



**WORLD BANK GROUP**



Funded by the European Union  
via the Structural Reform Support  
Programme and implemented  
by the World Bank, in cooperation  
with the European Commission



# Options for Including Functioning into Disability and Work Capacity Assessment in **Latvia**

---

Carolina Fellinghauer, Aleksandra Posarac, Jerome Bickenbach and Marijana Jasarevic

Report No: AUS0002809

April 8, 2022

© 2017 The World Bank  
1818 H Street NW, Washington DC 20433  
Telephone: 202-473-1000; Internet: [www.worldbank.org](http://www.worldbank.org)

Some rights reserved

This work is a product of the staff of The World Bank. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of the Executive Directors of The World Bank or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

### **Rights and Permissions**

The material in this work is subject to copyright. Because The World Bank encourages dissemination of its knowledge, this work may be reproduced, in whole or in part, for noncommercial purposes as long as full attribution to this work is given.

**Attribution**—Please cite the work as follows: “World Bank. 2022. Options for including Functioning into disability and work capacity assessment in Latvia. © World Bank.”

All queries on rights and licenses, including subsidiary rights, should be addressed to World Bank Publications, The World Bank Group, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: [pubrights@worldbank.org](mailto:pubrights@worldbank.org). “

The report should be cited as follows: Carolina Fellinghauer, Aleksandra Posarac, Jerome Bickenbach and Marijana Jasarevic. 2022. *Options for including Functioning into disability and work capacity assessment in Latvia*. © World Bank.

Report No: AUS0002809

## **Options for including functioning into disability and work capacity assessment in Latvia**

Carolina Fellinghauer, Aleksandra Posarac, Jerome Bickenbach and Marijana Jasarevic

April 8, 2022

SPL



## Table of Contents

Acknowledgements .....	1
Summary .....	2
1. About disability, disability assessment and WHODAS.....	2
2. Piloting the collection of information on functioning: the WHODAS 2.0 pilot .....	2
3. The analysis .....	3
i. Psychometric properties of WHODAS 2.0 in Latvia .....	3
ii. Performance of the current assessment of disability in Latvia.....	4
4. Options for including functioning in disability assessment in Latvia .....	7
i. The challenge .....	7
ii. Proposal for WHODAS disability severity bands .....	8
iii. Options for including functioning information into the current Latvian disability assessment system.....	9
Latvia: Options for including functioning into disability assessment .....	12
1. WHODAS: Technical details.....	12
2. Descriptive statistics of the WHODAS pilot sample .....	14
3. Psychometric analysis: Rationale and tests.....	16
4. Results - Metric properties of WHODAS .....	18
5. Summary: Psychometric properties of WHODAS.....	27
6. Exploring the metric properties of the WHODAS domains .....	28
7. Performance of the current assessment of disability in Latvia .....	31
7.1. Comparing disability status group determination and WHODAS scores by case studies: .....	32
7.2 Comparing discrimination between disability status groups and WHODAS scores .....	32
7.3 Comparing disability status groups with WHODAS scores in terms of correlations with ICD chapters .....	35
8. The self-assessment form and collection and use of data on functioning for the assessment .....	37
8.1 Descriptive statistics .....	37
8.2 Psychometric characteristics of the 21 ICF categories from the self-assessment form .....	41
9. Options for including functioning in disability assessment.....	43
9.1 Proposals for WHODAS disability severity bands .....	44
9.2 Distribution of WHODAS scores using the proposed Rasch based cut offs.....	46
9.3 Options for changing the current Latvian disability assessment system.....	49
References .....	54
Appendix 1 .....	56

## List of Figures

Figure 1: WHODAS-score density line by determined disability status groups .....	5
Figure 2: Row Score Distributions of the WHODAS .....	18
Figure 3: Person item map after the creation of testlet .....	21
Figure 4: Local item dependencies before the creation of testlets .....	22
Figure 5: Person item map after the creation of testlets .....	23
Figure 6: WHODAS-score density line by determined disability status groups .....	33
Figure 7: Percentage of self-reported problems in functioning by the CMS determined disability groups.....	39
Figure 8: Local item dependencies between the ICF-categories used for the self-assessment. ....	41

Figure 9: Person item map for the ICF-categories, without any adjustments of the data .....	42
Figure 10: Boxplot - Terminology.....	45
Figure 11: WHODAS-score distribution by disability status using WHODAS cut points for the Rasch-based scores (0-100) .....	50
Figure 12: WHODAS-score distribution by disability status for ICD-codes I6 – Cerebrovascular diseases using WHODAS cut points for the Rasch-based scores (0-100) .....	50

## List of Tables

Table 1: WHODAS 2.0 items for the 36-item long form .....	13
Table 2: Pilot sample descriptive statistics .....	15
Table 3: Prevalence of Health conditions in the study population by ICD-10 Health Condition Category. 16	
Table 4: Frequencies and Percentages of WHODAS Responses .....	19
Table 5: WHODAS fit, difficulties, threshold ordering, local item dependencies, and differential item functioning without any adjustments.....	24
Table 6: WHODAS Item Difficulties, fit, local item dependencies, and differential item functioning after adjustments .....	25
Table 7: Targeting and Reliability of WHODAS items .....	25
Table 8: Transformation Table.....	25
Table 9: Targeting, Item Parameter characteristics, Reliability, and Dimensionality of the WHODAS domains before and after adjustment.....	29
Table 10: Infit and outfit, differential item functioning, local item dependencies, recoding strategy by disordered thresholds.....	30
Table 11: WHODAS scores distribution for the pilot sample and by disability status group: mean, standard deviation and quartiles .....	33
Table 12: Analysis of variance of the WHODAS-scores by disability status groups.....	34
Table 13: Tukey Test for the WHODAS-scores by disability status groups .....	34
Table 14: Pilot sample descriptive statistics by disability status groups .....	34
Table 15: Frequency and Percentage of ICD chapters for the pilot sample and by disability status group as well as the mean and standard deviation (SD) of the corresponding WHODAS-scores.....	36
Table 16: ICF-based rating of levels of functioning problems .....	38
Table 17: ICF categories fit, difficulties, threshold disordering, local item dependencies, and differential item functioning without any adjustments .....	43
Table 18: Rasch-based 0-100 WHODAS-score ranges in Latvia, Lithuania, and Greece – suggested cut-points .....	45
Table 19: WHODAS-based disability severity – descriptive statistics .....	46
Table 20: WHODAS-based disability level versus medical disability status.....	47
Table 21: Frequency and percentage of ICD chapters for the pilot sample and by WHODAS-based functioning level as well as the mean and standard deviation (SD) of the corresponding WHODAS-scores .....	48

## List of Boxes

Box 1: The credibility of disability assessment .....	6
Box 2: The credibility of disability assessment .....	37

## Abbreviations

DG REFORM	Directorate-General for Structural Reform Support of the European Commission
DIF	Differential item functioning
ICD-10	International Classification of Diseases, Tenth Revision
LID	Local item dependency
MOW	Ministry of Welfare of the Republic of Latvia
SMC	State Medical Commission for the Assessment of Health Condition and Work Ability
WB	World Bank
WHO	World Health Organization
WHO ICF	World Health Organization's International Classification of Functioning, Disability and Health
WHODAS 2.0	World Health Organization Disability Assessment Schedule 2.0

## Acknowledgements

This Report was prepared under the Directorate-General for Structural Reform Support of the European Commission (DG REFORM) funded technical assistance project “Latvia - Disability assessment system development” implemented by the World Bank.

