indicate it has excellent internal consistency for the overall scale (Cronbach alpha = 0.98) and for all subscales (Cronbach alpha > 0.81) and also showed adequate structural and construct validity in the study population, where healthy workers and also workers with different health issues, job types, working hours, levels of education, sexes and ages have participated.

In a recent study conducted by our group [23], the Spanish version of the questionnaire showed adequate cross-cultural validity and test-retest reliability. Those results, together with our current results, support that the WRFQ-SpV is a reliable and valid instrument to measure work functioning.

Percentages of missing items in the responses for the overall scale did not exceed 3%, but for the subscale of physical demands we reported 30% missing items. This can be explained by the frequent use in this

subscale of the response option 'does not apply to my job' that, according to the questionnaire use instructions, had to be transformed to missing values. The subscale of physical demands contained 3,5% (n=16) real missing items.

The observed ceiling effects for the subscales work scheduling, mental and social demands indicate a lack of discriminative ability of certain items when differentiating workers with good health and working. These results are consistent with other studies that also found ceiling effects for the subscales of mental and social demands [7,20,23]. According to de Vet et al. [35] one explanation could be that, differently from the original version, that was developed for a working population with health problems [19], we included a percentage of participants (34%) who reported not having any health problem. To explore this possibility, we carried out a sensitivity analysis, restricting the sample to only those reporting health problems (n=299; 66% of the sample), and ceiling effects also appeared for the same subscales.

Our results on reliability were consistent with other studies. We found similar SEM and Cronbach alpha for the overall scale and for all subscales when compared to the validated Dutch version [7].

The CFA showed a better fit for the original 5 factors structure compared with the 4 factors structure, applying the method of "the robust categorical least squares", but Rhemtulla [36] suggests that the CFA could also be assessed applying "the method of the standard theory of maximum likelihood" treating these variables as if they were continuous. To verify the possible existence of differences depending on the method, calculations were performed applying both methods, obtaining similar results.

Based on participant suggestions and the literature, the Dutch version added a new subscale (flexibility demands) not present in the original version, modifying five items after an exploratory factor analysis [7]. This feature did not arise during the interviews conducted in the translation process and cross-cultural adaptation to Spanish spoken in Spain [23]. However, we do not know whether this could hinder comparisons between different countries or cultures.

The hypotheses that addressed health issues were confirmed, verifying that both workers reporting physical and mental health problems showed lower scores than those who reported no health problems;

workers with physical and mental health problems showed the lowest scores for physical and mental demands respectively and moderate to strong correlations were found with the three single WAI items, providing evidence of adequate construct validity. Again, these results are consistent with the validation study of the Dutch version [7], comparing the correlations between the WRFQ and other related constructs ('work ability', 'work productivity', 'work engagement' and 'work involvement').

This study shows that the WRFQ-SpV is able to distinguish among different groups of job types, health conditions and ages. For the overall scores, a significant trend for worse work functioning scores with increasing age ($p=0.026$) was found (hypothesis 4), consistent with other studies finding that both, chronological and functional age, are associated with a decrease in work ability and/or work outcomes [42-46]. This trend was clear for physical functioning, output and social demands, but was not so clearly present for work scheduling and mental demands, indicating the construct 'work functioning' is not entirely explained by chronological age [7,45]. This could mean that there are other qualities that older workers might bring into the workplace (e.g. efficiency, management and scheduling, expertise and experience and other more qualitative aspects) and other age concepts (e.g. functional age, personal perceived age or biological age) should be analyzed. The "healthy worker effect" could also be playing a role explaining why this trend of worse work functioning does not appear for certain subscales. Chronically ill and disabled workers are usually excluded from employment [48], and therefore the older workers who remain in their jobs are likely those with better health and therefore better work functioning.

Consistent with prior WRFQ analyses, correlations for item-subscale, item-total, among subscales and subscale-total were evaluated and found to be appropriate [7,20,21,23].

