Translation, cross-cultural adaptation and validation of the Work Role Functioning Questionnaire (WRFQ) to Spanish spoken in Spain

---

Traducción, adaptación cultural y validación del Work Role Functioning Questionnaire (WRFQ) al castellano hablado en España.

José María Ramada Rodilla

TESI DOCTORAL UPF / 2014

DIRECTORS DE LA TESI

Dra. Consol Serra Pujadas
CiSAL – Centro de Investigación en Salud Laboral, Universidad Pompeu Fabra. PRBB Building. Doctor Aiguader, 88. 08003- Barcelona, España.

Dr. George L Delclós Clanchet
Southwest Center for Occupational and Environmental Health, the University of Texas School of Public Health. 1200 Pressler Street. Houston, Texas 77030, USA.

DEPARTAMENT DE CIÈNCIES EXPERIMENTALS I DE LA SALUT

A mis tres hijos, que son la pasión de mi vida,
por su capacidad para comprender, aceptar y amar.
A la memoria de mi padre a quien tanto le debo.

AKNOWLEDGEMENTS (Agradecimientos – Agraïments)

Esta tesis llega cuando en mis gafas ya necesito una cierta adición de dioptrías para poder leer de cerca. Así que quiero comenzar por dar las gracias a todos los que toman decisiones sobre el futuro académico de las personas sabiendo que el empuje de la juventud no está en la edad sino en el espíritu.

Me viene a la cabeza la leyenda de San Agustín de Hipona y el niño en la playa y no sé cómo voy a meter, con mi capacidad limitada, todo el agua del mar en un hoyo tan pequeño. También ahora, en esta parte de mi tesis, se me ha pasado por la cabeza buscar el respaldo de mis Directores para controlar los sesgos, pero no he querido correr el riesgo de que me recomienden la realización de otra revisión sistemática. Así que con mis solas palabras asumo el reto en solitario.

Moltíssimes gràcies Consol. Gràcies mestra. La llista dels motius pel quals em sento tan agraït amb tu podria ser interminable. Comptes amb tota la meva admiració i respecte per la teva decència professional com a metgessa del treball i investigadora, per la teva particular i innovadora visió de la salut laboral, professionalitat i capacitat com a directora de tesi. Gràcies pel teu suport permanent en aquest projecte i per ser, en l'hospital on treballem, una cap i una companya de treball infatigable, incombustible. Mil gràcies per totes les oportunitats que m'has anat oferint al llarg d'aquestos darrers anys i que he intentat aprofitar sempre.

Gracias Jordi, por guiar mis primeros pasos en el mundo de la investigación en 2010, siendo mi tutor del Máster en Salud Laboral; en ese momento empezó a gestarse la posibilidad de seguir más allá, después del Máster. Gracias maestro, por enseñarme tanto en la Unidad de Patología Laboral, por la magnífica experiencia en la Escuela de Salud Pública de la Universidad de Texas, por poner a mi disposición tu enorme valía como clínico y como profesor. Gracias por tu humanidad, consideración, accesibilidad, ejemplaridad como investigador y

paciencia como director de tesis. También, gracias a Conchita por su hospitalidad y amabilidad durante mi estancia en Houston (Texas) en 2013.

Gracias Fernando, por abrirme las puertas del CiSAL, por tus constantes propuestas para involucrarme en proyectos y por tu apoyo desde el primer minuto para que mi estancia en la Universidad de Groningen (Holanda) en 2013 fuera posible.

Thank you Ute, Femke and Iris (University Medical Center Groningen, The Netherlands). Your warm welcome, support and help in this thesis at any time in Groningen were invaluable. Thank you, Roy, for making statistical analysis understandable and also for your kind and disinterested help and availability at all times.

Gràcies a la Direcció del PSMAR, on exerceixo com a metge del treball. Gràcies per fer realitat que aquesta organització sigui un dels pols més dinàmics de coneixement assistencial, docent i de recerca de la ciutat de Barcelona. Gràcies pel suport prestat per l'obtenció de la menció europea al títol de doctorat. Gràcies als/les companys/es metges/esses i infermers/es del PSMAR, que m'han ajudat amb tanta generositat en el treball de camp.

Gràcies als companys i companyes del Servei de Salut Laboral del PSMAR, pel suport donat a aquest projecte sempre que l'he necessitat: a Aida, a Carmen, a Chelo, a Julià i a Nuria. Mil gràcies a Fina Pi-Sunyer i a Joan Mirabent, per ser tan excel·lents infermers del treball i per tanta generositat en vostra col·laboració durant el treball de camp. Moltíssimes gràcies Dra. Villar (Rocio), sense el teu suport i companyonia de veritable col·lega tal vegada la estada a Groningen no hagués estat possible. De tot cor, moltes gràcies companys i companyes.

Gracias a los amigos y amigas del CiSAL-UPF por contar siempre conmigo, a pesar de no estar físicamente presente en el PRBB. Gracias a Montse Fernández y a Sandra Garrido por su ayuda siempre diligente en cualquier gestión con la

Universidad y con el CIBERSP. Gracias a María López por el ánimo ofrecido en todo momento y a Sergio Vargas por todos los favores realizados durante estos años.

Y para terminar gracias a mis hermanas y a mi madre, siempre disponibles. Yo siento por vosotras tres verdadera devoción; a Ram Dulthummon por su apoyo y por sus palabras *(always in English)* para poner algún límite a mi fantasía a veces desbordada y ayudarme a mantener los pies sobre la tierra; a mis tres hijos José, Borja y María Ángeles, por su generosa ayuda con la base de datos, por soportar mil veces los ensayos de mis exposiciones y por sus muestras de aliento permanente.

Y como no, gracias a la madre de mis tres hijos, Ángeles Calaforra, por su inmensa generosidad, por estar siempre ahí cuando ha sido necesario, por su paciencia inagotable, y por sus consejos siempre sensatos cuando han aparecido dudas sobre el sentido de este proyecto en este momento de mi vida.

A todos y todas, un millón de gracias.

SUMMARY

Background

Health and work mutually influence the working population. Health-related work functioning is the worker's ability to meet work demands for a given health status. Quality validated measurement tools are needed to assess how workers function at work along their professional life course and to evaluate interventions to accommodate job conditions to the worker's skills and health status.

The use of directly translated measurement tools may lead to unreliable or misleading results in research and practice, and could limit the exchange of information in the scientific community. Due to possible cultural differences in perception of work, health and disease, instruments developed in other languages or cultures should be systematically translated, adapted and validated for use in different target languages or cultures.

The Work Role Functioning Questionnaire (WRFQ) is an instrument designed to measure self-perceived difficulties to perform work, in active workers, given a certain health condition. Its results can be interpreted in terms of work functioning, work performance, work productivity, work disability and presenteeism, and they can be transformed into meaningful social and economic outcomes.

Objective

The aim of this thesis was to provide a high quality validated instrument in Spanish, able to assess the impact of health on "work functioning" and describe the extent to which workers improve or deteriorate their ability to meet the demands of the job in Spanish-speaking populations.

This overall objective was carried out through three specific objectives: 1) to review the literature on the methodology for cross-cultural adaptation and

validation (CCAV) of health questionnaires; 2) to estimate the degree of compliance with literature recommendations for CCAV in Spanish and Latin American scientific journals; and 3) to translate and cross-culturally adapt the WRFQ and validate it in a sample of a general working Spanish-speaking population.

Methods

An evidence-based decision was taken to select a generic measurement instrument that evaluates health-related work functioning. A comprehensive literature review was performed to identify and synthesize recommendations on the methodology of CCAV of health questionnaires. Five high impact journals in epidemiology and/or public health from Spain and Latin America were analyzed to estimate the degree of compliance with the methodological recommendations.

A systematic 5-step procedure (direct translation, synthesis, back-translation, consolidation by an expert committee and pre-test) described in the literature was followed to translate, cross-cultural adapt and validate the WRFQ. The applicability, readability and integrity of the Spanish version of the Work Role Functioning Questionnaire (WRFQ-SpV), together with its preliminary internal consistency, test-retest reliability and validity were assessed in a pre-test with 40 participants.

Next, a cross-sectional study was conducted among 455 active workers of a general working population to evaluate the reliability and validity of the WRFQ-SpV. A longitudinal survey was carried out to examine the responsiveness in a sample of 102 workers of this general working population. The consensus-based standards on measurement properties of health status measurement instruments (COSMIN) guided the design of the different studies.

Results

To identify and synthesize the literature recommendations on the methodology of CCAV of health questionnaires, 21 articles (out of 214 citations) and seven relevant books were selected for full text analysis. A high degree of consensus was found on the steps to follow to guarantee conceptual, semantic, idiomatic and experiential equivalence. Two steps were widely recommended to carry out the CCAV process: first, the cross-cultural adaptation process (following a systematic and rigorous procedure); and secondly, validation in the target language (evaluating reliability, validity and responsiveness). Only 6% of the retrieved articles followed all recommended steps.

The CCAV of the WRFQ was carried out without major difficulty. Idiomatic challenges were found and an expert committee provided a solution. The questionnaire showed adequate applicability and good face and content validity. Internal consistency was satisfactory (Cronbach alpha =0.98). The original five factor structure of the WRFQ reflected fair dimensionality of the construct (Chi-square, 1445.8; 314 degrees of freedom; root mean square error of approximation [RMSEA] =0.08; comparative fit index [CFI] >0.95 and weighed root mean residual [WRMR] >0.90). The test–retest reliability showed good reproducibility of the questionnaire outcomes ($0.77 \leq$ intraclass correlation coefficient [ICC]≤ 0.93 and standard error of measurement [SEM] =7.10). For construct validity assessment, all formulated hypotheses were confirmed differentiating groups with different jobs, health conditions and ages. Moreover, we verified that the WRFQ-SpV was able to detect (true) changes over time.

Conclusions:

The CCAV process should follow several well established steps. However, the degree of compliance of the scientific literature with the methodological recommendations for CCAV can be improved. The WRFQ-SpV is a reliable and valid instrument to measure health-related work functioning in day-to-day practice

and research in occupational health. Suggestive evidence about the possible use of the WRFQ-SpV in evaluative studies was found. More research is needed to examine the instrument responsiveness for groups who do not experience health improvement or deteriorate.

Key words:

Work functioning instrument; questionnaires; scales; health survey; measurement instrument; cross-cultural comparison; validation studies; psychometric properties; reliability; validity; responsiveness.

RESUMEN

Antecedentes

Salud y trabajo constituyen un binomio con una permanente influencia mutua. El desempeño del trabajo en relación con la salud se define como la capacidad de un/a trabajador/a para dar respuesta a las demandas del trabajo dado un determinado estado de salud. Se necesitan herramientas de medición validadas de calidad para evaluar los niveles de desempeño del trabajo a lo largo de la vida laboral y para evaluar las intervenciones destinadas a adaptar las condiciones de trabajo a las habilidades y el estado de salud de la población trabajadora.

El uso de instrumentos literalmente traducidos puede dar lugar a resultados poco fiables o engañosos en la práctica y en la investigación, pudiendo limitar el intercambio de información en la comunidad científica. Debido a las posibles diferencias culturales en la percepción del trabajo, la salud y la enfermedad, los instrumentos desarrollados en otros idiomas o culturas deberían ser traducidos de manera sistemática, adaptados y validados para su uso en idiomas o culturas diferentes.

El Cuestionario de Desempeño del Trabajo (del inglés *Work Role Functioning Questionnaire,* WRFQ) es un instrumento para medir las dificultades auto-percibidas para desempeñar el trabajo, en trabajadores en activo, dado un determinado estado de salud. Sus resultados pueden ser interpretados en términos de desempeño, rendimiento o productividad en el trabajo, discapacidad laboral y presentismo, pudiendo ser transformados en resultados con significación social y económica.

Objetivo

El objetivo de esta tesis fue poner a disposición un instrumento de calidad validado en español, capaz de evaluar el impacto de la salud en el desempeño del trabajo, y describir el grado en que los trabajadores mejoran o empeoran su capacidad para dar respuesta a las demandas del trabajo.

Este objetivo general se llevó a cabo por medio de tres objetivos específicos: 1) revisar la literatura sobre la metodología para la traducción, adaptación cultural y validación (TACV) de cuestionarios de salud; 2) estimar el grado de cumplimiento de las recomendaciones metodológicas en revistas científicas Españolas y de América Latina; 3) Traducir y adaptar el WRFQ y validarlo en una muestra de la población general trabajadora hispano-parlante.

Métodos

Se seleccionó un instrumento genérico para evaluar el desempeño del trabajo en relación con la salud en base a la evidencia. Se llevó a cabo una revisión bibliográfica exhaustiva para identificar y sistematizar las recomendaciones de la literatura sobre la TACV de cuestionarios de salud y adicionalmente se analizaron cinco revistas de epidemiología y/o salud pública de España y América Latina, con los factores de impacto más altos, para estimar el grado de cumplimiento con las recomendaciones metodológicas.

Se siguió un procedimiento en 5 pasos (traducción directa, síntesis, retrotraducción, consolidación por un comité de expertos y pre-test) descrito en la literatura para traducir, adaptar y validar el WRFQ. Se realizó un pre-test con 40 participantes para evaluar la aplicabilidad, legibilidad e integridad de la versión española del WRFQ (WRFQ-SpV), junto con su consistencia interna, fiabilidad test-retest y validez.

Posteriormente, se llevó a cabo un estudio transversal con una muestra de 455 trabajadores en activo para evaluar la fiabilidad y validez del WRFQ-SpV. Se llevó a cabo un estudio longitudinal en una muestra de 102 trabajadores en activo de una población general para examinar su sensibilidad al cambio. Se utilizaron los estándares de consenso para la evaluación de las propiedades de medición de los cuestionarios de salud (COSMIN) en el diseño de los diferentes estudios.

Resultados

Para identificar y sistematizar las recomendaciones metodológicas existentes en la literatura, se seleccionaron 21 artículos (de un total de 214 citas) y siete libros relevantes para su análisis. Se encontró un alto grado de consenso en la realización de dos pasos en la TACV para garantizar la equivalencia conceptual, semántica, idiomática y vivencial. El primero, el proceso de adaptación cultural (siguiendo un procedimiento sistemático y riguroso), y el segundo, la validación en el idioma de destino (evaluando la fiabilidad, validez y sensibilidad al cambio).

El grado de cumplimiento de las recomendaciones metodológicas para llevar a cabo la TACV puede ser mejorado. El 6% de los artículos recuperados siguieron todos los pasos recomendados en la literatura que les eran aplicables.

La TACV del WRFQ se llevó a cabo sin dificultades relevantes. Se encontraron desafíos idiomáticos y un comité de expertos proporcionó una solución. El cuestionario mostró una adecuada aplicabilidad, validez aparente o lógica así como de contenido. La consistencia interna fue satisfactoria (alfa de Cronbach =0.98). La estructura original de cinco factores del WRFQ refleja una adecuada dimensionalidad del constructo (Chi-cuadrado, 1445,8; 314 grados de libertad; error cuadrático medio de aproximación [RMSEA] =0,08, índice de ajuste comparativo [CFI] >0,95 y media ponderada de la raíz residual [WRMR] >0,90). La fiabilidad test-retest mostró una buena reproductibilidad de las puntuaciones del cuestionario (0.77 ≤ coeficiente de correlación intraclase [CCI] ≤ 0.93 y error estándar de la medida [SEM] =7.10). Para la evaluación de la validez de

constructo se confirmaron todas las hipótesis formuladas, diferenciando grupos con diferentes trabajos, problemas de salud y grupos de edad. Se verificó que el WRFQ-SpV fue capaz de detectar cambios (verdaderos) a lo largo del tiempo.

Conclusiones

El proceso de TACV debería seguir varios pasos bien establecidos. Sin embargo, el grado de cumplimiento de las recomendaciones metodológicas propuestas en la literatura científica para la TACV puede ser mejorado. El WRFQ-SpV es un instrumento fiable y válido para medir el desempeño del trabajo en relación con la salud tanto para la práctica diaria como para la investigación en salud laboral. Se ha encontrado evidencia sugerente sobre el posible uso de la WRFQ-SpV con fines evaluativos. Se necesita investigación adicional para examinar la sensibilidad al cambio del instrumento en grupos que no experimentan mejoría o que sufren deterioro de su salud.

Palabras clave:

Desempeño en el trabajo; cuestionarios, escalas; encuesta de salud; instrumento de medición; adaptación cultural; estudios de validación; propiedades psicométricas; fiabilidad; validez; sensibilidad al cambio.

PREFACE

The analysis of measurement instruments for use in occupational health research and practice is currently an area of research interest within the Center for Research in Occupational Health (CiSAL), and it is in this context that this doctoral thesis was undertaken. Its content is part of a CiSAL research project entitled "Evaluation of health-related work functioning and identification of preventive interventions with the Spanish version of the Work Role Functioning Questionnaire". This project is funded by the Instituto de Salud Carlos III, ISCIII (Ministry of Economy and Competitiveness, Spanish Government), FIS: PI12/02556 (Principal Investigator, Consol Serra Pujadas; co-investigators, José María Ramada and George Delclos).

This project arises from the need for validated instruments to assess the impact of health on "work functioning" in Spanish-speaking populations. There are a number of instruments to evaluate "health-related work functioning" in English, but these have not always been adapted and/or validated into the Spanish context. Thus, identifying and selecting an instrument to properly measure health-related work functioning and then translating, adapting and validating its measurement properties, for future use in research, was consistent with the goals of this project.

According to the policy of the Doctoral Program Committee in the Department of Experimental and Health Sciences at Pompeu Fabra University, this thesis is presented as a compendium of four scientific publications, derived from the literature review and field work conducted in the Parc de Salut Mar de Barcelona health system. The first publication was written in Spanish and the other three in English. All have been published recently in international occupational health peer-reviewed journals, indexed in PubMed, with the PhD candidate as first author.

The results have been presented in part at several scientific meetings, specifically: the First CiSAL Annual Scientific Meeting (1), the Second Scientific Conference on Work Disability Prevention and Integration (WDPI) (2), the XXII Diada of the Catalan Society of Safety and Occupational Medicine (3), the Third CiSAL Annual Scientific Meeting (4) and the First BiblioPRO Scientific Meeting (5).

In addition to the funding from the Instituto de Salud Carlos III (PI12/ 02556), this thesis received partial financial support from the The University of Texas School of Public Health at Houston (USA) and from the Network of Biomedical Research Centers in Epidemiology and Public Health (CIBERESP), for completion of short-term stays at international universities, in order to fulfill the requirements for a doctorate with European mention.

(1) Ramada JM, Serra C, Delclós J. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. 1ª Jornada Científica CISAL. Barcelona, 2011.

(2) Ramada JM, Serra C, Delclós GL. Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations. Second Scientific Conference on Work Disability Prevention and Integration 'Healthy ageing in a working society'. WDPI; Groningen, 2012.

(3) Ramada JM. Qüestionaris de salut de qualitat: requisits bàsics. XXII Diada de la Societat Catalana de Seguretat i Medicina del Treball. Barcelona, 2012.

