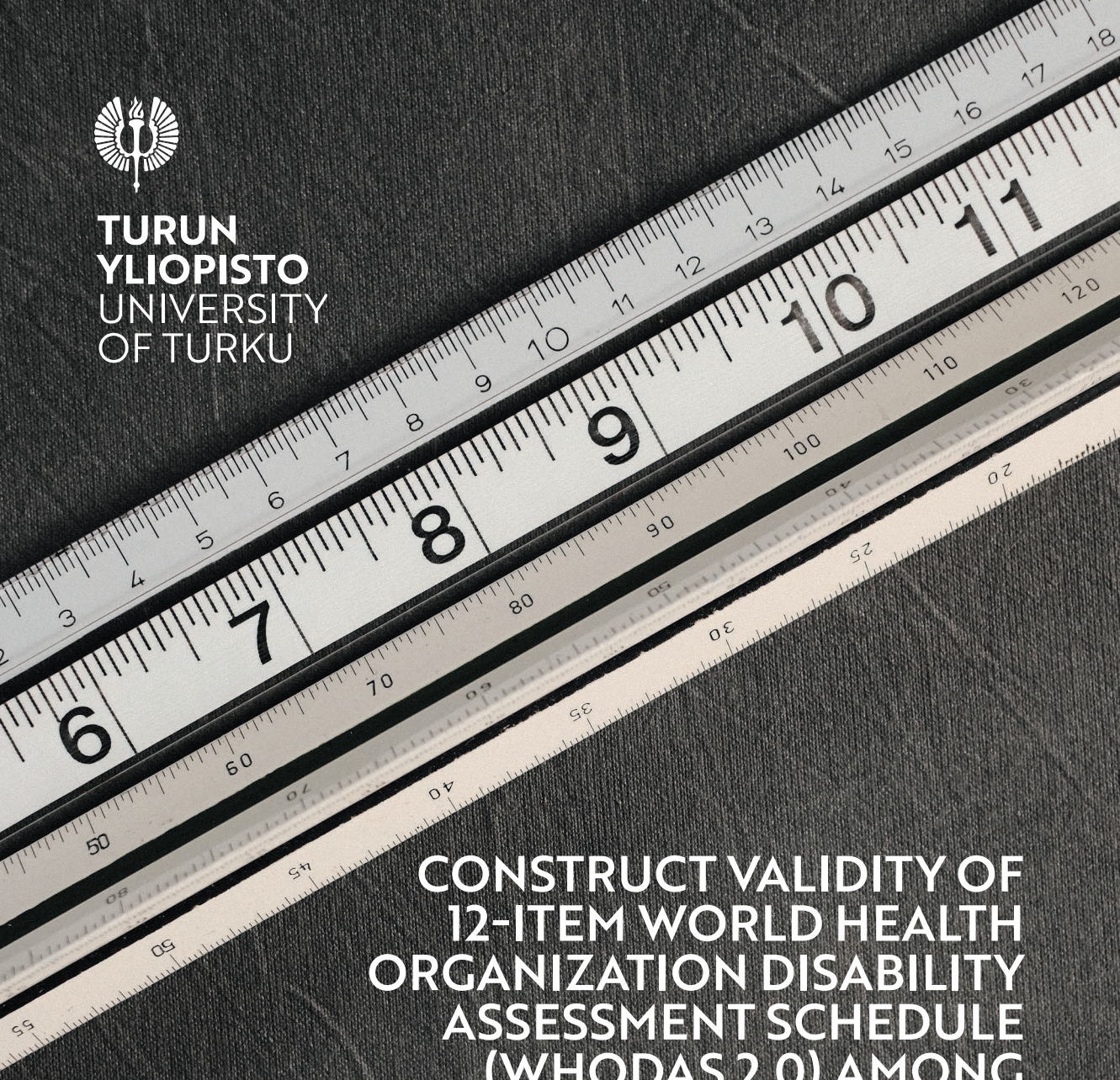

TURUN YLIOPISTO
UNIVERSITY OF TURKU

CONSTRUCT VALIDITY OF 12-ITEM WORLD HEALTH ORGANIZATION DISABILITY ASSESSMENT SCHEDULE (WHODAS 2.0) AMONG PEOPLE WITH CHRONIC LOW BACK PAIN

Esa Bärlund

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1687 | MEDICA – ODONTOLOGICA | TURKU 2023

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

CONSTRUCT VALIDITY OF 12-ITEM WORLD HEALTH ORGANIZATION DISABILITY ASSESSMENT SCHEDULE (WHODAS 2.0) AMONG PEOPLE WITH CHRONIC LOW BACK PAIN

Esa Bärlund

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. D OSA – TOM. 1687 | MEDICA – ODONTOLOGICA | TURKU 2023

University of Turku

Faculty of Medicine
Department of Physical and Rehabilitation Medicine
Doctoral Programme in Clinical Research DRCR

Supervised by

Professor Mikhail Saltychev, MD, PhD
Department of Physical and Rehabilitation Medicine
Department of Clinical Medicine
University of Turku
Turku, Finland

Adjunct Professor, Jaana Paltamaa, PhD, PT
JAMK University of Applied Sciences
School of Health and Social Studies
Jyväskylä, Finland

Reviewed by

Professor Anne Söderlund, PhD
School of Health, Care and Social Welfare
Mälardalen University
Mälardalen, Sweden

Adjunct Professor Aki Vainionpää, MD, PhD
Department of Physical and Rehabilitation Medicine
The Hospital District of South Ostrobothnia
Seinäjoki, Finland

Opponent

Professor, Thomas Maribo, PhD
Department of Public Health,
Department of Health Services Research
Aarhus University
Aarhus, Denmark

The originality of this publication has been checked in accordance with the University of Turku quality assurance system using the Turnitin OriginalityCheck service.

Cover Image: Esa Bärlund, Mika Arvola

ISBN 978-951-29-9135-8 (PRINT)
ISBN 978-951-29-9136-5 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)
Painosalama, Turku, Finland 2023

To Niina, Henri and Niklas.

A man's heart deviseth his way: but the LORD directeth his steps.
Proverbs 16:9

UNIVERSITY OF TURKU
Faculty of Medicine
Physical and Rehabilitation Medicine
ESA BÄRLUND: Construct validity of 12-item World Health Organization Disability Assessment Schedule (WHODAS 2.0) among people with chronic low back pain
Doctoral Dissertation, 125 pp.
Doctoral Programme in Clinical Research
March 2023

ABSTRACT

In order to ensure the effectiveness of the prevention, treatment and rehabilitation of low back pain, the assessment of the level of disability requires reliable scales, including patient-reported outcome measures. The main aim of this thesis was to explore the construct validity of the 12-item self-administered WHO Disability Assessment Schedule 2.0 (WHODAS 2.0).

This study evaluated internal consistency, factor structure and psychometric properties of the WHODAS 2.0. Additionally, the psychometric properties of the Oswestry Disability Index (ODI) were evaluated. The psychometric properties of WHODAS 2.0 were also assessed through a systematic review. The study used cross-sectional data from the Turku ICF Study, which covered 1,988 patients (65% women) with low back pain who had visited an outpatient Physical and Rehabilitation Medicine clinic between 2015 and 2017. The samples of the studies varied from 501 (Study I) up to 1,379 (Study V) participants. Internal consistency was assessed using Cronbach's alpha. Construct structure was evaluated by exploratory and confirmatory factor analyses. Psychometric properties were assessed using item response theory analysis. The convergent validity of WHODAS 2.0 was assessed by inspecting correlations between the WHODAS 2.0, the ODI and a pain numeric rating scale.

The internal consistency of WHODAS 2.0 was good. The exploratory factor analysis revealed two retained factors. The ability to discriminate between all the items was high or perfect with a shift towards higher levels of disability. The correlation between pain severity, measured by a numeric rating scale, and functioning level, measured by the WHODAS 2.0, was weak to moderate, with slightly stronger associations in the physical domains of functioning. When comparing the WHODAS 2.0 with the ODI, items defining physical functioning mostly showed stronger correlations than items defining social or psychological functioning.

It seems that the 12-item self-administered WHODAS 2.0 is an internally consistent and valid scale correlating well with other measures of disability among patients with chronic low back pain. However, due to a multidimensionality, its total score may represent the different combinations of several contributing factors. The WHODAS 2.0 and the ODI may complement each other when describing the functioning of people with chronic low back pain.

KEYWORDS: WHODAS 2.0, validity, psychometrics, rehabilitation, musculoskeletal disease, low back pain

TURUN YLIOPISTO
Lääketieteellinen tiedekunta
Fysiatria
ESA BÄRLUND: Maailman terveysjärjestö WHO:n terveyden ja toimintarajoitteiden arviointimenetelmän WHODAS 2.0 12 kysymyksen itse täytettävän mittarin rakennevaliditeetti kroonisilla alaselkäkipupotilailla
Väitöskirja, 125 s.
Turun kliininen tohtoriohjelma
Maaliskuu 2023

TIIVISTELMÄ

Alaselkäkivun tehokas ehkäisy, hoito ja kuntoutus edellyttävät luotettavia toimintakyvyn haitta-asteen mittareita, mukaan lukien potilaan itsearviointimittarit. Tämän väitöskirjan päätavoitteena oli tutkia Maailman terveysjärjestö WHO:n terveyden ja toimintarajoitteiden arviointimenetelmän (WHODAS 2.0) 12 kysymyksen itse täytettävän version rakennevaliditeettia kroonisilla alaselkäkipupotilailla.

Tutkimuksessa arvioitiin WHODAS 2.0 mittarin sisäistä yhtenevyyttä, faktorirakennetta ja mittarin psykometrisiä ominaisuuksia. Lisäksi arvioitiin Oswestry Disability Index (ODI):n psykometrisiä ominaisuuksia. Tietoa WHODAS 2.0:n psykometrisista ominaisuuksista arvioitiin systemaattisella katsauksella. Tutkimuksessa käytettiin fysiatrian poliklinikalla käyneiltä alaselkäkipupotilailta 2015 – 2017 kerättyä TURKU ICF-tutkimuksen poikkileikkausaineistoa (n= 1988) (65 % naisia). Osatutkimusten otoskoot vaihtelivat 501 (Tutkimus I) ja 1379 (Tutkimus V) tutkittavan välillä. Mittarin sisäistä yhtenevyyttä arvioitiin Cronbachin alfalla. Mittarin rakennetta arvioitiin eksploratiivisella ja konfirmatorisella faktorinanalyysilla. Psykometrisiä ominaisuuksia arvioitiin osiovasteteorian avulla. WHODAS 2.0:n rinnakkaisvaliditeettia arvioitiin tarkistamalla WHODAS 2.0:n, ODI:n ja kivun voimakkuuden (numeerinen asteikko) väliset korrelaatiot.

WHODAS 2.0:n sisäinen yhtenevyys oli hyvä. Eksploratiivinen faktorianalyysi tunnisti kaksi faktoria. Kyky erotella kaikki osiot olivat korkea tai täydellinen, painottuen kohti suurempaa toimintakyvyn haittaa. Kivun määrän, numeerisella asteikolla mitattuna, ja WHODAS 2.0:lla mitatun toimintakyvyn välinen korrelaatio oli heikko tai kohtalainen. Fyysisen toimintakyvyn aihealueilla yhteys oli vahvempaa. WHODAS 2.0 ja ODI:n fyysisen toimintatoimintakyvyn osiot korreloivat keskenään enimmäkseen vahvemmin verrattuna sosiaalista tai psyykkistä toimintakykyä määrittävien osioiden korrelaatioihin.

12 kysymyksen itse täytettävä WHODAS 2.0 vaikuttaa sisäisesti yhtenäiseltä ja luotettavalta asteikolta, joka korreloi hyvin muiden toimintakyvyn haittaa arvioivien mittareiden kanssa. Näyttää kuitenkin siltä, että WHODAS 2.0 on moniulotteinen asteikko, ja siten sen antama kokonaispistemäärä voi edustaa useiden vaikuttavien tekijöiden erilaisia yhdistelmiä. WHODAS 2.0 ja ODI saattavat täydentää toisiaan kuvatessaan kroonisten alaselkäkipua sairastavien ihmisten toimintakykyä.

AVAINSANAT: WHODAS 2.0, validiteetti, psykometriikka, kuntoutus, tuki- ja liikuntaelinsairaudet, alaselkäkipu

Table of Contents

Abbreviations

BMI	Body Mass Index
CFA	Confirmatory Factor Analysis
CI	Confidence interval
DIF	Differential item functioning
EFA	Exploratory Factor Analysis
ICD	International statistical classification of diseases and related health problems
ICF	International Classification of Functioning, Disability and Health
ICIDH	International Classification of Impairments, Disabilities and Handicaps
IQR	Interquartile range
IRT	Item response theory
LBP	Low back pain
MCID	Minimal clinically important difference
MDC	Minimal detectable change
NRS	Numeric or numerical ratings scale
ODI	Oswestry Disability Index
PROM	Patient-reported outcome measure
RCT	Randomized controlled trial
RMDQ	Roland-Morris Disability Questionnaire
RMSEA	Root Mean Square Error of Approximation
SF-36	36-Item Short-form Survey
TOIMIA	Functioning Measures Database
VAS	Visual analogue scale
WHO	World Health Organization
WHODAS	World Health Organization Disability Assessment Schedule

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Saltychev M, Bärlund E, Mattie R, McCormick Z, Paltamaa J, Laimi K. A study of the psychometric properties of 12-item World Health Organization Disability Assessment Schedule 2.0 in a large population of people with chronic musculoskeletal pain. *Clin Rehabil. 2017 Feb;31(2):262-272*. DOI: 10.1177/0269215516631385.

II Saltychev M, Mattie R, McCormick Z, Bärlund E, Laimi K. Psychometric properties of the Oswestry Disability Index. *Int J Rehabil Res*. 2017 Sep;40(3):202-208. DOI: 10.1097/MRR.0000000000000226.

III Saltychev M, Bärlund E, Laimi K. Correlation between the pain numeric rating scale and the 12-item WHO Disability Assessment Schedule 2.0 in patients with musculoskeletal pain. Int J Rehabil Res. 2018 Mar;41(1):87-91. DOI: 10.1097/MRR.0000000000000262.

IV Saltychev, M., Katajapuu, N., Bärlund, E., & Laimi, K. (2019). Psychometric properties of 12-item self-administered World Health Organization disability assessment schedule 2.0 (WHODAS 2.0) among general population and people with non-acute physical causes of disability – systematic review. *Disability and Rehabilitation, 43, 789-794.* DOI: 10.1080/09638288.2019.1643416.

V Bärlund E, Katajapuu N, Paltamaa J, Saltychev M. Correlation between Oswestry Disability Index and 12-item self-administered version of World Health Organization Disability Assessment Schedule (WHODAS 2.0) in patients with chronic low back pain. *Int J Rehabil Res. 2021, 44:170-172.* DOI: 10.1097/MRR.0000000000000465.

1 Introduction

Low back pain (LBP) affects people of all ages. Worldwide, its prevalence is 600 million. In 2019, according to the global age-adjusted Disability Adjusted Life Years (DALY) rates, LBP was among the ten leading causes of DALYs among both genders and all ages, also in Finland. The DALY LBP rates are higher among women and the highest DALY rates have been observed at ages between 45 and 49 years (Abrams et al., 2020; Wu et al., 2020).

LBP weakens functioning on an individual level, but it also involves significant financial losses on a societal level. In 2020, The Social Insurance Institution of Finland (2021) compensated 28,000 sick leaves due to dorsopathies, accounting for almost 3 830 000 working days and EUR 83 000 000.

Many different outcome measures have been suggested for standardizing the assessment of symptoms and disability caused by LBP. The International Consortium for Health Outcome Measurement has recommended such patient-reported outcome measures (PROMs) as a numeric rating scale (NRS) for measuring pain severity (Downie et al., 1978) and the Oswestry Disability Index (ODI) (Fairbank et al., 1980) for assessing disability level (Clement et al., 2015). The Initiative on Methods, Measurement, and Pain Assessment in Clinical Trials and Outcome Measures in Rheumatology (Taylor et al., 2016) has recommended several disease-specific measures of physical activity and functioning, including the Roland Morris Disability Questionnaire (RMDQ) (Roland & Morris, 1983) and the ODI (Fairbank, 1995) . The ODI, the RMDQ and NRSs have been recommended for use in clinical trials of nonspecific LBP (Chiarotto et al., 2018). Also in Finland, the Finnish Spine Register project has recommended using the ODI as one of the main outcome measures of functioning in cases of LBP (Pekkanen et al., 2019).

For forty years, the ODI has served as a gold standard for measuring disability caused by LBP. The ODI's validity and reliability has been studied in cross-cultural settings and found to be good (Denis & Fortin, 2012; Miekisiak et al., 2013). Although the ODI's test-retest reliability, convergent validity, and responsiveness to change (including its numerous translations) have been widely studied, knowledge of its psychometric properties, internal consistency, and factor structure is surprisingly scarce (Ramasamy et al., 2017; Eprovide, 2022). For example, only a

few previous studies have assessed the psychometric properties of the ODI using the Rasch analysis or the item response theory (IRT) (White & Velozo, 2002; Davidson, 2008). In other words, how well this widely used scale is able to measure what it is supposed to measure is still uncertain.

The intensity of LBP has been commonly assessed using a well-validated 11-point Likert-type NRS (Dworkin et al., 2005; Hjermstad et al., 2011; Clement et al., 2015; Chiarotto et al., 2019).

The connection between the level of disability and the severity of LBP is well established. Moderate to strong positive correlations have been observed between the ODI and LBP severity (Mannion et al., 2006; Vigatto et al., 2007; Lue et al., 2008; Monticone et al., 2009; Pekkanen et al., 2011; Payares et al., 2011; Solomon & Roopchand-Martin, 2011; Valasek et al., 2013; Algarni et al., 2014; Selva-Sevilla et al., 2019; Phedy et al., 2021).

The World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) (World Health Organization, 2010) has been extensively used since 2010 as a generic and standardized tool for assessing health and disability caused by different health conditions in diverse settings and cultures. The conceptual framework of WHODAS 2.0 is based on the International Classification of Functioning, Disability and Health (ICF) (World Health Organization, 2001) (Federici et al., 2017).

The 36-item version of WHODAS 2.0, either self-, interviewer- or proxy-administered, has been found to be a valid tool among people with chronic diseases (Garin et al. (2010). The 36-item interviewer-administered version of WHODAS 2.0 has shown excellent internal consistency and convergent validity among patients with LBP in primary care (Chwastiak & Von Korff, 2003).

Despite the popularity of WHODAS 2.0, there is some ambiguity concerning the factor structure, internal consistency, psychometric properties, minimal clinically important difference (MCID), and convergent validity of its 12-item version (Federici et al., 2017).

The correlations between WHODAS 2.0 scores and the results of other measurements of functioning among patients with LBP have mostly been studied using a 36-item version of WHODAS 2.0. The 36-item interviewer-administered version has demonstrated weak to strong positive associations with the ODI among patients with chronic LBP (Ćwirlej-Sozańska et al., 2020a). The 36-item interviewer-administered version of WHODAS 2.0 has also shown low to moderate positive associations with an NRS among patients with chronic LBP (Igwesi-Chidobe et al., 2020). Only a few previous studies have assessed the convergent validity of the 12-item WHODAS 2.0, and the results have been indecisive. Lee and Song (2013) demonstrated a strong positive association between the 12-item interviewer-administered WHODAS 2.0 and the ODI among chronic LBP patients.

The objectives of this thesis were to appraise available evidence on the psychometric characteristics of the 12-item self-administered WHODAS 2.0 and to evaluate these characteristics among patients with LBP. The study had a special focus on the convergent validity of the 12-item WHODAS 2.0, which was studied comparing the WHODAS 2.0 with the ODI and a pain NRS.

The ultimate goal of this thesis was to evaluate how well the 12-item self-administered version of WHODAS 2.0 defines the disability level of patients with LBP. In other words, the main objective was to investigate how strongly the disability level measured by WHODAS 2.0 correlated with the severity of LBP, and whether the information on disability level obtained from established PROMs such as the ODI or NRS, might be substituted by information obtained from more standardized measures such as WHODAS 2.0. Such knowledge may improve the precision of assessments of the functioning of patients with LBP and thus, ensure more cost-effective allocation of resources when managing LBP.

2 Review of the Literature

2.1 Low back pain

It has been estimated that 50% to 80% of all adults experience at least one episode of LBP during their lifetimes (Rubin, 2007).

In addition, the mean prevalence and incidence of LBP ranges from 1.4 up to 20% and 0.02 to 7%. Gender may be a risk factor for LBP as men were affected more than women (odds ratio from 1 to 17), and age, race have been identified as risk factors. Similarly, high-intensity physical activity, high spinal load, lifting, bending and twisting have also been acknowledged as risk factors for LBP (Fatoye et al., 2019).

In their study, Koponen et al. (2018) found that 44% of Finnish men and 48% of Finnish women had experienced back pain during the last 30 days and that the difference in this prevalence had increased between genders after the age of 60 years.

The Finnish Current Care Guideline for LBP (Low back pain: Current Care Guideline, 2017) recommends categorizing LBP on the basis of the duration of pain, with "acute" referring to pain lasting less than 6 weeks; "subacute" to pain lasting between 6 and 12 weeks; and "chronic" referring to pain lasting over 12 weeks. According to Deyo et al. (2014), chronic LBP is defined as pain that has persisted for a minimum of three months and appearing on at least half of the days in the last six months.

2.1.1 Patient-Reported Outcome Measures used for measuring pain and disability caused by low back pain

Research has endeavored to standardize outcome measures in people with chronic LBP. When defining standards for research related to self-report measures of chronic LBP, Deyo et al. (2014) concluded that physical functioning, depression, sleep disturbance, and catastrophizing should be included. Earlier, Deyo et al. (1998) and Bombardier (2000) recommended the ODI, and Dworkin et al. (2005) an 11-point NRS as outcome measures in trials for spinal disorders, LBP and chronic pain. In a systematic review, Froud et al. (2016) analyzed the PROMs used in LBP trials through 1980-2012 and found that four of the most commonly used primary or

secondary PROMs, in order of use, were a visual analogue scale (VAS) (Huskisson, 1974), the RMDQ, the ODI and an NRS. In addition, the four most used outcome measures have been classified into pain measures, disability measures, and psychological measures (Froud et al., 2016).

Similarly, when Chapman et al. (2011) reviewed outcome measures specifically used in randomized controlled studies on LBP, they found the most common functional outcome measures to be the ODI, the RMDQ and the range of motion. For pain, the most common measures were an NRS, the Brief Pain Inventory, the Pain Disability Index (Cleeland, 1989), the McGill Pain Questionnaire (Melzack, 1975) and a VAS. For psychosocial functioning, the most common measures were the Fear Avoidance Beliefs Questionnaire (Waddell et al., 1993), the Tampa Scale for Kinesiophobia (Miller et al., 1991) and the Beck Depression Inventory (Beck et al., 1961). Finally, the most common measures for the generic quality of life were the 12-item Short-form Health Survey (SF-12) (Ware & Sherbourne, 1992), the Nottingham Health Profile (Hunt et al., 1981), the 12-item Short-Form Health Survey (Ware et al., 1996) and the Sickness Impact Profile (Bergner et al., 1976).

2.2 The International Classification of Functioning, Disability and Health and World Health Organization Disability Assessment Schedule WHODAS 2.0

2.2.1 International Classification of Functioning, Disability and Health (ICF)

Whereas the International Classification of Impairments, Disabilities, and Handicaps (ICIDH) has been considered a "consequences of diseases" classification, the ICF has been constructed as a classification of "components of health" (ICIDH; World Health Organization, 1980; World Health Organization, 2001).