Our measurement property results are consistent with the Canadian French and Brazilian Portuguese WRFQ cross-cultural adaptation [20,21]. In both studies, the validation study was carried out in populations with musculoskeletal disorders, concluding that the adapted WRFQ versions were reliable. Only one study was found that assessed the responsiveness of the WRFQ, conducted in a relatively small, stable and healthy population, without any intervention between the first and the second test, to that the

number of workers reporting change was very small [7]. Other longitudinal studies with larger samples, designed to expect changes in participants, should be carried out to assess WRFQ responsiveness.

In conclusion, our study provides evidence of the reliability and validity of the WRFQ-SpV to measure health-related work functioning in day-to-day practice and research in occupational health care and the rehabilitation of disabled workers. It should be useful to monitor improvements in work functioning after implementing rehabilitation and/or accommodation programs. Longitudinal studies are needed to assess the responsiveness of the questionnaire.

Acknowledgements

We want to thank the nurses of the PSMAR Josefina Pi-Sunyer, Joan Mirabent, Victoria Abad, Silvia Rosado and Antonia Ruiz for their contribution engaging participants for the study. A word of gratitude to Roy Stewart from the University of Groningen for his kind helps with the statistical analysis. And gratefully acknowledge the help of David Domenech, Ángeles Calaforra, Ram Dulthummon, José, Borja and Ángeles Ramada for their gentle collaboration in all the logistics of the questionnaires and design of the database. This project has been supported by a grant from "Fondo de Investigaciones Sanitarias (FIS: PI12/02556), Instituto de Salud Carlos III, Subdirección General de Evaluación y Fomento de la Investigación, Ministerio de Ciencia e Innovación, Gobierno de España". None of the study authors have competing interests.

CONFLICT OF INTEREST

Author Jose M Ramada declares that he has no conflict of interest.

Author Consol Serra declares that she has no conflict of interest.

Author Benjamin C Amick III declares that he has no conflict of interest.

Author Femke I Abma declares that he has no conflict of interest.

Author Juan R Castaño declares that he has no conflict of interest.

Author Gemma Pidemunt declares that she has no conflict of interest.

Author Ute Bültmann declares that she has no conflict of interest.

Author George L Delclos declares that he has no conflict of interest.

This project has been supported by a grant from "Fondo de Investigaciones Sanitarias (FIS: PI12/02556), Instituto de Salud Carlos III, Subdirección General de Evaluación y Fomento de la Investigación, Ministerio de Ciencia e Innovación, Gobierno de España".

REFERENCES

1. Ross D. Ageing and work: an overview. *Occup Med (Lond)*. 2010;60:169-71.
2. Hairault JO, Langot F, Sopraseuth T. Distance to retirement and older workers employment: the case for delaying the retirement age. *J Eur Economic Assoc*. 2010;8:1034-76.
3. Macdonald EB, Sanati KA. Occupational health services now and in the future: the need for a paradigm shift. *J Occup Environ Med*. 2010;52:1273-7.
4. Sampere M, Gimeno D, Serra C, Plana M, Martínez JM, Delclos GL, Benavides FG. Organizational return to work support and sick leave duration: a cohort of Spanish workers with a long-term non-work-related sick leave episode. *J Occup Environ Med*. 2011; 53:674-9.
5. Wadell G, Burton T, Aylward M. Work and common health problems. *J Insur Med*. 2007;39:109-20.
6. Butterworth P, Leach LS, Strazdins L, Olesen SC, Rodgers B, Broom DH. The psychosocial quality of work determines whether employment has benefits for mental health: results from a longitudinal national household panel survey. *Occup Environ Med*. 2011;68:806-12.
7. Abma FI, van der Klink JJ, Bültmann U. The work role functioning questionnaire 2.0 (Dutch version): examination of its reliability, validity and responsiveness in the general working population. *J Occup Rehabil*. 2013;23:135-47.
8. Jette AM. The Functional Status Index: reliability and validity of a self-report functional disability measure. *J Rheumatol Suppl*. 1987;14 Suppl 15:15-21.
9. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. *Pharmacoeconomics*. 1993;4:353-65.