(4) Ramada JM, Serra C, Delclós J. Traducción, adaptación cultural y validación del "Work role functioning questionnaire (WRFQ-27)". 3ª Jornada Científica CISAL. Barcelona, 2013.

(5) Ramada JM, Serra C, Amick BC, Castaño JR, Delclós GL. Adaptación cultural del "Work Role Functioning Questionnaire (WRFQ)" al castellano hablado en España. I Jonada Científica BiblioPRO. IMIM-CIBERESP. Barcelona, 2013.

PRÓLOGO

El análisis de instrumentos de medición para su uso en la investigación y la práctica diaria en salud laboral es, en estos momentos, un área de interés para la investigación del Centro de Investigación en Salud Laboral (CiSAL), y es en este contexto en el que se ha desarrollado la presente tesis doctoral. El contenido de esta tesis forma parte del proyecto de investigación del CiSAL, titulado *"Evaluación de la capacidad para trabajar y posibilidades de intervención mediante el Work Role Functioning Questionnaire adaptado al castellano"*. Este proyecto ha sido financiado por el Instituto de Salud Carlos III, ISCIII (Ministerio de Economía y Competitividad, Gobierno de España), FIS: PI12/02556, (Investigadora principal Consol Serra Pujadas; co-investigadores José María Ramada y George Delclós).

Este proyecto surge de la necesidad de disponer de instrumentos en Español validados para evaluar el impacto de la salud sobre el "desempeño del trabajo" en poblaciones hispano-parlantes. Existe un número de instrumentos para evaluar el "desempeño del trabajo" en relación con la salud en Inglés, pero no siempre han sido adaptados y/o validados en el contexto Español. Por ello, la identificación y selección de un instrumento para medir adecuadamente el "desempeño del trabajo" en relación con la salud y proceder a su traducción, adaptación y validación de sus propiedades de medición, para su uso en futuras investigaciones, es consistente con los objetivos de este proyecto.

Conforme a la normativa dada por la Comisión de Dirección del Programa de Doctorado del Departamento de Ciencias Experimentales y de la Salud de la Universidad Pompeu Fabra, esta tesis doctoral se presenta como un compendio de cuatro publicaciones científicas en las que el doctorando es el primer autor, fruto de la revisión de la literatura y el trabajo de campo llevado a cabo en el sistema hospitalario del Parc de Salut Mar de Barcelona. La primera de las publicaciones fue escrita en español y las tres restantes en inglés. Tres de ellas

han sido publicadas recientemente en revistas internacionales de salud laboral, indexadas en PubMed y con revisión por pares. La cuarta se encuentra en el momento de la impresión de esta tesis en proceso de revisión por pares, en una revista internacional de salud laboral, asimismo indexada en Pubmed.

Los resultados han sido presentados parcialmente en la Primera Jornada Científica Anual del CiSAL (6); la Second Scientific Conference on Work Disability Prevention and Integration (WDPI) (7); la XXII Diada de la Societat Catalana de Seguretat i Medicina del Treball (8); la Tercera Jornada Científica Anual del CiSAL (9) y en la Primera Jornada Científica BiblioPRO (10).

Adicionalmente a la financiación del Instituto de Salud Carlos III (PI12/ 02556), esta tesis recibió apoyo económico parcial de la Escuela de Salud Pública de la Universidad de Texas (Estados Unidos de América) y del Centro de Investigación Biomédica en Red de Epidemiología y Salud Pública (CIBERESP).

(6) Ramada JM, Serra C, Delclós J. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. 1ª Jornada Científica CISAL. Barcelona, 2011.

(7) Ramada JM, Serra C, Delclós GL. Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations. Second Scientific Conference on Work Disability Prevention and Integration 'Healthy ageing in a working society'. WDPI; Groningen, 2012.

(8) Ramada JM. Qüestionaris de salut de qualitat: requisits bàsics. XXII Diada de la Societat Catalana de Seguretat i Medicina del Treball. Barcelona, 2012.

(9) Ramada JM, Serra C, Delclós J. Traducción, adaptación cultural y validación del "Work role functioning questionnaire (WRFQ-27)". 3ª Jornada Científica CISAL. Barcelona, 2013.

(10) Ramada JM, Serra C, Amick BC, Castaño JR, Delclós GL. Adaptación cultural del "Work Role Functioning Questionnaire (WRFQ)" al castellano hablado en España. I Jonada Científica BiblioPRO. IMIM-CIBERESP. Barcelona, 2013.

TABLE OF CONTENTS

1. INTRODUCTION

1.1. Statement of the problem

Health and work form an indivisible duality in which mutual influence is permanent. The World Health Organization (WHO) defines health as "a state of complete physical, mental and social well-being" and not merely the absence of disease. This definition is part of the Declaration of Principles of the WHO since its founding in 1948 (1).

Work is a health determinant and there is an increasing body of evidence showing that work has positive health effects when working conditions are reasonably acceptable (2,3). Decent work sums up the aspirations of people in their working lives. It involves opportunities for productive work, delivers a fair income, security in the workplace and social protection for families, opportunities for personal development and social integration, freedom for people to express their concerns, organize and participate in the decisions that affect their lives and equality of opportunity and treatment for all women and men. A community or a country improves population health status when everyone who is able to work can get a decent job (4).

Increased life expectancy and prolongation of retirement age are increasing the overall age of the workforce, and might result in an increasing number of employees working with chronic diseases (5-7). Interventions to keep these workers in the labor market and promote work participation are being increasingly developed to support a sustainable, active, and productive work life (7,8). Furthermore, rehabilitation programs and interventions to adapt or accommodate working conditions to the workers' health and skills are becoming more frequent, with the goal of achieving a safe return to work after a period of sick leave.

The effectiveness of these rehabilitation programs and interventions has usually been assessed using outcome measures such as work status (active, temporary or permanent disability), time to return to work, duration of functional disability and costs of incapacity to work (8-11). These outcomes have been useful but are limited, as they mainly assess whether workers are present or absent from their jobs. They do not offer information about the worker's participation in the job or the degree to which the worker is able to respond to the job's demands (12,13).

Quality validated measurement tools are needed to assess how workers function at work along their professional life course, and the existing continuum between working successfully at one extreme and work absence at the other (14).

Outcome measures able to describe the extent to which workers increase or decrease their ability to meet job demands and to fully assess rehabilitation programs and intervention effectiveness are needed in Spanish-speaking occupational health settings, yet there is a lack of quality validated instruments in Spanish for this purpose. Thus, the rationale for this thesis is to provide an evidence base for an instrument to evaluate health-related work functioning, and make it available to Spanish-speaking occupational health professionals and researchers for use in daily practice and research.

1.2. From work disability to health-related work functioning

Disability can be described as the environmentally determined effect of an impairment that, in interaction with other factors and within a specific social context, is likely to cause an individual to experience an undue disadvantage in his or her personal, social or professional life (15). Work disability could be defined as the effect of an illness or an accident in the ability of a person to perform a particular work activity.

Disability is not an absolute attribute of an individual; rather, it is a social construct. A person who is blind, or deaf, or needs a wheelchair to move can be completely

dependent in one setting, but fully autonomous and functional in a different one. Thus, the effect of impairment will always be referred to a given environment, and if we restrict disability to the functional effects of this impairment, regardless of the environment, we put the burden of the problem and the responsibility to find a solution on the individual.

From a social perspective, work disability should be understood as a manageable situation, where different stake holders (workers, employers, human resource managers, supervisors, unions and occupational health professionals) should be involved to respond to an individual's needs so that he/she can function successfully at work. Disability is, therefore, a social rather than a medical issue and from this perspective it is easier to understand that positive action towards integration and job participation is required, rather than merely passive measures to provide income support (15).

Once tucked into this paradigm, it is possible to analyze from a broader perspective the economic and social impact of removing barriers for integration of individuals with disabilities. Imaginative and economically viable solutions addressing a wider range of interventions may arise from this, varying from improving the workers' skills (through training and rehabilitation programs), to facilitating accommodation in suitable workplaces or intervening to adapt the workplace and/or working conditions to the specific needs of these individuals.

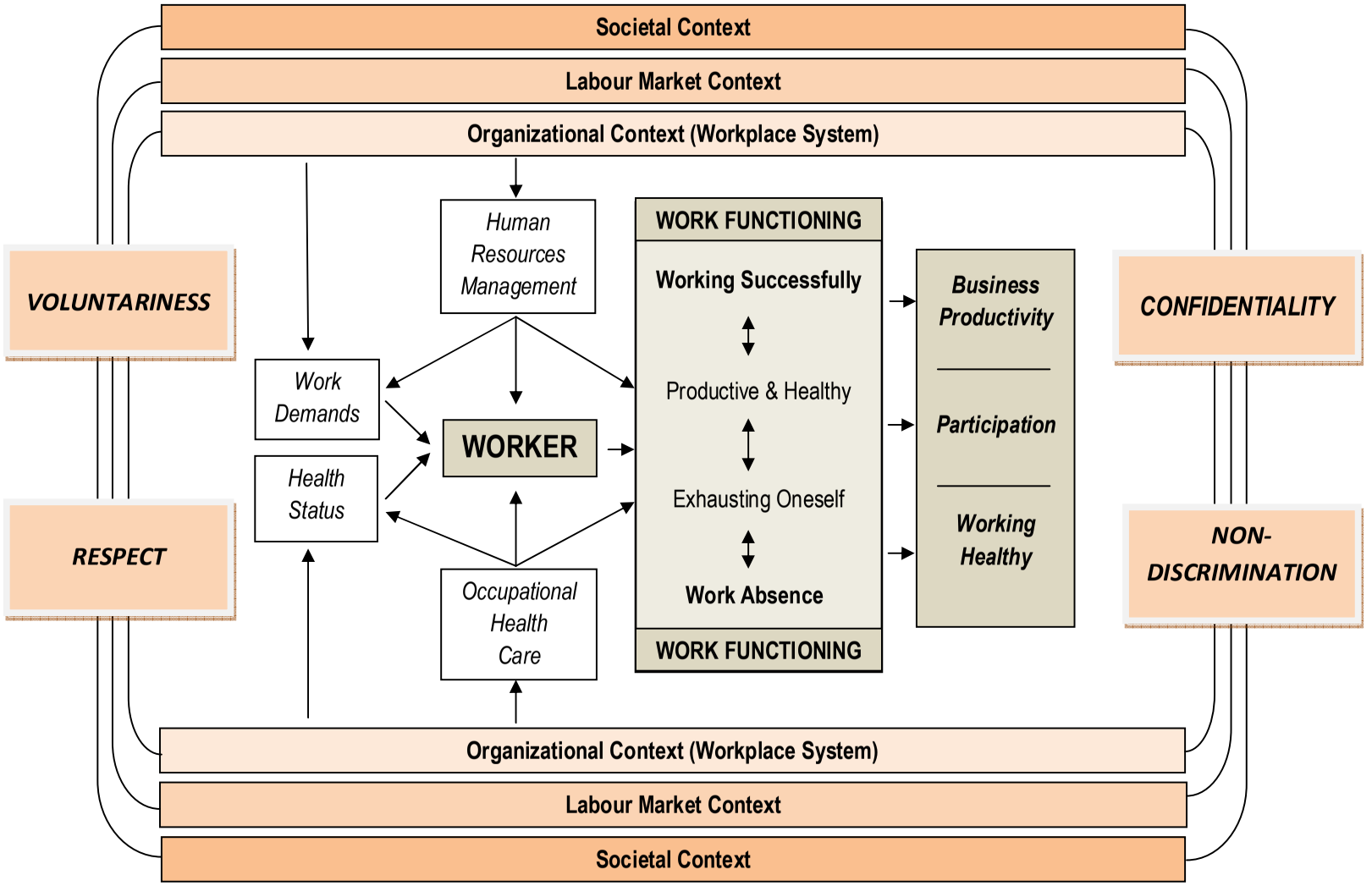
A significant number of research teams and occupational health services are increasingly designing and implementing rehabilitation and/or accommodation programs to adapt working conditions to worker skills and health to support an active working life (7,8,16,17). To fully assess intervention effectiveness requires outcome measures that describe the extent to which people increase their ability to meet the demands of the job.

Health-related work functioning is a comprehensive concept that incorporates the previously described paradigm shift, and can be defined as the ability of a worker to meet work demands for a given physical and emotional health status (18).

Theoretically, working conditions and demands are modifiable and health is a dynamic concept that can change over a lifetime. Hence, health-related work functioning constitutes a continuum rather than a dichotomy, with “working successfully” at one end and “work absence” at the other. Measuring the results of the impact of health on work in terms of "present” versus “absent" is not enough to understand what happens along this continuum (19). Based on the individual’s work performance and on-the-job productivity (Figure 1), and especially in the current European socio-economic context, it constitutes a phenomenon of great interest in occupational health care settings and research.

The rationale for this thesis arises from the need for quality validated measurement instruments to assess health-related work functioning in Spanish-speaking settings. This will serve to enhance the evaluation of rehabilitation, accommodation or adaptation programs. The emphasis is on the ability of the instrument to measure the worker's participation, and not only whether workers are present or absent from their jobs.

Figure 1. Conceptual frame of Health-Related Work Functioning based on Amick, Gimeno (18) and Abma (19) and ethical use of the questionnaire.



1.3. General overview of work outcome measurement tools

When reviewing the literature on work outcome measures it is possible to find different approaches to work outcome measurement and, in general, it is possible to retrieve four groups of work outcome measures (12). Several assess the labor force status (mainly time to return to work and duration of functional disability). Another group assesses the economic impact of work outcomes (especially lost time from work and self-reported effectiveness in performing the job). A third set of measures assesses the impact of health on role functioning (mixing work-role with other roles). And finally, there is a group of work-role specific functioning measurement instruments that measure health-related functioning at work. Several studies and reviews have analyzed both strengths and weaknesses of each group of measurement tools (12,18,20-26).

Focusing on the instruments that measure our phenomenon of interest (health-related work functioning), a number of health and/or job specific work functioning measurement instruments together with other generic instruments have been developed. The most relevant are shown in Table 1.

When measuring health-related work functioning in research and practice, evidence-based decisions should be made about which instrument to use. Evans recommends considering three areas when choosing a questionnaire: the psychometric properties of the instrument, administration complexity, and the setting of the evaluation (27). Firstly, it is essential to know the purpose for use of the instrument (in medicine, for example, it could be for diagnosis, evaluation or prediction) (28). Then, depending on this, it is necessary to find out whether the measurement properties of the instrument have been assessed with quality methodology.

If the instrument is going to be applied for diagnostic or prognostic purposes, such as to estimate work functioning status or to distinguish between different courses

(or outcomes) of work functioning, evidence of its discriminative ability should be provided; in this case, parameters of reliability are very important (including those of measurement error). But if the aim is to apply the instrument to evaluate interventions or to monitor work functioning in individuals, the instrument needs to provide evidence of its ability to detect (true) changes over time; in this case, parameters of responsiveness (on top of measurement error) are crucial (28).

It is also necessary to know in which language or culture the questionnaire was originally developed. If the intention is to use it in a different language, then it is necessary to determine whether the process of cross-cultural adaptation and validation in the target language employed quality evidence-based methods.

In the 2000s a series of specific work-role functioning questionnaires were developed; among them, the Work Limitations Questionnaire (WLQ) and the Work Role Functioning Questionnaire (WRFQ) (12,29) where developed as generic instruments to measure work functioning. These instruments provide an overall work functioning score, but also allow an estimation of work functioning in relation to each domain of work demands (work scheduling, output, physical, mental and social demands).

The WRFQ measures perceived difficulties to perform the job due to health problems. As mentioned above, it is a generic instrument conceptually developed to represent a wide range of health conditions and work demands and is freely available in the literature for professionals and researchers. The questionnaire has undergone various levels of validity and reliability testing and has displayed relevant levels of reliability and content, construct and criterion validity. Numerous studies have demonstrated the usefulness of this tool in English-speaking health care environments (30-32) and it has been successfully translated, adapted and validated in Canadian French (33), Brazilian Portuguese (34) and Dutch (14,19,35). No such version exists in Spanish.

Table 1. Specific and generic work functioning measurement instruments.

Acronym	Name of the Instrument	Type	Reference
HPQ	Health and Work Performance Questionnaire	Generic WF instrument. Single global rating.	(36)
WPAI	Work Productivity and Activity Impairment Questionnaire	Generic WF instrument. Single global rating.	(37)
WPSI	Work Productivity Short Inventory	Generic WF instrument. Single global rating.	(38)
WHI	Work and Health Interview	Specific WF instrument for lost productive time.	(39)
HLQ	Health and Labor Questionnaire	Generic WF instrument. Overall and subscales rating.	(40)
HRPQ-D	Health Related Productivity Questionnaire Dairy	Specific WF instrument for daily follow-up.	(41)
QQ	Quantity and Quality Instrument	Specific WFfor quantity and quality of work.	(42)
HAQ	Health Assessment Questionnaire	Specific WF for rheumatic conditions.	(43)
–	Angina-related Limitations at Work Questionnaire	Specific WF for angina pectoris.	(44)
WALS	Workplace Activity Limitations Scale	Specific WF or arthritic population.	(45)
LEAPS	Lam Employment Absence and Productivity Scale	Specific WF for clinically depressed population.	(46)
NWFQ	Nurses Work Functioning Questionnaire	Specific WF for nurses with common mental disorders	(47)
EWPS	Endicott Work Productivity Scale	Generic WF instrument. Overall rating.	(48)
SPS	Standford Presenteeism Scale	Generic WF instrument. Overall rating.	(49)
WLQ	Work Limitations Questionnaire	Generic WF instrument. Overall and subscales rating.	(29)
WRFQ	Work Role Functioning Questionnaire	Generic WF instrument. Overall and subscales rating.	(12)

1.4. Methodological quality in health-questionnaire validation

Since measurement is at the core of occupational health research and practice, access to quality measurement instruments is essential. Ensuring that it is well-designed and its content appropriate to measuring what it claims to measure should not be underestimated. In absolute terms, valid instruments do not exist. Validating a measuring instrument is a process, sometimes complex, in which a base of evidence has to be constructed to support that the instrument meets a number of measurement properties. When quality evidence is provided about the presence or absence of these properties, it is possible to assign a degree of quality to the instrument for a specific purpose. Hence, the methodology used to carry out a validation process becomes the most important determinant to accept or reject the quality of a measurement instrument.

This process becomes more challenging when using a measurement instrument developed in a particular language or culture with the aim of using it in a different one. In these cases, a simple (direct) translation of the questionnaire could be unreliable, because misinterpretation could appear due to language and cultural differences in the perception of work, health and/or disease. In these circumstances, it is necessary to perform a cross-cultural validation, following a systematic procedure. For several authors the cross-cultural validation is part of the construct validation and should be assessed to guarantee the validity of the instrument (28,50-52).

There are several approaches in the literature to address the validation process of a measuring instrument. Some approaches come from internationally renowned experts in the design and validation methodology of questionnaires (28,53-59). Others come from different research groups that have achieved international standards. Among the latter, the following stand out: the consensus-based standards on terminology and recommendations to assess the methodological quality of studies on measurement properties of health status measurements

instruments (COSMIN) (50-52);the standardized methodology for evaluating the measurement of patient-reported outcomes (EMPRO) to assist the choice of instruments (60); the methodology of the Health Technology Assessment Programme (HTA Programme) to evaluate patient-based outcome measures for use in clinical trials (61) and the criteria proposed by the Scientific Advisory Committee of the Medical Outcomes Trust (62).