The ICF is based on a biopsychosocial model and broadly integrating medical and social models of disability. Through this integrative biopsychosocial model of functioning, disability, and health, the ICF contextualizes functioning as a continuum, where an individual's functioning can be placed between the two endpoints: full functioning and complete loss of functioning (World Health Organization, 2001; Bickenbach et al., 2012).

The ICF is one of the reference classifications for the WHO Family of International Classifications (WHOFIC), together with ICD-11 and International Classification of Health Interventions (ICHI), and it aims to offer a common, standardized language, as well as a framework that describes health and the health-related domains. ICF is the WHO framework for measuring health and disability at both individual and population

levels. It also aims to increase scientific knowledge and understanding of functioning and disability in health-related studies by providing a common language for all healthcare professionals and policymakers. The common use of the ICF should make comparing data internationally, inter-professionally, and over time, easier and more efficient. Two versions of the ICF are available: a full and a short version. The full version has four levels of classification in contrast to the two levels in the short version (World Health Organization, 2001, 2021).

The ICF has two parts, each of which has two components, expressed in positive or negative terms. The first part covers Functioning and Disability (Body Functions and Structures, Activities and Participation) and the second covers Contextual Factors (Environmental Factors and Personal Factors). Functioning and disability dynamically interact with each other. Disability is defined as dysfunctioning at one or more levels: impairments, activity limitations and participation restrictions.

All the components of the ICF are defined in the context of health. The components are constructed from domains holding categories. The categories within the health and the health-related domains, e.g., transportation, education and social interactions, are the units of classification instead of persons. When recording an individual's health or health-related state, categorical coding and specific qualifiers can be used. "Capacity", as a construct, underlines the highest probable level of functioning that an individual can reach in the Activities and Participation part. It is measured in a standardized environment, whereas "performance" is seen as a construct that describes what an individual is able to do in their current environment. Also, if coding is done in the context of ICF, different levels of qualifiers can be used. The effect of environmental factors may be either positive (a facilitator) or negative (a barrier) (World Health Organization, 2001).

2.2.2 The International Classification of Functioning, Disability and Health and Core Sets

To facilitate the use and practicality of the ICF, the core sets have been developed for multiple conditions including LBP (Cieza et al., 2004). Developing these core set required an empirical multicenter study, a systematic literature review, and a Delphi study of experts and people suffering from LBP (Weigl et al., 2004). The cores sets include ICF items that describe the most common limitations of functioning in particular health conditions. They usually exist in two forms—a brief core set and a comprehensive core set.

The comprehensive ICF Core Set for LBP includes 78 categories, and the brief Core Set includes 35 categories (Table 1). They have been validated across different populations and cultures (Cieza et al., 2004; Bickenbach et al., 2012; Karlsson & Gustafsson, 2021).

Table 1. Comprehensive ICF Core Set for LBP. Modified from Cieza et al. (2004).

ICF code ICF Category Title, 2nd level	
Body Functions	
b126 Temperament and personality functions	b130 Energy and drive functions*
b134 Sleep functions*	b152 Emotional functions*
b180 Experience of self and time functions	b260 Proprioceptive function
b280 Sensation of pain*	b455 Exercise tolerance functions*
b620 Urination functions	b640 Sexual functions
b710 Mobility of joint functions*	b715 Stability of joint functions*
b720 Mobility of bone functions	b730 Muscle power functions*
b735 Muscle tone functions*	b740 Muscle endurance functions*
b750 Motor reflex functions	b770 Gait pattern functions
b780 Sensations related to muscles and movement functions	
Body Structures	
s120 Spinal cord and related structures*	s740 Structure of pelvic region
s750 Structure of lower extremity	s760 Structure of trunk*
s770 Additional musculoskeletal structures related to movement*	
Activities and Participation	
d240 Handling stress and other psychological demands*	d410 Changing basic body position*
d415 Maintaining a body position*	d420 Transferring oneself
d430 Lifting and carrying objects*	d445 Hand and arm use
d450 Walking*	d455 Moving around
d460 Moving around in different locations	d465 Moving around using equipment
d470 Using transportation	d475 Driving
d510 Washing oneself	d530 Toileting*
d540 Dressing*	d570 Looking after one's health
d620 Acquisition of goods and services	d630 Preparing meals
d640 Doing housework*	d650 Caring for household objects
d660 Assisting others	d710 Basic interpersonal interactions
d760 Family relationships*	d770 Intimate relationships
d845 Acquiring, keeping and terminating a job*	d850 Remunerative employment*
d859 Work and employment, other specified and unspecified*	d910 Community life
d920 Recreation and leisure	
Environmental Factors	
e110 Products or substances for personal consumption*	e120 Products and technology for personal indoor and outdoor mobility and transportation
e135 Products and technology for employment*	e150 Design, construction and building products and technology of buildings for public use
e155 Design, construction and building products and technology of buildings for private use*	e225 Climate
e255 Vibration	e310 Immediate family*
e325 Acquaintances, peers, colleagues, neighbours and community members	e330 People in positions of authority
e355 Health professionals*	e360 Other professionals
e410 Individual attitudes of immediate family members*	e425 Individual attitudes of acquaintances, peers, colleagues, neighbours and community members
e450 Individual attitudes of health professionals*	e460 Societal attitudes
e455 Individual attitudes of other professionals	e540 Transportation services, systems and policies
e465 Social norms, practices and ideologies	e570 Social security services, systems and policies*
e550 Legal services, systems and policies*	e580 Health services, systems and policies*
e575 General social support services, systems and policies	e590 Labour and employment services, systems and policies
e585 Education and training services, systems and policies	

* Belongs to Brief ICF Core Set for LBP.

2.2.3 The World Health Organization Disability Assessment Schedule WHODAS 2.0

WHODAS 2.0. is a generic and standardized tool developed to measure health and disability in different populations and cultures (Üstün et al., 2010). The WHODAS 2.0 is directly linked to the ICF (Üstün et al., 2010). Furthermore, it is grounded in the conceptual framework of the ICF (World Health Organization, 2021).

Of the seven available versions, the most detailed and comprehensive versions are the three versions of 36-item WHODAS 2.0. The 12-item WHODAS 2.0 is the shortest version, and it is useful for a quicker assessment, whereas the 12+24-item version is especially useful in computer-adaptive testing (CAT), in which the responses to a 12-item version may lead to additional 24 items. The 36- and 12-item WHODAS 2.0 are available in interviewer-administered, self-administered and proxy-administered versions, whereas the 12+24 version is only available in interviewer-administered or computerized adaptive testing versions (World Health Organization, 2010).

The WHODAS 2.0 covers the level of functioning of an individual in six domains of the ICF Activities and Participation including cognition, mobility, self-care, getting along, life activities, and participation. This direct link to the ICF is a very distinctive and unique feature of WHODAS 2.0. It also emphasizes that WHODAS 2.0 is etiologically neutral, meaning that it can be used directly to measure functioning and disability regardless of any preconditions. The six domains covered by the WHODAS 2.0 are (World Health Organization, 2010) are:

- Domain 1: Cognition – communicating and thinking
- Domain 2: Mobility – moving and getting around
- Domain 3: Self-care – bathing, dressing, eating, and staying alone
- Domain 4: Getting along – interacting with other people
- Domain 5: Life activities – day-to-day activities including household, work, and school activities
- Domain 6: Participation – problems participating in society because of the society in which they live

The recall period is the last 30 days. The respondents are asked questions about the degree of difficulty they have experienced while performing different activities in their usual way. The participants are also asked to provide an average score for good and bad days, based on how well they are usually performing when participating in different activities. They rate the difficulties on a five-point scale: “none” – “mild difficulties” – “moderate difficulties” – “severe difficulties”–

"extreme difficulties or cannot do". The difficulties include increased effort, discomfort or pain, slowness and/or changes in way the person does the activity.

The summary scoring for WHODAS 2.0 can be done in either a short or long way, i.e., by simple or complex scoring. In simple scoring, the scores of each item, from 0 to 4, are summed. Simple scoring is always sample-dependent and it cannot be used for comparing populations (World Health Organization, 2010).

Complex scoring, a method constructed based on the IRT, enables the comparison of different populations. Using that method, each item response is separately coded on the original scale and then computed by weighting severity. The summary score obtained by a complex scoring can be turned into a scale from 0 to 100, in which 0 points denotes no disability and 100 points full disability. A syntax algorithm is available for statistical analysis software from the WHO (World Health Organization, 2010).

2.2.4 Linking the International Classification of Functioning, Disability and Health items with the World Health Organization Disability Assessment Schedule, the Oswestry Disability Index and a pain numerical rating scale

The rules of linking have been described by Cieza et al. (Cieza et al., 2002; Cieza et al., 2005) to connect health measurements and such classifications as the ICF categories. Links between the WHODAS 2.0, the ODI and an NRS are summarized in Table 2.

Table 2. Linking the WHODAS 2.0 (36- and 12-item), the ODI and a NRS according to the one- or two-level ICF form. Modified from Sigl T et al. (2006); Üstün et al. (2010); Lu et al. (2020).

ICF category title	WHODAS 2.0 36-item	WHODAS 2.0 12-item	ODI	NRS
b134 Sleep functions	n/a	n/a	x	n/a
b144 Memory functions	X	n/a	n/a	n/a
b144 Memory functions	x	n/a	n/a	n/a
b152 Emotional functions	x	x	n/a	n/a
b280 Sensation of pain	n/a	n/a	x	x
d110 Watching	x	n/a	n/a	n/a
d115 Listening	x	n/a	n/a	n/a
d120 Other purposeful sensing	x	n/a	n/a	n/a
d129 Purposeful sensory experiences	x	n/a	n/a	n/a
d130 Copying	x	n/a	n/a	n/a
d135 Rehearsing	x	n/a	n/a	n/a
d140 Learning to read	x	n/a	n/a	n/a
d145 Learning to write	x	n/a	n/a	n/a
d150 Learning to calculate	x	n/a	n/a	n/a
d155 Acquiring skills	x	x	n/a	n/a

ICF category title	WHODAS 2.0 36-item	WHODAS 2.0 12-item	ODI	NRS
d159 Basic learning	x	n/a	n/a	n/a
d160 Focusing attention	x	x	n/a	n/a
d175 Solving problems	x	n/a	n/a	n/a
d210 Undertaking a single task	x	n/a	n/a	n/a
d220 Undertaking multiple tasking	x	n/a	n/a	n/a
d310 Communicating with—receiving—spoken messages	x	n/a	n/a	n/a
d350 Conversation	x	n/a	n/a	n/a
d410 Changing basic body position	x	n/a	x	n/a
d415 Maintaining a body position	x	x	n/a	n/a
d430 Lifting and carrying objects	n/a	n/a	x	n/a
d450 Walking	x	x	x	n/a
d460 Moving around in different locations	x	n/a	n/a	n/a
d498 Mobility, other specified	n/a	n/a	x	n/a
d510 Washing oneself	x	x	x	n/a
d540 Dressing	x	x	x	n/a
d550 Eating	x	n/a	n/a	n/a
d6 Domestic life	x	n/a	n/a	n/a
d640 Doing housework	x	x	n/a	n/a
d698 Domestic life, other specified	n/a	n/a	x	n/a
d720 Complex interpersonal interactions	x	n/a	n/a	n/a
d730 Relating with strangers	x	x	n/a	n/a
d750 Informal social relationships	x	x	n/a	n/a
d760 Family relationships	x	x	n/a	n/a
d770 Intimate relationships	x	x	n/a	n/a
d820 School education	x	n/a	n/a	n/a
d825 Vocational training	x	x	n/a	n/a
d830 Higher education	x	x	n/a	n/a
d850 Remunerative employment	x	x	n/a	n/a
d870 Economic self-sufficiency	x	n/a	n/a	n/a
d9 Community, social and civic life	x	n/a	n/a	n/a
d910 Community life	x	n/a	n/a	n/a
d920 Recreation and leisure	x	n/a	x	n/a
d940 Human rights	x	n/a	n/a	n/a
e110 Products or substances for personal consumption	n/a	n/a	x	n/a
e115 Products and technology for personal use in daily living	n/a	n/a	x	n/a
e120 Products and technology for personal indoor and outdoor mobility and transportation	n/a	n/a	x	n/a
TOTAL categories	43	15	12	1

x = available, n/a = not available; b = Body Functions, d = Activities and Participation; e = Environmental factors in ICF

2.2.5 Psychometric properties of WHO Disability Assessment Schedule 2.0

In the extensive research by Üstün et al. (2010), the WHODAS 2.0 was found to work well in different cultures and subgroups and among people with or without health problems such as mental disorders and addictions. Through a set of factor analyses, the unidimensionality of the questionnaire was approved. The internal consistency (Cronbach's alpha) was good varying from 0.8 to 1.0. Test-retest reliability was also good. The intraclass correlation coefficient at an item-level varied from 0.6 to 0.9, at a domain-level from 0.9 to 1.0, and at a total-score-level the coefficient was perfect 1.0. Concurrent validity has been found to be decent: the

total WHODAS 2.0 score was used to predict the scores for the WHO Quality of Life measure (r = 0.68), the London Handicap Scale (r = 0.75), The Functional Independence Measure (r = 0.68) and the 12- and 36-item Short-form Health Survey mental component scores (r = 0.17). One of the main findings was that the WHODAS 2.0 scores have highly correlated with the scores of the Functional Independence Measure's motor scale (r = 0.67) and the SF-36 Physical Component Score (r = 0.66) (Üstün et al., 2010).

Recently, when the psychometrics of WHODAS 2.0 (Federici et al., 2017) were reviewed, it was found that the WHODAS 2.0 was a valid and reliable scale for assessing disability. The review also suggested that the MCID of WHODAS 2.0 needs to be established.

The Finnish national Functioning Measures Database TOIMIA used by healthcare professionals, publishes frequently updates on the present-day data concerning the psychometrics of WHODAS 2.0. In Finland, WHODAS 2.0 has been recommended as a tool for screening restrictions to participation among adults and for self-assessment of adults' functional capacity in identifying the need for rehabilitation and monitoring rehabilitation (Kuntoutuksen tietopohja-hankkeen (KUTI) asiantuntijaryhmän jäsenet, 2020).

The 12-item interviewer-administered Finnish version of WHODAS 2.0 was translated according to the WHO translation guidelines by the Project of Health, Well-being, and Quality of Life of the Ageing Population in Europe (Perales et al., 2014). In 2014, the Finnish version of WHODAS 2.0 was piloted in the samples of people disabled due to neurological conditions, people undergoing cardiac rehabilitation and people with spinal cord injury (Paltamaa, 2014).

2.2.6 World Health Organization Disability Assessment Schedule 2.0 and chronic musculoskeletal pain patients

A single systematic review on the psychometric properties of WHODAS 2.0 has previously been published (Federici et al., 2017). Of the included papers, only one-third had concerned the 12-item version of WHODAS 2.0.

According to Federici et al. (2017), the 12-item version, whether self- or interviewer-administered, appeared to be unidimensional. In addition, the 12-item version of WHODAS 2.0 accounted for the 80% variance observing in a 36-item version. Test-retest reliability was reported as generally high in different populations. In summary, the correlations between the WHODAS 2.0 and other self-reporting measures, as for example, the World Health Organization Quality of Life (WHOQL: World Health Organization, 1997) or the 36-Item Short-form Survey (SF-36: Ware, 2000) have varied from weak to strong. In the review of both

WHODAS II and WHODAS 2.0, the authors also concluded that the MCID for WHODAS 2.0 needs to be defined (Federici et al., 2017).

The MCID and the minimal detectable change (MDC) for the self-administered 12-item WHODAS 2.0 were reported by Katajapuu et al. (2020) in a sample of 2,000 patients with chronic musculoskeletal pain derived from a population employed by the present thesis. The MCID was found to vary from 3.1 to 4.7 points, while the MDC was 8.6 points and the MDC% was 66% (Katajapuu et al., 2020). The same research group did not observe ceiling but significant floor effect of 15% to 79% for the items of WHODAS 2.0. Additionally, the same research group reported uniform gender-related differential item functioning (DIF) in seven out of twelve items (Katajapuu et al., 2019b; Katajapuu et al., 2020). Katajapuu (2021) stated that the WHODAS 2.0 score should be used with caution due to multidimensionality. Instead, creating the profile of limitations of functioning may be more reliable approach to use the WHODAS 2.0.

Ćwirlej-Sozańska et al. (2020a) assessed the psychometric properties of the Polish version of the 36-item interviewer-administered WHODAS 2.0 among 92 patients with LBP (women 62%, age 66 years). The convergent validity of WHODAS 2.0 was assessed comparing WHODAS 2.0 with the ODI and reporting positive correlations ranged from 0.35 to 0.87. The lowest correlation of 0.35 was detected between domain #4 “Getting along” and the total score for the ODI. Respectively, the highest correlation of 0.87 was seen between the total scores. Garin et al. (2010) suggested that the WHODAS 2.0 might fail to distinguish people with mild, moderate, or severe LBP.

In a recent review by Wong et al. (2022) all versions of the WHODAS 2.0 appeared to have adequate content validity, structural validity, internal consistency and reliability among people with LBP. However, the construct validity of both 36-item and 12-item versions was found to be indeterminate.

2.2.7 Psychometric properties of 12-item self-administered World Health Organization Disability Assessment Schedule 2.0 among people with chronic low back pain

Even though the WHODAS 2.0 has widely been evaluated among people with chronic musculoskeletal pain, only few previous studies have focused on the use of the WHODAS 2.0 among people with LBP particularly. Additionally, to the studies identified through the aforementioned conducted systematic review, some of the studies targeting primarily LBP should be mentioned.

In studying psychometric properties of 36-item interview-administered WHODAS 2.0 among Polish chronic low back pain patients, Ćwirlej-Sozańska et al.

(2020a) findings supported the use of WHODAS 2.0 as a valid endpoint in clinical trials. Also due to WHODAS 2.0 high feasibility its use was highly recommended in managing people with disability caused by LBP.

Although using a small sample size in their cross-cultural validation study of Nigerian Igbo version of 36-item interview-administered WHODAS 2.0 among chronic low back pain patients, Igwesi-Chidobe et al. (2020) noted WHODAS 2.0 to have good internal consistency. It was also noted to correlate moderately with performance-based disability, self-reported back pain-specific disability and pain intensity. Finally, the authors underlined the need for in-depth test item and factor structure studies of WHODAS 2.0 in the future.

In validating Indonesian version of the interview-administered WHODAS 2.0 among back pain patients, Rijanti et al. (2021) concluded the WHODAS 2.0 to be valid and reliable questionnaire in measuring disability among back pain patients. The total scores of WHODAS 2.0 were found to correlate with the total Oswestry Disability Index (ODI) score. Some problems of inconsistency by respondents in interpretating some WHODAS 2.0 questions were found, more guidance was needed in those cases. The specific item-version of the WHODAS 2.0 was not mentioned.

Similarly, in validating Korean version of WHODAS 2.0 12-item self-administered version, Hae-Jung and Da-Jeong (2011) conclude the measure to be reliable and valid measure for the disability among patients with low back and/or neck pain. Also the authors (Hae-Jung et al., 2011) found high level of internal consistency for WHODAS 2.0 and strong correlation (r=0.77) was observed too between WHODAS 2.0 and perceived disability measured by Functional Rating Index (Feise & Michael Menke, 2001).

In examining the relationship between health problems rated according Norwegian form of the Core Set for LBP and ODI and WHODAS II, Roe et al. (2008) found the Comprehensive ICF Core Set for LBP covering all of the items in ODI and ODI's all items linked to ICF. Also, ICF categories were found to be linked to several items in ODI and WHODAS II without complete overlapping with each others.

Chwastiak et al. (2003) found the 36-item interview-administered WHODAS 2.0 to be valid and reliable, including good internal consistency and convergent validity with some health status measures, among back pain and depression patients, its usefulness as health status instrument for measuring disability in primary care setting was emphasized too. Researchers underline the importance of independency of symptoms (pain and depression), activity limitations and participation and not seeing them as interchangeable although them being related in their study. Also, the total core of WHODAS 2.0 was found to be an accurate summary of subscale scores.

In studying generic self-report and performance based instruments of disability, depression and physical activity among patient with and without pain, including

LBP, Silva et al. (2016) found 12-item interview-administered version of WHODAS 2.0 covering the pain in association of disability, regardless of the pain location. Also, they recommended the use of WHODAS 2.0 as a routine measure in pain patients.

While the evidence suggest that the WHODAS 2.0 might be of great importance from points of views – clinical usefulness and scientific significance, the evidence is too scarce to make definite strong recommendations for the practical use of the WHODAS 2.0 among people with LBP. Additionally, it should be considered that even though the WHODAS 2.0 is a general scale, its practicality should be confirmed in different settings and translations. Even more, there is a great diversity in the WHODAS 2.0 versions used by previous research. In this particular case, the worth of the Finnish version of the WHODAS 2.0 should be re-checked without leaning on the assumption of its usefulness based on the data obtained from the research conducted in different settings using different translations.

2.2.8 Psychometric properties of 12-item self-administered WHODAS 2.0 among general population and people with non-acute physical causes of disability – a systematic review (Study IV)

To evaluate the psychometric properties of the 12-item self-administered WHODAS 2.0, we conducted a systematic review (Saltychev et al., 2021). Previous research on the topic have been predominated by studies with small sample sizes. The studies have often not reported exact version of WHODAS 2.0 describing the scale in general as "12-item WHODAS 2.0". The aim of the review was to explore the psychometric properties of the 12-item WHODAS 2.0 among both a general population and people with non-acute physical causes of disability.

2.2.8.1 Methods

The inclusion and exclusion criteria are shown in Table 3. Medline, Embase, Web of Science, Scopus, and PsycINFO databases were used as data sources. In order to avoid missing any potentially relevant studies, the search clauses were left as generic as possible, and the refining search was conducted manually. The references of identified articles and reviews were also checked for relevancy. The search strategy is presented in Table 4. The main outcomes were the psychometric properties of WHODAS 2.0 understood as any property of WHODAS 2.0 related to its validity, reliability, responsiveness, MCID, MDC or respective.

Table 3. Inclusion and exclusion criteria of the review. Modified from Original Publication IV.

Inclusion criteria	Papers (including short communications and letters to the editor, excluding conference proceedings, theses etc.) published in academic peer-reviewed journals. No restrictions on the time of publication or language.
Exclusion criteria	Major psychiatric diagnoses, acute traumas, other acute conditions (e.g., postpartum or pregnancy), hearing loss, progressive neurological disorders, age of <19 years.
Outcome	Psychometric properties of WHODAS 2.0 understood as any property of WHODAS 2.0 related to its validity, reliability, responsiveness, MCID or minimal detectable change.

Table 4. Search strategy of the review. Modified from Original Publication IV.

Databases	***Search clauses and filters***
Medline (PubMed)	(whodas [TI] OR "World Health Organization Disability Assessment Schedule" [TI] OR “who-das” [TI] OR “who das” [TI]) AND (“12” OR “twelve”) AND (hasabstract[text])
Embase	(whodas:ti OR "World Health Organization Disability Assessment Schedule":ti OR “who-das”:ti OR “who das”:ti) AND (“12” OR “twelve”)
Scopus	(ALL(("12" OR "twelve"))) AND (TITLE((whodas OR "World Health Organization Disability Assessment Schedule" OR "who-das" OR "who das"))) AND (LIMIT-TO(DOCTYPE,"ar") OR LIMIT-TO(DOCTYPE,"le") OR LIMIT-TO(DOCTYPE,"no")) AND (LIMIT-TO(SRCTYPE , "j"))
Web of Science	(TITLE: (((wholes OR "World Health Organization Disability Assessment Schedule") OR "who-das") OR "who das") AND ALL FIELDS: ("12" OR "twelve")) Timespan: All years. Indexes: SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH, BKCI-S, BKCI-SSH, ESCI, CCR-EXPANDED, IC. Refined by: DOCUMENT TYPES: (ARTICLE)
PsycINFO	TI ((whodas OR "World Health Organization Disability Assessment Schedule" OR “who-das” OR “who das”)) AND TX ((“12” OR “twelve”)) Source type: Academic Journals

After the search in January 2019, duplicates or multiple copies of the search results were removed using the Endnote® software. During the screening phase, two independent reviewer teams first checked all the search results according to the titles and abstracts of the articles, and then conducted a full text assessment. Disagreements between the reviewers were resolved by consensus or by a third reviewer. The screening process is shown as a PRISMA chart (Figure 1). The methodological quality of the included studies was not rated. One reviewer extracted all the relevant data from the selected articles using a form including the

title, first author, year of publication, country of origin, study settings, main diagnoses of participants if specified, sample size, gender distribution, participants' age, main psychometric measures, main quantitative results, and conclusions made the authors of original publications.