10. Van Roijen L, Essink-Bot ML, Koopmanschap MA, Bonsel G, Rutten FF. Labor and health status in economic evaluation of health care: the Health and Labor Questionnaire. *Int J Technol Assess Health Care*. 1996;12:405-15.
11. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. *Psychopharmacol Bull*. 1997;33:13-6.
12. Tuomi K, Ilmarinen J, Jahkola A, Katajarinne L, Tulkki A. Work Ability Index. In: Rautoja S, Pietiläinen R, editors. Finland: K-Print Oy Vantaa, Finnish Institute of Occupational Health; 1998.
13. Welbourne TM, Johnson DE, Erez A. Role-based Performance Scale: Validity Analysis of A Theory-Based Measure. *Acad Manage J*. 1998;41:5540-55.
14. Koopman C, Pelletier KR, Murray JF, Sharda CE, Berger ML, Turpin RS, et al. Stanford presenteeism scale: health status and employee productivity. *J Occup Environ Med*. 2002;44:14-20.
15. Gilworth G, Chamberlain MA, Harvey A, Woodhouse A, Smith J, Smyth MG, Tennant A. Development of a work instability scale for rheumatoid arthritis. *Arthritis Rheum*. 2003;49:349-54.
16. Gignac MA, Badley EM, Lacaille D, Cott CC, Adam P, Anis AH. Managing arthritis and employment: making arthritis-related work changes as a means of adaptation. *Arthritis Rheum*. 2004;51:909-16.
17. Schultz AB, Edington DW. Employee health and presenteeism: a systematic review. *J Occup Rehabil*. 2007;17:547-79.
18. Lerner D, Amick BC 3rd, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. *Med Care*. 2001;39:72-85.
19. Amick BC III, Lerner D, Rogers WH, Rooney T, Katz JN. A review of health-related work outcome measures and their uses and recommended measures. *Spine*. 2000;25:3152-60.

20. Durand MJ, Vachon B, Hong QN, Imbeau D, Amick BC III, Loisel P. The cross-cultural adaptation of the work role functioning questionnaire in Canadian French. *Int J Rehabil Res.* 2004;27:261-8.
21. Gallasch CH, Alexandre NM, Amick B 3rd. Cross-cultural adaptation, reliability, and validity of the work role functioning questionnaire to Brazilian Portuguese. *J Occup Rehabil.* 2007;17:701-11.
22. Abma FI, Amick III BC, Brouwer S, van der Klink JJ, Bültmann U. The cross-cultural adaptation of the work role functioning questionnaire to Dutch. *Work.* 2012;43:203-10.
23. Ramada JM, Serra C, Amick BC III, Castaño JR, Delclos GL. Cross-cultural adaptation of the work role functioning questionnaire to Spanish spoken in Spain. *J Occup Rehabil.* 2013; 23:566-75.
24. Terwee CB, Mokkink LB, Knol DL, Ostelo RW, Bouter LM, de Vet HC. Rating the methodological quality in systematic reviews of studies on measurement properties: a scoring system for the COSMIN checklist. *Qual Life Res.* 2011;21:651-7.
25. Williams RM, Schmuck G, Allwood S, et al. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: a systematic review. *J Occup Rehabil.* 2007;17:504-21.
26. Abma FI, van der Klink JJ, Terwee CB, Amick BC 3rd, Bültmann U. Evaluation of the measurement properties of self-reported health-related work-functioning instruments among workers with common mental disorders. *Scand J Work Environ Health.* 2012;38:5-18.
27. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *J Clin Epidemiol.* 2010;63:737-45.
28. Mokkink LB, Terwee CB, Knol DL, Stratford PW, Alonso J, Patrick DL, et al. The COSMIN checklist for evaluating the methodological quality of studies on measurement properties: A clarification of its content. *BMC Med Res Methodol.* 2010;10:22.