To state that a questionnaire has been validated, it is necessary to provide evidence about certain features: 1) whether an instrument measures what it purports to measure, 2) how it reflects the theory underlying the phenomenon being measured, 3) the degree to which the scores are an adequate reflection of a gold standard, 4) the extent to which the scores of the instrument are consistent with stated hypotheses,5) the degree of simplicity, feasibility and acceptability to patients, users and researchers, 4) the ability to measure free from error and, therefore, ability to provide reproducible results when applied to individuals who have not changed over time, and 5) the sensitivity to detecting true changes over time. All these features are related to three properties of the questionnaires: validity, reliability and responsiveness.

However, the terminology found in the literature can be confusing for several reasons. First, there are differences in terms used as synonyms for measurement properties (e.g. reliability, repeatability, stability, reproducibility and precision are used interchangeably). Second, there are different definitions given to the same concept (e.g. different authors give different definitions for responsiveness). Third, different research groups evaluate different properties and characteristics of the instruments when assessing their quality (e.g. evaluation of appropriateness, interpretability, acceptability or feasibility are recommended in some guides but not others). Fourth, there is a wide variety of classifications of measurement properties depending on authors and research groups (e.g. some authors, but not all, consider evaluating the cross-cultural adaptation as a part of construct validity;

some consider responsiveness to be an aspect of validity, and also that face validity is an aspect of content validity).

In this thesis a comprehensive review of the literature was conducted to systematize the steps involved in validating a health questionnaire, the *Work Role Functioning Questionnaire (WRFQ)*, following the methodological recommendations which found greater consensus. Next, the requirements for conducting a quality cross-cultural adaptation of health questionnaires were defined in detail and the properties evaluated were based on those most frequently recommended by experts and consensus groups, and then applied to this questionnaire.

REFERENCES

1. WHO: Constitution of the World Health Organization [Internet]. Geneva: World health Organization; 1948-2013. International Health Conference, 1948; [cited 2013 November 19]; Available from: http://apps.who.int/gb/bd/PDF/bd47/SP/constitucion-sp.pdf

2. Wadell G, Burton T, Aylward M. Work and common health problems. J Insur Med. 2007;39:109-20.

3. 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. J Occup Environ Med. 2011;68:806-12.

4. ILO: Promoting jobs, protecting people [Internet]. Geneva: International Labor Organization; 1996-2013. Decent Work; [cited 2013 July 19]; [about 2 screens]; Available from: http://www.ilo.org/global/topics/decent-work/lang--es/index.htm

5. Ross D. Ageing and work: an overview. Occup Med (Lond). 2010;60:169-71.

6. 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.

7. 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.

8. 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.

9. Squires H, Rick J, Carroll C, Hillage J. Cost-effectiveness of interventions to return employees to work following long-term sickness absence due to musculoskeletal disorders. J Public Health (Oxf).2012;34:115-24.

10. Noben CY, Nijhuis FJ, de Rijk AE, Evers SM. Design of a trial-based economic evaluation on the cost-effectiveness of employability interventions among work disabled employees or employees at risk of work disability: the CASE-study. BMC Public Health. 2012;18:12:43.

11. Arends I, Bültmann U, van Rhenen W, Groen H, van der Klink JJ. Economic evaluation of a problem solving intervention to prevent recurrent sickness absence in workers with common mental disorders. PLoS One. 2013;8:e71937.

12. 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.

13. Baldwin ML, Johnson WG, Butler RJ. The error of using returns-to-work to measure the outcomes of health care. Am J Ind Med. 1996;29:632-41.

14. 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.

15. ILO Encyclopaedia of Occupational health and Safety [Internet]. Part III. Management & Policy. Chapter 17: Disability and work [cited October 2013]; Available from: http://www.ilo.org/oshenc/part-iii/disability-and-work/item/170-disability-concepts-and-definitions

16. Arends I, Bruinvels DJ, Rebergen DS, Nieuwenhuijsen K, Madan I, Neumeyer-Gromen A et al. Interventions to facilitate return to work in adults with adjustment disorders. Cochrane Database Syst Rev. 2012, Dec 12;12.

17. Arends I, van der Klink JJ, Bültmann U. Prevention of recurrent sickness absence among employees with common mental disorders: design of a cluster-randomized controlled trial with cost-benefit and effectiveness evaluation. BMC Public Health. 2010;10:132.

18. Amick BC III, Gimeno D. Measuring work outcomes with a focus on health-related work productivity loss. In: Wittink H & Carr D Editors. Evidence, Outcomes & Quality of Life in Pain Treatment: A Handbook for Pain Treatment Professionals. London, UK: Elsevier, 2007. pp. 329-343.

19. Abma FI. Work functioning: development and evaluation of a measurement tool [PhD thesis]. Groningen, NL: University of Groningen; 2012. [Internet]. Available from: http://irs.ub.rug.nl/ppn/351176438

20. Lofland, J. H., Pizzi, L., & Frick, K. D. A review of health-related workplace productivity loss instruments. Pharmacoeconomics. 2004;22:165-84.

21. Prasad, M., Wahlqvist, P., Shikiar, R., & Shih, Y. T. A review of self-report instruments measuring health-related work productivity. Pharmacoeconomics. 2004;22:225-44.

22. Ozminkowski, R. J., Goetzel, R. Z., Chang, S., & Long, S. The application of two health and productivity instruments at a large employer. J Occup Environ Med. 2004;46:635-48.

23. Williams RM, Schmuck G, Allwood S, Sanchez M, Shea R, Wark G. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: a systematic review. J Occup Rehabil. 2007;17:504-21.

24. Beaton DE, Tang K, Gignac MA, 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 (Hoboken). 2010;62:28-37.

25. Nieuwenhuijsen K, Franche RL, van Dijk FJ. Work functioning measurement: tools for occupational mental health research. J Occup Environ Med. 2010;52:778-90.

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. Evans CJ. Health and work productivity assessment: state of the art or state of flux? J Occup Environ Med. 2004;46(6 Suppl):S3-11.

28. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: A practical guide. 1st ed. Cambridge, UK: The University Press Cambridge, 2011.

29. Lerner D, Amick BC 3rd, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Med Care. 2001;39:72-85.

30. Lerner DJ, Amick BC 3rd, Malspeis S, Rogers WH. A national survey of health-related work limitations among employed persons in the United States. Disabil Rehabil. 2000;22:225-32.

31. Schmidt LL, Amick BC 3rd, Katz JN, Ellis BB. Evaluation of an upper extremity student-role functioning scale using item response theory. Work. 2002;19:105-16.

32. Roy JS, MacDermid JC, Amick BC 3rd, Shannon HS, McMurtry R, Roth JH, et al. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. Phys Ther. 2011;91:254-66.

33. 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 2004;27:261-8.

34. 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.

35. 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.

36. Kessler R, Barber C, Beck A, Berglund P, Cleary PD, McKenas D et al. The World Health Organization health and work performance questionnaire (HPQ). J Occup Environ Med. 2003;45;156-74.

37. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4:353-65.

38. Goetzel RZ, Ozminkowski RJ, Long, SR. Development and reliability analysis of the Work Productivity Short Inventory (WPSI) instrument measuring employee health and productivity. J Occup Environ Med. 2003;45:743-62.

39. Stewart WF, Ricci JA, Leotta C, Chee E. Validation of the work and health interview. Pharmacoeconomics. 2004;22:1127-40.

40. 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.

41. Kumar RN, Hass SL, Li JZ, Nickens DJ, Daenzer CL, Wathen LK. Validation of the Health-Related Productivity Questionnaire Diary (HRPQ-D) on a sample of

patients with infectious mononucleosis: results from a phase 1 multicenter clinical trial. J Occup Environ Med. 2003;45:899-907.

42. Meerding WJ, IJzelenberg W, Koopmanschap MA, Severens JL, Burdorf A. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. J Clin Epidemiol. 2005;58:517-23.

43. Wolfe F, Michaud K, Pincus T. Development and validation of the health assessment questionnaire II: a revised version of the health assessment questionnaire. Arthritis Rheum. 2004;50:3296-305.

44. Lerner DJ, Amick BC 3rd, Malspeis S, Rogers WH, Gomes DR, Salem DN. The Angina-related Limitations at Work Questionnaire. Qual Life Res. 1998;7:23-32.

45. 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.

46. Lam RW, Michalak EE, Yatham LN. A new clinical rating scale for work absence and productivity: validation in patients with major depressive disorder. BMC Psychiatry. 2009;9:78.

47. Gärtner FR, Nieuwenhuijsen K, van Dijk FJ, Sluiter JK. Psychometric properties of the Nurses Work Functioning Questionnaire (NWFQ). PLoS One. 2011;6:e26565.

48. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. Psychopharmacol Bull. 1997;33:13-6.

49. 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.

50. 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.

51. 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.

52. 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. 2010;19:539-49.

53. Guillemin F. Cross-cultural adaptation and validation of health status measures. Scand J Rheumatol.1995;24:61-63.

54. Beaton DE, Bombardier C, Guillemin F, Bosi Ferraz M. Guidelines for the process of cross-cultural adaptation of self-reports measures. Spine. 2000; 25:3186-3191.

55. Aday LA, Cornelius LJ. Designing and conducting health surveys: a comprehensive guide. 3rd ed. San Francisco, CA: Jossey-Bass publisher; 2006.

56. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4thed. New York: Oxford University Press Inc.; 2008.

57. García de Yébenes MJ, Rodriguez-Salvanés F, Carmona-Ortells L. Validación de cuestionarios. Reumatol Clin. 2009;5:171-77.

58. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. J Psychosom Res. 2010;68:319-23.

59. 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.

60. Valderas JM, Ferrer M, Mendívil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008;11:700-8.

61. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assessment. 1998;2:1-74.

62. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11:193-205.

2. OBJECTIVES

2.1. Study I Objectives

To review the literature on the methodology for cross-cultural adaptation and validation (CCAV) of health questionnaires and to synthesize recommendations based on the scientific literature to facilitate this process.

To evaluate the degree of compliance with the methodological recommendations for the CCAV of health questionnaires in a selection of Spanish-language scientific journals.

2.2. Study II Objectives

To translate and adapt the Work Role Functioning Questionnaire to Spanish spoken in Spain.

To perform a preliminary evaluation of the Spanish version of the Work Role Functioning Questionnaire psychometric properties by means of a pre-test.

2.3. Study III Objective

To examine the reliability and validity of the Spanish version of the Work Role Functioning Questionnaire in a Spanish-speaking general working population.

2.4. Study IV Objective

To examine the responsiveness of the Spanish version of the Work Role Functioning Questionnaire in a Spanish-speaking general working population.

3. PAPER # 1

Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. ***Salud Pública de México. 2013; 55:57-66.***

Ramada-Rodilla JM, Serra-Pujadas C, Delclós-Clanchet GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. Salud Publica Mex. 2013; 55(1):57-66 http://dx.doi.org/10.1590/S0036-36342013000100009

3. PAPER # 2

Cross-cultural adaptation of the work role functioning questionnaire to Spanish spoken in Spain. ***Journal of Occupational Rehabilitation. 2013;23:566-75.***

Ramada JM, Serra C, Amick BC 3rd, Castaño JR, Delclos GL. Cross-cultural adaptation of the Work Role Functioning Questionnaire to Spanish spoken in Spain. J Occup Rehabil. 2013 Dec;23(4):566-75. doi: 10.1007/s10926-013-9420-6

4. PAPER # 3

Reliability and validity of the Work Role Functioning Questionnaire (Spanish version). [Submitted for peer-review].

Ramada JM, Serra C, Amick BC 3rd, Abma FI, Castaño JR, Pidemunt G, Bültmann U, Delclos GL. Reliability and validity of the work role functioning questionnaire (Spanish version). J Occup Rehabil. 2014 Dec;24(4):640-9. doi:10.1007/s10926-013-9495-0.

TITLE:

Reliability and validity of the Work Role Functioning Questionnaire (Spanish version).

AUTHORS:

José M Ramada Rodilla, MD, MSc[1,2,3]
Consol Serra Pujadas, MD, PhD[1,2,3]
Benjamin C Amick, PhD[4,5]
Femke I Abma, PhD[6]
Juan R Castaño Asins, MD[7]
Gemma Pidemunt Moli, MD, PhD[8]
Ute Bültmann, PhD[6]
George L Delclós Clanchet, MD, MPH, PhD[1,3,4]

AFFILIATIONS:

[1] Centro de Investigación en Salud Laboral (CiSAL), Universidad Pompeu Fabra, Barcelona, España.

[2] Servicio de Salud Laboral, Parc de Salut MAR, Barcelona, España.

[3] CIBER de Epidemiología y Salud Pública (CIBERESP).

[4] Southwest Center for Occupational and Environmental Health, The University of Texas School of Public Health. Houston, Texas, 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. Groningen, The Netherlands.

[7] Psychiatry Service. Parc de Salut MAR. Hospital del Mar. Barcelona, Spain.

[8] Orthopedic Surgery and Traumatology Service. Parc de Salut MAR. Hospital del Mar. Barcelona, Spain.

CORRESPONDING AUTHOR:

José Mª Ramada Rodilla
CiSAL - Universidad Pompeu Fabra
Dr. Aiguader, 88
08003-Barcelona
Correo electrónico: jramada@parcdesalutmar.cat
Tel. 932483066

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 alpha 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 (α=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 to 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 standard 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). 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 assessment:

Reliability is defined as the degree to which the measurement is free from measurement error [27], and can also be defined as the extent to which scores for participants who have not changed are the same for repeated measurement under several conditions [35]: 1) using different sets of items from the same muli-item measurement instrument (internal consistency); 2) over time (test-retest reliability); 3) by different raters on the same occasion (inter-rater reliability) or 4) by the same raters on different occasions (intra-rater reliability). The COSMIN taxonomy [27,35] also considers measurement error as an aspect of reliability.

Validity assessment:

Validity of a questionnaire is defined in the literature as the degree to which an instrument truly measures the construct it purposes to measure. In general, three different types of validity can be distinguished: content validity, criterion validity and construct validity, and within these three main types of validity there are some subtypes [35].

Content validity focuses on whether the content of the instrument corresponds with the construct that the instrument measures, with regard to relevance and comprehensiveness. This type of validity is frequently assessed by means of a systematic empiric procedure in which the authors of the questionnaire, a panel of experts and a sample of the target population participate. It was already assessed in our previous manuscript about the cross-cultural adaptation of the Work Role Functioning Questionnaire (WRFQ), following rigorously the recommendations of the literature [23].

Criterion validity can be assessed only in situations in which there is a gold standard for the construct to be measured, and refers to how well the scores of the measurement instrument agree with the scores obtained with the gold standard.

Since 'Work Functioning' is a construct that has not a gold standard, this type of validity cannot be assessed for the Work Role Functioning Questionnaire (WRFQ).

Construct validity should be evaluated in those situations in which there is no gold standard, and 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 evaluating the 'cross-cultural validity', which we already did in our previous manuscript [23]; the 'structural validity' which we carried out by means of a Confirmatory Factor Analysis (CFA) and 'hypotheses testing', which we carried out testing seven hypotheses.

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 Kruskall 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 using 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 to analyze the structural validity of the WRFQ-SpV, 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 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 ≤ 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 ≥ 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 ≥ 0.40 as evidence of moderate or strong correlations [40,41].

Construct validity was assessed by means of hypotheses testing. Significance of the differences among groups were tested using the non-parametric Kruskall 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 ≤ r ≤ 0.7= 'moderate; r > 0.7= 'strong' [41].

The basic principle of construct validation by means of hypotheses testing is that hypotheses are formulated about differences in the instrument scores between subgroups of participants or about the relationships of the scores of the instrument under study with scores on other similar or dissimilar measuring tools [35], therefore, seven hypotheses were formulated to asses construct validity:

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.

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 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)

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).

Table 2 shows the mean, SD and median scores for each WRFQ-SpV subscale and the overall scale. Higher values indicate better work functioning (less disability at work). 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.

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

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

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 / √2

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.

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.

	Health issue (median scores[a])				Job type (median scores[a])			
Work demands	**None (N= 156)**	**Mental (N= 125)**	**Physical (N= 139)**	**p[(b)] value**	**Manual (N= 109)**	**Non-manual (N= 123)**	**Mixed (N= 212)**	**p[(b)] value**
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.

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.

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.

	Median scores[a]											
	No health issues (N= 156)				Physical health issues (N= 139)				Mental health issues (N= 125)			
Work demands	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.[illegible]
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.

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

Age (Years)	18-35 (n=148)	36-45 (n=121)	p [(a)]	18-35 (n=148)	46-55 (N=123)	p [(a)]	18-35 (n=148)	56-65 (N=57)	p [(a)]	36-45 (n=121)	46-55 (N=123)	p [(a)]	36-45 (n=121)	56-65 (N=57)	p [(a)]	46-55 (N=123)	56-65 (N=57)	p [(a)]
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 score	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.

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.

Six hypotheses 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 he 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 he has no conflict of interest; author Ute Bültmann declares that he has no conflict of interest and 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. Psichopharmacol Bull. 1997;33:13-6.

12. Tuomi K, Ilmarien 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; [Epub ahead of print].

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/erilisjulkaisut/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. Measurement 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. 4thed. 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_ENCT.pdf

48. Shah D. Healthy worker effect phenomenon. Indian J Occup Environ Med. 2009;13:77-9.

5. PAPER # 4

Responsiveness of the Work Role Functioning Questionnaire (Spanish version). Journal of Occupational and Environmental Medicine. [In Press 2013].

Ramada JM, Delclos GL, Amick BC 3rd, Abma FI, Pidemunt G, Castaño JR, Bültmann U, Serra C. Responsiveness of the Work Role Functioning Questionnaire (Spanish version) in a general working population. J Occup Environ Med. 2014 Feb;56(2):189-94. doi: 10.1097/JOM.0000000000000074

TITLE:

Responsiveness of the Work Role Functioning Questionnaire (Spanish version) in a general working population.

Running title: Work Role Functioning Questionnaire Responsiveness

AUTHORS:

José M Ramada Rodilla, MD, MSc[1,2,3]
George L Delclós Clanchet, MD, MPH, PhD[1,3,4]
Benjamin C Amick, PhD[4,5]
Femke I Abma, PhD[6]
Juan R Castaño Asins, MD[7]
Gemma Pidemunt Moli, MD, PhD[8]
Ute Bültmann, PhD[6]
Consol Serra Pujadas, MD, PhD[1,2,3]

AFFILIATIONS:

[1] Centro de Investigación en Salud Laboral (CiSAL), Universidad Pompeu Fabra, Barcelona, España.
[2] Servicio de Salud Laboral, Parc de Salut MAR, Barcelona, España.
[3] CIBER de Epidemiología y Salud Pública (CIBERESP).
[4] Southwest Center for Occupational and Environmental Health, The University of Texas School of Public Health. Houston, Texas, 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. Groningen, The Netherlands.
[7] Psychiatry Service. Parc de Salut MAR. Hospital del Mar. Barcelona, Spain.
[8] Orthopedic Surgery and Traumatology Service. Parc de Salut MAR. Hospital del Mar. Barcelona, Spain.