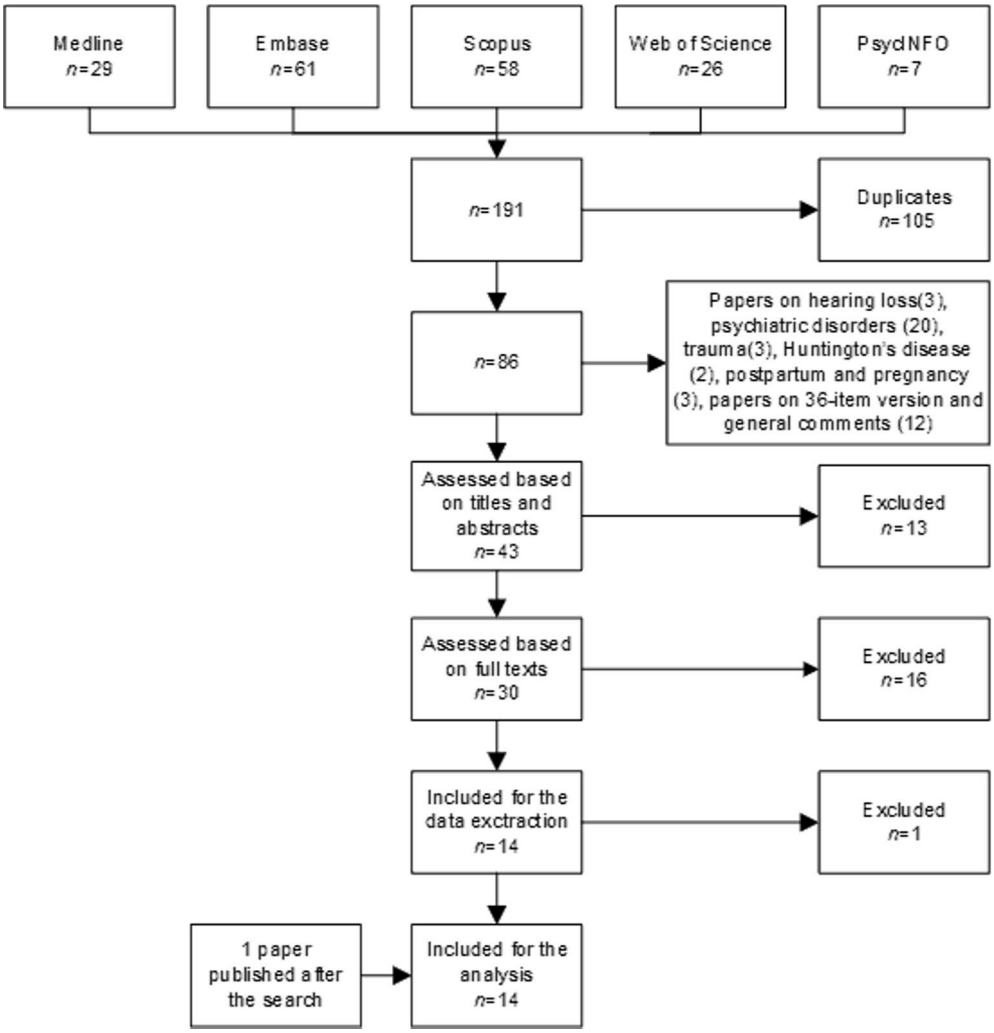

Figure 1. Search flow. From Original Publication IV.

2.2.8.2 Results

2.2.8.2.1 Search results

The search resulted in 191 records. Of them, 148 were excluded as duplicates and papers on hearing loss, psychiatric disorders, trauma, Huntington disease, postpartum and pregnancy, papers on 36-item version, and general commentaries. The remaining 43 records were screened based on their titles and abstracts and 13 irrelevant papers were excluded. The number of observed agreements between the reviewers was 30 (70% of the observations) and kappa was 0.23 (SE 0.16, 95% CI −0.09 to 0.55) considering the strength of agreement between the reviewers to be "fair." Thirty records were assessed based on their full-texts and 16 irrelevant papers were excluded comprising 14 relevant studies potentially fit for a qualitative analysis (Figure 1). Additionally, one study that was published after the search was considered relevant into further analysis (Katajapuu et al., 2019a). The attempt to extract relevant data regarding a 12-item self-administered WHODAS 2.0 version from the report by Tazaki et al. (2014) was unsuccessful and that study was excluded from further analysis. Tazaki et al. (2014) employed five different versions of WHODAS 2.0 (12- and 36-item interviewer-administered, 36-item proxy-administered, and 12- and 36-item self-administered versions) and there was a discrepancy in reporting a sample size (total n=126 but 62 men and 70 women). After the selection and data extraction phases, 14 records were included into further analysis.

2.2.8.2.2 Studied samples

All of the 14 remaining papers were published after 2013 (Table 5). All of them were observational studies. Two studies focused on the elderly (Galli et al., 2018; Silva et al., 2019) while the rest evaluated people of working age. The sizes of samples varied from 80 up to 31,251 participants. Except for one study with 98% women (Smedema et al., 2016), the proportions of female participants were between 47% and 65%. A great diversity was observed in the participants' health problems: patients waiting for an elective joint arthroplasty or neurosurgery (Schiavolin et al., 2014; Galli et al., 2018), general population or healthy volunteers (Kimber et al., 2015; Gaskin et al., 2017; Marom et al., 2017; Silva et al., 2019), patients with chronic musculoskeletal pain or fibromyalgia (Smedema et al., 2016; Saltychev et al., 2017a; Saltychev et al., 2017b), patients with spinal cord injury (Smedema et al., 2017; Tarvonen-Schröder et al., 2019), and people reimbursed for any disabilities (Xenouli et al., 2016) .

2.2.8.2.3 Psychometric properties

The most common psychometric properties reported by the included studies were Cronbach's alpha and convergent validity. The alpha estimates were usually high varying from 0.81 up to 0.96. Any pooling of the reported concurrent validity estimates was impossible as each study compared WHODAS 2.0 with different scales. However, the reported correlations between WHODAS 2.0 and other disability scales applied at the same time with WHODAS 2.0 were high in most of the studies.

Floor and ceiling effects were reported by three studies. One study (the biggest sample size of 31,251) reported a substantial floor effect up to 32% (on average 20%) without a ceiling effect (Gaskin et al., 2017). Another study conducted on a sample of 183 participants did not observe any floor or ceiling effects (Schiavolin et al., 2014) . Moreover, in that study, none of the participants—patients waiting for a neurosurgical procedure—reported a highest or lowest WHODAS 2.0 scores. The third study reported a significant floor effect up to 80% for all 12 items and for a total score without a ceiling effect (Katajapuu et al., 2019a).

Exploratory factor analysis or principal component analysis were employed by six studies (Smedema et al., 2016; Xenouli et al., 2016; Smedema et al., 2017; Gaskin et al., 2017; Saltychev et al., 2017a; Saltychev et al., 2017b). None of them reported a unidimensional structure of 12-item WHODAS 2.0. The number of factors varied from two up to five. Four studies employed confirmatory factor analysis. Only one of them reported a good model fit (Kimber et al., 2015). In one study, the hierarchical model with one common factor and six subfactors (as suggested by WHODAS 2.0 developers for 36-item version) was assessed resulting in poor fit (Smedema et al., 2017). One study reported a good fit of one-factor model, but the reported root mean square error of approximation (RMSEA) was insignificant 0.079 pointing at a poor fit (Schiavolin et al., 2014). Another study reported a good fit of a two-factor model (Saltychev et al., 2017b). In one study, the discriminative ability was assessed using Karnofsky Performance Status scale as indicator of disability (Schiavolin et al., 2014) reporting positive results. Another study assessed the discrimination ability using the item response theory (Saltychev et al., 2017a). That study reported discrimination of WHODAS 2.0 items being high to perfect, even though, the difficulty of items was shifted towards elevated disability rates. Such a shift implicates that a respondent should be experiencing slightly worse disability (compared with the average population rate) to achieve a 50/50 probability of giving an answer that would be interpreted by the WHODAS 2.0 as a "worse disability." Three studies assessed test–retest reliability of the 12-item WHODAS 2.0 (Moreira et al., 2015; Marom et al., 2017; Silva et al., 2019) reporting insignificant differences between repeated measures. In all three studies, the time interval between measures was 1 week.

Table 5. Basic characteristics of included studies. Modified from Original Publication IV.

Author and year	Country	Settings, participants, and main diagnoses	Sample size	Women	Age, mean (standard deviation), years	Psychometric properties
Galli et al. (2018)	Italy	Patients hospitalized for elective hip (60%) or knee (40%) arthroplasty (3 hospitals). After-surgery estimates excluded.	80	67%	70.1 (1.1)	Convergent validity with modified Barthel index: 0.335
Gaskin et al. (2017)	Australia	SAGA data[1] - general population >=50 years	31,251	54%	63.4 (9.5)	EFA[5]: 1 to 3 factors (mostly 2 or 3). Floor effect: 6% to 32% (overall 20%). Ceiling effect: none.
Kimber et al. (2015)	Canada	CCHS-MH data[2] - general population >=15-year youth group excluded.	23,798	51%	47.1 (0.2)	Alpha: 0.95 (95% CI 0.94 to 0.96) CFA[6] (assuming 1/6-factor structure).
Marom et al. (2017)	Israel	Volunteers - general working population. Group with acute trauma excluded.	155	51%	43.1 (15.0)	Alpha: 0.85. Reported including patients with acute trauma: Test/retest (1 week): ICC[7] 0.88 (95% CI 0.83 to 0.91). Convergent validity: PCS-12[8] -0.46 (95% CI -0.67 to -0.15), MCS-15[9] -0.62 (95% CI -0.78 to -0.36), QDASH[10] 0.53 (95% CI 0.33 to 0.69).
Moreira et al. (2015)	Portugal	General population using community support services.	144	64%	64.0 (6.7)	Alpha: 0.86. Test/retest (1 week): ICC 0.77(95% CI 0.69 to 0.83). Convergent validity: Barthel index: -0.27, LSNS[11] -0.19.
Saltychev et al. (2017a)[3]	Finland	University outpatient clinic. Chronic non-specific musculoskeletal pain.	501	65%	47.1 (13.9)	EFA: 2 factors. IRT[12]: discrimination - high to perfect for all items' difficulty - slight shift towards elevated disability rates.
Saltychev et al. (2017b)[4]	Finland	University outpatient clinic. Chronic non-specific musculoskeletal pain.	408	65%	47.0 (13.7)	EFA: 2 factors. CFA: 2-factor assumption.
Schiavolin et al. (2014)	Italy	Patients scheduled for different neurosurgical surgery.	183	50%	51.1 (13.1)	CFA: assuming 1-factor structure (insignificant RMSEA 0.079). Alpha 0.875. Convergent validity: EUROHIS-QOL[13] -0.52, PGWBI-S[14] -0.52. Discriminative validity: significant difference between KPS[15]>90 and KPS=<90. Floor and ceiling effects: none (0%).
Silva et al. (2019)	Portugal	Day Care Centers and Nursing Homes. Possibly including interviewer-administered version of WHODAS 2.0.	100	62%	82.3 (8.1)	Convergent validity: GST[16] –0.57 to –0.62, FTSTS[17] 0.41, TUG[18] 0.32 to 0.37. Test/retest (1 week): *p* value 0.32

Author and year	Country	Settings, participants, and main diagnoses	Sample size	Women	Age, mean (standard deviation), years	Psychometric properties
Smedema et al. (2017)	USA	Online survey. Patients with spinal cord injury.	247	50%	41.6 (12.4)	Alpha: 0.82. CFA: hierarchical and 1-factor with poor fit. EFA: 3 factors. Convergent validity reported for each of 3 factors separately. Convergent validity: SWLS[19] -0.16 to -0.36, CSES[20] -0.05 to -0.56, IPA[21] -0.16 to -0.47, SF-20[22] 0.0 to -0.62.
Smedema et al. (2016)	USA	Online survey. Patients with self-reported fibromyalgia.	302	98%	48.4 (10.4)	PCA[23]: 2 factors. 1st factor: alpha: 0.81; convergent validity: BFI[24] 0.35, pain intensity 0.28, MOS-Sleep[25] 0.33, GFQ[26] 0.53, CESD-10[27] 0.63, MSPSS[28] -0.37. 2nd factor: alpha 0.83; convergent validity BFI 0.43, pain intensity 0.43, MOS-Sleep 0.43, CFQ 0.31, CESD-10 0.42, MSPSS -0.20.
Tarvonen-Schröder et al. (2019)	Finland	University outpatient clinic. Patients with spinal cord injury.	142	47%	56.7 (16.9)	Alpha: 0.86. Convergent validity:7-item World Health Organization (WHO) minimal generic set 0.49.
Xenouli et al. (2016)	Greece	People without (A) and with (B) disabilities	109/101	65% / 63%	46.3 (13.0) / 51.5 (18.4)	EFA: group A – 5 factors, group B – 4 factors. Groups A + B: Alpha 0.85. Convergent validity: SF-PCS[29] -0.76, SF-MCS[30] -0.50, PSS-14[31] 0.55
Katajapuu et al. (2019a) [5]	Finland	University outpatient clinic. Chronic non-specific musculoskeletal pain.	1988	65%	47.6 (15.0)	Floor effect: 15% to 79%. Ceiling effect: none.

[1] World Health Organization's longitudinal Study on global ageing and adult health (6 countries); [2] Canadian Community Health Survey-Mental Health; [3] Subpopulation of Katajapuu et al. (2019a); [4] Subpopulation of Katajapuu et al. (2019a) ; [5] Exploratory factor analysis; [6] Confirmatory factor analysis; [7] Intraclass correlation coefficient; [8] Physical composite scores; [9] Mental composite scores; [10] Quick Disability of Arm, Shoulder, and Hand Outcome Measure; [11] Lubben Social Network Scale; [12] Item response theory analysis; [13] European Health Interview Survey-Quality of Life; [14] Psychological General Well-Being Index-Short; [15] Karnofsky Performance Status; [16] Gait speed test; [17] FTSST: Five-times-sit-to-stand-test; [18] Time Up & Go Test; [19] Satisfaction with Life Scale; [20] Core Self-Evaluations Scale; [21] Work and Education subscale of the Impact on Participation and Autonomy Questionnaire; [22] Medical Outcomes Study 20-Item Short-form Health Survey; [23] Principal Component Analysis; [24] Brief Fatigue Inventory; [25] Medical Outcomes Study – Sleep Scale; [26] Cognitive Failures Questionnaire; [27] Center for Epidemiological Studies Depression Scale-Short Form; [28] Multidimensional Scale of Perceived Social Support; [29] Short-form 36 Brief Physical Health Scale; [30] Short-form 36 Brief Mental Health Scale; [31] Perceived Stress Scale

2.2.8.2.4 Review conclusions

Based on the results of 14 observational studies, the self-administered 12-item WHODAS 2.0 was found to be an internally consistent and test-retest reliable questionnaire. The samples of the reviewed studies have varied from healthy adults to people suffering from chronic musculoskeletal pain. The WHODAS 2.0 did not show any significant ceiling effect. Opposingly, several studies have reported a substantial floor effect. The WHODAS 2.0 was able to discriminate especially well people with moderate or severe disability, but it did not perform equally well among people at the low end of the distribution of disability severity. Additionally, the analysis of data on the psychometrics of the WHODAS 2.0 pointed out to a possible inconsistency existed between the severity of disability people are experiencing in reality and the WHODAS 2.0 score – respondents might be tended to underestimate their functional limitations. While other properties of the WHODAS 2.0 have overall been reported as excellent, the factor structure of the scale has been questioned by several studies reporting the multidimensionality of the WHODAS 2.0 instead of a desired unidimensionality. A potential multidimensionality in the factor structure of the WHODAS 2.0 might weaken the value of a composite score produced by the WHODAS 2.0 as the different domains of functioning might contribute to that score differently while resulting in the same composite score (Ravaud et al., 1999).

The generalizability of the results of the review might be affected by a substantial heterogeneity seen in the studied populations, sample sizes and overall small number of the identified studies on the topic. When evaluating the convergent validity of the WHODAS 2.0, the reviewed studies have employed the very wide spectrum of scales. Combined with the overall smallness of study samples, this made a reliable pooling of the results concerning convergent validity of the WHODAS 2.0 impossible. The great diversity between designs, samples, settings and scales made the assessment of potential risk of systematic bias impossible as well.

While none of previous reviews on the subject has focused on a 12-item version of the WHODAS 2.0, the present results are somehow comparable with an international systematic review on the WHODAS 2.0 conducted in 2016 (Federici et al., 2017). Also, that review has observed the good convergent validity of the WHODAS 2.0 compared to other scales of functioning. The present review, however, could not confirmed the unidimensional structure of the WHODAS 2.0 reported by a previous review. This difference might be explained by the differences between selection criteria used in these two studies – only 12-item version vs. all possible versions and special focus on chronic musculoskeletal conditions. It is possible, and even expected, that different versions of

WHODAS 2.0 may produce different patterns in their factor structures especially when applied to more or less diverse populations. The differences between versions might result in dissimilar responses especially when comparing a self-rated form

with those filled out by professionals or proxies. Respondents may overstate or downgrade their responses in order to avoid some uncomfortable facts, guessing the expectations of professionals coding the given responses, or even trying to achieve some profitable outcome when applying for some social or insurance benefits. In turn, when filling interviewer- or proxy-based forms, a subjective component related to the interviewers could hardly been avoided. Additionally, many papers selected for the present review have been published after the previous systematic review, which has been submitted to a journal in the spring of 2016.

2.3 The Oswestry Disability Index

The ODI, a self-administered, free, easy to use, questionnaire (Fairbank et al., 1980) was developed at the end of the 1970s. It has since been widely used in research and clinical practice and is considered a gold standard in the field of spinal disorders. The ODI was revised in 2000, and the most recent version is the ODI_2.1b (Fairbank & Pynsent, 2000; Eprovide, 2022). According to Fairbank et al. (2000), it was agreed that the ODI would adopt the concept of disability defined by WHO, focusing on limitations of physical activities caused by LBP. The ODI has been translated and validated in 80 different languages (Eprovide, 2022). The ODI was translated into Finnish following the protocol of cross-cultural adaptation by Pekkanen et al. (2011). The methodological quality of cross-cultural validations of ODI has previously been criticized (Yao et al., 2016), although Sheahan et al. (2015) earlier found culturally adapted versions of ODIs to mostly have high psychometric properties. Over the years, the properties of the ODI have been extensively studied (Fairbank et al., 2000).

Considering the exceptionally broad use of ODI, the knowledge of its psychometrics is surprisingly scarce and inconsistent. There have been several studies on the internal consistency of the ODI and its factor structure, but only a few have used Rasch analysis or item response theory (IRT) to distinguish individuals with different levels on scales (White et al., 2002; Davidson, 2008).

Construct structure of ODI have been valued for its excellent psychometric properties some previous studies (Lauridsen et al., 2006; Pekkanen et al., 2011; Sheahan et al., 2015). Some other studies have been more careful in recommending ODI use (Davidson, 2008; Dawson et al., 2010; Lochhead & MacMillan, 2013). The internal consistency of ODI has been usually considered good. One and two factor structure models have been studied in order to explain factor structure of ODI (Chow & Chan, 2005; Monticone et al., 2009; Valasek et al., 2013; Algarni et al., 2014; van Hooff et al., 2015). There is no full agreement on ODI's factor structure, a latent variable or variables may co-exist with disability. Variation with results may be related to differences in small sample sizes. Only one previous study on ODI validity has been carried out on a large sample using the method of Rasch analysis, Brodke

et al. (2017) were not able to unquestionably confirm the unidimensionality of ODI. Additionally, although they reported good person and item reliability of the ODI, large floor and small ceiling effects were found, and this may have an effect interpreting results of scores. These observed limitations in the study may be related due the fact that study setting was a single-center study with relatively homogenous cohort of patients in terms of their demographics (Brodke et al., 2017). Authors also concluded that further study on evaluating extrinsic validity should be done.

2.4 Numeric Rating Scale

Very often, an 11-point NRS is an outcome measure of choice for LBP, due to its easiness and overwhelmingly widespread use (Chapman et al., 2011; Chiarotto et al., 2016; Froud et al., 2016; Chiarotto et al., 2019).

An NRS has been shown to be associated with the ODI among LBP patients even if the correlations may vary from weak to strong (Vigatto et al., 2007; Solomon et al., 2011; Hong & Shin, 2020). Also, an NRS has shown a low to moderate positive association with the RMDQ (Grotle et al., 2004; Moon et al., 2011; Asghari, 2011; Monticone et al., 2012; Raheem et al., 2021). Hush et al. (2010) noted that neither the RMDQ nor an NRS seemed to cover the whole complex personal experience of LBP, and may, therefore, lead to incorrect conclusions regarding a patient's condition. The MCID for pain NRS was established at 2.0–2.5 points (Farrar et al., 2001; Childs et al., 2005; Ostelo & de Vet, 2005; Maughan & Lewis, 2010; Suzuki et al., 2020). According to Shafshak and Elnemr (2021), a NRS and a pain visual analogue scale (VAS) correlated strongly and positively with each other, and both were able to predict disability among people with LBP. It has been suggested that such single-item scales as NRS or VAS might be insufficiently precise due to the difficulties experienced by respondents when quantifying such a complex concept as pain (Paungmali et al., 2012; Robinson-Papp et al., 2015; Chiarotto et al., 2019)

The summary of reliability and validity of the 12-item WHODAS 2.0, the ODI and a NRS in LBP is presented in Table 6.

Table 6. Common self-reported outcome measures of function, disability or pain on LBP research, their reliability and validity.