This Report is a result of a three-year effort that began in May 2019. It was prepared by a joint team from the World Bank (WB), DG REFORM and the Ministry of Welfare (MOW) of the Republic of Latvia. The team was led by Aleksandra Posarac, Lead Economist and Project Manager (WB) until September 2021 and then by Marijana Jasarevic, Social Protection Specialist (WB). The team also included Carolina Fellinghauer, Lead Statistical Analyst, University of Zurich, Department of Psychology, Chair for Psychological Methods, Evaluation and Statistics, Professor Jerome Bickenbach, Lead Technical Expert (Swiss Paraplegic Society and University of Luzern) and Guna Bērziņa, Assistant Professor and Senior Researcher, Faculty of Rehabilitation, Rīga Stradiņš University. Marc Vothknecht, Claudia Piferi and Oana Dimitrescu from DG REFORM provided valuable comments and suggestions. The team worked closely with, Elīna Celmiņa, Deputy Minister, (MOW), and Dace Kampenusa, Senior Expert in the Social Inclusion Policy Department (MOW).

The team is thankful to Cem Mete (World Bank Manager), Lars Sondergaard (World Bank Program Leader) and Geraldine Mahieu from DG REFORM for their continuous overall guidance and support. Finally, the team wishes to extend its deep gratitude to the Ministry of Welfare, without whose commitment and enormous engagement, this study would not have been possible.

## Summary

The Latvian Government is seeking to strengthen its disability assessment system by more effectively combining medical and functioning information.<sup>1</sup> Including functioning information into disability assessment requires empirical evidence from testing a functioning data collection instrument. To that end, a psychometric instrument for assessing disability, developed, and extensively tested by the World Health Organization (WHO), the Disability Assessment Schedule (WHODAS 2.0), was pilot tested in Latvia. This report presents details of the pilot, statistical analysis of the data collected and a proposal for including functioning information into the disability assessment procedure in Latvia.

### 1. About disability, disability assessment and WHODAS

In the World Health Organization's International Classification of Functioning, Disability and Health (WHO ICF) information about categories of Activities and Participation can be collected either from the ***perspective of capacity*** (reflecting exclusively the expected ability of a person to perform activities considering their health conditions and impairments) or the ***perspective of performance*** (reflecting the actual performance of activities in the real-world environmental circumstances in which the person lives). Information about capacity typically represents the results of a clinical inference or judgment based on medical information, while performance is a true description of what occurs in a person's life. The two perspectives are therefore very different, although capacity constitutes a determinant of performance.

A disability assessment is a summary measure of the level of a person's performance of an adequately representative set of behaviors and actions, simple to complex, in their actual environment, considering the person's state of health.

The WHO developed, tested and has consistently recommended the WHODAS as an instrument that can validly and reliably capture the performance of activities by an individual in his or her daily lives and actual environment. The 'actual environment' is represented in the ICF in terms of environmental factors that act either as environmental facilitators (e.g., assistive devices, supports, home modifications) or as environmental barriers (inaccessible houses, streets and public buildings, stigma, and discrimination). The WHODAS questionnaire, in short, is WHO's recommended, generic, performance-based disability assessment tool.

### 2. Piloting the collection of information on functioning: the WHODAS 2.0 pilot

WHODAS 36-question version administered by a trained professional was used to pilot test its reliability and validity in Latvia. The pilot was implemented from the beginning of July until the end of October 2021. A total of 2,202 persons who applied for a disability assessment in the referenced period were included in the pilot. The pilot overlapped with the CORONA pandemic. This resulted in a high number of interviews that could not take place in a face-to-face mode. In fact, only 14% of the interviews took place face-to face, while 83% were phone-interviews or interviews via WhatsApp (3%).

---

<sup>1</sup> In Latvia, disability and work capacity is assessed and certified by the State Medical Commission for the Assessment of Health Condition and Work Ability (SMC). For detailed discussion about the assessment system see: Aleksandra Posarac, Elina Celmina and Jerome Bickenbach. 2020. Disability Policy and Disability Assessment System in Latvia © World Bank.



The demographic characteristics of the participants in the survey were as follows: the proportion of female participants was higher (58.4% vs. 41.6% respectively); the average age was 58.46 years (SD = 13.57); most participants were married (45.5%), 13.1% were widowed, and 13.6% were cohabiting; most participants were living independently in the community (81.1%); the participants had an average of 12.53 (SD = 3.03) years of education; most reported either having a paid employment (31.2%), being retired (25.2%), and unemployed for health reasons (21.5%).

A total of N = 1,220 (55.4%) pilot participants were diagnosed with only one ICD-10 linked health condition, while N = 982, 44.6% reported one health condition with additional comorbidities. Diseases of the musculoskeletal system and connective tissue (N = 497, 23.01%) and diseases of the circulatory system (N = 372, 17.22%) were the most reported main diagnoses. Neoplasms were reported by N = 388 (17.96%) participants. ICD chapter XIX, external causes such as injuries, was seen as primary diagnosis in about 18% of the participants (N = 385, 17.82%). It is important to notice that **only 2.4% of the assessed population has an ICD-code from the chapter V Mental and behavioral disorders, and 50% of these group has some form of dementia.** In our experience this is a common phenomenon in disability assessment systems, probably caused by the fact that mental and behavioral disorders, other than dementias associated with ageing are under-diagnosed.

### 3. The analysis

Below, we summarize the results of the data analysis. First, we analyze metric and psychometric properties of the WHODAS pilot data. Second, we look at the current disability assessment method outcomes of the pilot participants as compared to WHODAS assessment and also analyze psychometric properties of the self-assessment form through which the current assessment system collects data on functioning.

#### i. Psychometric properties of WHODAS 2.0 in Latvia

A statistical analysis of psychometric properties of WHODAS pilot in Latvia that included seven essential statistical tests (described in the main text below) show that the data collected with WHODAS, under the Rasch analysis, display robust psychometric properties of validity and reliability. With a few adjustments, the scale is unidimensional and free of item dependencies with good targeting and with good reliability. Aggregating the items by domains solves observed local item dependencies and produces a unidimensional assessment metric. The domain-based testlets fit well, and a transformation table is obtained that translates the observed sum scores into an interval-scaled metric. It is important to keep in mind that the World Health Organization developed WHODAS explicitly to statistically capture the construct of functioning from the perspective of performance – namely the actual experience of performing activities by a person with an underlying health problem in their actual everyday life environment. There is an abundance of evidence from the scientific literature – supported by the results of this pilot – that WHODAS is a psychometrically sound instrument that reliably and validly collects information about levels of disability. ***In conclusion, WHODAS information is sufficiently robust and relevant, and we recommend that it is applied in the assessment of disability in Latvia in its shift from medical to a functioning based assessment.***

## ii. Performance of the current assessment of disability in Latvia

Detailed description of the disability assessment system in Latvia is provided in the 2020 World Bank Report *Disability Policy and Disability Assessment System in Latvia*.<sup>2</sup> In brief, in Latvia, the assessment is conducted by the State Medical Commission and mostly based on medical information. Information on functioning is included in the self-assessment form. It is our understanding that the functioning information is not systematically used in the disability status assessment. Persons with disability or reduced work capacity are categorized into these **disability status groups** by an ordinal scale:

- No disability - **no functioning restrictions** (a loss of general ability to work for up to 24.99 percent, not regarded as a disability for the purposes of the assessment).
- Group III disability - functioning restrictions are **moderate** (a loss of ability to work is assessed at 25.0-59.99 percent).
- Group II disability - functioning restrictions are **severe** (a loss of ability to work is 60.0-79.99 percent).
- Group I disability - functioning restrictions are **very severe** (a loss of ability to work is assessed as 80.0-100.0 percent).

While the disability status groups in the current system are described in terms of functioning limitations, in practice the only source of information used to assign an applicant to one of the four status groups is medical. Thus, the current assessment infers functioning or performance state from the medical information. From the perspective of the modern understanding of disability – the outcome of interactions between a person with a health condition and her or his environment - this is not correct and limits disability to medical state only.

By comparing disability status group determination and WHODAS scores by individual case studies and by comparing discrimination between disability status groups and WHODAS scores we demonstrate that the current system of determining disability status groups has problems to create disability groups, ordinally ranked by severity, when compared to WHODAS scores.