CORRESPONDING AUTHOR:

José Mª Ramada Rodilla

Occupational Health Service. Parc de Salut Mar.

Passeig Marítim, 25-29. 08003-Barcelona. Spain

Email address: jramada@parcdesalutmar.cat

Telephone number: +34 932483066;

Fax number: +34 933160410

CONFLICTS OF INTEREST AND SOURCE OF FUNDING:

This project was supported by a grant from the "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, Spanish Government".

Conflicts of interest: none declared.

CLINICAL SIGNIFICANCE

The study provides evidence that the WRFQ-SpV is an appropriate instrument to measure (true) changes in health-related work functioning over time. However, more research is needed to assess the ability of the instrument detecting (true) changes in groups whose health is stable or deteriorates in individualized health-related work functioning surveillance.

ABSTRACT

Objective: To examine the responsiveness of the Work Role Functioning Questionnaire Spanish-Version (WRFQ-SpV) so that it could be used in evaluative studies.

Methods: A longitudinal survey was performed. Combinations of distribution and anchor-based approaches were used. Five hypotheses were tested examining validity of change-scores. The consensus-based standards for the selection of health status measurement instruments (COSMIN) guided the study design.

Results: One hundred and two participants (mean age=47.3; SD=10.3) completed the WRFQ-SpV twice, within a mean interval of 3.7 months (SD=1.8). Four hypotheses were confirmed and one was rejected. It was verified that the WRFQ-SpV was able to detect (true) changes over time.

Conclusion: Suggestive evidence about the possible use of the WRFQ-SpV with evaluative purposes was provided. More research is needed to examine the instrument responsiveness for groups whose health is stable or deteriorates.

Key terms: responsiveness; measurement instrument; work-functioning instrument.

INTRODUCTION

Increasing life expectancy and delayed retirement age are creating an older active workforce. Many of the older workers likely have health issues due to chronic diseases that require some form of adaptation while working [1,2]. A paradigm shift is needed in occupational health care settings to sustain a productive labour force throughout a workers career. Helping workers to remain healthy at work should be an important target for occupational health research and practice [3].

Work functioning (WF) is determined by the joint influence of work and health and should be viewed as a continuum rather than a dichotomy [4]. Health is a dynamic concept that changes over time. Regarding sickness absence, effective workplace and/or individual interventions could contribute to an early and sustainable return to work.

Both, the interventions and the persons should be evaluated and monitored in this process. Traditional work outcome measures, such as “present versus absent” are no longer sufficient [5]. Validated instruments are required to evaluate the impact of health on WF and quality instruments are needed to examine “functioning at work” and to perform health-related WF surveillance.

A number of self-reported workplace productivity measurement instruments and work-role specific functioning questionnaires have been developed and compared [6,7,8,9]. Recently, the Work Role Functioning Questionnaire (WRFQ) [4], designed to measure health-related WF and freely available in the literature for professionals and researchers, has been successfully translated, adapted and validated to be used in different languages [10-14].

Recent reviews have found that health-related work outcome measures and health-related WF instruments need better validation to make them more meaningful for researchers, practitioners and patients [15,16]. If an instrument is only used to discriminate between patients at one point in time, responsiveness of

the tool is not usually an issue. But for evaluative purposes, when it is intended to measure change in longitudinal studies, this property is very relevant [17].

Responsiveness is the ability of an instrument to detect true change over time in the construct measured. It refers to the validity of a change score, which is the difference between two scores estimated on the basis of at least two measurements [17,18]. Responsiveness of the WRFQ has only been examined in the Dutch version. The Dutch researchers recommended further responsiveness assessments in samples expected to experience changes over time in either work conditions or health status [14].

The ultimate goal of many interventions in occupational health practice and research is to improve WF, and assessing whether a worker's WF status has changed over time is often one of the most important measurement purposes. Therefore, the objective of this study was to examine the responsiveness of the Spanish version of the WRFQ (WRFQ-SpV) so that it could be used in evaluative studies and/or as a monitoring or surveillance instrument.

METHODS

Procedures and sample characteristics

A longitudinal survey was conducted to examine the responsiveness of the WRFQ-SpV. The consensus-based standards on terminology and recommendations to assess the methodological quality of studies on measurement properties of health status measurements instruments (COSMIN) guided the design of the study [18-20].

All participants were recruited, before starting medical treatment, through the outpatient services of psychiatry, physical medicine and rehabilitation, orthopedic surgery and traumatology at a large public hospital system in Barcelona (Spain). Inclusion criteria were: 1) active workers of both sexes, age 18 years and over, 2)

attending his/her first hospital specialist visit, 3) working at least 10 hours per week in the past four weeks, and 4) able to read and understand Spanish. Participants were excluded if they had plans to stop working in the next six months.

Participants were invited to complete the WRFQ-SpV twice on paper (before and after the treatment). At the time of first completion, they provided information on age, gender, level of education (primary, secondary, higher), job type (manual, non-manual, mixed), working hours and health condition (none, musculoskeletal, mental health, others).

One to six months after finalizing treatment at the hospital all participants were invited to answer the WRFQ-SpV again. At that time, a single global perceived effect question (GPE-Q) was added asking respondents to rate their change in WF compared to their pre-treatment baseline, with response options for deterioration from -6 (much worse) to -1 (slightly worse), 0 for no change, and rating improvement from +1 (slightly better) to +6 (much better).

Measure

The WRFQ is a self-administered questionnaire that measures perceived difficulties to perform the job due to health problems [4]. Instructions to use the instrument have been described elsewhere [12]. The WRFQ-SpV consists of 27 items, grouped into 5 subscales reflecting different work demands (work scheduling, output, physical, mental and social). Each item is scored on a Likert five-point scale, anchored to percentages of working time with difficulty handling certain parts of the job. Response options 0, 2 and 4 are anchored to 0%, 50% and 100% respectively. Response option ’does not apply to my job’ is transformed into a missing value. Total scales and/or subscales containing more than 20% missing values are considered missing.

Statistical analysis

The WRFQ-SpV median scores, ranges, mean change scores and standard deviations of change (SD change) were determined. The standard error of measurement (SEM) and Cronbach alpha for the overall scale and each subscale were calculated to evaluate the reliability of the questionnaire. Floor and ceiling effects were explored. These effects occur when more than 15% of responses to a certain item cluster at the top or the bottom of the scale [21].

According to de Vet et al. [17] and the COSMIN panel [18-20] responsiveness is an aspect of validity, and its assessment should emphasize evaluating the validity of change scores. Therefore, analogous to validity, testing hypotheses formulated a priori (before data collection and analysis), concerning the expected correlations and/or expected relationships in different groups measured with the instrument and the GPE-Q, is considered an appropriate method.

A distribution approach was used to estimate the Minimal Important Change (MIC) and to evaluate the hypotheses based on SEM and Effect Size (ES). For those hypotheses addressing correlations and expected change scores in different subgroups of participants an anchor-based approach with the GPE-Q was used.

The following five hypotheses were formulated:

Hypothesis 1: Changes in WF were expected in participants as a result of the treatment, so it was hypothesized that changes in WF over the period of treatment, assessed by the GPE-Q correlate moderate to strongly with the change scores assessed by the WRFQ-SpV. Correlations were assessed using Pearson's correlation coefficient (r), interpreting $r \leq 0.4$ ='weak'; $0.4 \leq r \leq 0.7$ ='moderate' and $r > 0.7$= 'strong' [22].

Hypothesis 2: Treatment was expected to result in improvement, not deterioration, although both were possible. Hence, it was hypothesized a substantially greater effect size (ES) for improvement of WF than for deterioration.

To examine this hypothesis, Cohen's d effect size (ES = difference of means divided by the pooled SD) and the standardized response mean (SRM = mean change divided by SD of change) were calculated as an estimate of magnitude of change over time. Cohen's d ES thresholds were used for interpretation of Cohen's d effect size. Respondents were categorized as deteriorating group (-6 to -1) or improving group (+1 to +6), because statistical methods underlying the SRM assume that all participants change in the same direction [23].

Small ES values ($0.20 \leq ES \leq 0.50$) were hypothesized for participants reporting deterioration in WF, and large ES values ($ES \geq 0.80$) for those reporting improvement.

Hypothesis 3: Positive changes in WF reported on the GPE-Q, predict positive changes in the scores of the WRFQ-SpV. Therefore, it is expected that participants reporting improvement in WF on the GPE-Q, show an "important change" for the overall scale and each subscale of the WRFQ-SpV.

"Important change" was defined as a mean change score larger than both the MIC and the SEM, because responsiveness was being examined for evaluative purposes [21]. Statistical significance of the differences between the WRFQ-SpV scores found at baseline and at follow-up was assessed by means of paired t tests.

Since the WRFQ is a Patient Reported Outcome (PRO), the MIC was considered from the perspective of the patient, and was therefore defined using the GPE-Q as an anchor, as the smallest change of the WRFQ-SpV score which participants perceive as minimally important [17]. The MIC value for improvement was set at

the mean change score of participants reporting an improvement from +1 (slight improvement) to +2 (some improvement).

Hypothesis 4: Participants, who received treatment for physical issues, reporting improvement on the GPE-Q, show an "important change" for the WRFQ-SpV subscale of physical demands (defining "important change" as a change score larger than both, the MIC and the SEM). Statistical significance of change was assessed by means of paired t tests.

Hypothesis 5: Participants, who received treatment for mental health issues, reporting improvement on the GPE-Q, show an "important change" for the WRFQ-SpV subscale of mental demands ("important change" defined like in previous hypothesis). Statistical significance of change was assessed by means of paired t tests.

The contents of the study and the informed consent form were reviewed and approved by the Clinical Research Ethical Committee of the Parc de Salut Mar (Barcelona, Spain) and respect all the principles of the Declaration of Helsinki and Spanish legal regulations on protection of personal data. All analyses were performed with SPSS (Version 15.0. Chicago, IL; 2006).

RESULTS

Sample characteristics:

Table 1 shows the study sample characteristics. A total of 102 participants with a mean age of 47.3 years (SD=10.3) completed the WRFQ-SpV and were included in the analyses. All participants answered the questionnaire twice within a mean interval of 3.7 months (SD=1.8). All were active employees working an average of 39.6 hours per week (SD=7.2), with various levels of education, job types and health problems. Women and participants with high educational level were over

represented in this sample, compared to the general working Spanish population [24].

Median scores, ranges, mean change scores, SD change, SEM and Cronbach alpha for the overall scale and each subscale are presented in table 2.

Floor and ceiling effects:

The overall scale did not show floor or ceiling effects. Subscales did not show floor effects, but ceiling effects were found for the subscales of work scheduling (15%), mental (27%) and social demands (24%) [21] (table 2).

Responsiveness by means of hypothesis testing:

Hypothesis 1: Correlations between the GPE-Q scores and the overall change scores on the WRFQ-SpV for subgroups of participants reporting improvement or deterioration were r=0.5 (p=0.001) and r=0.6 (p=0.002), respectively. Correlations for all subscales and the overall scale in each subgroup were also above 0.4, except for the subscale of social demands in the subgroup of participants improving (table 3).

Hypothesis 2: ES values and SRM are presented in table 4. For the evaluation of change in WF, 34 participants reported deterioration by means of the GPE-Q (Mean= -3.53; SD=1.64), obtaining a mean change score in the WRFQ-SpV of -8.45 (SD=12.67). For those reporting deterioration, Cohen's d ES and SEM were -0.34 and -0.67, respectively. A total of 49 participants reported improvement (GPE-Q Mean=+3.73; SD=1.54) and an average increase in WRFQ-SpV of +22.35 (SD=20.83). For those reporting improvement, Cohen's d ES and SRM were 1.09 and 1.07 respectively, confirming hypothesis 2.

Table 1. Sample characteristics (n=102).

		Total n=102	Men n=45 (44%)	Women n=57 (56%)
Age in years, mean (SD)[a]		47.3 (10.3)	45.9 (9.4)	48.5 (10.8)
Education level, N (%)	Low	28 (27.5)	12 (26.7)	16 (28.1)
	Middle	38 (37.3)	20 (44.4)	18 (31.6)
	High	36 (35.2)	13 (28.9)	23 (40.3)
Working hours/week, mean (SD)[a]		39.6 (7.2)	41.9 (7.4)	37.7 (6.4)
Job type, N (%)	Manual	36 (35.3)	14 (31.1)	22 (38.6)
	Non-manual	27 (26.5)	10 (22.2)	17 (29.8)
	Mixed	39 (38.2)	21 (46.7)	18 (31.6)
Reported health issue type, N (%)	Physical	49 (48.0)	15 (33.3)	34 (59.7)
	Mental health	33 (32.4)	20 (44.4)	13 (22.8)
	Others	20 (19.6)	10 (22.3)	10 (17.5)
Disease duration in months:	Mean, (SD)	22.4 (36.1)	15.2 (16.8)	28.0 (45.3)
	Median, Range	12.0 1-300	12.0 1-60	12.00 1-300

(a): Standard deviation

Table 2. Spanish version of the Work Role Functioning Questionnaire: scores, floor and ceiling effects and Cronbach alpha at baseline. Standard error of measurement, (n=102).

	Valid n[a] (missing/not applicable)	Baseline[b] median scores	Baseline scores range	Baseline n at floor (0%)	Baseline n at ceiling (100%)	Mean change scores / (SD change)[c]	Standard error measurement (SEM)	Cronbach alpha at baseline
Work scheduling demands	98 (4)	80.6	0.00-100	1 (1.0)	15 (15.3)	5.8 (28.0)	19.8	0.93
Output demands	98 (4)	75.0	0.00-100	1 (1.0)	9 (9.2)	9.5 (24.8)	17.6	0.92
Physical demands	74 (28)	70.0	4.17-100	0 (0.0)	10 (13.5)	7.9 (26.3)	18.6	0.95
Mental demands	101 (1)	83.3	0.00-100	3 (3.0)	27 (26.7)	10.2 (24.3)	17.2	0.94
Social demands	88 (14)	83.3	0.00-100	2 (2.7)	21 (23.9)	10.0 (22.9)	11.5	0.93
Total scale	98 (4)	77.9	6.48-100	0 (0.0)	3 (3.1)	8.6 (21.7)	15.0	0.94

(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 mean better work functioning: difficulties all the time 0/100; difficulties no of the time 100/100.

(c) SD change: standard deviation of change.

Table 3. Bilateral correlations between the scores of the Global Perceived Effect Question scores and the WRFQ-SpV[a] change scores.

Participants declaring Improvement (n=49)	**Mean Change score WRFQ-SpV[a] (SD)**	**Pearson's (r)[b]**	**Bilateral significance**
Work Scheduling Demands	22.9 (25.6)	0.6	< 0.001
Output Demands	25.6 (23.4)	0.4	0.001
Physical Demands	23.3 (27.0)	0.4	0.165
Mental Demands	23.1 (26.6)	0.4	0.017
Social Demands	20.9 (27.3)	0.3	0.132
Overall Scale	22.3 (20.8)	0.5	0.001
Participants declaring Deterioration (n=31)	**Mean Change score WRFQ-SpV[a]**	**Pearson's (r)[b]**	**Bilateral significance**
Work Scheduling Demands	-24.0 (17.7)	0.6	0.006
Output Demands	-14.3 (10.5)	0.4	0.028
Physical Demands	-17.7 (19.4)	0.5	0.028
Mental Demands	-12.1 (10.6)	0.6	0.047
Social Demands	-16.7 (12.3)	0.4	0.484
Overall Scale	-10.5 (10.8)	0.6	0.002

(a) WRFQ-SpV: Work Role Functioning Questionnaire (Spanish version).

(b) Pearson's correlation coefficient.

Table 4. Assessment of responsiveness by means of the Standardized Response Mean, using Cohen's d Effect Size thresholds for interpretation, (n=102).

	Participants reporting deterioration in WF[a] (N=34)	**Participants reporting improvement in WF[a] (N=49)**
Mean score of the GPE-Q[b] (SD)	-3.53 (1.64)	3.73 (1.54)
WRFQ-SpV[c]mean score at baseline (SD)	71.08 (24.32)	66.09 (25.83)
WRFQ-SpVmean score at follow-up(SD)	62.63 (25.54)	88.44 (11.43)
Mean change (SD); p value[d]	-8.45 (12.60); p=0.001	22.35 (20.83); p<0.001
Standardized response mean (SRM)	-0.67	1.07
Cohen's d effect size (with pooled SD change)	-0.34	1.12

(a) WF: Work functioning (by means of the GPE-Q);
(b) GPE-Q: Global Perceived Effect Question.
(c) Spanish version of the Work Role Functioning Questionnaire.
(d) Paired t test.

Hypothesis 3: Participants reporting improvement in WF on the GPE-Q (n=49), showed mean change scores in all subscales and the overall scale above the MIC values for improvement and the SEM. All change scores in this group were statistically significant ($p < 0.001$) (table 5).

Hypothesis 4: Participants who received treatment for physical issues, reporting improvement on the GPE-Q (n=18), showed a significant change score in the subscale of physical demands (mean change score=19.2; SD=17.0; $p<0.001$), that was above the MIC value for improvement (11.3) and the SEM (12.0), confirming hypothesis 4 (table 5).

Hypothesis 5: Participants who received treatment for mental health issues, reporting improvement in the GPE-Q (n=24), showed a significant change score (mean change score=18.1; SD=28.9; $p=0.006$) that was above the MIC value for improvement (7.4) but below the SEM (20.5), rejecting hypothesis 5 (table 5).

Table 5. Scores on the Spanish version of the Work ole Functioning Questionnaire of participants under treatment, declaring improvement in the GPE-Q, by type of health issue (mental or physical).