Measurement	Reliability and validity	Benefits	Limitations
Oswestry Disability Index (ODI) (Fairbank et al., 1980) ★★★[a]	Reliability: - Test/Retest reliability: value for one day (0.94-0.99) (Fairbank et al., 1980; Payares et al., 2011) - Inter-rater reliability high (0.94) and internal consistency 0.84-0.90 (Payares et al., 2011) - Test/retest reliability has been assessed for excellent (sections of walking, sleep and total score) adequate (sections of pain intensity, lifting, sitting, standing, social life and travel) and poor (section of sex) (Dawson et al., 2010) Validity - Content and face validity; Adequate content validity is reported due to covering common LBP patients' activities of daily living - High internal consistency (Cronbach's α = 0.71-0.87), construct validity has been appraised excellent (Lauridsen et al., 2006; Pekkanen et al., 2011; Sheahan et al., 2015) - External construct and convergent validity; Correlation with other disability measures including RMDQ (correlation coefficient = 0.5) and also with pain scales (Fairbank et al., 2000; Roland & Fairbank, 2000)	Considered as a gold standard for measuring disability for adults with LBP Simple to use, it can be completed fast, similarly scored and also possible over the phone Good construct validity and internal consistency, test–retest reliability high over short intervals Good correlation with other tools like VAS Encouraged to be used also in association with pain-measuring unidimensional measures like NRS or VAS (Garg et al., 2020)	Poor test–retest reliability over long intervals. Some reports of being not so sound psychometrically(Garg et al., 2020) May be a multidimensional measure instead of one-factor-structure (Lee et al., 2017) In content validity misses lacks some common activities including work, leisure, recreation or sporting activities (Smeets et al., 2011)

Measurement	Reliability and validity	Benefits	Limitations
Visual Analogue Scale (VAS) (Huskisson, 1974) ★★[a]	Reliability - Test-retest reliability has been studied sparsely; acceptable test-retest reliability has been suggested (Paungmali et al., 2012) Validity - In studying content validity, difficulties noted in measuring pain, in defining “right” experiences and average pain (Robinson-Papp et al., 2015) - Structural validity not studied due to NRS being single-item scale and no studies available for cross-cultural validity (Chiarotto et al., 2019)	Fast, easy to use and score in research and clinical use	Patients may have difficulties in quantifying the pain (Robinson-Papp et al., 2015)
Numeric or numerical rating scale (NRS) (Downie et al., 1978) ★★[a]	Reliability - Single measurement is reliable and valid enough for basic research purposes (Jensen & McFarland, 1993) Validity - Structural validity not studied due to NRS being single-item scale and no studies available for cross-cultural validity (Chiarotto et al., 2019) - In studying content validity, NRS was suggested to be somewhat inadequate outcome measure among adults with the persistent LBP(Hush et al., 2010; Robinson-Papp et al., 2015)	Fast, easy to use and score in research and clinical use	May include a downward “response shift” over the time; Misses the dimension of personal experience of back pain (Hush et al., 2010) Reliability and validity may benefit from series of measurements over several days instead one single measurement (Jensen et al., 1993) Patients may have difficulties in quantifying the pain (Robinson-Papp et al., 2015) NRS may have difficulties capturing complexity of patient’s personal experience of pain (Hush et al., 2010)

Measurement	Reliability and validity	Benefits	Limitations
12-item self-administered WHODAS 2.0 (World Health Organization, 2010) ⋆[a]	Reliability - An adequate reliability and internal consistency of short versions of WHODAS 2.0 has been reported (Wong et al., 2022) Validity - An adequate structural validity of short versions of WHODAS 2.0 has been reported; Construct validity of the short versions of WHODAS 2.0 in LBP patient has been reported to be indeterminate (Wong et al., 2022) -	Short and easy to use; Useful in studying population with health problems (Ćwirlej-Sozańska et al., 2020b) Recommended to be used in measuring functioning in persons with LBP in variety of cultures and settings (Wong et al., 2022)	Limited studies of construct validations

[a] Psychometrics studied: ⋆= limitedly studied, ⋆⋆= moderately studied, ⋆⋆⋆=extensively studied

3 Aims of the Present Study

The aim of this study was to explore the convergent validity of the 12-item WHODAS 2.0 among people with chronic musculoskeletal pain. This study was part of the Turku ICF project.

The aims of this study were:

1. To assess the validity of the Finnish translation of the 12-item self-administered WHODAS 2.0. (Study I)
2. To investigate the psychometric properties of the ODI in a large cross-sectional cohort of individuals with chronic low back pain by defining its internal consistency, construct structure and validity, and its ability to discriminate between different degrees of functional limitation. (Study II)
3. To determine whether pain severity measured on an NRS is associated with restrictions in functioning, measured using the 12-item self-administered WHODAS 2.0, of chronic musculoskeletal-pain patients. (Study III)
4. To explore the available evidence on the psychometric properties of the 12-item self-administered WHODAS 2.0 among the general population and people with non-acute physical causes of disability. (Study IV)
5. To study whether the 12-item self-administered WHODAS 2.0 and the ODI are associated, to justify their interchangeability. (Study V)

In addition, this thesis discusses whether the results are in accordance with a priori set hypotheses:

1. The 12-item self-administered WHODAS 2.0 is an internally consistent scale identifying a single underlying latent trait, a “general disability factor”, in people with chronic musculoskeletal pain. (Study I)
2. The WHODAS 2.0 is able to distinguish different levels of functioning among people with chronic musculoskeletal pain. (Study I)
3. The ODI is an internally consistent scale identifying a single underlying latent trait, a “general disability factor”, in people with LBP. (Study II)

4. The ODI is able to distinguish different levels of functioning among people with LBP. (Study II)
5. Pain severity measured by using an NRS is positively and moderately or strongly associated with the 12-item self-administered WHODAS 2.0. (Study III)
6. The available evidence of the systematic review supports the use of the 12-item self-administered WHODAS 2.0 among the general population and among people with non-acute physical causes of LBP. (Study IV)
7. The 12-item self-administered WHODAS 2.0 is positively and moderately or strongly associated with the ODI in people with LBP. (Study V)

4 Materials and Methods

4.1 Study designs and participants in Studies I–III and V

The studies I–III and V employed the cross-sectional data from Turku ICF Study (Saltychev et al., 2013), which were collected between April 2015 and February 2017. Out of 3,150 patients who had visited Turku University Hospital's Physical and Rehabilitation Medicine Outpatient Clinic, 1,988 patients were included. All the patients have participated in the study voluntarily and all of them signed written informed consents. The study protocol and data collection were approved by the Ethics Committee of Hospital District of Southwest Finland (ETMK 60/180/2012).

The Data for the Turku ICF Study were collected through a patient-reported semi-structured questionnaire designed by the researchers. Patients received the questionnaire by mail prior to their first clinic visit and were required to fill it in before the physician's appointment. In addition to study-specific parts, the questionnaire also included information on the study, the patient's rights, and the written consent form.

The questionnaire consisted of several different parts, including a general part with questions on general health status, education and work history, medications, past medical history, living conditions, leisure time physical activities, anthropometrics (weight, height), and open questions on the symptom or symptoms, patient expectations of the visit, the origins of symptoms or pain, and their personal opinions on the possible origin of and changes in symptoms or pain. Also, the questionnaire contained pain drawings.

Patients responded to several PROMs, including (in the order they were presented in the questionnaire): a pain NRS, the 12-item self-administered WHODAS 2.0, the ODI 2.1a, and the Neck Disability Index (NDI).

The characteristics of the study samples are summarized in Tables 7 and 8 including the distributions of main diagnoses grouped by individual studies used by this thesis.

Table 7. Characteristics of population and samples in Studies I–III and V. Modified from Original Publications I-III, V.

Variable	Total	Women	Men	*P*-value
Total population, N	1,988	1,297/65%	691/35%	
Demographics				
Age, years, mean (SD)	47,6 (6.3)	47,5(15.1)	47,6(14.7)	0.927
High School, *n* (%)	1,258 (67)	473 (39)	136 (21)	<0.001
No high school, *n* (%)	609 (33)	743 (61)	515 (79)	<0.001
Health status				
BMI (kg/m^2), mean (SD)	27.4 (5.7)	27.0 (6.0)	28.2 (5.1)	<0.001
Pain, points 0–10, mean (SD)	6.3 (2.0)	6.4 (1.9)	6.2 (2.0)	0.005
Diagnoses				
Dorsalgia, n (%)	781 (39)			
Other soft tissue disorders, *n* (%)	202 (10)			
Functioning				
WHODAS 2.0 12-item total score[a], mean (SD), median (IQR)	13.1 (9.4), 12 (6,19)	13.1	13.0	0.843
Study I, n	501	325/64.9%	176/35.1%	
Age, years, mean (SD, range)	47.1(13.9, 16 to 84)			
Health status				
BMI (kg/m^2)[a], median(IQR, range)	26.5 (23.5 to 29.8, 15.6 to 84)			
Functioning				
WHODAS 2.0 12-item total score[a], median (IQR, range)	11.0 (6.0 to 18.0, 0 to 43)			
Study II, n	837	550/66%	287/34%	
Age, years, mean (SD)	47.6 (14.8)			
Functioning				
ODI, total score, median (IQR, range)	32 (22 to 44, 0 to 80)			
Study III, n	1207	798/66%	409/34%	
Age, years, mean (SD, range)	47.1 (14.5)			
No professional education, *n* (%)	173 (15)			
Married/cohabiting, *n* (%)[d]	836 (74)			
Single, *n* (%)[d]	289 (26)			
Functioning				
WHODAS 2.0 12-item total score[a], median (IQR, range)	25 (13 to 40, 0 to 79)			
Study V, n	1,379	881/64%	498/36%	
Age, years, mean (SD)	47.6 (14.9)			
Married/cohabiting, *n* (%)	588(73.6)			
High school education, *n* (%)	398 (30)			
Health status				
BMI (kg/m^2), mean (SD)	29.4 (76.4)			
Pain severity (0–10), mean (SD)	6.4 (1.9)			
ODI, total score, mean (SD)	34 (16.4)			
WHODAS 2.0 12-item total score[a], mean (SD)	14.4 (9.4)			

SD= Standard deviation, IQR = Interquartile range, [a]Simple scoring method

Table 8. Main diagnoses for visiting Physical and Rehabilitation Medicine Outpatient Clinic. Modified from Original Publications I-III, V.

ICD-10 diagnosis	n	%
Study I		
M54 Back pain	179	36.0
M51 Intervertebral disc disorder	55	11.0
M79 Soft tissue disorder	40	8.1
M75 Shoulder lesion	35	7.0
M50 Cervical disc disorder	23	4.6
M25 Joint disorder	20	4.0
M70 Soft tissue disorder related to use, overuse, and pressure	13	2.6
M43 Deforming dorsopathy	12	2.4
M53 Dorsopathy	11	2.2
G54 Nerve root and plexus disorder	7	1.4
M48 Spondylopathy	7	1.4
M47 Spondylosis	6	1.2
Other	89	17.9
Missing	4	0.8
Total	501	100
Study II		
M54 Dorsalgia	411	49.1
M51 Thoracic, thoracolumbar, and lumbosacral intervertebral disc disorders	93	11.1
M79 Other and unspecified soft tissue disorders, not elsewhere classified	53	6.3
M48 Other spondylopathies	25	3.0
M43 Other deforming dorsopathies	16	1.9
M47 Spondylosis 12 (1)	12	1.4
M53 Other and unspecified dorsopathies, not elsewhere classified	10	1.2
Other	213	25.4
Missing	4	0.5
Total	837	100.0
Study III		
Domain M, ICD-10 – Diseases of the musculoskeletal system and connective tissue	1061	88.0
Other	146	12.0
Total	1207	100.0
Study V		
Domain M, ICD-10 – Diseases of the musculoskeletal system and connective tissue	1258	92.0
Other	139	8.0
Total	1379	100.0

The sample sizes in Studies I–III and V varied between 501 and 1,379. Of the total population sample, 65% were women. The average age of the respondents varied from 46.1 to 47.6 years, and their body mass index (BMI) varied from 26.5 to

29.4 kg/m^2. The study participants were mostly married or cohabiting, and two-thirds of the original population had at least a high school-level education.

The gender- and age-based distributions of the participants were similar for all four original studies. Women formed about 65% of the study samples, and average age varied only slightly around 47-48 years. The severity of limitations in functioning was also similar for all the studies varying from 11 to 14 points, as measured by the WHODAS 2.0. For all four samples, around 90% of the respondents had diagnosis of the "M" domain of ICD-10 "Diseases of the musculoskeletal system and connective tissue" as a primary diagnosis for visiting a clinic. Accordingly, the severity of pain varied from six to seven points for usual pain and eight to nine points for the worst pain.

Figure 2 presents the design and settings, methods of statistical analyses, and conclusions for all four original studies.

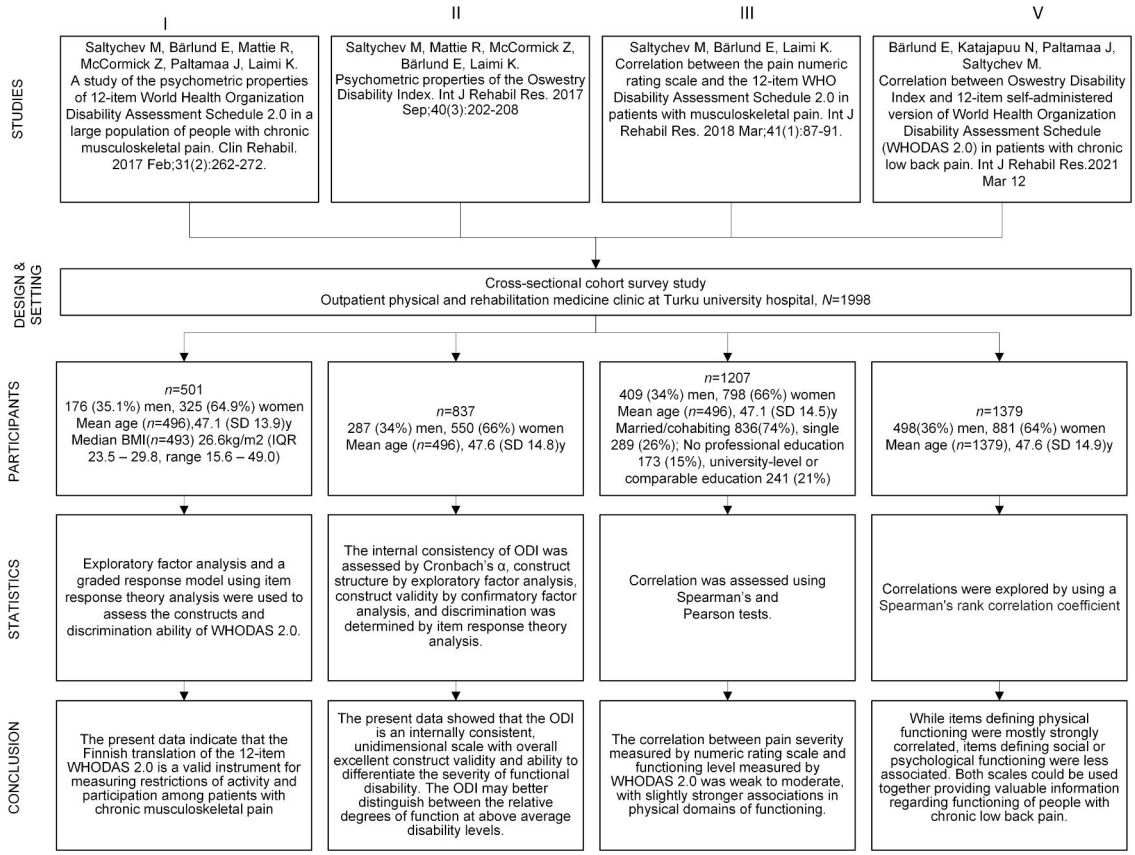

Figure 2. Short representation of workflow in Studies I–III, V. Modified from Original Publications I-III, V.

4.1.1 Variables in the original studies

The independent variables in the original studies were gender (Study IV), age (Studies I–III and V), educational level (Studies III and V), marital status (Study IV) and medical diagnosis (Studies I–III and V). Sex (man or woman) was classified according to personal identification numbers. Age was defined in full years at the time of visiting the clinic. Educational level was dichotomized as "high school", vs. "no high school". Marital status was dichotomized as single vs. married or cohabiting. Main diagnosis was determined by a physician according to the ICD-10.

The dependent variables were BMI (Studies I and IV), the 12-item self-administered WHODAS 2.0 (Studies I, III and V), the ODI (Studies II and V) and a pain NRS (Study V).

The 12-item self-administered WHODAS 2.0 covered 12 of the longer 36-item version's items, as follows:

1. Item: Standing for long periods,
2. Item: Taking care of household responsibilities,
3. Item: Learning a new task,
4. Item: Joining community tasks,
5. Item: Emotionally affected by health problems,
6. Item: Concentration on doing something for 10 minutes,
7. Item: Walking long distance such as 1km,
8. Item: Washing your whole body,
9. Item: Getting dressed,
10. Item: Dealing with people you do not know,
11. Item: Maintaining a friendship and
12. Item: Day-to-day work

The respondents used a five-point Likert-like scale in each of the items above to report the difficulties they had encountered due to their health conditions over the previous 30 days. The scale ratings were:

0. None
1. Mild limitation
2. Moderate limitation
3. Severe limitation
4. Extreme limitation or cannot do

A simple scoring was used. To calculate the total score, the levels were coded according to their numeric values of 0–4 and summed. The total score was the sum of all 12 items divided by the maximum possible total score (Equation 1). It described the degree of functional limitations.

$$Total\ score\ = \left(\frac{\sum \text{item scores}}{48}\right) \times\ 100 \quad (1)$$

BMI was defined as the patient-reported body mass in kilograms divided by the square of the body height in meters (Equation 2):

$$BMI\ =\ \frac{\text{Patient−reported bodymass in kilograms (kg)}}{\text{height in meters(m)}^2} \quad (2)$$

The ODI, which consist of ten sections covering pain, personal care, lifting, walking, sitting, standing, sleeping, sex life, social life, and traveling. For each section, the respondents chose from six statements reflecting disability without a defined recall period. The ODI uses a Likert-like scale from 0 (no limitation) to 5 (worst possible limitation).

The total score may vary from 0% to 100%, and the higher the score, the worse the disability. The percentage score was calculated by multiplying the sum of items' scores by 2 (Equation 3):

$$ODI\ score\ =\ \frac{\text{Total sum score of all answers}}{\text{Total possible score (50)}} \times 100 \quad (3)$$

The percentage scores were interpreted as follows:

0%–20% = minimal disability
21%–40% = moderate disability
41%–60% = severe disability
61%–80% = crippling back pain
81%–100% = bed-bound or exaggerated symptoms

If a missed answer was observed, it was not included in the total score and 5 was subtracted from the denominator for each missing answer per section (Equation 4):

$$ODI\ score = \frac{\text{Total sum score of all answers}}{\text{Total possible score (50)-5 for each missing answer in a section}} \times 100 \quad (4)$$

If a respondent had answered more than once in a section, the highest score was accepted. Scores were rounded up to the closest whole number (Fairbank et al., 2000).

The respondents also estimated their experience of pain, both average pain and worst pain over the last month using an NRS for pain. A horizontal 11-point scale of the numbers 0–10 was used, as follows:

☐ 0 ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ 10

0 denoted "no pain" and 10 the "worst pain ever".

4.2 Methods and statistical analyses

The main statistical method in the **Study I** was IRT, which was based on the results of the EFA conducted in prior to the IRT. The purpose of EFA was to evaluate the factor structure of the WHODAS 2.0 in order to confirm or to reject its multi- or unidimensionality. A conservative cut-off of >1.0 (Kaiser's rule) was chosen as a criterion for factor retaining.

Normally distributed data were presented as mean and standard deviations. Abnormally distributed or ordinal data were presented as medians along with interquartile ranges (IQR) and ranges. Two-tailed p-values were reported, considering values <0.05 to be statistically significant. The employed IRT graded-response model was a two-parameter model. The results were reported along with their 95% CIs. The EFA and the IRT were conducted using both numerical and graphical analyses.

In the IRT, the first parameter is called "difficulty", meaning here the severity of disability experienced by a person needed to achieve a particular response. E.g., how much disability in walking compared to the mean level observed in the sample should a person experience in order to mark grade "two" and not grade "one" etc. The scale used by the WHODAS 2.0 has five steps from zero to four. Thus, in an ideal situation, for each individual item, the difficulty of response close to the average level of disability (coefficient around 0) should be found near grade "two" – the middle point of a five-grade scale. If so, then disability more severe than the average in the sample should correspond to grades "three" and "four" (coefficients bearing a plus sign), while milder disability should correspond to grades "zero" and "one" (coefficients bearing a minus sign).

The second parameter used by the IRT is called "discrimination". Roughly explained, it is the steepness of curve describing the sensitivity of the WHODAS 2.0. Steeper curve means better ability to distinguish people with different severity of disability. Also, the object of interest is the location of the steepest interval in regard to the average level of disability in the studied sample. In an ideal situation, the curve is steep and the steepest interval is located around the average level of disability allowing good ability to distinguish people with disability, which is only mildly below or over the average level. Here, discrimination of 0.01 to 0.24 was considered "none" (curve approaching a horizontal line), 0.25 to 0.64 "low," 0.65 to 1.34 "moderate," 1.35 to 1.69 "high," and ≥1.7 was considered "perfect" (the steepest possible curve).

The **Study II** employed similar statistical techniques as the Study I did, but this time evaluating the ODI instead of the WHODAS 2.0. In addition to the EFA and the IRT, the Study II used a Cronbach's alpha as a measure of internal consistency and the CFA to describe the factor structure of ODI in way, which is more sophisticated than the EFA. A Cronbach's alpha >0.9 was considered "excellent",

>0.8 “good”, >0.7 “acceptable”, >0.6 “questionable”, >0.5 “poor”, and <0.5 “unacceptable”. Cronbach’s α one-side 95% CL was computed.

The CFA goodness of fit was assessed by using the Root Mean Square Error of Approximation (RMSEA). Covariances were added one at a time until the RMSEA achieved the level of <0.05.

The **Study III** investigated the convergent validity of the WHODAS 2.0 using a Spearman’s rank correlation test (correlation between ordinal scores of individual items) or a Pearson product–moment correlation test (correlations between continuous composite score and pain severity). A correlation of < 0.70 was considered “very strong”, 0.40–0.69 “strong”, 0.30–0.39 “moderate”, 0.20–0.29 “weak”, and <0.19 “none or negligible correlation”. The results were reported along with 95% CIs based on the Fisher’s transformation.

Similarly to the Study III, the **Study V** employed a Spearman’s rank correlation coefficient to investigate the convergent validity of the WHODAS 2.0 in relation to the ODI.

All the analyses were conducted using Stata/IC Statistical Software: Releases 14-16 (StataCorp LP, College Station, Texas, USA).

Table 9. Summary of statistical methods used in this thesis.

Method	Purpose	Definition, meaning or interpretation
Absolute numbers and percentages	To describe the distribution of estimates within the studied samples	
Mean and standard deviations	To describe normally distributed continuous or ordinal estimates	Mean (average) – sum of all estimates in the data set divided by the number of values
Median, range and interquartile range	To describe abnormally distributed continuous or ordinal estimates	Median – middle value when estimates are ordered from smallest to highest. Range – interval between minimum and maximum estimates. Interquartile range (IQR) – interval between 25% and 75% quartiles.
95% confidence interval/ limit	To describe with 95% confidence the possible range of mean estimates	Because the true population mean is unknown, 95% confidence interval (interval between lower and upper confidence limits) describes with 95% probability the range of possible mean values that may be measured if infinite number of similar samples are drawn from the studied population
p-value	To evaluate the probability of observing the estimated results, assuming that the null hypothesis is true	All the p-values were two-tailed with the level of statistical significance <=0.05

Method	Purpose	Definition, meaning or interpretation
Exploratory factor analysis (EFA) (Studies I and II)	To identify the number of main factors (latent variables) responsible for the composite score of a scale	In a perfect situation, a scale produces only one meaningful factor (unidimensionality); if a composite score describes more than one factor (multidimen-sionality) the score is unreliable
Confirmatory factor analysis (CFA)(Study II)	To identify correlations between individual items and an underlying (laten) trait (here disability)	CFA is a form of structured equations modelling (SEM) which is built based on the factor structure previously established by EFA.
Item Response Theory (IRT) (Studies I and II)	In this work, IRT describes two main abilities of the WHODAS 2.0 – difficulty and discrimination	Difficulty describes to what extend a respondent should perceive a trait of interest (e.g., disability) to give a particular graded response. In a perfect situation, the middle of a scale should be associated with the average severity of trait in the studied population. Discrimination describes how well a scale distinguish respondents with different severity of trait. In other words, it is the steepness of regression slope. In a perfect situation a scale is easily able to distinguish people a trait level a little above the average from those with a little more severe trait level.
Cronbach's ά (Study II)	A test of internal consistency	Internal consistency describes how well the items of a multiple-item scale are correlated with each other.
Spearman's rank correlation test and Pearson product-moment correlation test (Studies III and V)	Describe the strength and the direction of correlation between two variables	Pearson correlation test applies to continuous variables and Spearman's correlation test to ordinary variables

5 Results

5.1 Psychometric properties of 12-item self-administered WHO Disability Schedule 2.0 (Study I)

At the time of publishing the Study I, 501 consecutive patients with chronic musculoskeletal pain represented 67% of all the available responses. The patients were referred to an outpatient PRM clinic in 2014-2015 by their physicians (specialists or general practitioners) (Table 8). They responded to a survey usually within three days before an actual visit to a clinic. The mean age was 47 (14) years and 65% were women. Most of the respondents were normally weight with median BMI 26.5 kg/m^2. Most of the participants were experiencing only mild restrictions of functioning according to the WHODAS 2.0 low median estimate of 11 (IQR 6 to 18, range 0 to 43) points.