*Comparing disability status group determination and WHODAS scores by case studies:* The following four cases illustratively represent instances where the disability status group determined by the current medically based system does not align with the level of severity of functioning limitation identified by WHODAS score:

**A** was assessed in status group No disability. But the WHODAS functioning score is 60, indicating severe disability. She is a 59-year-old married and educated woman. She is in assisted living and cannot work because of health reasons. She was diagnosed with a malignant bladder cancer. She reports having had difficulties because of her health condition on every day of the last month and been unable to perform usual activities 2/3 of the time.

**B** was assessed in status group III or moderate functioning restriction. The WHODAS functioning score, however, is 55 or severe disability. He is a 40-year-old man with 14 years of education and is currently married and living independently in the community. He has a congenital malformation with a deformation of the spinal cord. He reports also having difficulties because of his health condition on every day of the last month. He is unable to perform usual activities or must reduce usual activities or work 1/3 of the time.

---

<sup>2</sup> Aleksandra Posarac, Elina Celmina and Jerome Bickenbach. *Disability Policy and Disability Assessment System in Latvia* © World Bank. 2020

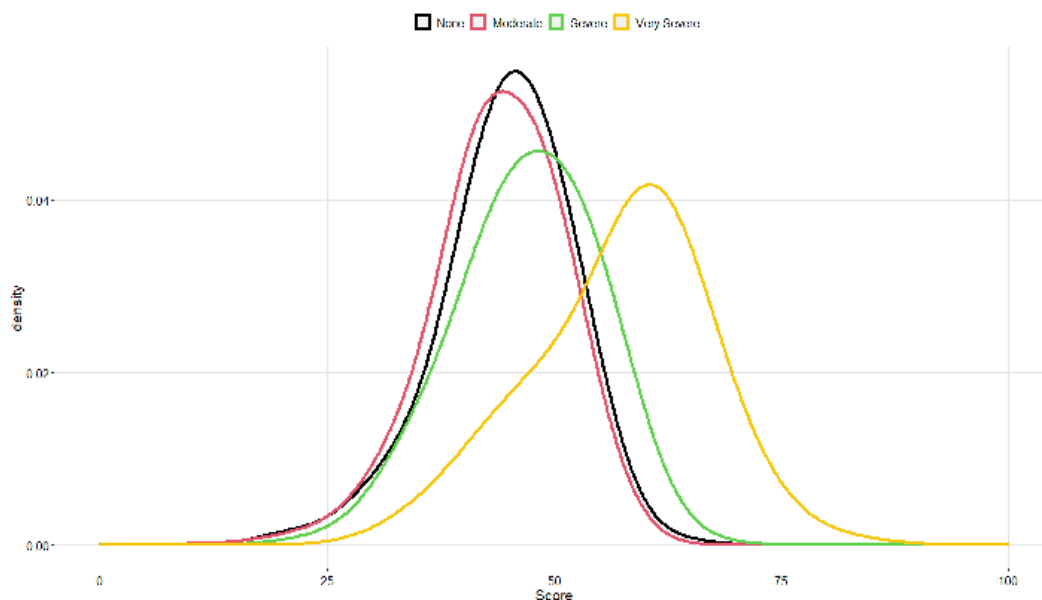
**C** is a 59-year-old married man with a WHODAS functioning score of 31 suggesting mild disability. He was determined to have status group II or severe functioning restriction due to an eye disease, specifically detachments and breaks of the retina, and a phonological disorder. He is still working and living independently in the community. His health condition is not limiting him in his daily life, and he can perform his usual activities normally without having to reduce them.

**D** is a 72-year-old woman with 15 years of education and lives with a partner. She was assessed as having disability status group 1 or severe functioning restriction. She was diagnosed with a malignant melanoma of the skin. She presents a WHODAS score of 32. Her health condition is not limiting her in performing daily life activities.

*Comparing discrimination between disability status groups and WHODAS scores:* a comparison between the discriminatory power of the medical assessment and the WHODAS score shows that **the current system has difficulties discriminating levels of functioning** (Figure 1).

For reference, in the pilot sample: 242 persons (11.2%) were certified as No disability, 831 persons (38.45%) were certified as Group III disability (moderate restrictions in functioning), 685 persons (31.7%) were certified as Group II disability (severe restrictions in functioning), and 403 persons (18.65%) were certified as Group I disability (very severe restrictions in functioning). Our analysis shows that the No disability and Group III (moderate) disability status groups have very similar mean WHODAS functioning scores: No 44.7(6.99) and Group III – 43.8(7.09) and differentiating them is not possible. Only the values of the medical categorization of very severe disability have WHODAS-quartile scores that are higher, indicating higher functioning problems.

Figure 1: WHODAS-score density line by determined disability status groups



This observed lack of discrimination between disability status groups is neither surprising nor unexpected, as the current disability status assessment is based on inference about disability from the medical information, while WHODAS captures the performance perspective on activities and participation.

*To conclude*, exploring the structure and properties of the disability status groups as they are currently determined with the metric standard that the WHODAS pilot data constructed, suggests that the status groups do not, in various ways, consistently represent a meaningful ordinal ranking of severity of disability across the participants in the pilot. Medical information is obviously relevant to the determination of disability. But these results shows that it is not suitable to rely on subjective judgement of medical professionals, based on diagnostic and other medical information alone to decide about the ordinal ranking of the severity of functioning problem of each applicant, represented by disability status groups.

*Functioning information collected through a self-assessment form.* In the current disability assessment system in Latvia, information on functioning is collected through a self-assessment form that is submitted with the application for the assessment. The self-assessment form contains, *inter alia*, 21 questions on functioning. It combines ICF items from the categories of body functions (b-codes) and activities and participation (d-codes). The items use the response options of ICF 0 = no, 1 = mild, 2 = moderate, 3 = severe, 4 = extreme problems in functioning. The self-assessment form was developed locally and, to the best of our knowledge, has not been tested psychometrically. This is very important as any functioning assessment instrument must meet the psychometric requirements to be valid and reliable and to allow that a Rasch score of functioning along the continuum 0-100 is derived.

It should be noted that information on functioning can be collected and used to inform the assessment in a qualitative manner. But this way calls for judgment on how to include it in the assessment, which may vary a lot from assessor to assessor. While ordinal scales and qualitative information are more suitable for the needs assessment where excellent psychometric properties of the instrument may not be *sine qua non* for an accurate assessment of needs, a disability assessment criteria and procedure, must minimize room for discretionary decision making (see Box 1).

#### Box 1: The credibility of disability assessment

The credibility and perceived legitimacy of a country's disability assessment procedure depends on a few fundamental considerations. First of all, the assessment s must be **valid** to minimize 'false positives' (people assessed as disabled and receiving benefits but are not disabled) or 'false negatives' (people who should be assessed as having a disability and receiving benefits, but they are not) – see four examples above. Second, the procedure must be **reliable**, in the sense that two assessors following the same rules and criteria should be able come to the same assessment of the same person (often called 'inter -rater reliability'). And lastly, the **decisions must be transparent and standardized**, so that the grounds for the decision-making are publicly known and their application in particular cases, when needed and applicable, independently evaluated. In short, the legitimacy of the disability assessment process depends on it being, and be seen to be, impartial, fair, and based on objective evidence.

Disability is complex and difficult to measure, and these credibility criteria are not easy to achieve in practice. Even in the most sophisticated and well -resourced countries time and other limitations mean that mistakes can be made. Assessors rely on the supporting evidence they are provided, which may contain errors, and there are invariably differences between assessors in how the evidence is evaluated and weighed. Yet the overall accuracy of disability assessment is crucial for the political sustainability, and perceived fairness of social security and other policies that rely on disability assessment. If expert disability assessors, following the rules they have been set down, often came to different judgments about the same applicant, then the process might be viewed as arbitrary and unjust.