	All participants, improvement (N=49)					
	WRFQ-SpV[(a)] Mean (SD)	WRFQ-SpV[(b)] Mean (SD)	Mean change score (SD)	**p value**	**MIC[(c)] improvement**	**SEM[(d)]**
Work scheduling demands	63.4 (30.2)	85.6 (17.3)	24.8 (27.0)*	<0.001	7.3	19.1
Output demands	61.9 (30.8)	86.8 (14.2)	23.9 (31.0)*	<0.001	11.1	21.9
Physical demands	70.1 (27.8)	87.7 (12.4)	20.4 (26.7)*	<0.001	11.3	18.9
Mental demands	68.7 (30.9)	91.8 (10.4)	23.1 (26.6)*	<0.001	7.4	18.8
Social demands	74.6 (26.7)	92.2 (10.8)	20.4 (28.8)*	<0.001	11.1	20.4
Total score	66.1 (25.8)	88.4 (11.4)	22.4 (20.8)*	<0.001	9.0	14.7
	Physical health issues, improvement (N=18)					
	WRFQ-SpV[(a)] Mean (SD)	WRFQ-SpV[(b)] Mean (SD)	Mean change score (SD)	**p value**	**MIC[(c)] improvement**	**SEM[(d)]**
Work scheduling demands	67.6 (28.3)	84.0 (20.6)	16.5 (19.0)	0.003	7.3	13.4
Output demands	64.1 (32.5)	85.0 (18.0)	20.9 (19.3)	<0.001	11.1	13.6
Physical demands	64.6 (26.8)	83.8 (13.7)	19.2 (17.0)*	<0.001	11.3	12.0
Mental demands	74.3 (35.4)	92.6 (12.6)	18.3 (25.3)	0.007	7.4	17.9
Social demands	82.1 (26.3)	89.9 (14.3)	7.7 (18.9)	0.150	11.1	13.4
Total score	68.5 (25.6)	86.7 (13.9)	18.3 (16.0)	<0.001	9.0	11.3
	Mental health issues, improvement (N=24)					
	WRFQ-SpV[(a)] Mean (SD)	WRFQ-SpV[(b)] Mean (SD)	Mean change score (SD)	**p value**	**MIC[(c)] improvement**	**SEM[(d)]**
Work scheduling demands	59.6 (30.3)	88.0 (13.6)	28.3 (30.0)	<0.001	7.3	21.2
Output demands	62.2 (25.9)	81.0 (18.4)	18.8 (28.8)	0.006	11.1	20.4
Physical demands	74.1 (28.1)	85.9 (22.3)	11.8 (17.7)	0.014	11.3	12.5
Mental demands	61.8 (29.3)	79.9 (25.4)	18.1 (28.9)*	0.006	7.4	20.5
Social demands	70.3 (25.5)	80.4 (20.7)	10.1 (27.6)	0.092	11.1	19.5
Total score	61.7 (26.1)	89.0 (9.6)	27.3 (24.2)	<0.001	9.0	17.1

(a) Mean scores at baseline; (b) Mean scores at follow-up (after treatment);

(c) Minimal Important Change (MIC); (d) Standard error of measurement

(*) Hypotheses 3 and 4 confirmed; hypothesis 5 rejected.

DISCUSSION

Four out of five hypotheses were confirmed to examine whether the WRFQ-SpV can be used for evaluative purposes. All hypotheses tested in this study were concerned with the expected changes in participants over time, based on the knowledge of the questionnaire structure and conceptual framework. Consequently, confirming such hypotheses supports the validity of the change scores and hence, the responsiveness of the instrument [17,18,20,21].

All correlations between the WRFQ-SpV and the GPE-Q were confirmed as expected in hypothesis 1, with the exception of the social demands subscale in participants reporting improvement, supporting the validity of the change scores and therefore providing insight about the WRFQ-SpV responsiveness.

Assessing responsiveness by means of ES values and SRM, as it has been carried out in this study to test the second hypothesis, has created some controversy in the literature [17,25]. While many authors have used ES as a measure of responsiveness [26-29] and SRM as one of the more valid measures for its estimation [30], others disagree as they consider ES values as a measure of the *magnitude* of the change scores, rather than its *validity* [17,20,25]. In this study, ES values and SRM were used to test a predetermined hypothesis (grounded in the construct of the questionnaire) about the expected magnitude of the ES for different groups of participants, providing additional understanding on the validity of the change scores, and hence, suggesting that the WRFQ-SpV is a responsive questionnaire.

Hypotheses 3 and 4 were confirmed, but hypothesis 5 was rejected because the mean change score of the WRFQ-SpV for the subscale of mental demands was above the MIC but below the SEM. Considering responsiveness as ‘the ability to detect change in general’ would have lead us to confirm hypothesis 5, but detecting ‘any’ change should not be considered responsiveness because ‘any’

change could refer to true change but also to measurement error or other bias [17,31], and the interest in the study was seeking true changes.

It could be debated what is the best approach to estimate the MIC because it depends on the baseline values of the instrument [32],the type of anchor from the patient or the clinician's perspective [33], the definition of ‘minimal important change’ and the direction of change [17,32]. According to de Vet et al. [17], it is reasonable to consider MIC values from the patient’s perspective for self-reported instruments. Consequently, it was considered appropriate to calculate MIC values from this perspective.

Using receiver operating characteristic (ROC) curves and area under ROC curve (AUC) values to evaluate responsiveness would have provided additional strength to the results. AUC values are interpreted as the probability that a measure correctly discriminates between participants who have improved and those who have not. However this method could not be used because no gold standard is available for WF. Many studies have used the GPE-Q as a gold standard, but it is doubtful the reliability and validity of such retrospective measures of change when used as a gold standard [17,31,34].

It could be argued whether confirming four hypotheses is enough to conclude that the WRFQ-SpV is a responsive instrument or not, or that other hypotheses could had been formulated regarding the expected scores of participants reporting deterioration. This study design was not appropriate to examine such hypotheses. As expected, the number of participants who deteriorated following treatment was small for analysis. Several authors agree that it is not possible to formulate standards on the number of hypothesis that need to be tested because testing responsiveness is a continuous process of accumulating evidence [17,21,31] and this study does provide evidence about the responsiveness of the questionnaire, mainly in participants reporting improvement.

The results of this study on responsiveness are consistent with those obtained for the Dutch WRFQ version, which showed moderate responsiveness. The difference between both studies is that in this study, responsiveness was assessed in a larger group of participants for whom a change was expected due to the treatment. For the Dutch version responsiveness was assessed in a relatively stable and healthy population, with no intervention between the first and the second test, and the number of those reporting change was very small [14].

Ceiling effects were identified for the subscales of work scheduling, mental and social demands. These results are consistent with other studies concerning the WRFQ subscales of mental and social demands [10,12-14]. It seems that certain items of the WRFQ have a lack of discriminative ability to differentiate workers at the higher range of WRFQ scores. It could be argued whether these ceiling effects could affect responsiveness, because participants scoring at the upper part of the scale at baseline cannot show further improvement.

To the authors' knowledge, this is the first study to evaluate responsiveness in a population of workers expected to show changes in WF over time.

A significant strength of this longitudinal study is that responsiveness was examined by focusing more on validity than on the magnitude of the change scores. Consistent with the literature [17,18,25,31], a valid instrument for evaluative purposes or to be used as surveillance instrument should provide evidence of its capacity to measure true changes.

Experts do not agree on a single preferred approach for responsiveness evaluation, and instead recommend combining several approaches, including both anchor and distribution based methods [23,35-38]. Hence, both approaches were combined to make the evaluation consistent with these recommendations.

This study also has limitations. It was carried out with a sample size that was initially considered adequate (n=102) [39]. However, stratification was necessary

and some subgroups were of smaller sample sizes. Further longitudinal studies with a larger number of participants and different health issues are needed, especially to examine responsiveness of the instrument in groups who do not experience improvement, or deteriorate.

The intervention was not uniform for all participants, with a wide range of treatments, from non-invasive (e.g., drug treatment, psychological support, occupational therapy or physical rehabilitation) to invasive (e.g., tendon or joint infiltration, minor outpatient or major surgery), which could cause different degrees of change in participants. As other authors have noted [17,31,40], the challenge in the study design was to find a population with expected changes over time. Therefore, for this study purpose, the main point was to secure a sample of participants in such a way that it would be expected for a proportion of participants to either improve or deteriorate, and then explore whether the questionnaire was able to detect the validity of the changes.

Quality validated instruments are needed for occupational health research and practice. Additionally, evidence on WRFQ responsiveness is needed if it is going to be used for evaluative purposes. This study provides suggestive evidence that the WRFQ is an adequate instrument to measure changes in health-related WF over time. However, more evidence is needed on the ability of the instrument to detect changes in groups who do not experience improvement, or deteriorate during individualized health-related WF surveillance.

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. 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.

5. Abma FI. Work functioning: development and evaluation of a measurement tool [PhD thesis]. Groningen, NL: University of Groningen; 2012. [Internet]. [cited 2013 May 28]; [about 160 p.]. Available from: http://irs.ub.rug.nl/ppn/351176438

6. Mattke S, Balakrishnan A, Bergamo G, Newbwrry SJ. A review of methods to measure health-related productivity loss. Am J Manag Care. 2007;17:547-79.

7. Ozminkowski RJ, Goetzel RZ, Chang S, Long S. The application of two health and productivity instruments at a large employer. J Occup Environ Med. 2004;46:635-48.

8. Prasad M, Wahlqvist P, Shikiar R, Shih YC. A review of self-report instruments measuring health-related work productivity: a patient-reported outcomes perspective. Pharmacoeconomics. 2004;22:225-44.

9. Lofland JH, Pizzi L, Frick KD. A review of health-related workplace productivity loss instruments. Pharmacoeconomics. 2004;22:165-84.

10. 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. 2004;27:261-8.

11. 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.

12. 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.

13. 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.

14. 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.

15. 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.

16. 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.

17. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Measurement in medicine: A practical guide. 1st ed. Cambridge, UK: The University Press Cambridge, 2011.

18. 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.

19. 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.

20. 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. 2010;19:539-49.

21. 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.

22. Evans JD. Straightforward statistics for the behavioral sciences. Pacific Grove, CA: Brooks/Cole Pub. Co.; 1996.

23. Liang MH. Longitudinal construct validity: establishment of clinical meaning in patient evaluative instruments. Med Care. 2000; 38(9 Suppl):II84-90.

24. Observatorio Estatal de Condiciones de Trabajo [Internet]. VI Encuesta Nacional de Condiciones de Trabajo. [cited 2007]; [about 117 screen, pages 20-23]; Available from:

ttp://www.oect.es/Observatorio/Contenidos/InformesPropios/Desarrollados/Ficheros/Informe_VI_ENCT.pdf

25. Angst F. The new COSMIN guidelines confront traditional concepts of responsiveness. BMC Med Res Methodol. 2011 Nov 18;11:152; author reply 152.

26. Hsueh IP, Hsieh CL. Responsiveness of two upper extremity function instruments for stroke in patients receiving rehabilitation. Clin Rehabil. 2002;16:617-24.

27. Lin JH, Hsu MJ, Sheu CF, Wu TS, Lin RT, Chen CH, et al. Psychometric comparisons of 4 measures for assessing upper-extremity function in people with stroke. Phys Ther. 2009;89:840-50.

28. Hsieh YW, Wu CY, Lin KC, Chang YF, Chen CL, Liu JS. Responsiveness and validity of three outcome measures of motor function after stroke rehabilitation. Stroke. 2009;40:1386-91.

29. Johansson C, Bodin P, Kreuter M. Validity and responsiveness of the spinal cord index of function: an instrument on activity level. Spinal Cord. 2009;47:817-21.

30. Sivan M. Interpreting effect size to estimate responsiveness of outcome measures. Stroke. 2009;40:e709; author reply e710-1.

31. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4th ed. New York, USA: Oxford University Press Inc., 2008.

32. Crosby RD, Kolotkin RL, Williams GR. Defining clinically meaningful change in health-related quality of life. J Clin Epidemiol. 2003;56:395-407.

33. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE Jr. Determining minimally important changes in generic and disease-specific health-related

quality of life questionnaires in clinical trials of rheumatoid arthritis. Arthritis Rheum. 2000;43:1478-87.

34. Norman GR, Stratford P, Regehr G. Methodological problems in the retrospective computation of responsiveness to change: the lesson of Cronbach. J Clin Epidemiol. 1997;50:869-79.

35. Dworkin RH, Turk DC, Wyrwich KW, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. J Pain. 2008;9:105-21.

36. Beaton DE, Hogg-Johnson S, Bombardier C. Evaluating changes in health status: reliability and responsiveness of five generic health status measures in workers with musculoskeletal disorders. J Clin Epidemiol. 1997;50:79-93.

37. Revicki D, Hays RD, Cella D, et al. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. J Clin Epidemiol. 2008;61:102-9.

38. Wright JG, Young NL. A comparison of different indices of responsiveness. J Clin Epidemiol. 1997;50:239-46.

39. 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.

40. Beaton DE, Tang K, Gignac MA, 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 (Hoboken). 2010;62:28-37.

6. GENERAL DISCUSSION

In this general discussion a short summary of the main findings will be presented, strengths and limitations will be discussed, together with implications for research and practice, and recommendations for future research.

The overall objective of this thesis was to provide an instrument to measure health-related work functioning, validated through a quality process, for use in Spanish-speaking populations. This objective arose from the need for instruments able to describe the extent to which workers improve or deteriorate in their ability to meet job demands, and to assess the impact of health on work functioning. This overall objective was structured into three main themes:

A. Definition and description of the concept of health-related work functioning, and selection of an instrument to measure this construct.

B. A literature review on the methodology for cross-cultural adaptation and validation (CCAV) of health questionnaires.

C. The translation, cross-cultural adaptation and validation of the Work Role Functioning Questionnaire.

6.1. The concept of health-related work functioning

Work functioning is a broad concept defined in the literature as the ability of a worker to meet job demands for a given health status. It can be viewed as a continuum of situations, varying from working successfully (with full participation or being able to respond to all the demands of the job) to sickness absence (absenteeism).

In the literature, it is possible to find different perspectives to define and describe this concept.

Amick and Gimeno (1) describe two dimensions of work functioning in relation to the impact of a health condition. The first dimension refers to the impact of this condition on the output of the worker in terms of economic consequences. The instruments to measure this dimension are mainly related to job performance and productivity loss, often measured by the costs associated with absenteeism and presenteeism, which are concepts also related to our phenomenon of interest. Work (health-related) absenteeism can be defined as an employee's non attendance at work due to illness, causing absences, short and long-term disability. Absenteeism would represent one of the extremes of work functioning, where work performance is absolutely impossible. Work presenteeism has different meanings depending on the context where it is applied. In several European countries the term (sickness) presenteeism is used in relation to those situations where the worker goes to work despite judging that one should have reported in sick (2-4). In the USA, presenteeism refers to the decreased on-the-job performance due to the presence of health problems that have not necessarily led to absenteeism (5,6).

This first dimension of work functioning is often measured as the costs associated with reduced work outputs, errors on the job, and failure to meet job demands or company production standards (6), and is frequently measured with instruments related to work functioning such as the Health and Work Performance Questionnaire (7), the Work Productivity and Activity Impairment Questionnaire (8), Work Productivity Short Inventory (9), the Work and Health Interview (10), the Health Related Productivity Questionnaire Dairy (11), the Lam Employment Absence and Productivity Scale (12), the Endicott Work Productivity Scale (13) and the Standford Presenteeism Scale (14). It is not clear whether this dimension should be understood as "work functioning" since it centers mainly on the economical meaning of work outcome, rather than on an individual's operational difficulties or limitations while working given certain work conditions (work demands) or due to the existence of a health issue.

The second dimension of work functioning by Amick and Gimeno (1) refers to the impact of the health condition on the process of work functioning, referring to different types of work demands. The instruments to measure this dimension are mainly related to self-reported work limitations or work difficulties to meet job demands, and include the Work Limitations Questionnaire (15) and the Work Role Functioning Questionnaire (16). These instruments are able to address the costs associated with reduced work outputs (and hence include the first dimension referred by Amick and Gimeno), but mainly address the worker's limitations or difficulties in relation to different work demands. Therefore, they point towards the possible need for interventions to modify, adapt o accommodate the process (work demands) of the work.

Other studies put the focus on a different dimension, the quality of work output (not defined in economic terms). In these studies, instruments like the Quantity and Quality Instrument (17) are used. These studies most frequently use more job-specific outcome tools to measure increased errors in the output or increased labor risks due to the health condition (18,19). It can be argued whether these measures of output quality should use the expression “work functioning” or a different one to describe what they are measuring, since their focus is not the individual’s performance, operational difficulties or limitations while working. Instead, they are measuring either the output (in terms of its quality) or the consequences of the work due to impaired work functioning, but not work functioning itself.

Dewa and Lin (20) explored what could be an additional dimension of work functioning, the “extra-effort required by the worker to remain productive”. This dimension could also relate to the notion of presenteeism because work functioning can be viewed as a continuum of situations varying from working successfully to sickness work absence. We may assume that there will be numerous intermediate situations where health issues and/or working conditions can reduce the worker’s abilities to meet job demands requiring an extra-effort to

the worker to avoid (sickness) absenteeism. In a recent study, Leineweber (2) supported this argument, concluding that (sickness) presenteeism is not just an alternative to sickness absence, reflecting only a part of the continuum from working successfully to work absence.

According to Nieuwenhuijsen (21), the concepts of “extra-effort to remain productive” and “quality of work output” have received almost no attention, and incorporation of both concepts in a more expanded view of work functioning would be valuable. It can be discussed whether needing more effort to be productive is sufficiently captured by the concept of generic work role limitations. Measuring the extra-effort to be productive separately enables study of its potential negative consequences over time, such as the impact on the worker's health, well-being and family life. Incorporating quality of work output into our measures of work functioning would allow employers to integrate information about economic consequences and others, such as safety risks, errors, and client satisfaction.

Other related (but different) concepts like work ability, work capacity, work impairment, work participation and work performance can be found in the literature. In our opinion these concepts need clarification and a more accurate definition to capture all the different angles of the concept work functioning.

In conclusion, work functioning in this thesis relates to the second dimension defined by Amick and Gimeno (1), referring to the "impact of health on the process of work functioning" and is defined as work difficulties to meet job demands given a health status. Understanding this was our phenomenon of interest, we decided to make an evidence-based decision for selection of an instrument able to capture this construct; the Work Role Functioning Questionnaire was identified as the most appropriate.

6.2. Selection of an instrument to measure health-related work functioning

Health-related work functioning questionnaires are frequently used in research and practice without clear evidence of the quality of their measurement properties (22). One could argue, therefore, which tool is the best possible to assess this concept.

Nieuwenhuijsen (21) recently conducted a systematic review of the literature to identify work functioning measurement tools for occupational mental health research, suggesting that there is a lack of studies testing these tools in common mental disorders, and that new work functioning instruments are needed to integrate work output and the effort required to remain productive. Other recent systematic reviews that evaluated the methodological quality of the studies on measurement properties using work functioning instruments in patients with musculoskeletal (23) and common mental disorders (24) concluded that the quality of these studies appeared to be limited.

Consequently, because of the low methodological quality of the validation studies on measurement properties using work functioning instruments, it could be concluded that we do not know enough about the measurement properties of the existing health-related work functioning tools. Therefore it is not possible to make evidence-based recommendations about which instrument should be used.

Recently, Abma et al (25,26) performed a quality assessment of the measurement properties of the Dutch version of the WRFQ. This research was conducted using rigorous consensus-based standards for the selection of health status measurement instruments (COSMIN) (27-29), and the authors concluded that the WRFQ is a reliable and valid instrument to measure health-related work functioning in the working population.

The Work Limitations Questionnaire (WLQ) (30) is a very closely related work functioning measurement tool, developed from the same item pool, and was also considered in our study. In fact, both questionnaires have the same introduction

(with minimal differences); the same response options (with differences in the wording for the middle category response option), the same scoring system and several items are common or have a very similar corresponding item. The main differences are the recall period (4 weeks for the WRFQ and 2-weeks for the WLQ) and the number of items (27 items for the WRFQ and 25 items for the WLQ). Although the WLQ has shown its usefulness in several studies in English-speaking health care environments (31-33), it was not freely available in the literature, and when applied to populations with common mental disorders, Abma (24) found limited evidence in relation to the internal consistency, hypotheses testing and no evidence regarding its responsiveness.