Of 12 possible factors, two factors were considered to have eigenvalues high enough to be retained. It has to be noted that one of these two factors had an eigenvalue far higher than another one –5.15 vs. 1.04 (Table 10 and Figure 3). In total, these two factors accounted for 97% of the variation (Table 11).

The factor with higher eigenvalue was connected to all 12 items of the WHODAS 2.0 with loadings (correlation coefficients) varying between 0.55 and 0.77. In turn, another factor was relatively loosely (loadings over 0.40) associated with three items: “washing”, “dressing”, and “dealing with people you do not know”. It should be pointed at the surprising observation that that second factor was correlated with “dealing with people you do not know” even negatively.

Due to such a wide gap between the eigenvalues of two retained factors, the WHODAS 2.0 composite score was considered to be that close to unidimensionality that the IRT analysis could be applied. However, to confirm this assumption, a sensitivity analysis was applied running the IRT without items S8-S10. The results were similar with the results obtained from the IRT analysis including all 12 items. The IRT analysis among 489 responses achieved a good fit after five iterations pointing to a sufficient sample size. The discrimination ability of all the items was high or perfect, all being statistically significant and all demonstrating narrow 95% CIs (Table 12 and 13).

The difficulty levels of seven items – #S3 Learning a new task, #S4 Joining in community activities, #S6 Concentrating, #S7 Walking long distances, #S8 Washing,

#S10 Dealing with people you do not know, #S11 Maintaining a friendship – were shifted towards the elevated disability level (compared with the average disability level in the studied population) (Table 13). In other words, these items better describe people with higher levels of restrictions compared to those with only mild restrictions or without them at all. This also was true for the composite score when evaluating a test characteristic curve and a test information function curve (Figures 4 and 5).

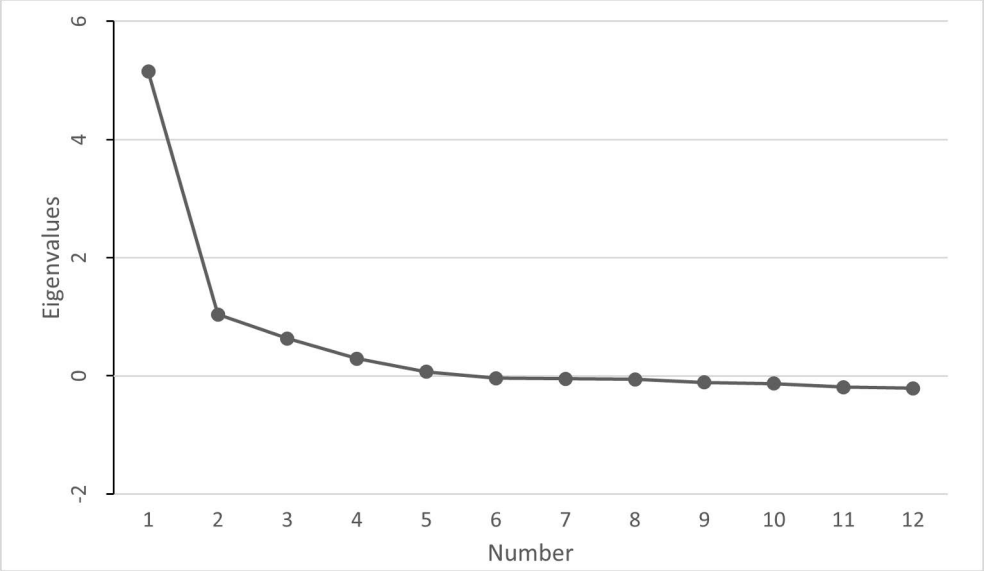

Figure 3. Scree plot of exploratory factor analysis of WHODAS 2.0. From Original Publication I.

Table 10. Exploratory factor analysis of 12-item WHODAS 2.0. Modified from Original Publication I.

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	5.15	4.11	0.81	0.81
Factor 2	1.04	0.42	0.16	0.97
Factor 3	0.63	0.34	0.10	1.07
Factor 4	0.29	0.21	0.05	1.11
Factor 5	0.07	0.11	0.01	1.12
Factor 6	- 0.04	0.01	-0.01	1.12
Factor 7	- 0.05	0.01	-0.01	1.11
Factor 8	- 0.06	0.05	-0.01	1.10
Factor 9	- 0.11	0.03	-0.02	1.08
Factor 10	- 0.13	0.06	-0.02	1.06
Factor 11	- 0.19	0.03	-0.03	1.03
Factor 12	- 0.21	0.00	-0.03	1.00

Table 11. Factor loadings (pattern matrix) and unique variances of two retained factors. Modified from Original Publication I.

WHODAS 2.0 item	Retained factor 1	Retained factor 2	Unique variance
S1 Standing for long periods	0.63	0.10	0.60
S2 Household responsibilities	0.75	0.28	0.36
S3 Learning a new task	0.55	-0.32	0.59
S4 Joining in community activities	0.77	-0.12	0.39
S5 Emotionally affected by health problems	0.66	-0.14	0.54
S6 Concentrating	0.64	-0.23	0.54
S7 Walking long distances	0.65	0.12	0.56
S8 Washing	0.64	0.44	0.40
S9 Dressing	0.55	0.49	0.46
S10 Dealing with people you do not know	0.59	-0.44	0.46
S11 Maintaining a friendship	0.70	0.32	0.40
S12 Day-to-day work	0.66	-0.13	0.55

Table 12. Discrimination ability of WHODAS 2.0 questionnaire items. Modified from Original Publication I.

WHODAS 2.0 item	Discrimination	Lower 95% CI	Upper 95% CI
S1 Standing for long periods	1.47[a]	1.20	1.73
S2 Household responsibilities	2.28[a]	1.92	2.65
S3 Learning a new task	1.79[a]	1.37	2.21
S4 Joining in community activities	2.82[a]	2.32	3.32
S5 Emotionally affected by health problems	1.85[a]	1.55	2.16
S6 Concentrating	2.02[a]	1.65	2.40
S7 Walking long distances	1.51[a]	1.24	1.79
S8 Washing	1.59[a]	1.28	1.90
S9 Dressing	1.22[a]	0.97	1.47
S10 Dealing with people you do not know	2.26[a]	1.72	2.80
S11 Maintaining a friendship	2.45[a]	1.98	2.93
S12 Day-to-day work	1.81[a]	1.51	2.11

[a] *p* value <0.0001

Table 13. Difficulty of WHODAS 2.0 questionnaire items. Modified from Original Publication I.

			95% CI					95% CI	
Item	Response	Disability	Lower Limit	Upper Limit	Item	Response	Disability	Lower Limit	Upper Limit
S1	≥1	-0.69	-0.89	-0.49	S7	≥1	-0.31	-0.48	-0.13
	≥2	-0.06	-0.22	0.11		≥2	0.52	0.34	0.69
	≥3	0.74	0.54	0.93		≥3	1.02	0.81	1.24
	4	1.78	1.46	2.1		4	1.8	1.49	2.12
S2	≥1	-0.98	-1.15	-0.8	S8	≥1	0.18	0.02	0.34
	≥2	0.05	-0.08	0.19		≥2	1.11	0.89	1.33
	≥3	1.01	0.83	1.18		≥3	2.1	1.74	2.46
	4	2.03	1.74	2.32		4	3.06	2.48	3.63
S3	≥1	1.02	0.81	1.23	S9	≥1	-0.24	-0.43	-0.04
	≥2	1.59	1.3	1.88		≥2	1.24	0.97	1.52
	≥3	2.18	1.77	2.59		≥3	2.4	1.93	2.87
	4	3.41	2.63	4.19		4	4.32	3.29	5.35
S4	≥1	0.01	-0.11	0.14	S10	≥1	1.1	0.91	1.29
	≥2	0.66	0.52	0.8		≥2	1.59	1.33	1.85
	≥3	1.13	0.96	1.3		≥3	1.92	1.61	2.24
	4	1.76	1.52	1.99		4	2.77	2.26	3.28
S5	≥1	-1.3	-1.53	-1.08	S11	≥1	0.42	0.29	0.56
	≥2	-0.21	-0.36	-0.06		≥2	1.07	0.89	1.25
	≥3	0.75	0.57	0.92		≥3	1.66	1.41	1.9
	4	2.63	2.2	3.05		4	2.51	2.12	2.91
S6	≥1	0.3	0.16	0.45	S12	≥1	-1.3	-1.53	-1.07
	≥2	1.02	0.83	1.21		≥2	-0.3	-0.46	-0.14
	≥3	1.87	1.58	2.15		≥3	0.57	0.4	0.73
	4	2.8	2.32	3.28		4	1.26	1.04	1.48

CI: confidence interval; WHODAS 2.0: World Health Organization Disability Assessment Schedule.

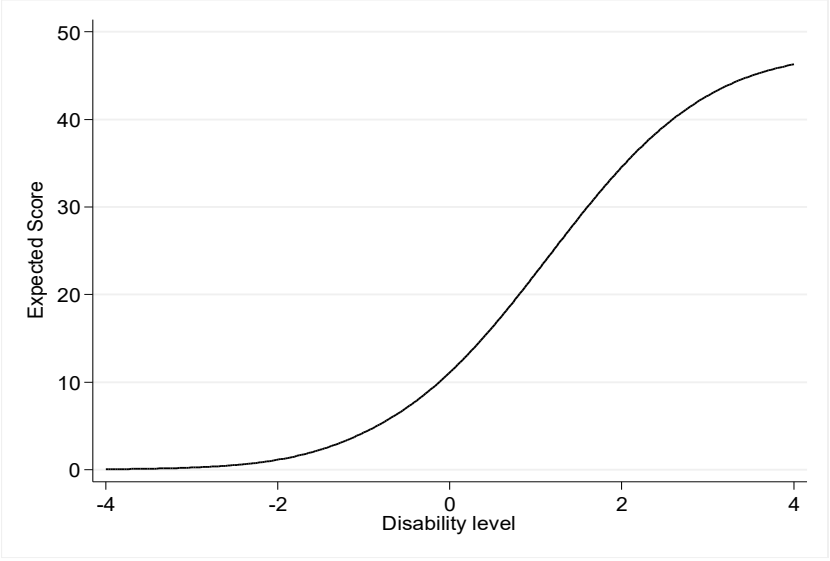

Figure 4. Test characteristic curve for 12-item WHODAS 2.0. From Original Publication I.

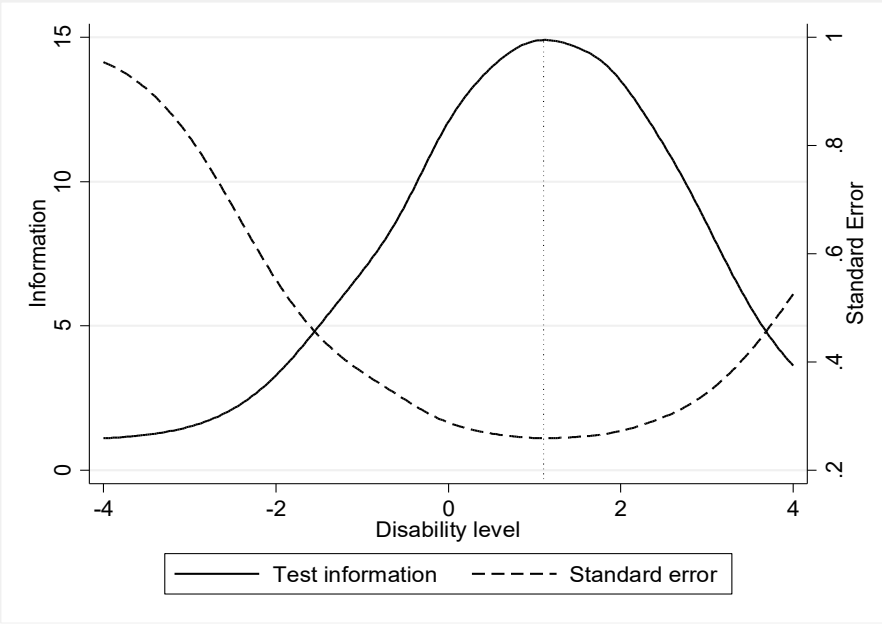

Figure 5. Test information function curve for 12-item WHODAS 2.0. From Original Publication I.

5.2 Psychometric properties of Oswestry Disability Index (Study II)

Of 1,774 people, who had visited a clinic, 1,246 responded to a survey (response rate 70%). Of them, 837 filled out the ODI. The majority of the respondents had some diagnosis of LBP or comparable as a primary diagnosis (Table 7). The mean age was 48 (15) and 66% were women. Most of the respondents valued their functional restrictions caused by LBP as moderate, the ODI median score being 32% (IQR 22–44%). Also, most of the respondents were experiencing severe pain – median usual pain scored 7 (IQR: 5–8) points, whereas the worst pain scored 9 (IQR: 8–9) points.

The internal consistency of the ODI was good with Cronbach's α of 0.85 (lower 95%CI 0.84). The EFA analysis resulted in a unidimensional factor structure. Even if there seemed to be three factors above the parallel analysis line, two factors had very small eigenvalues of 0.2 and 0.1 and they were very close to the parallel analysis line (Figure 6 and Table 14).

This unidimensional factor structure was used as a base for the CFA, which was conducted among 523 respondents after deleting records with any missing ODI item response. To achieve a good model fit (RMSEA significance level ≤ 0.05), covariance between the 'sitting' and 'traveling' items was added. After that, the relative χ^2 value was 2.3 (df. 34) being close enough to the cutoff point of 2.0, and the model was accepted (Figure 8). The estimates related to the goodness of model

fit are shown in Table 15. The standardized regression weights were relatively high for all 10 items, varying from 0.5 (“sex life”) to 0.7 (“social life”) (Figure 8).

Only five iterations were needed to achieve the acceptable IRT model (n = 837), which was pointing to the sufficient sample size for such an analysis. For eight items #1, #3-#7, #9, and #10, the estimates of item difficulty presented an ideal situation – centrally placed responses of two or three points were associated with average level of disability in the studied population (Table 16). In other words, these eight items were well able to measure the entire spectrum of disability severity. In turn, it seemed that items #2 and #8 might underestimate the severity of disability. For these two items, considerably more severe disability was needed to choose centrally located responses of two or three points. The discrimination abilities of seven items – “Pain”, “Personal care”, “Lifting”, “Walking”, “Standing”, “Social life” and “Travelling” – were high or perfect, all being statistically significant with narrow 95% CIs (Table 17). Three items, #5 “Sitting”, #7 “Sleeping”, and #8 “Sex life”, demonstrated a moderate strength of discrimination ability. Both the test characteristic curve and the test information function curve of the composite ODI score were slightly shifted toward higher levels of disability (Figure 9 and Figure 10).

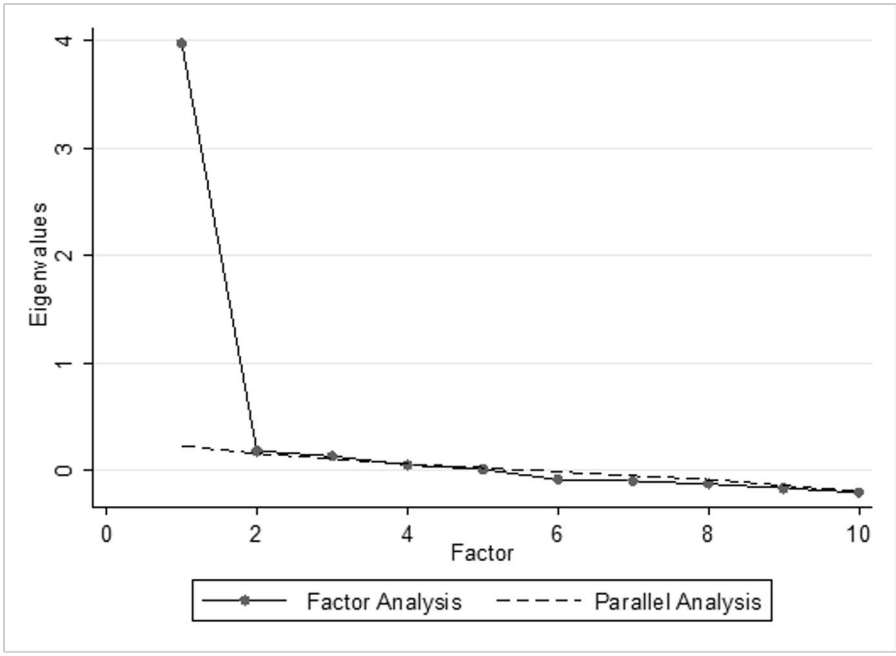

Figure 6. Exploratory factor analysis of Oswestry Disability Index. From Original Publication II.

Table 14. Factor loadings (pattern matrix) and unique variances for exploratory factor analysis of Oswestry Disability Index. Modified from Original Publication II.

ODI items	Loading	Unique variance
Item 1: Pain intensity	0.58	0.66
Item 2: Personal care	0.64	0.59
Item 3: Lifting	0.68	0.54
Item 4: Walking	0.63	0.61
Item 5: Sitting	0.60	0.64
Item 6: Standing	0.66	0.57
Item 7: Sleeping	0.57	0.68
Item 8: Sex life	0.52	0.73
Item 9: Social life	0.70	0.51
Item 10: Traveling	0.71	0.50

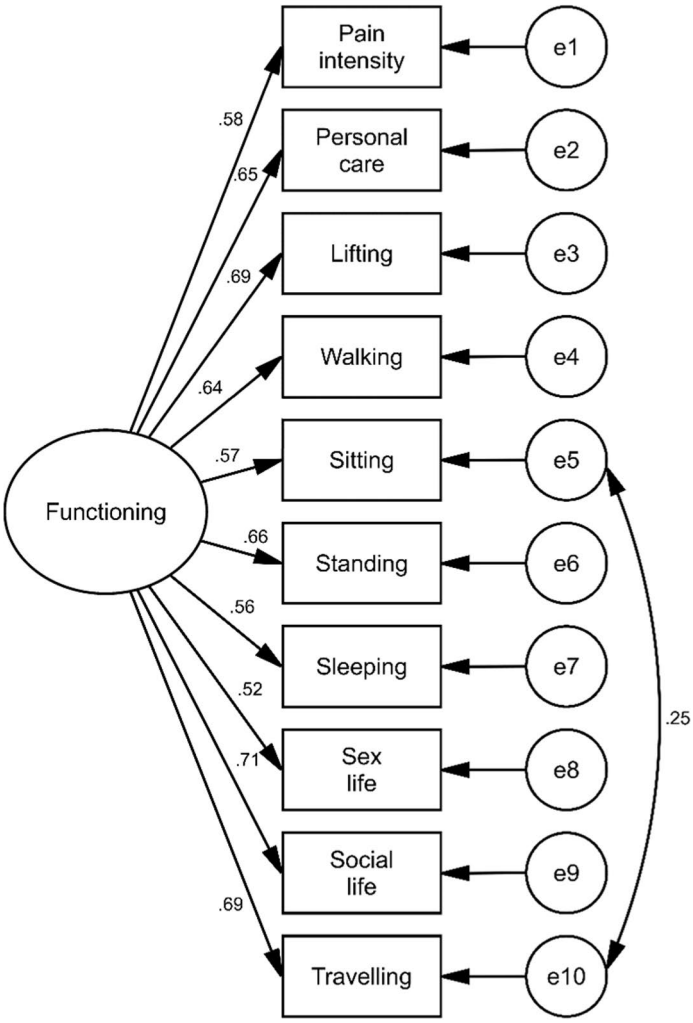

Figure 8. Confirmatory factor analysis of Oswestry Disability Index. Round shapes e1–e10 represent possible measurement errors. One-headed arrows represent correlations, whereas the two-headed arrow represents covariance. From Original Publication II.

Table 15. Testing goodness of model fit. Modified from Original Publication II.

Methods	Value
Standardized root mean square residual (SRMR)	0.32
Bentler–Bonett normed fit index (NFI)	0.96
Bollen's relative fit index (RFI)	0.94
Bollen's incremental fit index (IFI)	0.98
Tucker–Lewis coefficient (TLI)	0.97
Bentler's comparative fit index (CFI)	0.97
Parsimony ratio (PRATIO)	0.76
Parsimony adjustment to the NFI (PNFI)	0.72
Parsimony adjustment to the CFI (PCFI)	0.74
Akaike information criterion (AIC)	139.4
Browne–Cudeck criterion (BCC)	140.7
Hoelter's 'critical N' for a significance level of.05 (HOELTER.05)	328
Hoelter's 'critical N' for a significance level of.01 (HOELTER.01)	379

Methods	Value	90% confidence interval	
Noncentrality parameter (NCP)	43.4	21.6	72.9
Minimum value of the discrepancy (FMIN)	0.15	0.04	0.14
Root mean square error of approximation (RMSEA)	0.049	0.035	0.064
Except for a constant scale factor (ECVI)	0.27	0.23	0.32

Method	Value	*d.f.*	*P*-value	CMIN/*d.f.*
Minimized value of the discrepancy function (CMIN)	77.4	34	<0.0001	2.3

Table 16. Difficulty of Oswestry Disability Index items. Modified from Original Publication II.

			95% CI					95% CI	
Item	Response	Difficulty	Lower limit	Upper limit	Item	Response	Difficulty	Lower Limit	Upper Limit
1	≥1	−2.81	−3.21	−2.41	6	≥1	−1.6	−1.81	−1.4
	≥2	−1.1	−1.29	−0.92		≥2	−0.3	−0.42	−0.17
	≥3	0.59	0.44	0.74		≥3	0.29	0.17	0.42
	≥4	1.85	1.59	2.11		≥4	1.42	1.23	1.61
	5	3.65	3.07	4.23		5	3.52	2.98	4.06
2	≥1	−0.58	−0.72	−0.45	7	≥1	−2.77	−3.2	−2.34
	≥2	1.17	1.0	1.35		≥2	0.38	0.22	0.53
	≥3	2.5	2.18	2.82		≥3	1.6	1.34	1.85
	≥4	3.43	2.89	3.97		≥4	3.09	2.61	3.57
	5	4.81	3.51	6.1		5	4.2	3.45	4.96
3	≥1	−2.06	−2.3	−1.81	8	≥1	0.43	0.24	0.63
	≥2	−0.7	−0.84	−0.57		≥2	1.57	1.22	1.91
	≥3	−0.04	−0.16	0.08		≥3	1.77	1.4	2.14
	≥4	1.05	0.89	1.2		≥4	2.62	2.09	3.15
	5	2.56	2.25	2.88		5	3.93	3.07	4.78
4	≥1	−0.37	−0.5	−0.24	9	≥1	−0.68	−0.81	−0.55
	≥2	0.97	0.82	1.12		≥2	−0.47	−0.59	−0.45
	≥3	1.98[a]	1.73	2.22		≥3	0.18	0.07	0.29
	≥4	2.72	2.37	3.08		≥4	1.44	1.26	1.62
	5	4.29	3.36	5.21		5	3.45	2.91	4.0
5	≥1	−2.23	−2.55	−1.92	10	≥1	−1.53	−1.72	−1.35
	≥2	−0.84	−1.01	−0.68		≥2	−0.22	−0.33	−0.11
	≥3	0.82	0.65	0.98		≥3	1.12	0.97	1.27
	≥4	2.42	2.09	2.76		≥4	1.98	1.76	2.21
	5	4.66	3.73	5.59		5	2.76	2.42	3.11

*P = 0.484, all others P<0.0001

Table 17. Discrimination ability of Oswestry Disability Index (n = 837). Modified from Original Publication II.