See: Bickenbach, Jerome; Posarac, Aleksandra; Cieza, Alarcos; Kostanjsek, Nenad. 2015. *Assessing Disability in Working Age Population: A Paradigm Shift from Impairment and Functional Limitation to the Disability Approach*. World Bank, Washington, DC. © World Bank.

<https://openknowledge.worldbank.org/handle/10986/22353>  
License: CC BY 3.0 IGO.”

We tested the psychometric properties of the functioning categories in the self-assessment form. The objective was to investigate whether these ICF categories could be validly and reliably summarized in one summary score as in the case of WHODAS. A Rasch analysis showed that using these ICF categories to build a functioning scale would be problematic. Only two items are not dependent on others, there are significant dependencies between ICF-categories, and they only correspond to some extent to the structure of the ICF. These dependencies also affect the dimensionality and show that they define a multidimensional construct. The reliability is at an acceptable level, but only until adjusting for the item dependencies (item dependencies are known to inflate reliability estimates). After adjustment, the reliability dropped to an unacceptable level. To some extent this low reliability can also be explained by the poor targeting of the ICF-categories to the sample's functioning level. This indicates, that taken together, the selection of ICF-categories would overestimate the levels of functioning problems. Regarding the item fit, only two items showed some misfit. The person item map shows that the rating scale with 5 options (1= no problem, 2 = mild, 3 = moderate, 4 = severe, 5 = extreme) is not working properly. All items present disordered thresholds.

**In summary**, the selection of the ICF categories from the self-assessment form fails to achieve the essential statistical properties required to measure functioning. As it stands, therefore, the self-assessment form cannot be used to create a summary score of functioning. There may, nonetheless, be reasons to keep the self-assessment form, as we will discuss in the next section.

#### 4. Options for including functioning in disability assessment in Latvia

This section presents options for including functioning into disability assessment in Latvia in a systematic, formalized, and transparent mode. For the source of functioning information, we use WHODAS because its good psychometric characteristics observed in other studies were confirmed in the Latvian pilot.

##### i. The challenge

A disability assessment system that is valid and reliable must, at a minimum, consistently map disability status groups that are ordinaly arranged by severity onto a functioning severity scale. WHODAS has been empirically shown to produce such a scale. As mentioned above, the only reliable way in which functioning information can be collected is, first by using a psychometrically robust and scientifically validated data collection instrument that collects functioning information from the perspective of performance, and secondly by avoiding as many of the potential biases and distortions associated with self-report as possible, by administering the instrument by a trained professional in a face-to-face interview. WHODAS is statistically and psychometrically valid and robust instrument to use – the results of the analysis of the WHODAS pilot in Latvia confirm that it is psychometrically strong and collects functioning information from the perspective of performance.

The challenge in Latvia is how best to combine medical and functioning information. Although functioning information is directly relevant to disability, purely medical information is also important to support a valid and fair assessment of disability in individuals applying for benefits. Medical

information, in the ICF terms, is information about the “intrinsic” capacity of the body and mind. In many instances, the biomedical problems people have can make all the difference to what they experience in their lives. A person in chronic pain, a person missing a limb, or experiencing severe depression is experiencing disability, and it may not matter much what his or her environment is. The body makes a difference in disability, and ignoring the body, or downplaying the importance of the body distorts the concept of disability. Moreover, medical information gives us a longitudinal perspective: we know what to expect as the disease progresses, what complications or secondary conditions may arise in the future. This too is relevant to the overall determination of the degree of disability that a person experiences.

Unlike other countries, Latvia has chosen to define disability in terms of disability status groups rather than in terms of a percentage of disability that an individual experiences. The disability status groups are ordinal in nature and represent four levels of severity: no disability, moderate (group III), severe (group II) and very severe (group I). Somewhat arbitrarily, these ordinal groups are made artificially linear by assigning each group to a range, or band, of percentage: <24.99% - no disability; 25-59.99% - moderate; 60-79.99% severe and 80-100% very severe).

We have no information where these percentage numbers came from and whether they are based on some scientific evidence. We assume that they were decided on based on expert deliberations. As they are part of the current system, however, we need to take them into account as we develop and propose options for moving forward with our recommendations.

That being said, our first task is to try to define bands using WHODAS Rasch transformed data that can parallel the disability status group percentage bands. Only by doing so is it possible to suggest ways in which the medically based disability status group procedure can be integrated with the WHODAS disability linear metric.

## **ii. Proposal for WHODAS disability severity bands**

We are not aware of any previous attempt to create ranges or bands of percentages of disability severity using WHODAS data. There are suggestions, however, for cut-off points in terms of which these bands could be constructed. Based on the available information in literature, from WHODAS officially published norm-values and information on WHODAS score distributions collected in Latvia, Lithuania, and Greece we suggest the following cut points for the Rasch-based 0-100 WHODAS-score: WHODAS Rasch scores < 25 indicate no disability, scores 26-45 indicate (mild to) moderate disability, scores of 46-60 indicate severe disability, and scores of 61 to 100 very severe disabilities. The suggested cut-offs are not computed based on a sound statistical methodology and will need revision, once the WHODAS is implemented and more data points are collected over time on the continuum from low to high levels of disability.

Applying the proposed cut-offs to the pilot data, the following results were obtained: First, a very low percentage of the population would have no disability (1.18%); this is much less than the number of persons having no disability status (11.2%) in the pilot population. The percentage of very severe disability is 8.6%, significantly lower than what the current system determines as having a very severe disability (18.65%). The major shift is thus from very severe to severe and no disability to mild to moderate group. Consequently, the mild to moderate and very severe are larger than in the current medical assessment of disability (41.5% vs 38.45% and 48.8% vs. 31.7%). There is an effect of gender, age, marital status, living situation and the working situation characteristics on the observed values of WHODAS-based functioning level. In comparison, for example, gender was not significant when stratifying by the SMC determined disability status groups.

### iii. Options for including functioning information into the current Latvian disability assessment system

Below we present three options for changing the current Latvia disability assessment system, which in effect are modes for integrating functioning information. All options require that WHODAS be used in the assessment process at some point; but the difference is how the information from WHODAS is used, and when. All options also keep the disability status groups as they are now, there is no reason to change this approach to disability determination; it serves its purposes.

#### A. *Flagging mechanism*

The least disruptive change to the current system would be to keep the medical assessment and disability status groupings as is, including the self-assessment form but to use the WHODAS Rasch score derived bands in the determination decision. A systematic procedure can be devised so that any individual whose WHODAS score places them in a band that is different from the disability status group medically assessed is 'flagged'. Whether WHODAS rates the disability percentage level higher or lower than the range for the status group assigned, that individual's case needs to be reconsidered so that the divergence is explained. The explanation might point to the nature of the health condition the person has – e.g., a condition that will inevitably worsen may need to be in a higher status group than WHODAS will indicate, or WHODAS may be indicating more functioning problems than are typically experienced by a person with that health condition.

With this strategy, the WHODAS scores are used as indicators of the extent of problems in functioning, although the cut-offs that create the percentage bands may need to be refined based on a larger set of cases and more insight into the relationship between the population's lived experience and the reported scores.

#### B. *Priority to WHODAS summary scores*

A more radical suggestion is to, in effect, reverse the sequencing of disability assessment in Latvia by using the WHODAS summary score, and the percentage bands described above, to determine for each applicant the disability status group to which they belong – from No Disability to Very Severe Disability. The argument in favor of doing so has been made several times in this document: WHODAS is a psychometrically powerful and scientifically robust questionnaire that has been explicitly designed to capture precisely what disability assessment is about: namely, the overall, 'whole person' level of functioning problems that people actually experience in their daily lives, taking full account of all environmental barriers and facilitators. We have repeatedly described this as ICF's perspective of performance arguing that disability assessment is a matter of validly assessing the actual lived experience of people with health conditions. This is exactly what WHODAS does.