In conclusion, to our knowledge, the WRFQ is the only instrument measuring health-related work functioning that has been rigorously evaluated (using expert consensus-based standards), performing appropriately in terms of its reliability and validity in a general working population. Additionally, is freely available for practitioners and researchers. Hence, we decided to translate, cross-cultural adapt to Spanish and validate this instrument. However, it is not known for sure whether the WRFQ is the best instrument to measure our phenomenon of interest since the available evidence on the quality of the measurement properties of the other existing tools is limited.

6.3. Cross-cultural adaptation and validation process (reliability, validity and responsiveness)

Cross-culturally adapted and validated instruments are needed to measure health-related work functioning in Spanish-speaking populations. A rigorous, stepwise procedure for translation and cross-cultural adaptation of the WRFQ led to the development of a Spanish spoken in Spain version equivalent to the original English version. A cross-sectional study was performed to evaluate the reliability and validity of the WRFQ-SpV, and a longitudinal survey was carried out to assess the responsiveness of the instrument. Internal consistency was good or excellent

for the overall scale and all subscales. Test-retest reliability showed good or very good repeatability for the overall scores and all subscales. The original five factor structure reflected adequately the different dimensions of the questionnaire. The questionnaire showed adequate construct validity confirming all tested hypothesis, and suggestive evidence was provided about the possible use of the WRFQ for evaluative purposes.

In relation to the cross-cultural adaptation, minor changes were made to maximize questionnaire understandability and it was necessary to adjust the wording of certain parts of the instrument as had happened when the questionnaire was adapted into Canadian French (34), Brazilian Portuguese (35) and Dutch (26).

The results of the test–retest reliability were very similar to those obtained by Gallasch (35), but partially different from those obtained by Abma (25), who found moderate test-retest reliability for three subscales and low test-retest reliability for the subscale of flexibility demands. It is relevant to notice that neither our validation study nor the study performed by Gallasch found it necessary to modify the subscales. The Dutch version (25) added a new subscale (flexibility demands), based on participant suggestions and the literature and therefore, the test-retest reliability assessed in this study was performed using a slightly different version of the WRFQ. This subscale was not present in the original version, and required the modification of five items after an exploratory factor analysis, resulting in a four factor structure (work scheduling & output, physical, mental & social, and flexibility demands).

In our study, we found fair dimensionality of the WRFQ-SpV when maintaining the original five factor structure (work scheduling, output, physical, mental and social demands). It could be argued whether the subscales of the WRFQ, developed more than 20 years ago, should be updated. Labor flexibility demands commonly refer to flexible work (in relation to skills, workplace or working hours) or flexible employment (in relation to the nature of the employment contract) (36). Also, labor

flexibility could be understood as possible variation in the number of employees needed to perform a specific duty or the need for the same worker to provide different tasks or services. Essentially, it is related to the capacity to adapt, or the personal disposition to change (37). Although it seems feasible that, after these years, certain demands of the jobs could have changed or new ones could have appeared, this was not apparent in the interviews conducted during the pre-test for the adaptation of the questionnaire into Spanish. There may be several reasons for this. First, participants were asked (in a general way) whether the questionnaire adequately reflected their work demands and flexibility demands did not appear as a spontaneous reply. Second, the Dutch and the Spanish samples differed significantly in the number of participants with non-manual jobs (70% and 28% for the Dutch and the Spanish samples respectively), and it is possible that non-manual workers had a higher perception of the demands of flexibility than manual workers. Third, it is reasonable to think that, conceptually, work scheduling demands are part of the flexibility demands since the latter include (not exclusively) time flexibility that is reflected in the worker's work schedule. And fourth, it is possible to think that there are country-specific differences on work demands.

It can be argued whether the modification of the original WRFQ subscales could hinder comparisons between different countries or cultures. Adapting the questionnaire, including a modification of the questionnaire subscales, will probably provide an equivalent overall measure of the construct but may not allow comparisons among different countries or cultures at the subscale level.

Consistent with prior WRFQ analyses, our measurement property results and correlations for item-subscale, item-total, among subscales and subscale-total were found to be appropriate. The validation for the Canadian French (34) and Brazilian Portuguese (35) versions were performed in populations with musculoskeletal disorders, and in the Dutch version (25) in a general working population, concluding that the adapted WRFQ versions were reliable.

The WRFQ-SpV construct validity was assessed by means of hypotheses testing, confirming all seven formulated hypotheses, showing that the instrument was 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 was found, consistent with other findings that both, chronological and functional age, are associated with a decrease in work functioning (38-42).

Hypotheses about correlations with another related construct (WAI) were confirmed. Again, these results were consistent with the validation study of the Dutch version (25), comparing the correlations between the WRFQ and other related constructs ('work ability', 'work productivity', 'work engagement' and 'work involvement').

To assess the WRFQ-SpV responsiveness, four out of five hypotheses were confirmed, examining whether the instrument can be used in evaluative studies. All hypotheses tested in this study were related to expected changes in participants over time, based on the knowledge of the questionnaire structure and its conceptual framework. Consequently, confirming these hypotheses supports the validity of the change scores and the responsiveness of the instrument (27,29,43,47).

The results on responsiveness are consistent with those obtained for the Dutch WRFQ version, which showed moderate responsiveness. The difference is that in this thesis, responsiveness was assessed in a larger group of participants for whom a change was expected due to the treatment applied to the participants in the study. For the Dutch version, responsiveness was assessed in a relatively stable and healthy population, with no intervention between the first and the second test, and the number of those reporting change was relatively small (25). However, since stratification was necessary during the analyses in our study, some subgroups had small sample sizes and therefore, further longitudinal studies with a larger number of participants and different health issues are needed,

especially to examine the instrument responsiveness in workers whose health remain stable or deteriorate.

In conclusion, this thesis 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, and the first results on its responsiveness point suggest the WRFQ is adequate for measuring changes in health-related work functioning over time.

6.4. Standards to be used for methodological quality in health-questionnaire validation

The "consensus-based standards for the selection of health status measurement instruments" (COSMIN) guided the design of the different studies in this thesis. The COSMIN checklist is a useful instrument when designing a study on measurement properties. It includes explicit criteria for satisfying these standards, to assess the quality of the validation studies of health measurement instruments.

Since there are other approaches available in the literature to evaluate the methodological quality of studies on measurement properties, whether any other approach would have been a better method can be discussed.

In the literature review we identified two sources of methodological recommendations: one from internationally recognized experts in the design and validation of questionnaires (43-50) and the other from different research groups that achieved different standards (51,52), sometimes by expert consensus (27-29,53).

The main advantage of using research group standards is that they are not based on the criterion of a single specialist but the judgment of a group of experts. This should make research group standards more reliable when assessing methodological quality of studies or evaluating quality of instruments. In our

opinion, consistent with Terwee (54), some (e.g. the criteria proposed by the Scientific Advisory Committee of the Medical Outcomes Trust or the method proposed by the Health Technology Assessment Programme) (51,52) have some relevant disadvantages because they do not have user-friendly checklists and/or were not developed as methodological quality assessment tools for systematic reviews.

Other checklists were developed for similar, but different objectives. One relevant instrument is the "evaluating the measurement of patient-reported outcomes" (EMPRO) checklist (53), developed for the standardized assessment of patient-reported outcomes (PROs) to assist the choice of instruments. This tool allows rating the quality of an instrument rather than the methodological quality of a validation study. Although in this checklist several methodological quality aspects of the study are taken into account, the “evaluation of the methodological quality of a validation study” and the “evaluation of the quality of an instrument” are conceptually different issues.

Thus, different methodologies are available in the literature for different evaluation purposes. The most appropriate choice will depend on the objective of the study, the motivation to use consensus-based standards instead of the criterion of a single expert, and the degree of user-friendliness that is needed.

Since the studies of this thesis were carried out to validate the WRFQ-SpV, it was ensuring the quality of the validation studies according to international consensus-based standards was crucial. In this sense, both the COSMIN and the EMPRO checklists could have been used. The final decision on the COSMIN standards was taken on the basis that the main use of this checklist is to assess the quality of the validation studies rather than the quality of the instrument (which is the main purpose of the EMPRO checklist).

6.5. Implications for research and practice

The WRFQ-SpV is a brief and user-friendly instrument to measure health-related work functioning, consisting of 27 items grouped in five subscales related to different work demands. It offers scores as percentages from 0 to 100 for the overall scale and each subscale, and is easily interpretable and linkable to meaningful social and economic outcomes.

The age of workers is increasing, and an early, safe and productive reincorporation to work after a sick leave absence is necessary in the current social context. Individualized monitoring of health-related work functioning is becoming a necessity in occupational health services and the WRFQ-SpV might be a helpful instrument for Spanish-speaking occupational health professionals in day-to-day practice and research. Also, several interventions to accommodate work conditions to workers' skills and health status are being implemented to facilitate an active working life and achieve better work functioning. The effectiveness of these interventions needs to be evaluated using quality validated instruments, and the WRFQ-SpV could be a useful instrument for these purposes.

The WRFQ-SpV might also be helpful for human resource managers directly involved in those processes. Since the instrument provides information on job demands difficult to meet for the worker, it could be used as a starting point to: 1) diagnose current functioning of workers without over-medicalizing the situation; 2) monitor work functioning to provide guidance on work accommodations or preventive actions; and 3) to evaluate (with caution) the results of an intervention.

This study provides suggestive evidence that the WRFQ-SpV is an adequate instrument to measure changes in health-related work functioning over time. However, more evidence is needed on the ability of the instrument to detect changes in individualized health-related work functioning surveillance in groups whose health remains stable or deteriorates, and more evidence on

responsiveness is needed if it is to be used for individual surveillance or in evaluative studies.

De Vet (43) describes three important uses of instruments: diagnosis (the discriminative ability of the instrument is very important), evaluation (responsiveness of the instrument is crucial) and prediction of future course. The studies included in this thesis showed that the WRFQ is able to differentiate between several subgroups (mental and physical health issues, type of job and groups of age). Abma (25) showed that the instrument could also differentiate groups with different need for recovery and fatigue, indicating that the instrument has discriminative ability. Although one of the studies in this thesis provides suggestive evidence of the responsiveness of the questionnaire, this measurement property has yet not been sufficiently examined for individual surveillance and intervention assessment. Therefore, the WRFQ-SpV is not yet recommended for these purposes (individual surveillance and intervention assessment) and additional research is needed to further examine this measurement property.

Finally, it is important to ensure that the questionnaire is used in a manner that respects the conceptual framework to prevent unethical use. For this purpose, occupational health professionals and company managers should provide public and clear commitment to the basic principles of voluntariness, respect, confidentiality and non-discrimination.

6.6. Future research

Validation of a measuring instrument is an ongoing process through which a body of evidence is created about the different features of the tool. There is no agreement in the literature regarding standards on the number and/or the degree of quality of validation processes that an instrument should follow to confirm that is a feasible (understandable, practical, easily interpretable) and validated instrument (reliable, valid and responsive).

The WRFQ has been adapted and validated in different cultural contexts and populations (22,34,35,55-57). Additional validation processes in other languages and cultures, as well as in other occupational settings, work conditions and populations with different health issues would be desirable. It is particularly necessary to assess the responsiveness of the questionnaire in groups whose health remain stable or deteriorate and also to examine the responsiveness of the instrument assessing individual surveillance and evaluating workplace interventions.

Various perspectives of work functioning are being used in the literature and in occupational health settings, and other concepts referring to “optimal” or “satisfactory” work functioning are arising in the recent literature (25,58-63). A comprehensive analysis of these concepts for a deeper insight on the meaning of the managed terms is lacking, and an international consensus-based agreement of experts to unify cut-offs and definitions is needed.

National and international comparisons on work functioning trajectories over time, effectiveness of interventions to enable sustainable work participation and analyses on barriers and facilitators for productive work functioning over the working life are not available. However, they are necessary to enhance the Spanish policies pursuing the European Commission’s 2020 strategy and targets for employment (for Spain, 74% of 20-64 year-olds to be employed) (63) by means of improving worker health and participation.

REFERENCES

1. Amick BC III, Gimeno D. Measuring work outcomes with a focus on health-related work productivity loss. In: Wittink H & Carr D Editors. Evidence, Outcomes & Quality of Life in Pain Treatment: A Handbook for Pain Treatment Professionals. London, UK: Elsevier, 2007. pp. 329-343.

2. Leineweber C, Westerlund H, Hagberg J, Svedberg P, Alexanderson K. Sickness presenteeism is more than an alternative to sickness absence: results from the population-based SLOSH study. Int Arch Occup Environ Health. 2012;85:905-14.

3. Aronsson G, Gustafsson K, Dallner M. Sick but yet at work. An empirical study of sickness presenteeism. J Epidemiol Community Health. 2000;54:502-9.

4. Dew K, Keefe V, Small K. 'Choosing' to work when sick: workplace presenteeism. Soc Sci Med. 2005;60:2273-82.

5. Mattke S, Balakrishnan A, Bergamo G, Newberry SJ. A review of methods to measure health-related productivity loss. Am J Manag Care. 2007 Apr;13(4):211-7.

6. Schultz AB, Edington DW. Employee health and presenteeism: a systematic review. J Occup Rehabil. 2007;17:547-79.

7. Kessler R, Barber C, Beck A, Berglund P, Cleary PD, McKenas D et al. The World Health Organization health and work performance questionnaire (HPQ). J Occup Environ Med. 2003;45;156-74.

8. Reilly MC, Zbrozek AS, Dukes EM. The validity and reproducibility of a work productivity and activity impairment instrument. Pharmacoeconomics. 1993;4:353-65.

9. Goetzel RZ, Ozminkowski RJ, Long, SR. Development and reliability analysis of the Work Productivity Short Inventory (WPSI) instrument measuring employee health and productivity. J Occup Environ Med. 2003;45:743-62.

10. Stewart WF, Ricci JA, Leotta C, Chee E. Validation of the work and health interview. Pharmacoeconomics. 2004;22:1127-40.

11. Kumar RN, Hass SL, Li JZ, Nickens DJ, Daenzer CL, Wathen LK. Validation of the Health-Related Productivity Questionnaire Diary (HRPQ-D) on a sample of patients with infectious mononucleosis: results from a phase 1 multicenter clinical trial. J Occup Environ Med. 2003;45:899-907.

12. Lam RW, Michalak EE, Yatham LN. A new clinical rating scale for work absence and productivity: validation in patients with major depressive disorder. BMC Psychiatry. 2009;9:78.

13. Endicott J, Nee J. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. Endicott Work Productivity Scale (EWPS): a new measure to assess treatment effects. Psychopharmacol Bull. 1997;33:13-6.

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. Lerner D, Amick BC 3rd, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Med Care. 2001;39:72-85.

16. 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.

17. Meerding WJ, IJzelenberg W, Koopmanschap MA, Severens JL, Burdorf A. Health problems lead to considerable productivity loss at work among workers with high physical load jobs. J Clin Epidemiol. 2005;58:517-23.

18. Suzuki K, Ohida T, Kaneita Y, et al. Mental health status, shift work, and occupational accidents among hospital nurses in Japan. J Occup Health. 2004;46:448-54.

19. Haslam C, Atkinson S, Brown S, Haslam RA. Perceptions of the impact of depression and anxiety and the medication for these conditions on safety in the workplace. Occup Environ Med. 2005;62:538-45.

20. Dewa CS, Lin E. Chronic physical illness, psychiatric disorder and disability in the workplace. Soc Sci Med. 2000;51:41-50.

21. Nieuwenhuijsen K, Franche RL, van Dijk FJ. Work functioning measurement: tools for occupational mental health research. J Occup Environ Med. 2010;52:778-90.

22. Abma FI. Work functioning: development and evaluation of a measurement tool [PhD thesis]. Groningen, NL: University of Groningen; 2012. [Internet]. Available from: http://irs.ub.rug.nl/ppn/351176438

23. Williams RM, Schmuck G, Allwood S, Sanchez M, Shea R, Wark G. Psychometric evaluation of health-related work outcome measures for musculoskeletal disorders: a systematic review. J Occup Rehabil. 2007;17:504-21.

24. 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.

25. 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.

26. 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.

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. 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. 2010;19:539-49.

30. Lerner D, Amick BC 3rd, Rogers WH, Malspeis S, Bungay K, Cynn D. The Work Limitations Questionnaire. Med Care. 2001;39:72-85.

31. Lerner DJ, Amick BC 3rd, Malspeis S, Rogers WH. A national survey of health-related work limitations among employed persons in the United States. Disabil Rehabil. 2000; 22:225–32.

32. Roy JS, MacDermid JC, Amick BC 3rd, Shannon HS, McMurtry R, Roth JH, et al. Validity and responsiveness of presenteeism scales in chronic work-related upper-extremity disorders. Phys Ther. 2011;91:254-66.

33. Schmidt LL, Amick BC 3rd, Katz JN, Ellis BB. Evaluation of an upper extremity student-role functioning scale using item response theory. Work. 2002;19:105-16.

34. 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. 2004;27:261-8.

35. 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.

36. MacEachen E, Polzer J, Clarke J. "You are free to set your own hours": governing worker productivity and health through flexibility and resilience. Soc Sci Med. 2008;66:1019-33.

37. Aronsson G. Influence of worklife on public health. Scand J Work Environ Health. 1999;25:597-604.

38. 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.

39. Costa G, Sartori S. Ageing, working hours and work ability. Ergonomics. 2007;50:1914-30.

40. 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.

41. 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.

42. 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.

43. de Vet HCW, Terwee CB, Mokkink LB, Knol DL. Mesurement in medicine: A practical guide. 1st ed. Cambridge, UK: The University Press Cambridge, 2011.

44. Guillemin F. Cross-cultural adaptation and validation of health status measures. Scand J Rheumatol.1995;24:61-63.

45. Beaton DE, Bombardier C, Guillemin F, Bosi Ferraz M. Guidelines for the process of cross-cultural adaptation of self-reports measures. Spine. 2000; 25:3186-3191.

46. Aday LA, Cornelius LJ. Designing and conducting health surveys: a comprehensive guide. 3rd ed. San Francisco, CA: Jossey-Bass publisher; 2006.

47. Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 4thed. New York: Oxford University Press Inc.; 2008.

48. García de Yébenes MJ, Rodriguez-Salvanés F, Carmona-Ortells L. Validación de cuestionarios. Reumatol Clin. 2009;5:171-77.

49. Keszei AP, Novak M, Streiner DL. Introduction to health measurement scales. J Psychosom Res. 2010;68:319-23.

50. 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.

51. Aaronson N, Alonso J, Burnam A, Lohr KN, Patrick DL, Perrin E, et al. Assessing health status and quality-of-life instruments: attributes and review criteria. Qual Life Res. 2002;11:193-205.

52. Fitzpatrick R, Davey C, Buxton MJ, Jones DR. Evaluating patient-based outcome measures for use in clinical trials. Health Technol Assessment. 1998;2:1-74.