Oswestry Disability Index items	Discrimination coefficient	Standard error	Z-score [a]	95% CI	
				Lower limit	Upper limit
1: Pain intensity	1.35	0.1	13.48	1.16	1.55
2: Personal care	1.66	0.12	13.64	1.42	1.9
3: Lifting	1.73	0.11	15.11	1.51	1.96
4: Walking	1.73	0.13	13.82	1.48	1.97
5: Sitting	1.32	0.1	13.47	1.12	1.51
6: Standing	1.68	0.11	14.77	1.46	1.91
7: Sleeping	1.18	0.1	12.13	0.99	1.37
8: Sex life	1.08	0.12	9.11	0.84	1.31
9: Social life	2.01	0.14	14.45	1.74	2.28
10: Traveling	2.00	0.13	14.99	1.74	2.26

[a]All p<0.0001

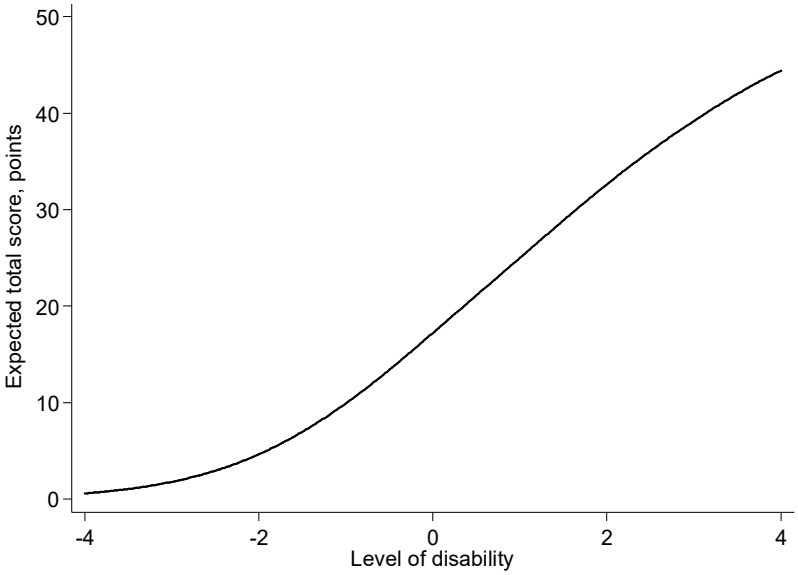

Figure 9. Test characteristics curve of Oswestry Disability Index. From Original Publication II.

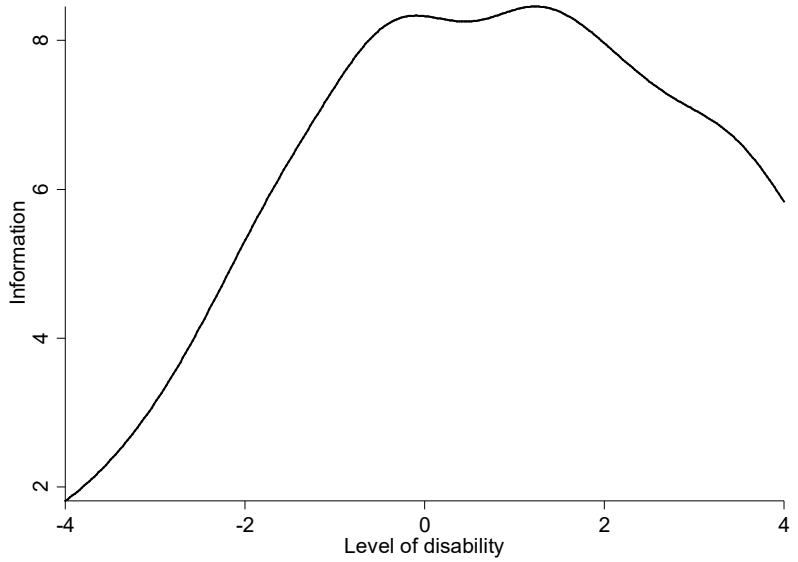

Figure 10. Test information function curve of Oswestry Disability Index. From Original Publication II.

5.3 Correlation between pain numeric rating scale and 12-item self-administered WHO Disability Assessment Schedule 2.0 (Study III)

The initial sample for this study was the same as in the Study II. After excluding those with missing responses to a pain NRS or the WHODAS 2.0, the analyses were conducted among 1,207 respondents. The average age was 47 (15) years and 66% were women. Most of the respondents had only mild of functional restrictions as measured by the median WHODAS 2.0 score of 25% (IQR: 13–40%). The overall level of pain was, however, at least moderate for the majority of the respondents – usual pain severity perceived was 7 (IQR: 5–8) points, whereas the worst pain severity scored 8 (IQR: 7–9) points.

All WHODAS 2.0 items were statistically significantly correlated pain severity. However, the associations were mainly weak, except for moderate correlations between pain severity and the WHODAS 2.0 items "S2 household responsibilities", "S8 washing", "S9 dressing", and "S12 day-to-day work" (Table 18). Also, the composite score of the WHODAS 2.0 demonstrated a moderate association of 0.40 with both usual and the worst pain severity. As a sensitivity test, the correlation between the composite score of WHODAS 2.0 and pain severity was checked also by using a Pearson's product–moment correlation, yielding figures that were similar to Spearman's correlation coefficients (Figure 11 and Table 19).

Table 18. Spearman's rank correlation between WHODAS 2.0 items and average and worst pain experienced during the last month measured on a numeric rating scale of 0 to 10 (n=1 158). Modified from Original Publication III.

	Average pain			Worst pain		
		95% confidence interval			95% confidence interval	
WHODAS 2.0 items	**Correlation**	**Lower limit**	**Upper limit**	**Correlation**	**Lower limit**	**Upper limit**
S1 Standing for long periods	0.28	0.23	0.33	0.24	0.18	0.29
S2 Household responsibilities	0.36	0.31	0.41	0.38	0.33	0.43
S3 Learning a new task	0.17	0.11	0.22	0.14	0.08	0.19
S4 Joining in community activities	0.29	0.23	0.34	0.29	0.24	0.34
S5 Emotionally affected by health problems	0.24	0.19	0.30	0.25	0.20	0.31
S6 Concentrating	0.24	0.18	0.29	0.25	0.19	0.30
S7 Walking a long distance	0.28	0.22	0.33	0.24	0.18	0.29
S8 Washing	0.33	0.27	0.37	0.34	0.29	0.39
S9 Dressing	0.30	0.25	0.35	0.36	0.31	0.41
S10 Dealing with people you do not know	0.14	0.08	0.20	0.12 [a]	0.06	0.17
S11 Maintaining a friendship	0.18	0.13	0.24	0.23	0.18	0.28
S12 Day-to-day work	0.25	0.20	0.31	0.32	0.27	0.37
Total score	0.41	0.36	0.45	0.42	0.37	0.46

[a] p=0.0001, all other p values <0.0001

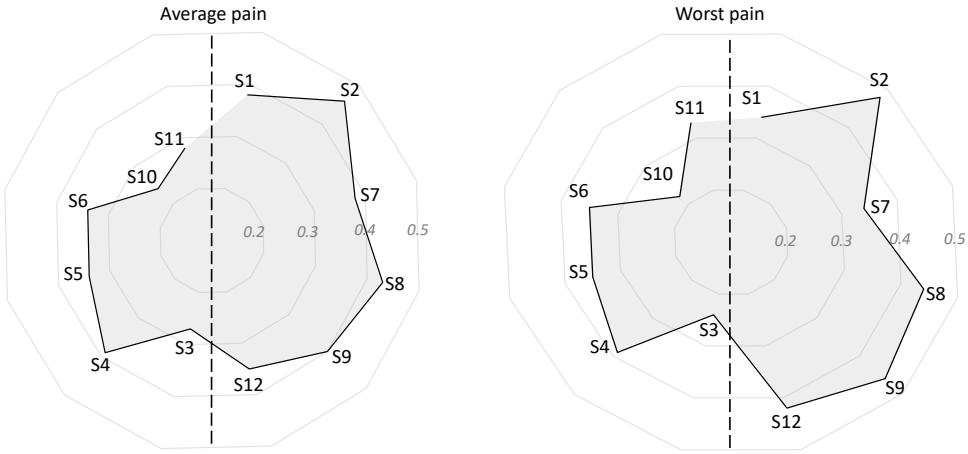

Figure 11. Radar chart showing correlations between pain severity measured on a numeric rating scale and functioning limitations measured by 12 items of the WHO Disability Assessment Schedule (WHODAS) 2.0. WHODAS 2.0 items representing 'physical' domain of functioning. From Original Publication III.

Table 19. Spearman's correlations between WHODAS 2.0 items (n = 1207). Modified from Original Publication III.

WHODAS 2.0 items	S1 Standing	S2 Household	S7 Walking	S8 Washing	S9 Dressing	S12 work	S3 Learning	S4 Community activities	S5 Emotional affectedness	S6 Concentratin	S10 Dealing with people you do not	S11 Friendship
S1 Standing	1.00											
S2 Household	0.47	1.00										
S7 Walking	0.61	0.42	1.00									
S8 Washing	0.30	0.52	0.34	1.00								
S9 Dressing	0.29	0.47	0.27	0.75	1.00							
S12 Work	0.30	0.51	0.30	0.41	0.37	1.00						
S3 Learning	0.27	0.29	0.25	0.21	0.16	0.18	1.00					
S4 Community activities	**0.44**	**0.54**	**0.43**	0.38	0.31	**0.43**	0.45	1.00				
S5 Emotional affectedness	0.32	**0.49**	0.31	0.35	0.29	**0.41**	0.35	0.52	1.00			
S6 Concentrating	0.33	**0.41**	0.31	0.35	0.27	0.38	0.42	0.47	0.50	1.00		
S10 Dealing with people you do not know	0.22	0.29	0.22	0.25	0.19	0.25	0.49	0.49	0.36	0.43	1.00	
S11 Friendship	0.27	**0.41**	0.32	0.31	0.24	0.38	0.42	0.57	0.54	0.46	0.58	1.00

Area within dash line borders represents correlations between items belonging to "physical" and "psychological" domains of WHODAS 2.0.
In this area, strong correlations $\geq$0.40 are shown in bold.
p<0.0001 for all correlations in the table.

5.4 Correlation between Oswestry Disability Index and 12-item self-administered WHO Disability Assessment Schedule 2.0 (Study V)

The sample was formed of those 1,379 people, who responded to both the ODI and the WHODAS 2.0. Their mean age was 48 (SD 15) years and 64% were women. The respondents were slightly overweight with BMI of 29.4 (76.4) kg/m^2. The severity of usual pain was moderate 6.4 (1.9) points, while the worst pain severity was a little worse with 8.2 (1.6) points. On average, the respondents were experiencing moderate restrictions of functioning as measured by the ODI – 34% (16%). A similar situation was observed when measuring functional restriction using the WHODAS 2.0 – the severity of restrictions was moderate 14.4 (SD 9.4) points. Of all the possible 143 correlations, 46 (32%) were strong and nine (6%) were very strong (Table 20). All the correlation coefficients were positive and significant, with *P* values of <0.0001. Of the ODI items, items #9 "Social life" and #10 "Travelling" correlated with the individual items of the WHODAS 2.0 especially well. Also, clear positive association between the composite scores of two scales was observed.

Table 20. Spearman's rank correlations between 12-item WHO Disability Assessment Schedule 2.0 and Oswestry disability index. Modified from Original Publication V.

ODI items		WHODAS 2.0 items												
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**Total**
1	*ρ*	0.38	**0.40**	0.15	0.33	0.30	0.23	0.35	0.33	0.30	0.16	0.23	0.36	**0.43**
	n	1300	1298	1265	1288	1287	1287	1296	1313	1313	1300	1302	1204	1341
2	*ρ*	0.38	**0.48**	0.21	**0.40**	0.30	0.26	0.39	**0.60**	**0.64**	0.17	0.25	0.37	**0.54**
	n	1300	1298	1266	1288	1285	1284	1294	1314	1314	1300	1302	1204	1341
3	ρ	**0.43**	0.50	0.22	0.41	0.28	0.25	0.45	**0.41**	0.38	0.18	0.26	0.47	**0.52**
	n	1304	1300	1269	1292	1289	1288	1296	1316	1317	1303	1305	1208	1344
4	*ρ*	**0.54**	**0.45**	0.23	**0.42**	0.24	0.24	**0.77**	**0.40**	0.34	0.19	0.25	0.38	**0.56**
	n	1274	1271	1242	1263	1257	1257	1268	1285	1284	1275	1276	1177	1308
5	*ρ*	0.30	0.31	0.16	0.30	0.27	0.27	0.24	0.33	0.32	0.10	0.20	0.37	0.38
	n	1286	1282	1252	1274	1269	1270	1278	1298	1298	1286	1288	1190	1324
6	*ρ*	0.75	**0.46**	0.21	**0.40**	0.26	0.29	**0.51**	0.36	0.31	0.16	0.25	0.38	**0.55**
	n	1284	1278	1246	1269	1266	1267	1274	1294	1294	1281	1284	1188	1318
7	*ρ*	0.30	0.33	0.18	0.31	0.25	0.21	0.28	0.30	0.30	0.18	0.22	0.30	0.37
	n	1280	1278	1248	1270	1268	1265	1273	1295	1295	1282	1284	1189	1321
8	*ρ*	0.30	0.37	0.16	0.37	0.32	0.30	0.34	**0.42**	0.37	0.17	0.28	0.36	**0.46**
	n	1003	998	986	1001	994	996	996	1011	1009	1004	1005	944	1028
9	*ρ*	**0.44**	**0.55**	0.31	**0.56**	**0.44**	0.38	**0.50**	**0.45**	**0.40**	0.31	**0.48**	**0.48**	**0.63**
	n	1268	1266	1233	1258	1256	1257	1261	1281	1281	1270	1271	1180	1306
10	*ρ*	**0.45**	**0.47**	0.23	**0.46**	0.31	0.33	**0.44**	**0.42**	**0.41**	0.22	0.29	**0.49**	**0.54**
	n	1264	1259	1232	1254	1254	1249	1256	1276	1276	1266	1268	1174	1300
Total	*ρ*	**0.62**	**0.62**	0.29	**0.58**	**0.42**	**0.40**	**0.61**	**0.57**	**0.52**	0.25	0.39	**0.56**	**0.71**
	n	1333	1330	1297	1319	1319	1318	1327	1348	1349	1335	1337	1236	1379

All the *P* values are <0.0001. The strength of shading corresponds to the strength of the correlation. WHODAS 2.0 items: S1 standing, S2 household responsibilities, S3 learning a new task, S4 joining in community activities, S5 emotionally affected, S6 concentrating, S7 walking long distances, S8 washing whole body, S9 getting dressed, S10 dealing with people you do not know, S11 maintaining friendships and S12 work/school activities. Oswestry disability index items: 1, pain intensity; 2, personal care (washing, dressing etc.); 3, lifting; 4, walking; 5, sitting; 6, standing; 7, sleeping; 8, sex life (if applicable); 9, social life; and 10 traveling. WHODAS, WHO Disability Assessment Schedule.

6 Discussion

This thesis provides evidence on both construct and convergent validity for the 12-item self-administered WHODAS 2.0 among patients with LBP. Six out of seven priori-set hypotheses were confirmed. One of the major strengths of this study was large sample size, which was able to provide strong enough power for the statistical analysis methods that were used, particularly in the factor analyses and IRT models.

6.1 Psychometric properties of the World Health Organization Disability Assessment Schedule 2.0

The main results of first study were that the 12-item self-administered WHODAS 2.0 had good psychometric properties in the studied population.

The results did not support the conclusions regarding unidimensionality for the measure as have previously been described (Schiavolin et al., 2014; Kimber et al., 2015), but they did support the findings of multidimensionality observed by other researchers (Luciano et al., 2010; Sousa et al., 2010; Saltychev et al., 2021). These differences may be explained by the differences in the samples of the studies – disease-specific cohorts vs. more general cohorts; acute vs. chronic LBP; different gender and age distributions and different versions of the WHODAS 2.0 used in relation to the types of administration (interview-administered vs. self-administered vs. proxy-administered) and number of total items (12, 12+24, 36).

As noted in this thesis' review of the literature (Table 2), 60% of the ICF categories included in the original 36-item version were lost when a 12-item version was derived of the original 36-item version. Among these were such items as d410, Changing basic body positions; d430, Lifting and carrying objects; and d460, Moving around in different locations. These categories all include normal activities of daily living i.e., bending, rolling while lying, lifting, carrying objects, and maintaining positions such as sitting or standing. They are all included in the ICF's core set for LBP. Cieza et al. (2004) concluded that overall, the limitations and restrictions to activities and participation may be most relevant to patients with LBP. Fatoye et al. (2019) also found that some of the risk factors for LBP were lifting and bending.

The finding that the items of WHODAS 2.0 were able to discriminate between respondents with a higher disability level and those with a lower disability level was

in accordance with the results of previous studies (Buist-Bouwman et al., 2008; Hudson et al., 2008; Luciano et al., 2010; Schiavolin et al., 2014).

The novel finding was that the 12-item self-administered WHODAS 2.0 was more sensitive in discriminating between disability levels at the higher end of disability continuum ("much more" vs. "more" than average) disability than between "more" and "less" than average disability.

6.2 Psychometric properties of the Oswestry Disability Index

The second study found the ODI being internally consistent and unidimensional scale. It also showed good construct validity and the ability to differentiate between the severity of functional disability.

This result concurs with findings of previous studies (Lauridsen et al., 2006; Valasek et al., 2013; Sheahan et al., 2015; van Hooff et al., 2015), and contradicts the multidimensionality described by others (Osthus et al., 2006; Payares et al., 2011; Pekkanen et al., 2011; Algarni et al., 2014; Lee et al., 2017). The differences may be explained by the differences in sample sizes—all the mentioned research sample sizes varied between 96 and 244 participants, which might be too limited for stable estimates in factor analyses. For the ODI, that a sample size of 500–1,000 has been suggested to ensure a reliable factor analysis. Previously, the unidimensionality of ODI was confirmed by a large sample size of 35,264 respondents with LBP (Gabel et al., 2017). Some previous studies have reported weaknesses in the ODI's psychometric properties (Chow et al., 2005; Davidson, 2008; Dawson et al., 2010; Lochhead et al., 2013; Algarni et al., 2014), which could possibly be explained by their limited sample sizes (79 to 214 participants).

The good reliability of ODI found in this study has been supported by a similar sample size in the study by Brodke et al. (2017) with exception of weaker evidence of unidimensionality. The differences could be related to the use of different statistical methods i.e., Rasch analysis vs. IRT. This study supported the ODI's good psychometric properties among Finnish chronic LBP in terms of internal consistency, unidimensionality, excellent construct validity and an ability to differentiate the severity of functional disability.

6.3 Correlation between pain numerical ratings scale and the World Health Organization Disability Assessment Schedule 2.0

The results of the third study supported a weak to moderate positive association between pain severity measured by an NRS and limitations of functioning measured by WHODAS 2.0.

Likewise, some previous studies have reported a weak or moderate correlation between pain severity and WHODAS 2.0 (Baron et al., 2008; Garin et al., 2010; Silva et al., 2013)

Contrary to the findings of this study, van Tubergen et al. (2003) reported a strong positive correlation between an NRS and the WHODAS 2.0 among people with ankylosing spondylitis. Several factors may explain this difference. Ankylosing spondylitis is an inflammatory disease, while a correlation between pain and functioning may be stronger which provokes the pain, as opposed to this study's patients, who had non-inflammatory chronic musculoskeletal pain. Secondly, some evidence suggests that the correlation between pain and functioning may be stronger in longitudinal studies than in cross-sectional studies (McGorry et al., 2011).

6.4 Correlation between the Oswestry Disability Index and the World Health Organization Disability Assessment Schedule 2.0

The study observed a strong positive association between one-third of the WHODAS 2.0 items and the ODI. The association were stronger among items defining physical rather than social or psychological functioning. Only a few studies have focused on the comparison between the WHODAS 2.0 and the ODI. The findings agreed with the results of Ćwirlej-Sozańska et al. (2020a), who observed a strong association between all the domains of the 36-item interviewer-administered WHODAS 2.0 and the ODI. An association has been detected between the WHODAS 2.0 and the RMDQ, a scale close to the ODI (Igwesi-Chidobe et al., 2020). The generalization of the study (Bärlund et al., 2021) findings may be affected by several factors. CLBP as diagnosis was not specifically addressed but LBP, women were mainly presented in age group around 50 years and WHODAS 2.0 and ODI having some similar items possibly affecting psychometrics of responses. Also, some of the correlations may not have any real-life importance, e.g., correlation between WHODAS 2.0 item #S6 'Concentrating on doing something for 10 min' and ODI's item #10 'Travelling'. Regardless of these addressed weaknesses, this study was the first with large enough sample size for proper statistical significance levels in correlations.

While being an LBP-specific scale, the ODI describes solely the impact of pain on physical functioning and participation, it misses psychological aspect. One out of then items describes the effect of pain on participation, i.e., #9 'Social life'. WHODAS 2.0, on the contrary, describes more psychological and social aspects of functioning with total of eight items out of twelve. More common restrictions in activities of daily living in LPB, like difficulties in sitting, sleeping, lifting, and travelling, are left undefined. Similarly, the WHODAS 2.0 fails to address pain intensity.

Based on the evidence presented previously, it is suggested that the WHODAS 2.0 and the ODI may supplement combined, at the same time expanding the view to functioning.

6.5 Evidence from systematic review of using 12-item self-administered WHO Disability Assessment Schedule 2.0 among general population and people with non-acute physical causes of disability

The results of the fifth study showed evidence of a substantial floor effect in the 12-item WHODAS 2.0, which may limit the use of the measure as a screening tool in a general population with mild severity. Evidently, the measure produces reliable repeated measurements, and over time this encourages the screening of functioning. Multidimensionality, as opposed to unidimensionality, affects the use of the total score as a comparison, and is not recommended in this population.

The findings were in accordance with the only similar study previously conducted by Federici et al. (2017). A floor effect was found in some populations, strong correlations with some activity limitation measures, and moderate to strong correlations with disease-specific measures of various conditions including back pain. This review's findings of good concurrent validity were supported.

Contrary to the multidimensionality found in this review, Federici et al. (2017) emphasized the unidimensional nature of the measure. These differences may be explained by the reviews' different views of measuring selection: This review merely narrowed the search down to a 12-item self-administered version, whereas the other review chose all the different versions of WHODAS 2.0 (36, 12, 12–24; interview-administered, self-administered and proxy-administered), including the children's version. The majority of the papers included in this review were published after the review by Federici et al. (2017), and this may also have influenced the conclusions.

As a novel finding, this review underlined the evidence supporting the multidimensionality of the 12-item self-administered WHODAS 2.0, the use of the scores of individual items instead of a total score, and the possible presentation of the score as a profile.

6.6 Summary of confirmation of a priori set hypotheses for validation of 12-item self-administered WHODAS 2.0

It has been suggested that in the assessment of the construct validity of a measure, the results should support a minimum of 75% of an a priori set hypothesis (Terwee

et al., 2007; Prinsen et al., 2016). In this thesis, 85% of priori-set hypotheses were supported by the combined results of this thesis. Only one priori-set hypothesis was not supported evidence from this study. The association between pain severity measured by using an NRS was expected to be positively and moderately or strongly correlated with WHODAS, evidence showed weak to moderate correlations.
Overall, the good and adequate construct and convergent validity of the Finnish version of the12-item self-administered WHODAS 2.0 in this population was supported (Table 21).