29. Mokking 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.* 2010;19:539-49.
30. Ahlstrom L, Grimby-Ekman A, Hagberg M, Dellve L. The work ability index and single-item question: associations with sick leave, symptoms, and health – a prospective study of women on long-term sick leave. *Scand J Work Environ Health.* 2010;36:404-12.
31. El Fassi M, Bocquet V, Majery N, Lair ML, Couffignal S, Mairiaux P. Work ability assessment in a worker population: comparison and determinants of Work Ability Index and Work Ability score. *BMC Public Health.* 2013;13:305.
32. Gould R, Ilmarinen J, Järvisalo J, Koskinen S. Dimensions of work ability. Results of the Health 2000 Survey [Internet]. Finnish Centre for Pensions (ETK). 2008 [cited 2013 Sep 13]; [pp. 25-34]; Available from: http://www.etk.fi/fi/gateway/PTARGS_0_2712_459_440_3034_43/http%3B/content.etk.fi%3B7087/publishedcontent/publish/etkfi/fi/julkaisut/tutkimusjulkaisut/erillisjulkaisut/dimensions_of_work_ability_7.pdf
33. Kuijer PP, Gouttebarge V, Wind H, van Duivenbooden C, Sluiter JK, Frings-Dresen MH: Prognostic value of self-reported work ability and performance based lifting tests for sustainable return to work among construction workers. *Scand J Work Environ Health.* 2012;38:600-3.
34. 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;60:34-42.
35. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. *Mesurement in medicine: A practical guide.* 1st ed. Cambridge, UK: The University Press Cambridge, 2011.

36. Rhemtulla M, Brosseau-Liard PÉ, Savalei V. When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychol Methods*. 2012;17:354-73.
37. Brown TA. *Confirmatory factor analyses for applied research*. 1st ed. New York, USA: Guilford Press, 2006.
38. Thompson, B. *Exploratory and confirmatory factor analysis: Understanding concepts and applications*. Washington, DC, USA: American Psychological Association, 2004.
39. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modeling and confirmatory factor analysis results: a review. *J Educ Research*. 2010;99:323-38.
40. Streiner DL, Norman GR. *Health measurement scales: a practical guide to their development and use*. 4th ed. New York: Oxford University Press Inc.; 2008.
41. Evans JD. *Straightforward statistics for the behavioral sciences*. Pacific Grove, CA: Brooks/Cole Pub. Co.; 1996.
42. Ilmarinen J, Tuomi K, Klockars M. Changes in the work ability of active employees over an 11-year period. *Scand J Work Environ Health*. 1997;23 Suppl 1:49-57.
43. Costa G, Sartori S. Ageing, working hours and work ability. *Ergonomics*. 2007;50:1914-30.
44. van den Berg TI, Elders LA, de Zwart BC, Burdorf A. The effects of work-related and individual factors on the Work Ability Index: a systematic review. *Occup Environ Med*. 2009;66:211-20.
45. Koolhaas W, van der Klink JJ, Groothoff JW, Brouwer S. Towards a sustainable healthy working life: associations between chronological age, functional age and work outcomes. *Eur J Public Health*. 2012;22:424-9.

46. Soer R, Brouwer S, Geertzen JH, van der Schans CP, Groothoff JW, Reneman MF. Decline of functional capacity in healthy aging workers. *Arch Phys Med Rehabil.* 2012;93:2326-32.
47. Observatorio Estatal de Condiciones de Trabajo [Internet]. VI Encuesta Nacional de Condiciones de Trabajo. [cited 2013]; [about 117 screen, pages 21-23]; Available from: http://www.oect.es/Observatorio/Contenidos/InformesPropios/Desarrollados/Ficheros/Informe_VI_E NCT.pdf
48. Shah D. Healthy worker effect phenomenon. *Indian J Occup Environ Med.* 2009;13:77-9.