53. Valderas JM, Ferrer M, Mendívil J, Garin O, Rajmil L, Herdman M, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008;11:700-8.

54. 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. 2012;21:651-7.

55. Ramada JM, Serra C, Amick Iii BC, 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.

56. Ramada JM, Serra C, Amick BC III, Abma FI, Pidemunt G, Castaño JR, et al. Reliability and Validity of the Work Role Functioning Questionnaire (Spanish version) in a general working population. [Submitted for peer-review].

57. Ramada JM, Delclos GL, Amick BC III, Abma FI, Pidemunt G, Castaño JR, et al. Responsiveness of the Work Role Functioning Questionnaire (Spanish version) in a general working population. J Occup Environ Med. 2013 [In Press].

58. Koolhaas W, Brouwer S, Groothoff JW, van der Klink JJ. Enhancing a sustainable healthy working life: design of a clustered randomized controlled trial. BMC Public Health. 2010;10:461.

59. Arends I, van der Klink JJ, Bültmann U. Prevention of recurrent sickness absence among employees with common mental disorders: design of a cluster-randomized controlled trial with cost-benefit and effectiveness evaluation. BMC Public Health. 2010;10:132.

60. de Vries HJ, Brouwer S, Groothoff JW, Geertzen JH, Reneman MF. Staying at work with chronic nonspecific musculoskeletal pain: a qualitative study of workers' experiences. BMC Musculoskelet Disord. 2011;12:126.

61. Viikari-Juntura E, Kausto J, Shiri R, Kaila-Kangas L, Takala EP, Karppinen J, et al. Return to work after early part-time sick leave due to musculoskeletal disorders: a randomized controlled trial. Scand J Work Environ Health. 2012;38:134-43.

62. Abma FI, Amick BC 3rd, van der Klink JJ, Bültmann U. Prognostic factors for successful work functioning in the general working population. J Occup Rehabil. 2013;23:162-9.

63. Holwerda A, van der Klink JJ, de Boer MR, Groothoff JW, Brouwer S. Predictors of work participation of young adults with mild intellectual disabilities. Res Dev Disabil. 2013;34:1982-90.

64. European Commission (EU). Europe 2020: Targets [Internet]. Brussels; 2010 [updated 2013 Jan 21; cited 2013 Nov 23]. Available from: http://ec.europa.eu/europe2020/europe-2020-in-a-nutshell/targets/index_en.htm

7. GENERAL CONCLUSION

As a result of the need for a tool to assess health-related work functioning in daily practice and research in occupational health, the WRFQ was selected based on the best available evidence in the literature. A comprehensive literature review on the methodology for translation, cultural adaptation and validation of health questionnaires was performed to facilitate a proposal to carry out this process. This proposal was applied rigorously to the WRFQ, obtaining a version adapted to Spanish (spoken in Spain), equivalent to the original English version.

The WRFQ-SpV was validated in different studies and appears to be a reliable and valid instrument to measure health-related work functioning in Spanish-speaking populations. Also, the results provide evidence of the responsiveness of the instrument, suggesting the possibility of its use in evaluative studies since it was able to detect (true) changes over time.

However, more research is needed to confirm the responsiveness of the instrument for use in individual surveillance and to assess the effectiveness of work interventions to improve work functioning.

8. APPENDICES

Appendix I: WRFQ (English version)

WL-27

YOUR WORK AND HEALTH

These questions ask you to rate the amount of time during the **past four weeks** that you had difficulty handling certain parts of your job.

Mark the "Does Not Apply to My Job" box only if the question describes something that is not part of your job.

In the past 4 weeks, how much of the time did your physical health or emotional problems make it difficult for you to do the following?

DIFFICULT	All of the Time (100%)	Most of the Time	Half of the Time (50%)	Some of the Time	None of the Time (0%)	Does Not Apply to My Job
1. Work the required number of hours	☐1	☐2	☐3	☐4	☐5	☐0
2. Get going easily at the beginning of the workday	☐1	☐2	☐3	☐4	☐5	☐0
3. Start on your job as soon as you arrive at work	☐1	☐2	☐3	☐4	☐5	☐0
4. Do your work without stopping to take extra breaks or rests	☐1	☐2	☐3	☐4	☐5	☐0
5. Stick to a routine or schedule	☐1	☐2	☐3	☐4	☐5	☐0
6. Handle the workload	☐1	☐2	☐3	☐4	☐5	☐0
7. Work fast enough	☐1	☐2	☐3	☐4	☐5	☐0
8. Finish work on time	☐1	☐2	☐3	☐4	☐5	☐0
9. Do your work without making mistakes	☐1	☐2	☐3	☐4	☐5	☐0
10. Satisfy the people who judge your work	☐1	☐2	☐3	☐4	☐5	☐0
11. Feel a sense of accomplishment in your work	☐1	☐2	☐3	☐4	☐5	☐0
12. Feel you have done what you are capable of doing	☐1	☐2	☐3	☐4	☐5	☐0
13. Walk or move around different work locations (for example, going to meetings)	☐1	☐2	☐3	☐4	☐5	☐0
14. Lift, carry, or move objects at work weighing more than 10 pounds	☐1	☐2	☐3	☐4	☐5	☐0

DIFFICULT	All of the Time (100%)	Most of the Time	Half of the Time (50%)	Some of the Time	None of the Time (0%)	Does Not Apply to My Job
15. Sit, stand, or stay in one position for longer than 15 minutes while working	☐1	☐2	☐3	☐4	☐5	☐0
16. Repeat the same motions over and over again while working	☐1	☐2	☐3	☐4	☐5	☐0
17. Bend, twist, or reach while working	☐1	☐2	☐3	☐4	☐5	☐0
18. Use hand-held tools or equipment (for example, a phone, pen, keyboard, computer mouse, drill, hairdryer or sander)	☐1	☐2	☐3	☐4	☐5	☐0
19. Keep your mind on your work	☐1	☐2	☐3	☐4	☐5	☐0
20. Think clearly when working	☐1	☐2	☐3	☐4	☐5	☐0
21. Do work carefully	☐1	☐2	☐3	☐4	☐5	☐0
22. Concentrate on your work	☐1	☐2	☐3	☐4	☐5	☐0
23. Work without losing your train of thought	☐1	☐2	☐3	☐4	☐5	☐0
24. Easily read or use your eyes when working	☐1	☐2	☐3	☐4	☐5	☐0
25. Speak with people in person, in meetings or on the phone	☐1	☐2	☐3	☐4	☐5	☐0
26. Control your temper around people when working	☐1	☐2	☐3	☐4	☐5	☐0
27. Help other people to get work done	☐1	☐2	☐3	☐4	☐5	☐0

8. APPENDICES

Appendix II: WRFQ (Spanish version)

WRFQ-27

SU TRABAJO Y SU SALUD

En las siguientes preguntas le pedimos que nos indique, **para las últimas cuatro semanas**, la cantidad de tiempo en que tuvo dificultad para realizar ciertos aspectos de su trabajo.

Marque la casilla "No Aplicable a Mi Trabajo" sólo en caso de que la pregunta se refiera a algo que no es parte de su trabajo.

En las últimas 4 semanas, ¿durante cuánto tiempo de su trabajo le fue difícil realizar las siguientes actividades por motivos de su salud física o problemas emocionales?

	Fue difícil todo el tiempo (100%)	**Fue difícil la mayor parte del tiempo**	**Fue difícil la mitad del tiempo (50%)**	**Fue difícil una parte del tiempo**	**Nunca fue difícil (0%)**	**No aplicable a mi Trabajo**
1. Trabajar el número de horas requeridas	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
2. Empezar la jornada de trabajo con facilidad	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
3. Ponerse a trabajar nada más llegar al trabajo	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
4. Hacer su trabajo sin parar a hacer descansos adicionales	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
5. Ajustarse a una rutina u horario	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
6. Manejar su carga de trabajo	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
7. Trabajar lo suficientemente rápido.	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
8. Acabar el trabajo a tiempo	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
9. Hacer su trabajo sin cometer errores	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
10. Satisfacer a las personas que evalúan su trabajo	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$
11. Tener sensación de trabajo bien hecho	☐$_0$	☐$_1$	☐$_2$	☐$_3$	☐$_4$	☐$_5$

	Fue difícil todo el tiempo (100%)	Fue difícil la mayor parte del tiempo	Fue difícil la mitad del tiempo (50%)	Fue difícil una parte del tiempo	Nunca fue difícil (0%)	No aplicable a mi Trabajo
12. Sentir que ha hecho lo que es capaz de hacer	☐0	☐1	☐2	☐3	☐4	☐5
13. Caminar o desplazarse a distintos lugares de trabajo	☐0	☐1	☐2	☐3	☐4	☐5
14. Levantar, cargar o mover objetos de más de 5 kg de peso, en el trabajo.	☐0	☐1	☐2	☐3	☐4	☐5
15. Permanecer sentado, de pie o en una misma posición durante más de 15 minutos, mientras trabaja	☐0	☐1	☐2	☐3	☐4	☐5
16. Repetir los mismos movimientos una y otra vez mientras trabaja	☐0	☐1	☐2	☐3	☐4	☐5
17. Doblarse, girarse o alcanzar un objeto mientras trabaja	☐0	☐1	☐2	☐3	☐4	☐5
18. Usar equipos o herramientas de mano	☐0	☐1	☐2	☐3	☐4	☐5
19. Mantener la mente en su trabajo	☐0	☐1	☐2	☐3	☐4	☐5
20. Pensar con claridad mientras trabaja	☐0	☐1	☐2	☐3	☐4	☐5
21. Hacer el trabajo con cuidado	☐0	☐1	☐2	☐3	☐4	☐5
22. Concentrarse en su trabajo	☐0	☐1	☐2	☐3	☐4	☐5
23. Trabajar sin perder el hilo (de las ideas).	☐0	☐1	☐2	☐3	☐4	☐5

	Fue difícil todo el tiempo (100%)	Fue difícil la mayor parte del tiempo	Fue difícil la mitad del tiempo (50%)	Fue difícil una parte del tiempo	Nunca fue difícil (0%)	No aplicable a mi Trabajo
24. Leer o usar los ojos con facilidad mientras trabaja	☐0	☐1	☐2	☐3	☐4	☐5
25. Hablar con la gente cara a cara, en reuniones o por teléfono	☐0	☐1	☐2	☐3	☐4	☐5
26. Controlar su genio delante de otras personas mientras trabaja	☐0	☐1	☐2	☐3	☐4	☐5
27. Ayudar a otras personas a acabar el trabajo	☐0	☐1	☐2	☐3	☐4	☐5

8. APPENDICES

Appendix III: Single items of the work ability index (WAI)

Por favor, responda a las siguientes preguntas:

1. Suponga que su capacidad para el trabajo en el mejor momento de su vida tiene un valor de 10 puntos. ¿Cuántos puntos le daría a su capacidad para el trabajo actualmente? (Un 0 significa que actualmente es totalmente incapaz de trabajar).

0	1	2	3	4	5	6	7	8	9	10
Totalmente Incapaz de trabajar										**Capacidad de trabajar en su mejor momento**

2. ¿Cómo calificaría su capacidad para el trabajo actual **con respecto a las exigencias físicas** de su trabajo?

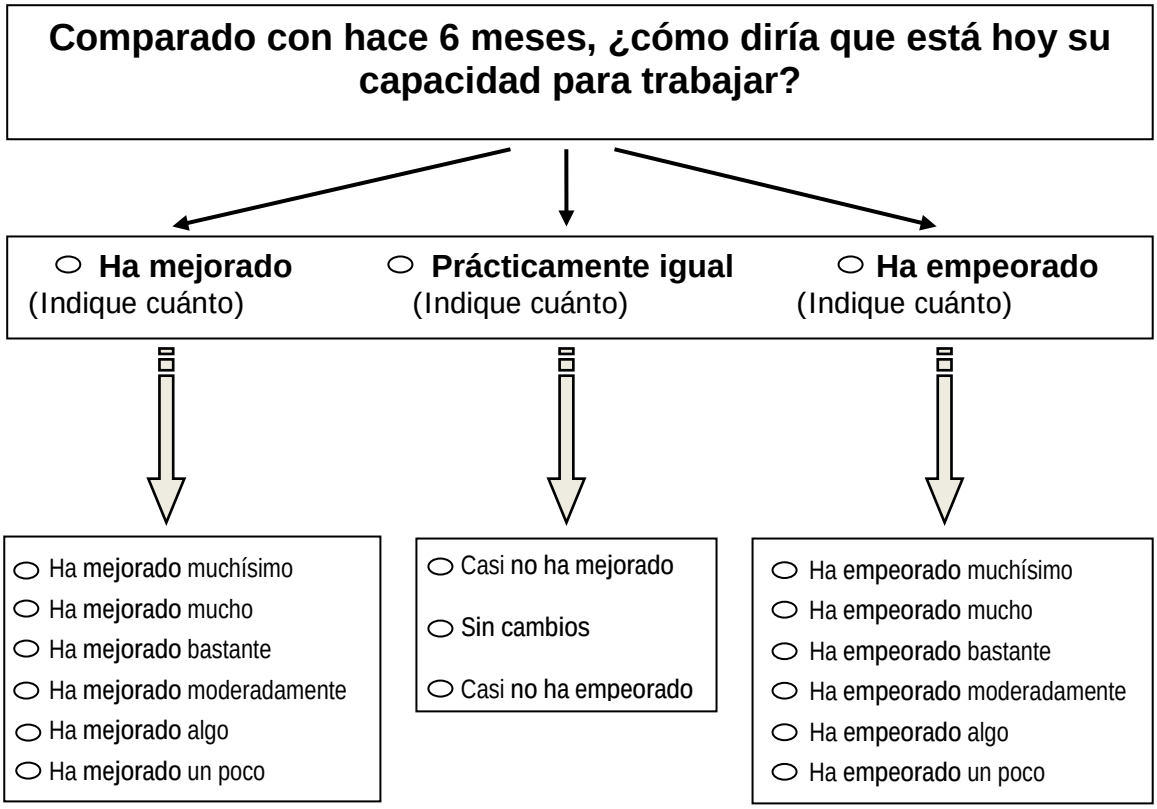
0 Buena
1 Bastante buena
2 Moderada
3 Bastante pobre
4 Muy pobre

3. ¿Cómo calificaría su capacidad para el trabajo actual **con respecto a las exigencias mentales** de su trabajo?

0 Buena
1 Bastante buena
2 Moderada
3 Bastante pobre
4 Muy pobre

8. APPENDICES

Appendix IV: Global perceived effect question (GPE-Q)



8. APPENDICES

Appendix V: Clinical Research Ethical Committee approval

Informe del Comité Ético de Investigación Clínica

Doña Mª Teresa Navarra Alcrudo Secretaria del Comité Ético de Investigación Clínica Parc de Salut MAR

CERTIFICA

Que éste Comité ha evaluado el proyecto de investigación clínica nº 2011/4168/I titulado *"Traducción, adaptación cultural y validación del "Work role functioning questionnaire" (WRFQ) castellano hablado en España"* propuesto por el Dr. José María Ramada Rodilla del Servicio de Salud Laboral del Hospital del Mar.

Que adjunta documento de consentimiento informado.

Y que considera que:

Se cumplen los requisitos necesarios de idoneidad del protocolo en relación con los objetivos del estudio y están justificados los riesgos y molestias previsibles para el sujeto.

La capacidad del investigador y los medios disponibles son apropiados para llevar a cabo el estudio.

El alcance de las compensaciones económicas que se solicitan está plenamente justificado.

Y que éste Comité acepta que dicho proyecto de investigación sea realizado en el Hospital del Mar por el Dr. José María Ramada Rodilla como investigador principal tal como recoge el ACTA de la reunión del día 8 de Marzo de 2011.

Lo que firmo en Barcelona, a 1 de Abril de 2011

Firmado: .....
Doña Mª Teresa Navarra Alcrudo

CEIC – Parc de Salut MAR
Dr. Aiguader, 88 I 08003 Barcelona I Telèfon 93 316 06 77 I Fax 93 316 06 36
ceic-psmar@imim.es | www.parcdesalutmar.cat

8. APPENDICES

Appendix VI: Informed consent

CONSENTIMIENTO INFORMADO PARA LA PARTICIPACIÓN VOLUNTARIA EN LA INVESTIGACIÓN QUE LLEVA POR TÍTULO:

Traducción, adaptación cultural y validación del "Work Role Functioning Questionnaire (WRFQ)"

Yo _____, con DNI/Pasaporte _____

He sido informado sobre el principal objetivo y alcance del estudio que para la traducción, adaptación cultural y validación del cuestionario "Work Role Functioning Questionnaire" (WRFQ). He tenido oportunidad de efectuar preguntas sobre el estudio, he recibido respuestas satisfactorias y he recibido suficiente información sobre el mismo.

Entiendo que la participación es voluntaria. Entiendo que puedo abandonar el estudio cuando lo desee, sin que tenga que dar explicaciones y sin que ello afecte a MIS cuidados médicos.

También he sido informado de forma clara, precisa y suficiente de que los datos serán tratados y custodiados con respeto a mi intimidad y a la vigente normativa de protección de datos. Sobre estos datos me asisten los derechos de acceso, rectificación, cancelación y oposición que podré ejercitar mediante solicitud ante el investigador responsable en la dirección de contacto que figura en este documento. Estos datos no podrán ser cedidos sin mi consentimiento expreso y no lo otorgo en este acto.

Declaro que he leído y conozco el contenido del presente documento. Y, por ello, firmo este consentimiento informado de forma voluntaria para manifestar mi deseo de participar en este estudio de INVESTIGACIÓN hasta que decida lo contrario.

Al firmar este consentimiento no renuncio a ninguno de mis derechos. Recibiré una copia de este consentimiento para guardarlo y poder consultarlo en el futuro.

Barcelona, _____ de _____ de _____

Dr. José Mª Ramada Rodilla
DNI 22.675.983
Dirección de contacto:
Parc de Salut Mar. Servicio de Salud Laboral.
Passeig Marítim, 25-29 08003-Barcelona

Firma de la persona que consiente en colaborar en el estudio.

8. APPENDICES

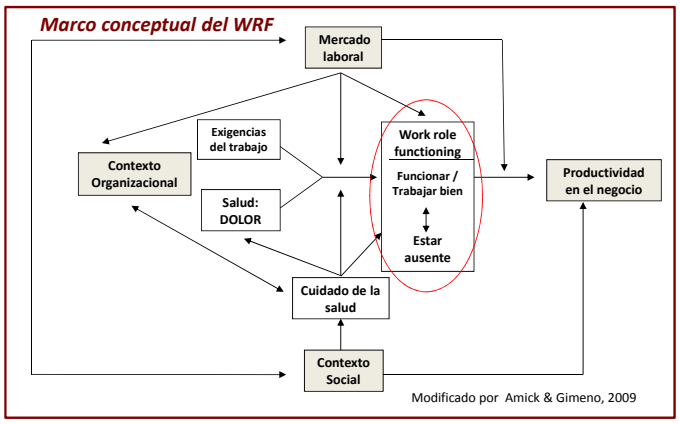
Appendix VII: Poster

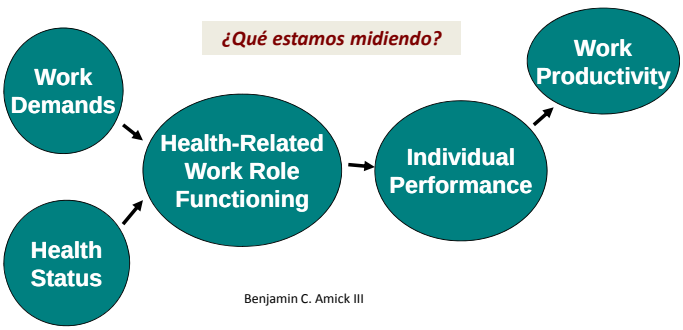
Ramada JM, Serra C, Delclós J. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. Primera Jornada Científica CISAL. Barcelona, 2011.