Table 21. Summary of hypotheses confirmations.

Originally set a priori hypotheses	Results supporting hypotheses
1. The 12-item self-administered WHODAS 2.0 is an internally consistent a unidimensional identifying only a single underlying latent trait, a "general disability factor", in people with chronic musculoskeletal pain. (Study I)	✔ **Results support the hypothesis**. Although some slight uncertainty of unidimensionality, one-dimensionality, one single latent trait, was agreed. Further research was encouraged.
2. The 12-item self-administered WHODAS 2.0 is able to distinguish different levels of functioning among people with chronic musculoskeletal. (Study I).	✔ **Results support the hypothesis.** The ability to distinguish all the items was shown to be high or perfect under all the different response options and among people with both less and greater disability.
3. The ODI is an internally consistent and unidimensional scale identifying only a single underlying latent trait, a "general factor" in people with LBP. (Study II)	✔ **Results support the hypothesis.** The ODI is an internally consistent, unidimensional scale with overall excellent construct validity.
4. The ODI is able to distinguish different levels of functioning among people with LBP. (Study II)	✔ **Results support the hypothesis**. Among LBP, the ODI was found capable to discriminate the severity of functional disability, better with above-average disability levels.
5. Pain severity measured by using an NRS is positively and moderately or strongly associated with the 12-item self-administered WHODAS 2.0. (Study III)	✘ **Results do not support the hypothesis.** The association between pain severity measured by using an NRS with 12-item self-administered WHODAS 2.0 was found to be weak to moderate.
6. The available evidence of the systematic review supports the use of the 12-item self-administered among the general population and among people with non-acute physical causes of LBP. (Study IV)	✔ **Results support the hypothesis.** The 12-item WHODAS 2.0 was found to be internally consistent and reliable scale, good correlation with other disability measures was supported too.
7. The 12-item self-administered WHODAS 2.0 is positively and moderately or strongly associated with the ODI in people with LBP. (Study V)	✔ **Results support the hypothesis.** Positive and moderate or strong correlations were identified. Evidence supports using both measures together in gathering information from the functioning of people with CLBP

6.7 Limitations of the studies

As the original studies were closely related to the original sample of a cross-sectional cohort of physician-selected patients, selection bias may have occurred, although it was not observed. Furthermore, no information could be attained on the possible changes in the scores over time. The samples in the original studies were somewhat over-represented by female participants and limited to certain age groups. These factors may both affect the generalization of the results.

The ODI's low response rates may increase selection bias and negatively impact generalization. In addition, the patients entering the PMR clinic were all experiencing longer, more chronic periods of musculoskeletal pain. As such, these samples lacked patients with acute or subacute pain experiences. This imbalance may also affect the generalization of the results.

Also, it is important to understand that 1988 patients, who answered to questionnaire in this cohort cross-sectional study, were rather heterogeneous study population. Exact confirmation of those patient who were acute instead of being chronic was missing. It is possible that large total sample sizes alleviates this problem. In addition, the specific Physical and Rehabilitation Medicine Clinic did not have a 24/7 clinic for acute patients. Often in Finland, for patients to get an appointment with the University hospital specialist takes several months time. In this case, 1-2 months.

One obvious limitation was that the original planning and collection of data for this cohort started before the official Finnish translation of WHODAS 2.0 was finalized. On some of the difficulty levels, the Finnish terminology varied in the different versions. This semantic misalignment may have influenced how respondents understood and answered some of the levels. It may also have affected the interpretation of the final results.

In the systematic review, the research papers included in the final level were rather heterogeneous and were also limited by small sample sizes. In addition, there was a shortage of relevant papers from the same patient group, and the joint sample sizes remained small. This limited any comprehensive analysis and conclusions. In the review, uncertainty regarding the methodological quality of the included papers meant it was not possible to generalize any of the results. This uncertainty about psychometrics due to the lack of methodological quality was a disturbing finding and is supported by the systematic review by Chiarotto et al. (2019) of the measuring properties of the VAS, NRSs and the Pain Severity Subscale of the Brief Pain Inventory for LBP.

Although today, self-reporting of outcome measures is highly recommended and have their strengths, PROMs have some possible limitations (Fowler, 2009; Cella et al., 2015). In patient-reported outcomes, the obvious strength is that the respondent provides an expert opinion, as it based on their personal experience. Moreover, as no interviewer is involved, the sensitive data collected may be more valid (Fowler, 2009). Difficulties in cognition and communication skills, and possible defects

related to age or developmental level may affect the reliability of responses. In some cases, good reading and writing skills play a role (Cella et al., 2015). This type of response is one of the most cost-effective ways of reporting, as it also encourages respondents to answer at their own pace. Nonetheless, the downside to self-reporting is that in the event of problems, no assistive questions can be provided, and the likelihood of missing data increases. Moreover, the questionnaires need to be simple enough to suit the level of all the respondents. Methods of administration may also vary. A paper-and-pencil method may create errors in the data entry, and scoring is often more time consuming. The setting of the data collection can also make a difference. The clinical setting may increase respondents' anxiety, and possible interruptions may lead to missing data. Although answering at home may increase overall patient safety and data security, and lead to more accessible responses, the real respondent cannot be controlled (Fowler, 2009). Also, some evidence (Rasmussen et al., 2018) state that there may be a recall bias among LBP individuals, high variability in retrospective measurements of LBP.

Finally, in convergent evidence studies, the instruments being tested directly should be maximally different. The instruments in this thesis make defining the meaning of "maximally different" difficult. It has been suggested that a self-report scale versus an observer-report scale could solve this. It is possible that the different measurements used in this thesis, although all self-report versions, met the requirement to be maximally different, partly also because they did not share one and the same conceptual framework (Streiner et al., 2015).

6.8 Practical implications and suggestions for further research

The findings of the construct validation supported the hypotheses in the described patient sample.

The use of the 12-item self-administered WHODAS 2.0 is recommended, as it is a fast, convenient, and practical tool for measuring disability. Due to the observed multidimensionality of the measure, the use of WHODAS to create a profile of functioning seems more appropriate than using a total sum score. The joint use of the ODI, an NRS, and the WHODAS 2.0 among chronic LBP patients is recommended as these tools do not completely substitute each other. Finally, a considerable floor effect of the WHODAS 2.0 was observed.

Further investigation may include similar validation studies across the specific subtypes of LBP (acute vs. subacute vs. chronic), different age groups and different versions of the 12-item WHODAS 2.0 (interview-administered vs. self-administered vs. proxy). The reason for the observed shift to higher difficulty levels in certain items with higher disability levels could also be of great interest.

7 Conclusions

The internal consistency of WHODAS 2.0 was good. The EFA revealed two retained factors, suggesting that the WHODAS 2.0 may be of better use to construct a profile of functioning rather than to calculate a single total score. The ability to discriminate among all the items was high or perfect with a shift towards higher levels of disability. This means that the WHODAS 2.0 may perform better among patients with more severe disability levels. The correlation between pain severity measured by an NRS and functioning levels measured by WHODAS 2.0 was weak to moderate, with slightly stronger associations in the physical domains of functioning. In comparison to the ODI, items defining physical functioning demonstrated stronger correlations than items defining social or psychological functioning. The Finnish version of the 12-item WHODAS 2.0 is a valid tool for measuring restrictions to activity and participation among patients with LBP.

Acknowledgements

This work was carried out in the Doctoral Programme in Clinical Research at the University of Turku during the years 2017-2022.

I am deeply grateful to my supervisors, Professor Mikhail Saltychev and Docent Jaana Paltamaa for guiding my steps into the wonders of science and academia. You were gentle, yet strong and passionate, fueling this study with your extensive skills, knowledge, and encouragement.

I would like to thank my reviewers Anne Söderlund and Aki Vainionpää for your excellent and in-depth suggestions and corrections on my thesis. Your positive and thoughtful feedback and reviewing helped me to reach my goal

I am truly thankful to Docent Katri Laimi for originally inviting me to the Turku ICF study group to learn more about science and academic writing. I warmly remember your encouraging, kind words: "Esa, in this group, we do no harm to each other." This I will carry with me in my further adventures in science. I thank professor emeritus Ari Heinonen for serving in my supervisory board.

I want to express my deepest gratitude to my co-authors: Mikhail Saltychev, Katri Laimi, Jaana Paltamaa, Ryan Mattie, Zack McCormick, and Niina Katajapuu for their invaluable collaboration. I am humbled by all your experience, skills and ease in scientific thinking and writing.

I am thankful for Alice Lehtinen (Altexta) and Kelly Raita for revising in-depth the English language of my thesis manuscript.

I sincerely thank you the Finnish Association of Physiotherapists for granting my thesis language revision process.

I also want to express my gratitude to all the physiotherapy bachelor's degree students I have taught since 2000. You have challenged my knowledge and skills; you have pushed me to go further, and you have helped me to understand.

Also, I thank you, my once-upon-a-time closest colleagues at Kankaanpää Rehabilitation Center: Leena, Seija, Virpi, Sari, Kati, Nella, Maria, and Marko. Working daily with you, shoulder-to-shoulder, treating neurological patients was such a joy, and it is the foundation of my physiotherapy work, knowledge, and interest in science. By tolerating my practical jokes, you made a difference.

I thank all my teacher colleagues on the bachelor's degree programme in Physiotherapy at the Satakunta University of Applied Sciences. Thank you for letting me slip so smoothly into the teaching world in 2000 and for encouraging me to further pursue science and improve my critical thinking.

My fellow teaching comrades at the Turku University of Applied Science, in both the physiotherapy and occupational therapy programs, I salute you and take a bow. I cherish our collegiality. Working with you every day preserved my sanity in the midst of my PhD studies and all the writing.

My friends, former colleagues and annual canastanians: Pekka, Teppo, Jouni, Jari, and Kimmo. Thank you for allowing me occasional wins, which I of course always earned. Those wins, everlasting talks, and your bad jokes have given me much needed relief during all the writing and studying. I value your friendship, and I am lucky to have you and, I guess, all that Elvis.

I am forever grateful to my friend and mentor in life and science, James Selfe. You have provided me with a valuable perspective of academia, science, and life. I have enormously enjoyed every step of our journey, whether on the Isle of Skye or on Säppi. I look forward to whatever there is left for us in this life. Slàinte mhath, mo charaid!

I also want to express my deepest gratitude to my parents Timo and Aila, to my sister Heidi, and to my brothers Kari and Ismo for all their support, opportunities, and encouragement on my path through life as a child and brother. We have come a long way since the snowball fights and snow castle-building.

Finally, and most importantly, my warmest thanks go to my wife Niina for her friendship, love, understanding, and scientific support. And for spurring me on. Thank you for walking, and sometimes running with me on the path that lay ahead. My sons, Henri and Niklas, the joys of my life, thank you for our wonderful life together.

And last but definitely not least, there is so much more than wisdom, knowledge and understanding in science:

> "*The fear of the Lord is the beginning of wisdom: and the knowledge of the holy is understanding*." Proverbs 9:10.

March 2023
Esa Bärlund

8 References

Abrams, E. M., Akombi, B., Alam, S., Alcalde-Rabanal, J. E., Allebeck, P., Amini-Rarani, M., . . . Yip, P. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet (British edition), 396*(10258), 1204-1222.

Algarni, A. S., Ghorbel, S., Jones, J. G., & Guermazi, M. (2014). Validation of an Arabic version of the Oswestry index in Saudi Arabia. *Ann Phys Rehabil Med, 57*(9-10), 653-663.

Asghari, A. (2011). Psychometric properties of a modified version of the Roland-Morris disability questionnaire (M-RMDQ). *Arch Iran Med, 14*(5), 327-331.

Baron, M., Schieir, O., Hudson, M., Steele, R., Kolahi, S., Berkson, L., . . . Zummer, M. (2008). The clinimetric properties of the World Health Organization Disability Assessment Schedule II in early inflammatory arthritis. *Arthritis and rheumatism, 59*(3), 382-390.

Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & Erbaugh, J. (1961). An inventory for measuring depression. *Arch Gen Psychiatry, 4*, 561-571.

Bergner, M., Bobbitt, R. A., Kressel, S., Pollard, W. E., Gilson, B. S., & Morris, J. R. (1976). The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. *Int J Health Serv, 6*(3), 393-415.

Bickenbach, J., Alarcos, C., Rauch, A., & Stucki, G. (Eds.). (2012). *ICF core sets : manual for clinical practice*. Cambridge, MA: Hogrefe.

Bombardier, C. (2000). Outcome assessments in the evaluation of treatment of spinal disorders: summary and general recommendations. *Spine (Phila Pa 1976), 25*(24), 3100-3103.

Brodke, D. S., Goz, V., Lawrence, B. D., Spiker, W. R., Neese, A., & Hung, M. (2017). Oswestry Disability Index: a psychometric analysis with 1,610 patients. *Spine J, 17*(3), 321-327.

Buist-Bouwman, M. A., Ormel, J., De Graaf, R., Vilagut, G., Alonso, J., Van Sonderen, E., & Vollebergh, W. A. (2008). Psychometric properties of the World Health Organization Disability Assessment Schedule used in the European Study of the Epidemiology of Mental Disorders. *Int J Methods Psychiatr Res, 17*(4), 185-197.

Bärlund, E. T., Katajapuu, N. K., Paltamaa, J. P., & Saltychev, M. M. (2021). Correlation between Oswestry disability index and 12-item self-administered version of World Health Organization Disability Assessment Schedule (WHODAS 2.0) in patients with chronic low back pain. *Int J Rehabil Res, 44*(2), 170-172.

Cella, D., Hahn, E., Jensen, S., Butt, Z., Nowinski, C., Rothrock, N., & Lohr, K. (2015). *Patient-Reported Outcomes in Performance Measurement.*: Research Triangle Park (NC): RTI Press.

Chapman, J. R., Norvell, D. C., Hermsmeyer, J. T., Bransford, R. J., DeVine, J., McGirt, M. J., & Lee, M. J. (2011). Evaluating common outcomes for measuring treatment success for chronic low back pain. *Spine (Phila Pa 1976), 36*(21 Suppl), S54-68.

Chiarotto, A., Boers, M., Deyo, R., Buchbinder, R., Corbin, T., Costa, L., . . . Ostelo, R. (2018). Core outcome measurement instruments for clinical trials in nonspecific low back pain. *Pain, 159*(3), 481-495.

Chiarotto, A., Maxwell, L., Ostelo, R., Boers, M., Tugwell, P., & Terwee, C. (2019). Measurement Properties of Visual Analogue Scale, Numeric Rating Scale, and Pain Severity Subscale of the

Brief Pain Inventory in Patients With Low Back Pain: A Systematic Review. *J Pain, 20*(3), 245-263.

Chiarotto, A., Terwee, C. B., & Ostelo, R. W. (2016). Choosing the right outcome measurement instruments for patients with low back pain. *Best Pract Res Clin Rheumatol, 30*(6), 1003-1020.

Childs, J. D., Piva, S. R., & Fritz, J. M. (2005). Responsiveness of the numeric pain rating scale in patients with low back pain. *Spine (Phila Pa 1976), 30*(11), 1331-1334.

Chow, J. H., & Chan, C. C. (2005). Validation of the Chinese version of the Oswestry Disability Index. *Work, 25*(4), 307-314.

Chwastiak, L. A., & Von Korff, M. (2003). Disability in depression and back pain. *Journal of Clinical Epidemiology, 56*(6), 507-514.

Cieza, A., Stucki, G., Weigl, M., Disler, P., Jäckel, W., van der Linden, S., . . . de Bie, R. (2004). ICF Core Sets for low back pain. *J Rehabil Med*(44 Suppl), 69-74.

Cleeland, C. (1989). Measurement of pain by subject report. In J. Chapman & J. Loeser (Eds.), *Issues in Pain Measurement* (pp. 391-403). New York: Raven Press.

Clement, R. C., Welander, A., Stowell, C., Cha, T. D., Chen, J. L., Davies, M., . . . Fritzell, P. (2015). A proposed set of metrics for standardized outcome reporting in the management of low back pain. *Acta Orthop, 86*(5), 523-533.

Ćwirlej-Sozańska, A., Bejer, A., Wiśniowska-Szurlej, A., Wilmowska-Pietruszyńska, A., de Sire, A., Spalek, R., & Sozański, B. (2020a). Psychometric Properties of the Polish Version of the 36-Item WHODAS 2.0 in Patients with Low Back Pain. *Int J Environ Res Public Health, 17*(19), 7284.

Ćwirlej-Sozańska, A., Sozański, B., Kotarski, H., Wilmowska-Pietruszyńska, A., & Wiśniowska-Szurlej, A. (2020b). Psychometric properties and validation of the polish version of the 12-item WHODAS 2.0. *BMC Public Health, 20*(1), 1203.

Davidson, M. (2008). Rasch analysis of three versions of the Oswestry Disability Questionnaire. *Man Ther, 13*(3), 222-231.

Dawson, A. P., Steele, E. J., Hodges, P. W., & Stewart, S. (2010). Utility of the Oswestry Disability Index for studies of back pain related disability in nurses: evaluation of psychometric and measurement properties. *Int J Nurs Stud, 47*(5), 604-607.

Denis, I., & Fortin, L. (2012). Development of a French-Canadian version of the Oswestry Disability Index: cross-cultural adaptation and validation. *Spine (Phila Pa 1976), 37*(7), E439-444.

Deyo, R. A., Battie, M., Beurskens, A. J., Bombardier, C., Croft, P., Koes, B., . . . Waddell, G. (1998). Outcome measures for low back pain research. A proposal for standardized use. *Spine (Phila Pa 1976), 23*(18), 2003-2013.

Deyo, R. A., Dworkin, S. F., Amtmann, D., Andersson, G., Borenstein, D., Carragee, E., . . . Weiner, D. K. (2014). Report of the NIH Task Force on research standards for chronic low back pain. *J Pain, 15*(6), 569-585.

Downie, W. W., Leatham, P. A., Rhind, V. M., Wright, V., Branco, J. A., & Anderson, J. A. (1978). Studies with pain rating scales. *Ann Rheum Dis, 37*(4), 378-381.

Dworkin, R. H., Turk, D. C., Farrar, J. T., Haythornthwaite, J. A., Jensen, M. P., Katz, N. P., . . . Witter, J. (2005). Core outcome measures for chronic pain clinical trials: IMMPACT recommendations. *Pain, 113*(1-2), 9-19.

Eprovide. (2022). Oswestry Disability Index (ODI). *Fairbank, J.* Date Accessed 27.12.2022 <https://eprovide.mapi-trust.org/instruments/oswestry-disability-index>.

Fairbank, J. (1995). Use of Oswestry Disability Index (ODI). *Spine (Phila Pa 1976), 20*(13), 1535-1537.

Fairbank, J., Couper, J., Davies, J., & O'Brien, J. (1980). The Oswestry low back pain disability questionnaire. *Physiotherapy, 66*(8), 271-273.

Fairbank, J., & Pynsent, P. (2000). The Oswestry Disability Index. *Spine, 25*(22), 2940-2952.

Farrar, J., Young, J. P., Jr., LaMoreaux, L., Werth, J. L., & Poole, M. R. (2001). Clinical importance of changes in chronic pain intensity measured on an 11-point numerical pain rating scale. *Pain, 94*(2), 149-158.

Fatoye, F., Gebrye, T., & Odeyemi, I. (2019). Real-world incidence and prevalence of low back pain using routinely collected data. *Rheumatol Int, 39*(4), 619-626.

Federici, S., Bracalenti, M., Meloni, F., & Luciano, J. V. (2017). World Health Organization disability assessment schedule 2.0: An international systematic review. *Disabil Rehabil, 39*(23), 2347-2380.

Feise, R. J., & Michael Menke, J. (2001). Functional rating index: a new valid and reliable instrument to measure the magnitude of clinical change in spinal conditions. *Spine (Phila Pa 1976), 26*(1), 78-86; discussion 87.

Fowler, F. (2009). Methods of Data Collection. *Survey Research Methods* (4th ed., pp. 68-85.). Thousand Oaks, California: SAGE Publications, Inc.

Froud, R., Patel, S., Rajendran, D., Bright, P., Bjørkli, T., Buchbinder, R., . . . Underwood, M. (2016). A Systematic Review of Outcome Measures Use, Analytical Approaches, Reporting Methods, and Publication Volume by Year in Low Back Pain Trials Published between 1980 and 2012: Respice, adspice, et prospice. *PLoS One, 11*(10), e0164573.

Gabel, C. P., Cuesta-Vargas, A., Qian, M., Vengust, R., Berlemann, U., Aghayev, E., & Melloh, M. (2017). The Oswestry Disability Index, confirmatory factor analysis in a sample of 35,263 verifies a one-factor structure but practicality issues remain. *Eur Spine J, 26*(8), 2007-2013.

Galli, T., Mirata, P., Foglia, E., Croce, D., Porazzi, E., Ferrario, L., . . . Banfi, G. (2018). A comparison between WHODAS 2.0 and Modified Barthel Index: which tool is more suitable for assessing the disability and the recovery rate in orthopedic rehabilitation? *Clinicoecon Outcomes Res, 10*, 301-307.

Garg, A., Pathak, H., Churyukanov, M. V., Uppin, R. B., & Slobodin, T. M. (2020). Low back pain: critical assessment of various scales. *Eur Spine J, 29*(3), 503-518.

Garin, O., Ayuso-Mateos, J. L., Almansa, J., Nieto, M., Chatterji, S., Vilagut, G., . . . Ferrer, M. (2010). Validation of the "World Health Organization Disability Assessment Schedule, WHODAS-2" in patients with chronic diseases. *Health Qual Life Outcomes, 8*(51), 1477-7525.

Gaskin, C. J., Lambert, S. D., Bowe, S. J., & Orellana, L. (2017). Why sample selection matters in exploratory factor analysis: implications for the 12-item World Health Organization Disability Assessment Schedule 2.0. *BMC Med Res Methodol, 17*(1), 40.

Grotle, M., Brox, J. I., & Vøllestad, N. K. (2004). Concurrent comparison of responsiveness in pain and functional status measurements used for patients with low back pain. *Spine (Phila Pa 1976), 29*(21), E492-501.

Hae-Jung, L., & Da-Jeong, K. (2011). Internal Consistency and Concurrent Validity of Korean Language Version of WHODAS 2.0: 12 Item-Self Administered *The Journal of Korean Physical Therapy, 23*(6), 23-29.

Hjermstad, M. J., Fayers, P. M., Haugen, D. F., Caraceni, A., Hanks, G. W., Loge, J. H., . . . Kaasa, S. (2011). Studies comparing Numerical Rating Scales, Verbal Rating Scales, and Visual Analogue Scales for assessment of pain intensity in adults: a systematic literature review. *J Pain Symptom Manage, 41*(6), 1073-1093.

Hong, S., & Shin, D. (2020). Relationship between pain intensity, disability, exercise time and computer usage time and depression in office workers with non-specific chronic low back pain. *Med Hypotheses, 137*, 109562.

Hudson, M., Steele, R., Taillefer, S., & Baron, M. (2008). Quality of life in systemic sclerosis: psychometric properties of the World Health Organization Disability Assessment Schedule II. *Arthritis Rheum, 59*(2), 270-278.

Hunt, S. M., McKenna, S. P., McEwen, J., Williams, J., & Papp, E. (1981). The Nottingham Health Profile: subjective health status and medical consultations. *Soc Sci Med A, 15*(3 Pt 1), 221-229.

Hush, J. M., Refshauge, K. M., Sullivan, G., De Souza, L., & McAuley, J. H. (2010). Do numerical rating scales and the Roland-Morris Disability Questionnaire capture changes that are meaningful to patients with persistent back pain? *Clin Rehabil, 24*(7), 648-657.

Huskisson, E. C. (1974). Measurement of Pain. *The Lancet, 304*(7889), 1127-1131.