However, this is not to say, nor is this option defining a process that ignores the essential medical information that we have repeatedly said is essential to disability assessment. What this option suggests is that WHODAS summary score, based on Rasch-derived metric scale, can provide the first estimate of which disability status group the applicant appropriately belongs to. Medical information can then be used to adjust or refine this first estimate to reflect the nature of the applicant's underlying health condition.

#### C. *Comprehensive disability assessment (World Bank recommended option)*

This last option allows for incrementally introduced reforms without being disruptive to existing procedures and practices, but nonetheless constitutes an important revision that systematically brings functioning information into the disability assessment system in Latvia. ***It is our recommended option,***

so we describe it in more detail. A comprehensive disability assessment depends on three sources of information:

1. *Functioning information* presented as a summary score for a 'whole person' level of functioning. The WHODAS pilot has confirmed, and we have reported, that the WHODAS 36-question version is both feasible to introduce into the Latvian system and has the desired psychometric properties of validity and reliability.
2. *Health status information* coming from the health sector. We suggest that the collection and quality of medical information can be improved. The medical referral form should be revised to explicitly require the primary health condition, as well as secondary conditions and other co-morbidities to be listed (and their ICD codes), and description of diagnostic test results, the state of health and proposed therapies. This information should be typed and electronically submitted to the SMC (this will likely require the assistance of a nurse or other assistant to transfer the physicians' handwritten notes to more legible format).
3. *Information about the applicant's environment* – family, home, school, workplace, community. The most direct way of getting this information is from the self-assessment form, which can be revised to include questions on personal and demographic data, household composition, living arrangements, housing situation, education and employment. This should be the primary function of this form: the functioning questions are not necessary because they are already included in WHODAS. If possible, the form could also include questions about sick leave and the benefits and services the person is already receiving, including assistive technology. The form might also have questions on what benefits and services the person believes they need, as well as their wishes and plans. In short, we propose that a well-structured needs assessment section is included in the self-assessment form as a way of collecting information that the SMC could use to assess needs and propose existing benefits and services to improve disability experience.

As already explained, a significant challenge in including functioning into disability assessment in Latvia is the fact that the current system does not assign individual percentages of disability but uses an ordinal scale (with underlying percentage bands of "difficulties in functioning"). The disability assessment method proposed here avoids this issue by bringing together three sources of information for a compressive and individualized assessment that is grounded in functioning information. The assessment procedure should be the same for all applicants irrespective of whether the health condition or impairment was caused by age, accident or occupational disease or work accident.

Procedurally, we suggest the following process (many existing steps in the current disability assessment procedure would remain the same, we are only listing the changes):

1. A person submits an application along with the revised medical referral and revised self-assessment form. Application, medical referral, and self-assessment should all be in electronic format and part of the applicant's electronic file. When the file is composed, a cross-check/verification of data should be run, and inconsistencies or missing information flagged.
2. The appointment interview is scheduled electronically, and the person is informed.
3. A SMC employee prepares the file for the face-to-face interview. The assessors should not have any connection to the applicant (if there is even a remote connection, she/he should be recused from the assessment).
4. Prior to the interview, a trained assessor not participating in the face-to-face meeting, administers WHODAS. The answers should be immediately marked in an electronic file, so that an automated algorithm can generate the WHODAS Rasch score immediately.
5. During the interview – recommended is presence of two assessors, an administrative assistant and the applicant and possibly one more person close to the applicant, the assessors evaluate disability experience of the applicant. They should use medical information, self-assessment



information (revised content) and information from WHODAS (but without the WHODAS Rasch score). The referring physician might be present as well and present the medical case for disability. The assessors should be trained in interview techniques and a guidance on what and how to ask should be prepared. The interview should be recorded but only with the consent of the applicant.

6. Based on the interview and the documents, the assessors prepare evaluation and propose the group of disability with a comprehensive justification using the assessment guidelines. The proposal should also include a section on proposed benefits and support measures. The evaluation form is sent to the supervisor.
7. The supervisor reviews the evaluation and compares the proposed disability group with the WHODAS Rasch score. If the two overlap, the evaluation is completed, and a certificate is issued along with the proposed interventions. The person may be automatically referred to the benefits and service administrators without the need to apply for them separately. If the proposed disability group is different from the WHODAS Rasch score-based group, the supervisor should appoint a different assessor to review all documents about the case and make her/his proposal to the supervisor. Should it be needed, this assessor may talk to the applicant, her/his physician or any other person who may provide additional information. The case should then be discussed in the case meeting chaired by the supervisor and in the presence of all assessors. Optimally, the decision on the disability group would be reached unanimously. If the consensus is not possible, the chair decides.

This proposal moves disability assessment system toward holistic, comprehensive assessment of disability. The systematic and transparent inclusion of functioning will not require dramatic changes in the organization of disability assessment. The proposed changes in the self-assessment form will enhance the assessment of needs, which will make decisions about benefits and interventions to accompany the certification easier for all parties concerned.

The new system will require adjustments in the information management system, and the assessment instructions and guidelines. It is very important to establish a statistical and analytical unit at the Ministry or SMC to analyze WHODAS and other disability related data to (i) fine tune the WHODAS Rasch cut-offs, (ii) analyze disability trends; (iii) conduct analytical and statistical research needed for the development of evidence-based disability policies, and system, including disability assessment. An alternative would be for the Ministry and SMC to establish a formalized collaboration with one of the premier universities in Latvia.

## Latvia: Options for including functioning into disability assessment

The Latvian Government is seeking to strengthen its disability assessment system by more effectively combining medical and functioning information.<sup>3</sup> Including functioning information into disability assessment requires empirical evidence from testing a functioning data collection instrument. To that end, a psychometric instrument for assessing disability, developed and extensively tested by the World Health Organization (WHO), the Disability Assessment Schedule (WHODAS 2.0), was pilot tested in Latvia. This report presents details of the pilot, statistical analysis of the data collected and a proposal for including functioning information into the disability assessment procedure in Latvia.

### 1. WHODAS: Technical details

In the International Classification of Functioning, Disability and Health (ICF), information about categories of Activities and Participation can be collected either from the ***perspective of capacity*** (reflecting exclusively the expected ability of a person to perform activities considering their health conditions and impairments) or the ***perspective of performance*** (reflecting the actual performance of activities in the real-world environmental circumstances in which the person lives). Information about capacity typically represents the results of a clinical inference or judgment based on medical information, while performance is a true description of what occurs in a person's life. The two perspectives are therefore very different, although capacity constitutes a determinant of performance.

As the administrative act of establishing eligibility for services and supports, disability should be assessed as the overall lived experience of an individual living with one or more health problems – or in ICF terms, it is the level of a person's performance in light of their intrinsic health capacity and environmental facilitators or barriers. Disability assessment is a 'whole person' or global assessment of the extent or level of person's disability. This is important because a disability assessment should be a summary measure of functioning levels across domains of actions, simple and complex, from walking, taking care of children to working at a job. A disability assessment is an assessment of the overall level of disability that a person experiences in his or her life. A summary or global assessment of disability, of necessity, must be based both on the individual health state and on specific assessments of specific activities. Yet a summary assessment of disability is valid only if the specific assessments can be statistically summarized into a single assessment score. ***A disability assessment is a summary measure of the level of a person's performance of an adequately representative set of behaviors and actions, simple to complex, in their actual environment, considering the person's state of health.***

The ICF understands 'disability' to be any level of problem or difficulty in functioning in some domain, from the perspective of performance. The WHO developed, tested and has consistently recommended the WHODAS as a questionnaire that can capture the performance of activities by an individual in his or her daily lives and actual environment. The 'actual environment' is represented in the ICF in terms of environmental factors that act either as environmental facilitators (e.g., assistive devices, supports, home modifications) or as environmental barriers (inaccessible houses, streets and public buildings, stigma and discrimination). The WHODAS questionnaire, in short, is WHO's