TABLES

Table 1. Participants' characteristics.

		Total n=455	Participants with health issues (n=299)	Participants without health issues (n=156)
Age in years, mean (SD)		42.1 (11.1)	43.7 (10.8)	39.0 (11.0)
Education level, n (%)	Low	73 (16.0)	61 (20.4)	12 (7.7)
	Middle	157 (34.5)	121 (40.5)	36 (23.1)
	High	225 (49.5)	117 (39.1)	108 (69.2)
Job type, n (%)	Manual	111 (24.4)	81 (27.1)	30 (19.2)
	Non-manual	125 (27.5)	82 (27.4)	43 (27.6)
	Mixed	218 (47.9)	136 (45.5)	83 (53.2)
Working hours/week, mean (SD)		38.7 (8.5)	38.8 (7.8)	38.7 (9.7)
Health issue type, n(%)	None	156 (34.3)	0 (0.0)	156 (100.0)
	Physical	139 (30.5)	139 (46.5)	0 (0.0)
	Mental health	125 (27.5)	125 (41.8)	0 (0.0)
	Others	35 (7.7)	35 (11.7)	0 (0.0)
Disease duration in months, mean (SD)		13.0 (27.7)	19.9 (32.2)	0 (0.0)

Table 2. Reliability, floor and ceiling effects of the Spanish version of the Work Role Functioning Questionnaire, (n=455).

Work demands	Valid n	Missing / Not applicable ^a		Mean scores (SD) ^b	Range	Median scores	n at floor n (%)	n at ceiling n (%)	Cronbach alpha
		n	(%)						
Work scheduling	445	10	(2.2)	75.65 (26.51)	0-100	85.00	7 (1.54)	90 (19.78)	0.91
Output	448	7	(1.5)	74.79 (24.25)	0-100	82.14	6 (1,32)	53 (11.65)	0.92
Physical	317	138	(30.3)	75.66 (25.62)	0-100	84.17	5 (1,10)	63 (13.85)	0.92
Mental	452	3	(0.7)	79.53 (26.12)	0-100	91.67	10 (2,20)	130 (28.57)	0.95
Social	408	47	(10.3)	82.90 (22.88)	0-100	91.67	8 (1.76)	142 (31.21)	0.81
Overall scale	443	12	(2.6)	77.05 (22.35)	0-100	84.90	4 (0.88)	18 (3.96)	0.98

(a) Subscales with more than 20% of items scoring "does not apply to my job" or missing values were excluded.

(b) Each subscale is scored from 0 - 100. Higher scores indicate better work functioning: difficulties all the time 0/100; difficulties no of the time 100/100.

Table 3. Pearson’s correlations of the WRFQ^a and the three single item question of the WAI^b.

Work demands	Work Scheduling	Output	Physical	Mental	Social	Total scale
Work scheduling	-	0.849	0.747	0.724	0.701	0,918
Output	0.849	-	0.730	0.752	0.731	0,935
Physical	0.747	0.730	-	0.494	0.576	0,821
Mental	0.724	0.752	0.494	-	0.766	0,861
Social	0.701	0.731	0.576	0.766	-	0,825
Overall scale	0.918	0.935	0.821	0.861	0.825	-
WAI overall-item	0.707	0.661	0.649	0.517	0.531	0.713 ^c
WAI physical demands	0.586	0.586	0.615 ^c	0.406	0.419	0.594
WAI mental demands	0.659	0.629	0.448	0.665 ^c	0.627	0.682

(a) Work Role Functioning Questionnaire; (b) Work Ability Index;

(c) Hypothesis 3a and 3b confirmed.

Table 4. Scores obtained by health condition and job type on the WRFQ-SpV.