Traducción, adaptación cultural y validación del "Work Role Functioning Questionaire, (WRFQ)" al castellano hablado en España

José M. Ramada [a,b]; Jordi Delclós [a,c,d]; Consol Serra [a,b,d]

(a)Centro de Investigación en Salud Laboral (CISAL), Universidad Pompeu Fabra. Barcelona; (b)Servicio de Salud Laboral, Parc de Salut MAR. Barcelona; (c) Epidemiology, Human Genetics and Environmental Sciences Division, The University of Texas School of Public Health, Houston, Texas, EE.UU; (d) CIBER de Epidemiología y Salud Pública (CIBERESP).





ANTECEDENTES

El "Work Role Functioning Questionnaire" (WRFQ) es una herramienta que permite medir el grado de discapacidad laboral y evaluar el impacto percibido de un problema de salud sobre la capacidad del trabajador para realizar su trabajo[1].

La comunidad científica internacional viene realizando grandes esfuerzos para evaluar las condiciones de trabajo y salud de las poblaciones usando parámetros comunes a nivel internacional.

Los investigadores que aborden un proceso de traducción, adaptación y validación de una herramienta ya existente en la literatura científica internacional deben tener en cuenta las diferencias en la percepción de la salud y la enfermedad de aquellas poblaciones en las cuales se desea aplicar[2].

Este tipo de adaptaciones culturales exige la utilización de un método estandarizado que garantice la equivalencia entre la herramienta original y la versión adaptada, así como la máxima preservación de las propiedades psicométricas del mismo[3-5].

OBJETIVO

El objetivo del presente estudio es la traducción y adaptación del "Work Role Functioning Questionnaire, WRFQ" al castellano hablado en España así como la evaluación de la validez y fiabilidad de las propiedades psicométricas de la versión adaptada.

CARACTERÍSTICAS DEL WRFQ-27

- Cuestionario autoadministrado que contiene 27 Items.
- Estructurado en 5 subescalas relacionadas con las exigencias del trabajo.
- Subescalas: exigencias de gestión del tiempo; físicas; psicológicas; sociales y de producción.
- Las respuestas se vinculan con el % de tiempo de la jornada de trabajo en que se es capaz de desempeñar productivamente las exigencias del trabajo.
- Las respuestas se refieren siempre a las 4 últimas semanas.
- Es aplicable a diferentes sectores de la actividad económica, utilizando la categoría de respuesta 'Does Not Apply to My Job' .
- Se puede aplicar a trabajdores con diferentes tipos de enfermedad y estados de salud.

FASES DE LA ADAPTACIÓN CULTURAL AL CASTELLANO DEL WRFQ-27

FASE 1: ***Traducción directa*** → ***FASE 2:*** ***Síntesis de las traducciones*** → ***FASE 3:*** ***Traducción inversa (retro-traducción)*** → ***FASE 4:*** ***Consolidación por comité de expertos*** → ***FASE 5:*** ***Pre-test***

--- Obtención del cuestionario final adaptado ---

EVALUACIÓN DE LA FIABILIDAD Y LA VALIDEZ DE LAS PROPIEDADES PSICOMÉTRICAS

EVALUACIÓN 1: ***Fiabilidad y homogeneidad*** — *Coef. α de Cronbach*; *Test-retest de fiablidad*

EVALUACIÓN 2: ***Validez de contenido*** — *Comité de expertos*

EVALUACIÓN 3: ***Validez de constructo*** — *Técnicas de validación para grupos conocidos*

BIBLIOGRAFÍA:

1. 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 Dec; 25(24):3152-60.
2. Beaton DE, Bombardier C, Guillemin F, Bosi Ferraz M. Guidelines for the process of cross- cultural adaptation of self-reports measures. Spine. 2000;25(24):3186–91.
3. Hutchinson A, Bentzen N, Konig-Zanhn C. Cross cultural health outcome assessment: a user's guide. The Netherlands: ERGHO; 1996.
4. Guillemin F. Cross-cultural adaptation and validation of health status measures. Scand J Rheumatol 1995;24(2):61–3.
5. Alexandre NMC, Guirardello EB. Adaptación cultural de instrumentos utilizados en salud ocupacional. Rev Panam Salud Pública. 2002;11(2):109–11.

8. APPENDICES

Appendix VIII: Poster

Ramada JM, Serra C, Delclós GL. Cross-cultural adaptation and health questionnaires validation: revision and methodological recommendations. Second Scientific Conference on Work Disability Prevention and Integration 'Healthy ageing in a working society'. WDPI; Groningen, 2012.

Cross-cultural adaptation and validation of the Work Role Functioning Questionnaire to Spanish spoken in Spain.

Jose Maria Ramada[1,2,3], Consol Serra [1,2,3], George L Delclos[2,3,4]

(1) Occupational Health Service. Parc de Salut MAR. Barcelona, Spain. (2) Center for Research in Occupational Health (CiSAL), University Pompeu Fabra. Barcelona, Spain. (3) CIBER of Epidemiology and Public Health (CIBERESP). (4) Epidemiology, Human Genetics and Environmental Sciences Division, The University of Texas School of Public Health, Houston, Texas, USA.

Introduction

- **Translation, cultural adaptation and validation (TCAV) of existing questionnaires to other languages and cultures facilitates international and multicultural research projects promoting information exchange and reducing time and costs to develop similar instruments[1].**
- **The simple translation of a questionnaire may lead to misinterpretation due to language and cultural differences. When using questionnaires developed in other countries and languages in scientific studies it is necessary, besides the translation, to carry out a cross-cultural adaptation and validation[2].**
- **The Work Role Functioning Questionnaire (WRFQ) is a self-completed work disability tool used to measure the perceived impact of a health problem on the worker's ability to perform his job[3].**

1. García de Yébenes MJ, Rodriguez-Salvanés F, Carmona-Ortells L. Validación de cuestionarios. Reumatol Clin. 2009;5:171-177.
2. Beaton DE, Bombardier C, Guillemin F, Bosi Ferraz M. Guidelines for the process of cross-cultural adaptation of self-reports measures. Spine. 2000;25:3186-3191.
3. 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.

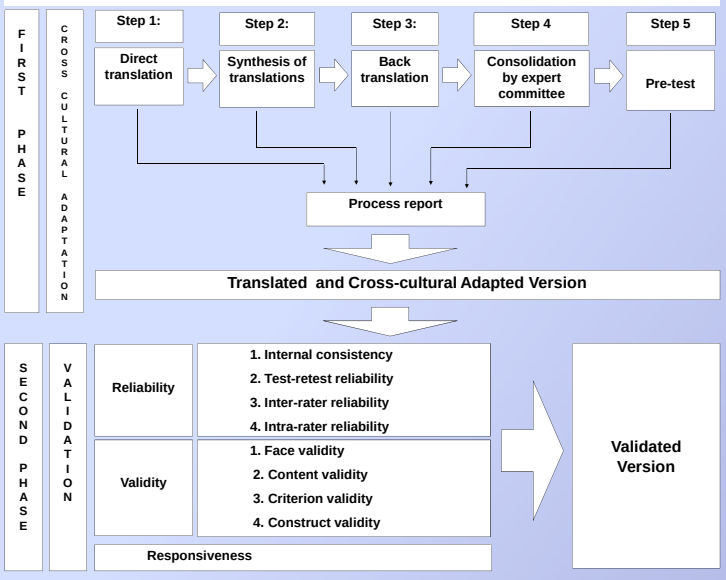
Objective

The aim of this study is to translate and adapt the WRFQ to Spanish spoken in Spain and evaluate its validity and reliability.

Methods

- **Systematic 5-step procedure: direct translation, synthesis, back-translation, expert committee and pre-test.**
- **Structured interviews (n=40) with workers with musculoskeletal and prevalent mental health disorders were performed to evaluate the comprehensibility, usability, applicability and completeness of the adapted questionnaire.**
- **Panel of experts to assess content validity.**
- **Techniques for analysis of known groups to assess construct validity assessed.**
- **Internal consistency (Item-to-subscale and item-to-total correlations), and checking of floor and ceiling effects were examined.**
- **Test-retest reliability at 15 days was assessed by calculating the Intraclass Correlation Coefficient (ICC) after administering the WRFQ twice to the same sample of 40 workers.**

Figure 1. Translation, cultural adaptation and validation (Adapted from Beaton et al. 2000). Barcelona, November 2011.



Results

Table 1. Participants' socio-demographic characteristics. Pre-test with the adapted version of the WRFQ* to Spanish spoken in Spain (n=40). April-May, 2012.

		Total n=40	Men n=15 (37,5%)	Women n=25 (62,5%)
Age in years, mean (SD)		49,1 (10,0)	47,9 (8,9)	49,8 (10,7)
Education level, n (%)	Low	13 (32,5)	7 (46,7)	6 (24,0)
	Middle	15 (37,5)	6 (40,0)	9 (36,0)
	High	12 (30,0)	2 (13,3)	10 (40,0)
Job type, n (%)	Manual	17 (42,5)	6 (40,0)	11 (44,0)
	Non-manual	11 (27,5)	5 (33,3)	6 (24,0)
	Mixed	12 (30,0)	4 (26,7)	8 (32,0)
Working hours/week, mean (SD)		40,23 (10,7)	46,1 (9,6)	36,7 (9,8)
Disease type, n(%)	Physical	17 (42,5)	6 (40,0)	11 (44,0)
	Mental	17 (42,5)	8 (53,3)	9 (36,0)
	Both	6 (15,0)	1 (6,7)	5 (20,0)
Disease duration in months, mean (SD)		34,68 (51,1)	23,1 (22,4)	41,6 (61,8)

*WRFQ: Work Role Functioning Questionnaire.

Table 2. Pre-test results with the Spanish version of the Work Role Functioning Questionnaire, (n=40). April-May, 2012.

	Valid n (missing/not applicable)*	Mean§ (SD)	Range	Median	n at floor (0%) n (%)	n at ceiling (100%) n (%)	Cronbach's alpha	Subscale-total correlations
Work scheduling demands	39 (1)	67,7 (27,8)	5-100	75,0	0 (0,0)	3 (7,5)	0,88	0,95
Output demands	39 (1)	64,4 (25,8)	14,3-100	67,9	0 (0,0)	1 (2,5)	0,90	0,94
Physical demands	36 (4)	59,0 (32,3)	4,17-100	62,5	0 (0,0)	5 (12,5)	0,95	0,88
Mental demands	40 (0)	73,9 (26,1)	0-100	79,2	1 (2,5)	9 (22,5)	0,96	0,81
Social demands	35 (5)	76,9 (21,1)	25-100	83,3	0 (0,0)	5 (12,5)	0,56	0,83
Total score	40 (0)	67,6 (22,7)	21,3-98,1	74,5	0 (0,0)	0 (0,0)	0,97	--

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

(§) 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. Test-retest reliability. Intraclass Correlation Coefficients (ICC). Pre-test of the Spanish version of the WRFQ*, April-May 2012.

Subscales	Test-retest CCI	IC 95%**
Work scheduling demands	0,92	(0,85-0,96)
Output demands	0,89	(0,78-0,94)
Physical demands	0,93	(0,84-0,97)
Mental demands	0,85	(0,72-0,92)
Social demands	0,77	(0,58-0,88)
Total score	0,94	(0,83-0,98)

* WRFQ: Work Role Functioning Questionnaire
** IC 95%: Interval of confidence to 95%

Table 4. Subscale description by type of health problem (mental or physical). Pre-test with the adapted version of the WRFQ* to Spanish spoken in Spain (n=40). April-May, 2012.

	Median**		Test U of Mann Whitney
	Mental health problem	Physical health problem	Asymptotic significance (bilateral)
Work scheduling demands	85,0	65,0	0,478
Output demands	78,6	82,1	0,850
Physical demands	85,0	55,0	0,007
Mental demands	75,0	95,8	0,018
Social demands	83,3	87,5	0,917

(*) Work Role Functioning Questionnaire.

(**) 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.

Conclusions

Our results confirm that the process used for translation and cross-cultural adaptation of the WRFQ was carried out successfully and indicate the existence of a good preservation of its psychometric properties[1,2].

BIBLIOGRAFIA:

1. **Ramada JM, Serra C, Delclós GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. Salud Publica Mex. 2013;55:57-66.**
2. **Ramada JM, Serra C, Amick III BC, Castaño J, Delclos GL. Cross-Cultural Adaptation of the Work Role Functioning Questionnaire to Spanish Spoken in Spain. J Occup Rehabil. 2013; [Epub ahead of print].**

8. APPENDICES

Appendix VIII: Poster

Ramada JM, Serra C, Delclós J. Traducción, adaptación cultural y validación del "Work role functioning questionnaire (WRFQ-27)". 3ª Jornada Científica CISAL. Barcelona, 2013.

Cross-cultural adaptation and validation of the Work Role Functioning Questionnaire to Spanish spoken in Spain.

José M Ramada[1,2,3], Consol Serra [1,2,3], Benjamin Amik III[4], Juan R Castaño[5,6], George L Delclos[2,3,4]

(1) Occupational Health Service. Parc de Salut MAR. Barcelona, Spain. (2) Center for Research in Occupational Health (CiSAL), University Pompeu Fabra. Barcelona, Spain. (3) CIBER of Epidemiology and Public Health (CIBERESP). (4) Epidemiology, Human Genetics and Environmental Sciences Division, The University of Texas School of Public Health, Houston, Texas, USA. (5) Psychiatry Service. Parc de Salut MAR, Barcelona. (6) Neuropsychiatry and Addictions Institute (INAD), Parc de Salut MAR, Barcelona.

Introduction

- Translation, cultural adaptation and validation (TCAV) of existing questionnaires to other languages and cultures facilitates international and multicultural research projects promoting information exchange and reducing time and costs to develop similar instruments[1].
- The simple translation of a questionnaire may lead to misinterpretation due to language and cultural differences. When using questionnaires developed in other countries and languages in scientific studies it is necessary, besides the translation, to carry out a cross-cultural adaptation and validation[2].
- The Work Role Functioning Questionnaire (WRFQ) is a self-completed work disability tool used to measure the perceived impact of a health problem on the worker's ability to perform his job[3].

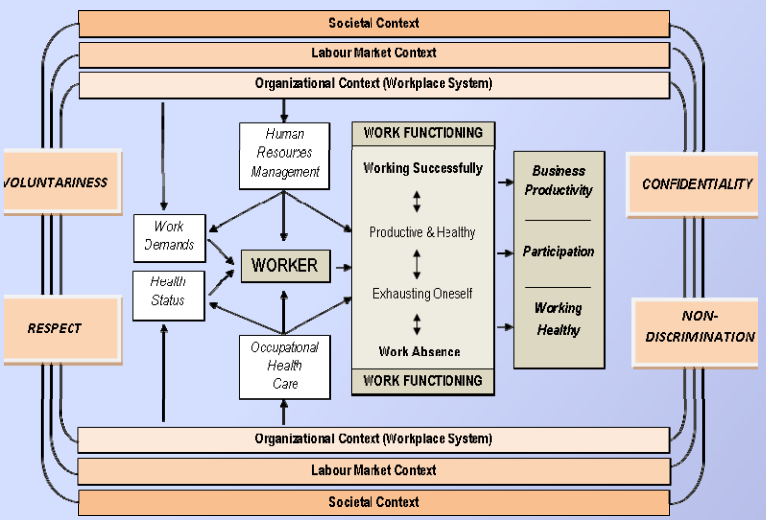
1. García de Yébenes MJ, Rodriguez-Salvanés F, Carmona-Ortells L. Validación de cuestionarios. Reumatol Clin. 2009;5:171-177.
2. Beaton DE, Bombardier C, Guillemin F, Bosi Ferraz M. Guidelines for the process of cross-cultural adaptation of self-reports measures. Spine. 2000;25:3186-3191.
3. Amick BC 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.

Objective

The aim of this study is to translate and adapt the WRFQ to Spanish spoken in Spain and evaluate its validity and reliability.

Methods

- Systematic 5-step procedure: direct translation, synthesis, back-translation, expert committee and pre-test.
- Structured interviews (n=40) with workers with musculoskeletal and prevalent mental health disorders were performed to evaluate the comprehensibility, usability, applicability and completeness of the adapted questionnaire.
- Panel of experts to assess content validity.
- Techniques for analysis of known groups to assess construct validity assessed.
- Internal consistency (Item-to-subscale and item-to-total correlations), and checking of floor and ceiling effects were examined.
- Test-retest reliability at 15 days was assessed by calculating the Intraclass Correlation Coefficient (ICC) after administering the WRFQ twice to the same sample of 40 workers.



Results

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] n	(%)	Mean scores	(SD)[b]	Range	Median scores	n at floor n (%)	n at ceiling n (%)	Cronbach alpha
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.

	Health issue (median scores[a])				Job type (median scores[a])			
Work demands	None (N= 156)	Mental (N= 125)	Physical (N= 139)	p[(b)] value	Manual (N= 109)	Non-manual (N= 123)	Mixed (N= 212)	p[(b)] value
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
Work demands	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	value
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.

Conclusions

1. The WRFQ-SpV was validated in different studies and appears to be a reliable and valid instrument to measure health-related work functioning in Spanish-speaking populations.
2. The results provide evidence of the responsiveness of the instrument, suggesting the possibility of its use in evaluative studies since it was able to detect (true) changes over time.

BIBLIOGRAFIA:
1. Ramada JM, Serra C, Delclós GL. Adaptación cultural y validación de cuestionarios de salud: revisión y recomendaciones metodológicas. Salud Publica Mex. 2013;55:57-66.
2. Ramada JM, Serra C, Amick III BC, Castaño J, Delclos GL. Cross-Cultural Adaptation of the Work Role Functioning Questionnaire to Spanish Spoken in Spain. J Occup Rehabil. 2013; [Epub ahead of print].
3. Ramada JM, Serra C, Amick III BC, Abma FI, Castaño JR, Pidemunt G, Bültmann U, Delclos GL. Reliability and validity of the Work Role Functioning Questionnaire (Spanish version). [Submitted for peer-review].
4. Ramada JM, Delclos GL, Amick III BC, Abma FI, Castaño JR, Pidemunt G, Bültmann, Serra C.Responsiveness of the Work Role Functioning Questionnaire (Spanish version). J Occup Environ Med. [In Press 2013].