Igwesi-Chidobe, C. N., Kitchen, S., Sorinola, I. O., & Godfrey, E. L. (2020). World Health Organisation Disability Assessment Schedule (WHODAS 2.0): development and validation of the Nigerian Igbo version in patients with chronic low back pain. *BMC Musculoskelet Disord, 21*(1), 755.

Jensen, M. P., & McFarland, C. A. (1993). Increasing the reliability and validity of pain intensity measurement in chronic pain patients. *Pain, 55*(2), 195-203.

Karlsson, E., & Gustafsson, J. (2021). Validation of the international classification of functioning, disability and health (ICF) core sets from 2001 to 2019 - a scoping review. *Disabil Rehabil*, 1-13.

Katajapuu, N. (2021). *Psychometric properties of 12-item World Health Organization disability assessment schedule 2.0 (WHODAS 2.0) amongst people with chronic musculoskeletal pain.* University of Jyväskylä, Jyväskylä. Date Accessed 27.12.2022 <http://urn.fi/URN:ISBN:978-951-39-8587-5>.

Katajapuu, N., Heinonen, A., & Saltychev, M. (2020). Minimal clinically important difference and minimal detectable change of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) amongst patients with chronic musculoskeletal pain. *Clin Rehabil, 34*(12), 1506-1511.

Katajapuu, N., Laimi, K., Heinonen, A., & Saltychev, M. (2019a). Floor and ceiling effects of the World Health Organization Disability Assessment Schedule 2.0 among patients with chronic musculoskeletal pain. *Int J Rehabil Res, 42*(2), 190-192.

Katajapuu, N., Laimi, K., Heinonen, A., & Saltychev, M. (2019b). Gender-related differences in psychometric properties of WHO Disability Assessment Schedule 2.0. *Int J Rehabil Res, 42*(4), 316-321.

Kimber, M., Rehm, J., & Ferro, M. (2015). Measurement Invariance of the WHODAS 2.0 in a Population-Based Sample of Youth. *PLoS One, 10*(11), e0142385.

Koponen, P., Borodulin, K., Lundqvist, A., Sääksjärvi, K., & Koskinen, S. (2018). Terveys, toimintakyky ja hyvinvointi Suomessa : FinTerveys 2017 -tutkimus *Report 4/2018* (pp. 236). Helsinki: National Institute for Health and Welfare (THL).

Kuntoutuksen tietopohja-hankkeen (KUTI) asiantuntijaryhmän jäsenet. (2020). Aikuisten toimintakyvyn itsearviointi kuntoutusmistarpeen tunnistamisessa ja kuntoutumisen seurannassa. Guideline in TOIMIA Functioning Measures Database. Date Accessed 27.12.2022. <https://www.terveysportti.fi/apps/dtk /tmi/article/tms00056>.

Lauridsen, H. H., Hartvigsen, J., Manniche, C., Korsholm, L., & Grunnet-Nilsson, N. (2006). Danish version of the Oswestry Disability Index for patients with low back pain. Part 1: Cross-cultural adaptation, reliability and validity in two different populations. *Eur Spine J, 15*(11), 1705-1716.

Lee, C. P., Fu, T. S., Liu, C. Y., & Hung, C. I. (2017). Psychometric evaluation of the Oswestry Disability Index in patients with chronic low back pain: factor and Mokken analyses. *Health Qual Life Outcomes, 15*(1), 192.

Lee, H. J., & Song, J. M. (2013). Analysis of the characteristics of Patients with Chronic Low Back Pain Using the ICF Concept. *The Journal of Korean Physical Therapy, 25*(5), 282-287.

Lochhead, L. E., & MacMillan, P. D. (2013). Psychometric properties of the Oswestry disability index: Rasch analysis of responses in a work-disabled population. *Work, 46*(1), 67-76.

Low back pain. (2017). Current Care Guidelines. Working group set up by the Finnish Medical Society and the Finnish Society of Physical and Rehabilitation Medicine. Helsinki: The Finnish Medical Society Duodecim. Date Accessed 27.12.2022 <www.kaypahoito.fi/hoi20001>.

Lu, Z., MacDermid, J. C., & Rosenbaum, P. (2020). A narrative review and content analysis of functional and quality of life measures used to evaluate the outcome after TSA: an ICF linking application. *BMC Musculoskelet Disord, 21*(1), 228.

Luciano, J. V., Ayuso-Mateos, J. L., Fernández, A., Serrano-Blanco, A., Roca, M., & Haro, J. M. (2010). Psychometric properties of the twelve item World Health Organization Disability Assessment Schedule II (WHO-DAS II) in Spanish primary care patients with a first major depressive episode. *J Affect Disord, 121*(1-2), 52-58.

Lue, Y. J., Hsieh, C. L., Huang, M. H., Lin, G. T., & Lu, Y. M. (2008). Development of a Chinese version of the Oswestry Disability Index version 2.1. *Spine (Phila Pa 1976), 33*(21), 2354-2360.

Mannion, A. F., Junge, A., Fairbank, J., Dvorak, J., & Grob, D. (2006). Development of a German version of the Oswestry Disability Index. Part 1: cross-cultural adaptation, reliability, and validity. *Eur Spine J, 15*(1), 55-65.

Marom, B. S., Carel, R. S., Sharabi, M., & Ratzon, N. Z. (2017). Cross-cultural adaptation of the World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) for Hebrew-speaking subjects with and without hand injury. *Disabil Rehabil, 39*(12), 1155-1161.

Maughan, E. F., & Lewis, J. S. (2010). Outcome measures in chronic low back pain. *Eur Spine J, 19*(9), 1484-1494.

McGorry, R. W., Shaw, W. S., & Lin, J. H. (2011). Correlations between pain and function in a longitudinal low back pain cohort. *Disabil Rehabil, 33*(11), 945-952.

Melzack, R. (1975). The McGill Pain Questionnaire: major properties and scoring methods. *Pain, 1*(3), 277-299.

Miekisiak, G., Kollataj, M., Dobrogowski, J., Kloc, W., Libionka, W., Banach, M., . . . Gierlotka, M. (2013). Validation and cross-cultural adaptation of the Polish version of the Oswestry Disability Index. *Spine (Phila Pa 1976), 38*(4), E237-243.

Miller, R. P., Kori, S. H., & Todd, D. D. (1991). The Tampa Scale: a Measure of Kinisophobia. *The Clinical Journal of Pain, 7*(1), 51.

Monticone, M., Baiardi, P., Ferrari, S., Foti, C., Mugnai, R., Pillastrini, P., . . . Zanoli, G. (2009). Development of the Italian version of the Oswestry Disability Index (ODI-I): A cross-cultural adaptation, reliability, and validity study. *Spine (Phila Pa 1976), 34*(19), 2090-2095.

Monticone, M., Baiardi, P., Vanti, C., Ferrari, S., Pillastrini, P., Mugnai, R., & Foti, C. (2012). Responsiveness of the Oswestry Disability Index and the Roland Morris Disability Questionnaire in Italian subjects with sub-acute and chronic low back pain. *Eur Spine J, 21*(1), 122-129.

Moon, J., Kim, Y. C., Park, S. Y., Lee, S. C., Choi, S. P., Nahm, F. S., . . . Kang, J. M. (2011). Psychometric characteristics of the Korean version of the Roland-Morris Disability Questionnaire. *J Korean Med Sci, 26*(10), 1364-1370.

Moreira, A., Alvarelhão, J., Silva, A., Costa, R., & Queirós, A. (2015). Validation of a Portuguese version of WHODAS 2.0: 12 items in people aged 55 or more. *Revista Portuguesa de Saúde Pública, 33*, 179-182.

Ostelo, R. W., & de Vet, H. C. (2005). Clinically important outcomes in low back pain. *Best Pract Res Clin Rheumatol, 19*(4), 593-607.

Osthus, H., Cziske, R., & Jacobi, E. (2006). Cross-cultural adaptation of a German version of the Oswestry Disability Index and evaluation of its measurement properties. *Spine (Phila Pa 1976), 31*(14), E448-453.

Paltamaa, J. (2014). *WHODAS 2.0 : terveyden ja toimintarajoitteiden arvioinnin käsikirjan osat 2 & 3*. Jyväskylä: Jyväskylän ammattikorkeakoulu.

Paungmali, A., Sitilertpisan, P., Taneyhill, K., Pirunsan, U., & Uthaikhup, S. (2012). Intrarater reliability of pain intensity, tissue blood flow, thermal pain threshold, pressure pain threshold and lumbo-pelvic stability tests in subjects with low back pain. *Asian J Sports Med, 3*(1), 8-14.

Payares, K., Lugo, L. H., Morales, V., & Londoño, A. (2011). Validation in Colombia of the Oswestry disability questionnaire in patients with low back pain. *Spine (Phila Pa 1976), 36*(26), E1730-1735.

Pekkanen, L., Kautiainen, H., Ylinen, J., Salo, P., & Hakkinen, A. (2011). Reliability and validity study of the Finnish version 2.0 of the oswestry disability index. *Spine, 36*(4), 332-338.

Pekkanen, L., Salmenkivi, J., Salo, J., Haapakoski, J., Järvelin, J., Keinänen, M., . . . Malmivaara, A. (2019). Selkärekisterin loppuraportti. Date Accessed 26.2.2021. <https://thl.fi/documents/2616650/4353715/Selkarekisterin loppu raportti.pdf/6427dde5-213b-bdd4-ccdc-6e645526d7a7?t=1607518843199>.

Perales, J., Martin, S., Ayuso-Mateos, J. L., Chatterji, S., Garin, N., Koskinen, S., . . . Haro, J. M. (2014). Factors associated with active aging in Finland, Poland, and Spain. *Int Psychogeriatr, 26*(8), 1363-1375.

Phedy, P., Djaja, Y. P., Tobing, S., Gatam, L., Librianto, D., Fachrisal, . . . Hardiansyah, N. P. (2021). Cross-cultural adaptation and psychometric validation of the Indonesian version of the Oswestry Disability Index. *Eur Spine J, 30*(4), 1053-1062.

Prinsen, C. A., Vohra, S., Rose, M. R., Boers, M., Tugwell, P., Clarke, M., . . . Terwee, C. B. (2016). How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. *Trials, 17*(1), 449.

Raheem, S., Ibrahim, A. A., Ganiyu, S. O., Faruk, A. U., & Akindele, M. O. (2021). Translation, Cross-cultural Adaptation and Psychometric Evaluation of the Hausa Roland-Morris Disability Questionnaire in Mixed Rural and Urban Nigerian Populations with Low Back Pain. *Spine (Phila Pa 1976), 46*(11), E639-e647.

Ramasamy, A., Martin, M. L., Blum, S. I., Liedgens, H., Argoff, C., Freynhagen, R., . . . Patrick, D. L. (2017). Assessment of Patient-Reported Outcome Instruments to Assess Chronic Low Back Pain. *Pain Med, 18*(6), 1098-1110.

Rasmussen, C. D. N., Holtermann, A., & Jørgensen, M. B. (2018). Recall Bias in Low Back Pain Among Workers: Effects of Recall Period and Individual and Work-Related Factors. *Spine (Phila Pa 1976), 43*(12), E727-e733.

Rijanti, K. A., Subadi, I., & Kurniawati, P. M. (2021). VALIDITY AND RELIABILITY OF WHO DISABILITY ASSESSMENT SCHEDULE 2.0 (WHODAS 2.0) TYPE 12 QUESTIONS INDONESIAN VERSION ON BACK PAIN. *Folia Medica Indonesiana (2355-8393), 57*(3), 220-225.

Robinson-Papp, J., George, M. C., Dorfman, D., & Simpson, D. M. (2015). Barriers to Chronic Pain Measurement: A Qualitative Study of Patient Perspectives. *Pain Med, 16*(7), 1256-1264.

Roe, C., Sveen, U., & Bautz-Holter, E. (2008). Retaining the patient perspective in the International Classification of Functioning, Disability and Health Core Set for low back pain. *Patient Prefer Adherence, 2*, 337-347.

Roland, M., & Fairbank, J. (2000). The Roland-Morris Disability Questionnaire and the Oswestry Disability Questionnaire. *Spine (Phila Pa 1976), 25*(24), 3115-3124.

Roland, M., & Morris, R. (1983). A study of the natural history of back pain. Part I: development of a reliable and sensitive measure of disability in low-back pain. *Spine (Phila Pa 1976), 8*(2), 141-144.

Rubin, D. I. (2007). Epidemiology and risk factors for spine pain. *Neurol Clin, 25*(2), 353-371.

Saltychev, M., Bärlund, E., Mattie, R., McCormick, Z., Paltamaa, J., & Laimi, K. (2017a). A study of the psychometric properties of 12-item World Health Organization Disability Assessment Schedule 2.0 in a large population of people with chronic musculoskeletal pain. *Clin Rehabil, 31*(2), 262-272.

Saltychev, M., Katajapuu, N., Bärlund, E., & Laimi, K. (2021). Psychometric properties of 12-item self-administered World Health Organization disability assessment schedule 2.0 (WHODAS 2.0) among general population and people with non-acute physical causes of disability - systematic review. *Disabil Rehabil, 43*(6), 789-794.

Saltychev, M., Kinnunen, A., & Laimi, K. (2013). Associations between rehabilitants' functional limitations and conclusions made by a multidisciplinary team during vocational rehabilitation evaluation: Turku ICF Study. *Open Journal of Therapy and Rehabilitation, 1*(2), 36-39.

Saltychev, M., Mattie, R., McCormick, Z., & Laimi, K. (2017b). Confirmatory factor analysis of 12-Item World Health Organization Disability Assessment Schedule in patients with musculoskeletal pain conditions. *Clin Rehabil, 31*(5), 702-709.

Schiavolin, S., Ferroli, P., Acerbi, F., Brock, S., Broggi, M., Cusin, A., . . . Raggi, A. (2014). Disability in Italian neurosurgical patients: validity of the 12-item World Health Organization Disability Assessment Schedule. *Int J Rehabil Res, 37*(3), 267-270.

Selva-Sevilla, C., Ferrara, P., & Gerónimo-Pardo, M. (2019). Psychometric Properties Study of the Oswestry Disability Index in a Spanish Population With Previous Lumbar Disc Surgery: Homogeneity and Validity. *Spine (Phila Pa 1976), 44*(7), E430-e437.

Shafshak, T. S., & Elnemr, R. (2021). The Visual Analogue Scale Versus Numerical Rating Scale in Measuring Pain Severity and Predicting Disability in Low Back Pain. *J Clin Rheumatol, 27*(7), 282-285.

Sheahan, P. J., Nelson-Wong, E. J., & Fischer, S. L. (2015). A review of culturally adapted versions of the Oswestry Disability Index: the adaptation process, construct validity, test-retest reliability and internal consistency. *Disabil Rehabil, 37*(25), 2367-2374.

Sigl T, Cieza A, Brockow T, Chatterji S, Kostanjsek N, & Stucki G. (2006). Content comparison of low back pain-specific measures based on the International Classification of Functioning, Disability and Health (ICF). *Clin J Pain, 22*(2), 147-153.

Silva, A., Cerqueira, M., Santos, A., Ferreira, C., Alvarelhão, J., & Queirós, A. (2019). Inter-rater reliability, standard error of measurement and minimal detectable change of the 12-item WHODAS 2.0 and four performance tests in institutionalized ambulatory older adults. *Disabil Rehabil, 41*(3), 366-373.

Silva, A. G., Queirós, A., & Rocha, N. P. (2016). Generic self-reported and performance based instruments: How to capture pain associated disability. *Revista portuguesa de saúde pública, 34*(2), 125-133.

Silva, C., Coleta, I., Silva, A. G., Amaro, A., Alvarelhão, J., Queirós, A., & Rocha, N. (2013). Adaptation and validation of WHODAS 2.0 in patients with musculoskeletal pain. *Revista de saude publica, 47*(4), 752-758.

Smedema, S., Ruiz, D., & Mohr, M. (2017). Psychometric Validation of the World Health Organization Disability Assessment Schedule 2.0—Twelve-Item Version in Persons With Spinal Cord Injuries. *Rehabilitation Research, Policy, and Education, 31*(1), 7-20.

Smedema, S., Yaghmaian, R., Ruiz, D., Muller, V., Emre, U., & Fong, C. (2016). Psychometric Validation of the World Health Organization Disability Assessment Schedule 2.0 - 12-Item Version in Persons with Fibromyalgia Syndrome. *Journal of Rehabilitation, 82*(3), 28-35.

Smeets, R., Koke, A., Lin, C.-W. C., Ferreira, M. L., & Demoulin, C. (2011). Measures of Function in Low Back Pain/Disorders Low Back Pain Rating Scale (LBPRS), Oswestry Disability Index (ODI), Progressive Isoinertial Lifting Evaluation (PILE), Quebec Back Pain Disability Scale (QBPDS), and Roland-Morris Disability Questionnaire (RDQ). *Arthritis care & research (2010), 63*(S11), S158-S173.

Social Insurance Institution of Finland. (2021). *Statistical Yearbook of the Social Insurance Institution 2020.* Helsinki: The Social Insurance Institution of Finland. Date Accessed 3.12.2022 <https://helda.helsinki.fi/handle/10138/336533>.

Solomon, D., & Roopchand-Martin, S. (2011). How well do pain scales correlate with each other and with the Oswestry Disability Questionnaire? *International Journal of Therapy and Rehabilitation, 18*(2), 108-113.

Sousa, R. M., Dewey, M. E., Acosta, D., Jotheeswaran, A. T., Castro-Costa, E., Ferri, P., . . . Prince, M. J. (2010). Measuring disability across cultures--the psychometric properties of the WHODAS II in older people from seven low- and middle-income countries. The 10/66 Dementia Research Group population-based survey. *Int J Methods Psychiatr Res, 19*(1), 1-17.

Streiner, D., Norman, G., & Cairney, J. (2015). Construct validity. *Health measurement scales : a practical guide to their development and use* (5th ed. ed., pp. 240). New York, New York: Oxford University Press.

Suzuki, H., Aono, S., Inoue, S., Imajo, Y., Nishida, N., Funaba, M., . . . Sakai, T. (2020). Clinically significant changes in pain along the Pain Intensity Numerical Rating Scale in patients with chronic low back pain. *PLoS One, 15*(3), e0229228.

Tarvonen-Schröder, S., Kaljonen, A., & Laimi, K. (2019). Utility of the World Health Organization Disability Assessment Schedule and the World Health Organization minimal generic set of domains of functioning and health in spinal cord injury. *J Rehabil Med, 51*(1), 40-46.

Taylor, A. M., Phillips, K., Patel, K. V., Turk, D. C., Dworkin, R. H., Beaton, D., . . . Witter, J. (2016). Assessment of physical function and participation in chronic pain clinical trials: IMMPACT/OMERACT recommendations. *Pain, 157*(9), 1836-1850.

Tazaki, M., Yamaguchi, T., Yatsunami, M., & Nakane, Y. (2014). Measuring functional health among the elderly: development of the Japanese version of the World Health Organization Disability Assessment Schedule II. *Int J Rehabil Res, 37*(1), 48-53.

Terwee, C. B., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., . . . de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol, 60*(1), 34-42.

Valasek, T., Varga, P. P., Szoverfi, Z., Kumin, M., Fairbank, J., & Lazary, A. (2013). Reliability and validity study on the Hungarian versions of the oswestry disability index and the Quebec back pain disability scale. *Eur Spine J, 22*(5), 1010-1018.

van Hooff, M. L., Spruit, M., Fairbank, J., van Limbeek, J., & Jacobs, W. C. (2015). The Oswestry Disability Index (version 2.1a): validation of a Dutch language version. *Spine (Phila Pa 1976), 40*(2), E83-90.

van Tubergen, A., Landewé, R., Heuft-Dorenbosch, L., Spoorenberg, A., van der Heijde, D., van der Tempel, H., & van der Linden, S. (2003). Assessment of disability with the World Health Organisation Disability Assessment Schedule II in patients with ankylosing spondylitis. *Annals of the rheumatic diseases, 62*(2), 140-145.

Vigatto, R., Alexandre, N. M., & Correa Filho, H. R. (2007). Development of a Brazilian Portuguese version of the Oswestry Disability Index: cross-cultural adaptation, reliability, and validity. *Spine (Phila Pa 1976), 32*(4), 481-486.

Waddell, G., Newton, M., Henderson, I., Somerville, D., & Main, C. J. (1993). A Fear-Avoidance Beliefs Questionnaire (FABQ) and the role of fear-avoidance beliefs in chronic low back pain and disability. *Pain, 52*(2), 157-168.

Ware, J. (2000). SF-36 health survey update. *Spine (Phila Pa 1976), 25*(24), 3130-3139.

Ware, J., Jr., Kosinski, M., & Keller, S. D. (1996). A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care, 34*(3), 220-233.

Ware, J. E., Jr., & Sherbourne, C. D. (1992). The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care, 30*(6), 473-483.

Weigl, M., Cieza, A., Andersen, C., Kollerits, B., Amann, E., & Stucki, G. (2004). Identification of relevant ICF categories in patients with chronic health conditions: a Delphi exercise. *J Rehabil Med*(44 Suppl), 12-21.

White, L. J., & Velozo, C. A. (2002). The use of Rasch measurement to improve the Oswestry classification scheme. *Arch Phys Med Rehabil, 83*(6), 822-831.

Wong, J. J., DeSouza, A., Hogg-Johnson, S., De Groote, W., Southerst, D., Belchos, M., . . . Côté, P. (2022). Measurement properties and minimal important change of the World Health Organization Disability Assessment Schedule 2.0 in persons with low back pain: A systematic review (In Press). *Arch Phys Med Rehabil*, 1-15. Date Accessed 29.12.2022. <https://doi.org/10.1016/j.apmr.2022.06.005>.

World Health Organization. (1980). *International Classification of Impairments, Disabilities, and Handicaps. A manual of classification relating to the consequences of disease.* Geneva: World Health Organization.

World Health Organization. (1997). WHOQOL : measuring quality of life. In H. Division of Mental & A. Prevention of Substance (Eds.). Geneva: World Health Organization.

World Health Organization. (2001). International classification of functioning, disability and health : ICF. Geneva: World Health Organization.

World Health Organization. (2021). WHO Family of International Classifications Network. Date Accessed 13.5.2021 <https://www.who.int/groups/who-family-of-international-classifications-network>.

World Health Organization (Ed.). (2010). *Measuring health and disability : manual for WHO Disability Assessment Schedule WHODAS 2.0*. Geneva: World Health Organization.

Wu, A., March, L., Zheng, X., Huang, J., Wang, X., Zhao, J., . . . Hoy, D. (2020). Global low back pain prevalence and years lived with disability from 1990 to 2017: estimates from the Global Burden of Disease Study 2017. *Ann Transl Med, 8*(6), 299.

Xenouli, G., Xenoulis, K., Sarafis, P., Niakas, D., & Alexopoulos, E. C. (2016). Validation of the World Health Organization Disability Assessment Schedule (WHO-DAS II) in Greek and its added value to the Short Form 36 (SF-36) in a sample of people with or without disabilities. *Disabil Health J, 9*(3), 518-523.

Yao, M., Wang, Q., Li, Z., Yang, L., Huang, P. X., Sun, Y. L., . . . Cui, X. J. (2016). A Systematic Review of Cross-cultural Adaptation of the Oswestry Disability Index. *Spine (Phila Pa 1976), 41*(24), E1470-e1478.

Üstün, T. B., Chatterji, S., Kostanjsek, N., Rehm, J., Kennedy, C., Epping-Jordan, J., . . . Pull, C. (2010). Developing the World Health Organization Disability Assessment Schedule 2.0. *Bulletin of the World Health Organization, 88*, 815-823.

TURUN
YLIOPISTO
UNIVERSITY
OF TURKU

Painosalama, Turku, Finland 2023

ISBN 978-951-29-9135-8 (PRINT)
ISBN 978-951-29-9136-5 (PDF)
ISSN 0355-9483 (Print)
ISSN 2343-3213 (Online)