Work demands	Health issue (median scores ^a)			p ^(b) value	Job type (median scores ^a)			p ^(b) value
	None (N= 156)	Mental (N= 125)	Physical (N= 139)		Manual (N= 109)	Non-manual (N= 123)	Mixed (N= 212)	
Work scheduling	95.00	70.00	82.00	0.000	80.00	90.00	85.00	0.011
Output	89.29	67.86	78.57	0.000	78.57	82.14	83.33	0.137
Physical	95.83	83.33	63.75	0.000	66.67	91.67	85.00	0.000
Mental	95.83	62.50	91.67	0.000	89.59	83.33	91.67	0.406
Social	91.67	75.00	91.67	0.000	91.67	83.33	91.67	0.208
Overall scale	92.05	68.52	81.73	0.000	77.78	83.33	87.04	0.027

(a) Subscales with more than 20% of items scoring "does not apply to my job" or missing values were excluded. Hypotheses 1a, 1b and 1c confirmed. (b) Kruskal Wallis H test.

Table 5. Differences between mean scores of known age groups in WRFQ^a (ANOVA).

Age (Years)	18-35 (N=148)	36-45 (N=121)	46-55 (N=123)	56-65 (N=57)	p value
Work demands	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	
Work scheduling	78.86 (22.45)	77.84 (25.92)	75.19 (25.63)	76.90 (25.95)	0.120
Output	77.84 (20.50)	76.00 (24.33)	72.53 (24.78)	70.34 (28.83)	0.130
Physical	81.07 (20.19)	77.36 (23.16)	71.33 (26.41)	61.36 (33.99)	0.000
Mental	81.17 (21.73)	78.55 (26.47)	77.82 (26.68)	79.61 (28.50)	0.719
Social	88.22 (16.48)	83.10 (23.05)	81.96 (21.99)	79.79 (27.73)	0.043
Overall scale	81.11 (17.40)	78.41 (21.23)	74.96 (22.70)	72.34 (27.13)	0.026

(a) WRFQ, Work Role Functioning Questionnaire. Hypothesis 4 (partially) confirmed.

SUPPLEMENTARY MATERIALS

Supplementary materials (1). Work Ability Index (WAI) scores obtained in a convenience subsample of participants (n=181).

Extended survey with WAI	Total n=181	Men n=71 (39.2%)	Women n=110 (60.8%)
WAI ^a overall-item, mean (SD)	7.6 (2.1)	7.6 (2.1)	7.7 (2.0)
WAI ^b physical demands, mean (SD)	3.8 (1.0)	3.7 (1.0)	3.8 (1.0)
WAI ^b mental demands, mean (SD)	3.9 (1.2)	3.9 (1.2)	3.8 (1.2)

(a) Single item question of the work ability index (scale 0-10)

(b) Single item question of the work ability index (scale 0-5).

Supplementary materials (2). Calculation of SEM with scores of the first 40 participants answering the WRFQ twice under similar conditions in a period of 7-15 days (N=40).

	WRFQ-SpV (a) Mean (SD)	WRFQ-SpV (b) Mean (SD)	Mean change score (SD)	SEM (c)
Work scheduling demands	66.4 (28.7)	65.8 (29.6)	0.7 (12.1)	8.5
Output demands	64.9 (25.8)	68.6 (26.3)	3.7 (12.6)	8.9
Physical demands	59.8 (31.2)	66.8 (29.6)	6.9 (12.1)	8.6
Mental demands	73.9 (26.1)	73.1 (27.8)	0.9 (14.9)	10.5
Social demands	77.0 (20.8)	76.8 (25.4)	0.2 (18.8)	13.3
Overall scale	67.6 (22.7)	69.5 (24.2)	1.9 (10.0)	7.1

(a) Mean Scores and Standard Deviation at first time.

(b) Mean Scores and Standard Deviation at second time.

(d) Standard Error of Measurement = SD of change scores / $\sqrt{2}$

Supplementary materials (3). Post-hoc paired analyses comparing median scores between participants with none health issues and participants with mental and physical health issues.