---

<sup>3</sup> In Latvia, disability and work capacity is assessed and certified by the State Medical Commission for the Assessment of Health Condition and Work Ability (SMC). For detailed discussion about the assessment system see: Aleksandra Posarac, Elina Celmina and Jerome Bickenbach. 2020. Disability Policy and Disability Assessment System in Latvia © World Bank.

recommended, generic, performance-based disability assessment tool. It is structured around six basic functioning domains:

- D1: Cognition – understanding & communicating
- D2: Mobility– moving & getting around
- D3: Self-care– hygiene, dressing, eating & staying alone
- D4: Getting along– interacting with other people
- D5: Life activities– domestic responsibilities, leisure, work & school
- D6: Participation– joining in community activities

Conducted by a skilled and trained professional, a WHODAS interview collects information about functioning and problems in functioning – i.e., disability – by asking standardized questions – and if necessary, follow-up probing questions – and in light of the WHODAS’s 5-level responses scale (1 = None, 2 = Mild, 3 = Moderate, 4 = Severe, 5 = Extreme or Cannot do) to rate each question for that individual. It should be clear that, as used in this pilot, WHODAS is not a self-administered questionnaire; it is rather a questionnaire administered in a face-to-face or telephone interview by a trained professional. Respondents are informed that their answers about each domain of functioning should adopt the perspective of performance – that is, they should describe what they actually do, taking into account their actual experience in their daily life and specifically in light of all environmental barriers and facilitators that they experience. The WHODAS 36-item, professionally administered version was chosen for the pilot to collect information about a substantial range of functioning domains in order to create a full picture of the disability experienced by the respondent in their everyday life. The 36 items are shown in Table 1 by functioning domain.

Table 1: WHODAS 2.0 items for the 36-item long form

<b>In the past 30 days, how much difficulty did you have in:</b>	
<b>Understanding and communicating</b>	
D1.1	Concentrating on doing something for ten minutes?
D1.2	Remembering to do important things?
D1.3	Analyzing and finding solutions to problems in day-to-day life?
D1.4	Learning a new task, for example, learning how to get to a new place?
D1.5	Generally understanding what people say?
D1.6	Starting and maintaining a conversation?
<b>Getting around</b>	
D2.1	Standing for long periods such as 30 minutes?
D2.2	Standing up from sitting down?
D2.3	Moving around inside your home?
D2.4	Getting out of your home?
D2.5	Walking a long distance such as a kilometer [or equivalent]?
<b>Self-care</b>	
D3.1	Washing your whole body?
D3.2	Getting dressed?
D3.3	Eating?
D3.4	Staying by yourself for a few days?
<b>Getting along with people</b>	
D4.1	Dealing with people you do not know?
D4.2	Maintaining a friendship?
D4.3	Getting along with people who are close to you?
D4.4	Making new friends?

D4.5	Sexual activities?
	<b>Life activities</b>
D5.1	Taking care of your household responsibilities?
D5.2	Doing most important household tasks well?
D5.3	Getting all the household work done that you needed to do?
D5.4	Getting your household work done as quickly as needed?
D5.5	Your day-to-day work/school?
D5.6	Doing your most important work/school tasks well?
D5.7	Getting all the work done that you need to do?
D5.8	Getting your work done as quickly as needed?
	<b>Participation in society:</b>
D6.1	How much of a problem did you have in joining in community activities in the same way as anyone else can?
D6.2	How much of a problem did you have because of barriers or hindrances in the world around you?
D6.3	How much of a problem did you have living with dignity because of the attitudes and actions of others?
D6.4	How much time did you spend on your health condition or its consequences?
D6.5	How much have you been emotionally affected by your health condition?
D6.6	How much has your health been a drain on the financial resources of you or your family?
D6.7	How much of a problem did your family have because of your health problems?
D6.8	How much of a problem did you have in doing things by yourself for relaxation or pleasure?

## 2. Descriptive statistics of the WHODAS pilot sample

A total of 2,202 persons who applied for a disability assessment in 2020-2021 were included in the pilot. The pilot overlapped with the Corona pandemic. This resulted in a high number of interviews that could not take place in a face-to-face mode. In fact, only 14% of the interviews took place face-to-face, while 86% were phone-interviews or interviews via WhatsApp (3%). Table 2 presents descriptive statistics about the persons included in the pilot. The proportion of female participants was higher (58.4% vs. 41.6% respectively). The average age was 58.46 years (SD = 13.57). Most participants were married (45.5%), 13.1% were widowed, and 13.6% were cohabiting. Most participants were living independently in the community (81.1%). The participants had an average of 12.53 (SD = 3.03) years of education. Most reported either having a paid employment (31.2%), being retired (25.2%), and unemployed for health reasons (21.5%).

Table 2: Pilot sample descriptive statistics

<b>Sample Size</b>	<b>2202</b>
Gender = Male N (%)	916 (41.6)
Age - mean (SD)	58.46 (13.57)
Years of Education - mean (SD)	12.53 (3.03)
Marital Status N (%)	
Never married	207 (9.4)
Currently married	1003 (45.5)
Separated	121 (5.5)
Divorced	284 (12.9)
Widowed	288 (13.1)
Cohabiting	299 (13.6)
Living Condition N (%)	
Independent in the community	1785 (81.1)
Assisted living	413 (18.8)
Hospitalized	4 (0.2)
Work Status N (%)	
Paid work	687 (31.2)
Self-employed	56 (2.5)
Non-paid work	3 (0.1)
Student	16 (0.7)
Keeping house	16 (0.7)
Retired	555 (25.2)
Unemployed (health reasons)	474 (21.5)
Unemployed (other reasons)	61 (2.8)
Other	334 (15.2)

A total of N = 1,220 (55.4%) pilot participants were diagnosed with only one ICD-10 linked health condition, while N = 982, 44.6% reported one health condition with additional comorbidities.

Table 3 presents the most frequently observed ICD-10 diagnostic chapters for the participants' main health condition. Diseases of the musculoskeletal system and connective tissue (N = 497, 23.01%) and diseases of the circulatory system (N = 372, 17.22%) were the most reported main diagnoses. Neoplasms were reported by N = 388 (17.96%) participants. ICD chapter XIX, external causes such as injuries, was seen as primary diagnosis in about 18% of the participants (N = 385, 17.82%).

Table 3: Prevalence of Health conditions in the study population  
by ICD-10 Health Condition Category

ICD-Chapter	N	%
I Certain infectious and parasitic diseases	28	1.3 %
II Neoplasms	388	17.96 %
III Diseases of the blood	3	0.14 %
IV Endocrine, nutritional, and metabolic diseases	47	2.18 %
IX Diseases of the circulatory system	372	17.22 %
V Mental and behavioral disorders	44	2.04 %
VI Diseases of the nervous system	124	5.74 %
VII Diseases of the eye and adnexa	158	7.31 %
VIII Disease of the ear and mastoid process	5	0.23 %
X Diseases of the respiratory system	23	1.06 %
XI Diseases of the digestive system	27	1.25 %
XII Diseases of the skin and the subcutaneous tissue	11	0.51 %
XIII Diseases of the musculoskeletal system and connective tissue	497	23.01 %
XIV Diseases of the genitourinary system	19	0.88 %
XIX Injury, poisoning and certain other consequences of external causes	385	17.82 %
XVII Congenital malformations, deformations, and chromosomal abnormalities	16	0.74 %
XVIII Symptoms, signs, and abnormal clinical and laboratory findings	1	0.05 %
XXI Factors influencing health status and contact with health services	12	0.56 %

### 3. Psychometric analysis: Rationale and tests

The pilot was implemented using the 36-item version of WHODAS 2.0. However, items D5.5 to D5.8 (see Table 1 above), which were responded to only by persons working or in education, had more than 70% of missing values and were excluded from the analysis (Table 4). The WHODAS 2.0 Manual (Ustun 2009) indicates that the 32-item score is comparable to the 36-item score. The psychometric analysis included the remaining 32 items and investigated whether a WHODAS total score with interval scaled properties can be derived.