Work demands	None health issues (n=109)	Mental health issues (n=125)	p^(a) value	None health issues (n=109)	Physical health issues (n=139)	p^(a) value
Work scheduling	95.00	70.00	<0.001	95.00	82.00	<0.001
Output	89.29	67.86	<0.001	89.29	78.57	<0.001
Physical	95.83	83.33	<0.001	95.83	63.75	<0.001
Mental	95.83	62.50	<0.001	95.83	91.67	0.050
Social	91.67	75.00	<0.001	91.67	91.67	0.004
Overall score	92.05	68.52	<0.001	92.05	81.73	<0.001

(a) Mann-Whitney test for two independent samples

Supplementary materials (4). Scores of the Spanish version of the Work Role Functioning Questionnaire, distributed by existing health condition and job type.

Work demands	Median scores ^a											
	No health issues (N= 156)				Physical health issues (N= 139)				Mental health issues (N= 125)			
	Manual (N= 30)	Non-manual (N=43)	Mixed (N= 83)	P values ^b	Manual (N= 52)	Non-manual (N=26)	Mixed (N=61)	P values ^b	Manual (N=20)	Non-manual (N=46)	Mixed (N=59)	P values ^b
Work scheduling	95.00	100.00	95.00	0.030	72.50	90.00	85.00	0.015	72.50	75.00	60.00	0.053
Output	96.13	85.71	89.29	0.496	67.86	83.93	82.14	0.131	71.43	71.43	59.52	0.247
Physical	93.33	100.00	95.83	0.006	54.17*	66.67	70.83	0.021	72.92	87.50	80.00	0.016
Mental	100.00	91.67	95.83	0.031	93.75	93.75	91.67	0.939	60.42	68.75	58.33	0.145
Social	91.67	91.67	100.00	0.306	91.67	100.00	91.67	0.020	83.33	75.00	75.00	0.943
Overall scale	93.75	91.00	92.59	0.791	72.17	87.58	83.65	0.007	68.99	78.64	60.65	0.083

(a) Subscales with more than 20% of items scoring "does not apply to my job" or missing values were excluded. Hypothesis 2 confirmed.

(b) Kruskal Wallis H tests.

Supplementary materials (5). Post-hoc paired analyses comparing differences in Mean Scores between different groups of age.

Age (Years)	18-35	36-45	p ^(a)	18-35	46-55	p ^(a)	18-35	56-65	p ^(a)	36-45	46-55	p ^(a)	36-45	56-65	p ^(a)	46-55	56-65	p ^(a)
	(n=148)	(n=121)		(n=148)	(N=123)		(n=148)	(N=57)		(n=121)	(N=123)		(n=121)	(N=57)		(N=123)	(N=57)	
Work demands	Mean	Mean	value	Mean	Mean	value	Mean	Mean	value	Mean	Mean	value	Mean	Mean	value	Mean	Mean	value
Work scheduling	78.86	77.84	0.444	78.86	75.19	0.206	78.86	76.90	0.025	77.84	75.19	0.687	77.84	76.90	0.151	75.19	76.90	0.232
Output	77.84	76.00	0.474	77.84	72.53	0.097	77.84	70.34	0.027	76.00	72.53	0.424	76.00	70.34	0.154	72.53	70.34	0.402
Physical	81.07	77.36	0.213	81.07	71.33	0.017	81.07	61.36	0.001	77.36	71.33	0.307	77.36	61.36	0.002	71.33	61.36	0.015
Mental	81.17	78.55	0.529	81.17	77.82	0.277	81.17	79.61	0.721	78.55	77.82	0.695	78.55	79.61	0.888	77.82	79.61	0.646
Social	88.22	83.10	0.051	88.22	81.96	0.027	88.22	79.79	0.008	83.10	81.96	0.960	83.10	79.79	0.380	81.96	79.79	0.355
Overall scale	81.11	78.41	0.267	81.11	74.96	0.032	81.11	72.34	0.005	78.41	74.96	0.377	78.41	72.34	0.097	74.96	72.34	0.324

(a) t-Test.