In principle, ordinally scaled values do not allow for the calculation of averages or variances. An ordinal-to-interval transformation becomes essential to make the information collected through questionnaires usable for measurement and so for parametric and inferential statistical testing. To determine the measurement properties of the WHODAS items, a Rasch analysis (Rasch 1960) was performed on the entire sample of N = 2,202 applicants for disability assessment included in the pilot data collection.

Rasch analysis is a statistical method from the field of probabilistic measurement. It is a modern test theory approach first introduced in the 1960s by the Danish mathematician George Rasch (Rasch 1960). To analyze the WHODAS 2.0 items, a polytomous version of the Rasch model called the Partial Credit Model (PCM) (Masters 1982) was applied.

A Rasch analysis allows us to test core measurement assumptions (Bond and Fox 2001; Tennant and Conaghan 2007), specifically (1) targeting of the scale, (2) model reliability, (3) the ordering of the items' response options, (4) the absence of strong associations between items (or local item independence – LID), (5) the fit of the items to the Rasch model, (6) the absence of effects of person factors such as gender and age on item responses (or differential item functioning – DIF), and (7) the unidimensionality of the questionnaire. If these measurement assumptions can be met, a

questionnaire is psychometrically sound with interval-scaled total scores that are operative for measurement. This is the gold standard for instruments of this sort.

For a well-functioning questionnaire, the items' difficulty has to match the population's level of ability, i.e., the questionnaire must not be too easy or too difficult. Statistically, good targeting implies that the mean item difficulty and mean person ability approximate 0 and that the items' difficulties match the ability of the population. This would mean that the items included in the questionnaire capture the disability range of a population.

A Person Separation Index (PSI) above 0.8 indicates good reliability of the scale, values above 0.9 very good reliability. The PSI indicates how well the scale can discriminate levels of functioning in the population. The Cronbach  $\alpha$ , which is typically also reported, is a measure of the data's internal consistency, i.e., how well the items work to describe one construct (Nunnally and Bernstein 1994).

In the presence of disordered response options an analysis of response probability curves allows us to determine which response options cause a problem and to decide on strategies for collapsing i.e., aggregating, adjacent response options. For example, if an item's response options 2 and 1 appear reversed and indicate that an expected increase of difficulty is not observed in the data, the item responses can be recoded so that these options represent only one level of response.

Local item dependency (see point 4 below) often occurs when items are redundant and measure approximately the same aspect of a construct. The most widely reported statistic for item dependencies is the Q3 matrix which is just another name for the Rasch residual's correlation matrix (Yen 1984). Marais (2013) recommends considering LID relative to the residual correlations' average because the residual correlation's magnitude depends on the number of items. Christensen, Makransky, and Horton (2017) formalized this, illustrating that if the largest Q3 value is more than 0.2 above the average, it would indicate an anomaly. A way to address local item dependency without deleting items is to aggregate (i.e., sum up) the correlated items into so-called testlets (Yen 1993). In item testlets, the ordering of the thresholds is not expected anymore.

With good item fit (point 5 below), the Infit and Outfit values are below 1.2 (R. M. Smith, Schumacker, and Bush 1998). The Outfit statistic is a more outlier sensitive alternative to the Infit statistic, meaning that the Outfit statistic can sometimes indicate misfit, while the Infit does not.

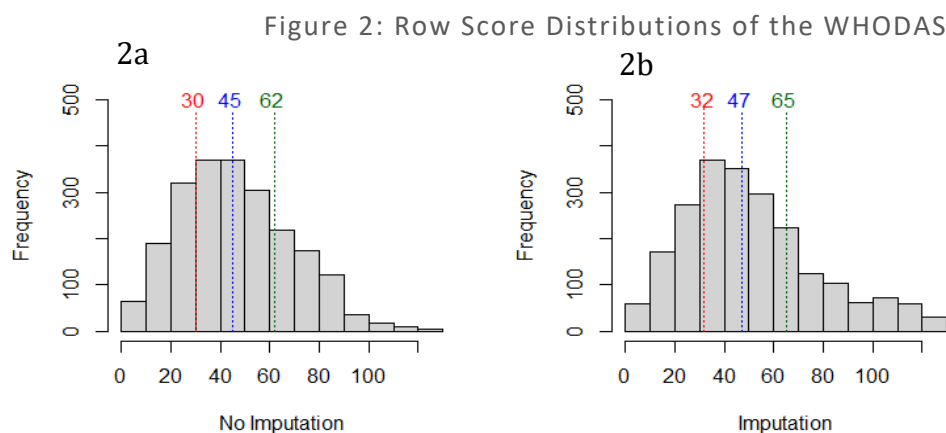
Ideally, items of a questionnaire should not favor sample subgroups. The analysis of (point 6 below) DIF with ANOVA flags exogenous variables, or DIF variables, which cause a lack of invariance of the item difficulty estimates (Holland and Wainger 1993). It is worth noting that a DIF analysis does not always indicate a metric bias but can also represent subgroups with an unequal ability (Boone 2016). A two-way ANOVA is used to test for uniform (DIF variable) and non-uniform (DIF variable x score level) DIF. The questionnaire was tested for DIF by gender and age groups.

Finally, a questionnaire should measure only one construct (point 7 below). If the questionnaire presents several separate dimensions, the validity of the summary total score is undermined. A principal component analysis of the residuals determined the questionnaire's degree of unidimensionality (E. V. Smith 2002). Typically, a first eigenvalue  $< 1.8$  is deemed indicative of unidimensionality. Based on simulation analyses, R. M. Smith and Miao (1994) suggest considering the second component's size, with values below 1.4 as indicative of unidimensionality. The metric analyses were performed with the software R (Team 2016), specifically, the package mirt for the Rasch analysis (Mair, Hatzinger, and Maier 2019) and iarm for the DIF analysis.

#### 4. Results - Metric properties of WHODAS

First, the work and education related items D5.5 to D5.8 resulted in >70% of missing values, where the persons not responding to these items were on average significantly older (60.93, SD = 13.86) compared to the average of the sample (52.49, SD = 10.69). They were either mostly retired (35.7%), unemployed (34.4%), or responded with 'other' (21.5%) to the employment question. Also, 94.2% of those in assisted living were not working or going to school. The psychometric analysis did not include these items. A few other items resulted in higher percentages of missing values: *D3.4 Staying by yourself for a few days* (19.94%), *D4.4 Making new friends* (46.64%), *D4.5 Sexual activities* (44.69%), and *D6.1 How much of a problem did you have in joining community activities (for example festivities, religious or other activities) in the same way as anyone else can?* (25.61%).

While the Rasch model can handle percentages of missing values below 15% without introducing detrimental bias in the estimates (Fellinghauer, Proding, and Tennant 2018), percentages above 50% are not acceptable. (Note that these same items also resulted in higher proportions of missing values in the World Bank's Disability Assessment Pilot in Lithuania, but not to that extent. World Bank 2021) It is likely that the higher proportion of missing values in these questions is a consequence of the confinement or social distancing measures in place to manage the Corona pandemic in 2020-2021. Based on the WHODAS 2.0 Manual (Ustun 2009), the simple approach to impute missing values is to use the mean of a person's score but only if they have 1 or 2 items with missing values. In the actual data collection, N = 759 (34.47%) of the pilot participants had more than two missing values. Only N = 706 (32.06%) of applicants answered all of the 32 items used in the psychometric analysis of the WHODAS. A statistical imputation approach including the socio-demographic information could therefore be used for the missing values in the WHODAS items.



Figures 2a and 2b show the distribution of the WHODAS-score without or with data imputation. The red lines show the 25% distribution quantile, the blue dotted line the median, and the green line the 75% distribution quantile.

Table 4 shows descriptive statistics for all of the 36 WHODAS items, including the number and percentage of missing values.



Table 4: Frequencies and Percentages of WHODAS Responses

Item	No	Mild	Moderate	Severe	Extreme, cannot do	Missing
D1.1	718 (32.61%)	494 (22.43%)	646 (29.34%)	196 (8.9%)	144 (6.54%)	4 (0.18%)
D1.2	548 (24.89%)	554 (25.16%)	604 (27.43%)	346 (15.71%)	147 (6.68%)	3 (0.14%)
D1.3	667 (30.29%)	479 (21.75%)	560 (25.43%)	264 (11.99%)	211 (9.58%)	21 (0.95%)
D1.4	466 (21.16%)	407 (18.48%)	521 (23.66%)	360 (16.35%)	282 (12.81%)	166 (7.54%)
D1.5	1029 (46.73%)	603 (27.38%)	383 (17.39%)	122 (5.54%)	63 (2.86%)	2 (0.09%)
D1.6	1223 (55.54%)	554 (25.16%)	246 (11.17%)	99 (4.5%)	76 (3.45%)	4 (0.18%)
D2.1	169 (7.67%)	180 (8.17%)	484 (21.98%)	549 (24.93%)	810 (36.78%)	10 (0.45%)
D2.2	291 (13.22%)	434 (19.71%)	776 (35.24%)	423 (19.21%)	277 (12.58%)	1 (0.05%)
D2.3	600 (27.25%)	697 (31.65%)	494 (22.43%)	173 (7.86%)	234 (10.63%)	4 (0.18%)
D2.4	372 (16.89%)	433 (19.66%)	696 (31.61%)	355 (16.12%)	331 (15.03%)	15 (0.68%)
D2.5	155 (7.04%)	251 (11.4%)	413 (18.76%)	456 (20.71%)	744 (33.79%)	183 (8.31%)
D3.1	619 (28.11%)	415 (18.85%)	563 (25.57%)	299 (13.58%)	305 (13.85%)	1 (0.05%)
D3.2	722 (32.79%)	573 (26.02%)	515 (23.39%)	169 (7.67%)	221 (10.04%)	2 (0.09%)
D3.3	1620 (73.57%)	245 (11.13%)	166 (7.54%)	77 (3.5%)	93 (4.22%)	1 (0.05%)
D3.4	870 (39.51%)	233 (10.58%)	225 (10.22%)	124 (5.63%)	311 (14.12%)	439 (19.94%)
D4.1	958 (43.51%)	349 (15.85%)	447 (20.3%)	210 (9.54%)	101 (4.59%)	137 (6.22%)
D4.2	1308 (59.4%)	400 (18.17%)	202 (9.17%)	70 (3.18%)	151 (6.86%)	71 (3.22%)
D4.3	1513 (68.71%)	399 (18.12%)	161 (7.31%)	61 (2.77%)	55 (2.5%)	13 (0.59%)
D4.4	543 (24.66%)	243 (11.04%)	163 (7.4%)	123 (5.59%)	103 (4.68%)	1027 (46.64%)
D4.5	592 (26.88%)	245 (11.13%)	186 (8.45%)	103 (4.68%)	92 (4.18%)	984 (44.69%)
D5.1	218 (9.9%)	328 (14.9%)	869 (39.46%)	395 (17.94%)	166 (7.54%)	226 (10.26%)
D5.2	323 (14.67%)	391 (17.76%)	666 (30.25%)	417 (18.94%)	178 (8.08%)	227 (10.31%)
D5.3	268 (12.17%)	330 (14.99%)	605 (27.48%)	540 (24.52%)	233 (10.58%)	226 (10.26%)
D5.4	180 (8.17%)	289 (13.12%)	589 (26.75%)	631 (28.66%)	284 (12.9%)	229 (10.4%)
D5.5	54 (2.45%)	224 (10.17%)	249 (11.31%)	101 (4.59%)	24 (1.09%)	1550 (70.39%)
D5.6	153 (6.95%)	212 (9.63%)	209 (9.49%)	63 (2.86%)	13 (0.59%)	1552 (70.48%)
D5.7	103 (4.68%)	172 (7.81%)	256 (11.63%)	102 (4.63%)	18 (0.82%)	1551 (70.44%)
D5.8	129 (5.86%)	173 (7.86%)	232 (10.54%)	98 (4.45%)	18 (0.82%)	1552 (70.48%)
D6.1	297 (13.49%)	401 (18.21%)	443 (20.12%)	303 (13.76%)	194 (8.81%)	564 (25.61%)
D6.2	774 (35.15%)	507 (23.02%)	408 (18.53%)	253 (11.49%)	107 (4.86%)	153 (6.95%)
D6.3	1519 (68.98%)	295 (13.4%)	221 (10.04%)	89 (4.04%)	36 (1.63%)	42 (1.91%)
D6.4	154 (6.99%)	513 (23.3%)	808 (36.69%)	420 (19.07%)	293 (13.31%)	14 (0.64%)
D6.5	259 (11.76%)	317 (14.4%)	605 (27.48%)	555 (25.2%)	454 (20.62%)	12 (0.54%)
D6.6	368 (16.71%)	280 (12.72%)	493 (22.39%)	539 (24.48%)	510 (23.16%)	12 (0.54%)
D6.7	591 (26.84%)	392 (17.8%)	534 (24.25%)	334 (15.17%)	324 (14.71%)	27 (1.23%)
D6.8	356 (16.17%)	460 (20.89%)	563 (25.57%)	403 (18.3%)	330 (14.99%)	90 (4.09%)

More than 50% of the applicants indicated severe or extreme problems in D2.1 Standing for long periods such as 30 minutes and D2.5 Walking a long distance such as one kilometer; more than 40% in D6.5 How much have you been emotionally affected by your health condition? and D6.6 How much has your health been a drain on the financial resources for you or your family? A third of the sample reported severe to extreme problems in D1.4 Learning a new task, for example, learning how to get

to a new place, D2.2 Standing up from sitting down, D2.4 Getting out of your home, D5.3 Getting all the household work done that you needed to do, D5.4 Getting your household work done as quickly as needed, D6.4 How much time did you spend on your health condition or its consequences? On the other hand, more than 50% reported not having any problems in D1.6 Starting and maintaining a conversation (55.54%), D3.3 Eating (73.57%), D4.2 Maintaining a friendship (59.4%), D4.3 Getting along with people who are close to you (68.71%), and D6.3 How much of a problem did you have living with dignity because of the attitudes and actions of others (68.98%).

Table 5 and Table 6 provide the essential fit statistics for the WHODAS items at the start and after the metric adjustments based on the outcomes of the Rasch-based analysis with the Partial Credit Model. The whole scale showed multidimensionality with a strong tendency of the items to load by WHODAS domains. However, a few items cross-loaded, and a few items were free of dependencies (Figure 4). The multidimensionality caused by the WHODAS item dependencies was adjusted first by aggregating the items into testlets based on the WHODAS domain structure.

A significant local dependency remained between D2 *Getting around* and D3 *Self-care*, as well as D1 *Understanding and Communicating* and D4 *Getting along with people*. These domains were also aggregated. This adjustment strategy worked well, with good reliability. Only the domain D5 *Life Activities* presented some outlying values that somewhat impacted on the more sensitive fit statistic. All testlets presented good Infit values. Analysis was undertaken with non-imputed data; however, the model fit statistics are also shown for imputed data to indicate that imputation would not alter the model targeting and fit. The missing value imputation was performed with MissForest, a robust multiple imputation method for mixed-type data (Stekhoven and Buhlmann 2012).

Table 7 below shows the starting approach's reliability statistics after aggregating the items for all dependencies among domains, without imputation, and with data imputation.

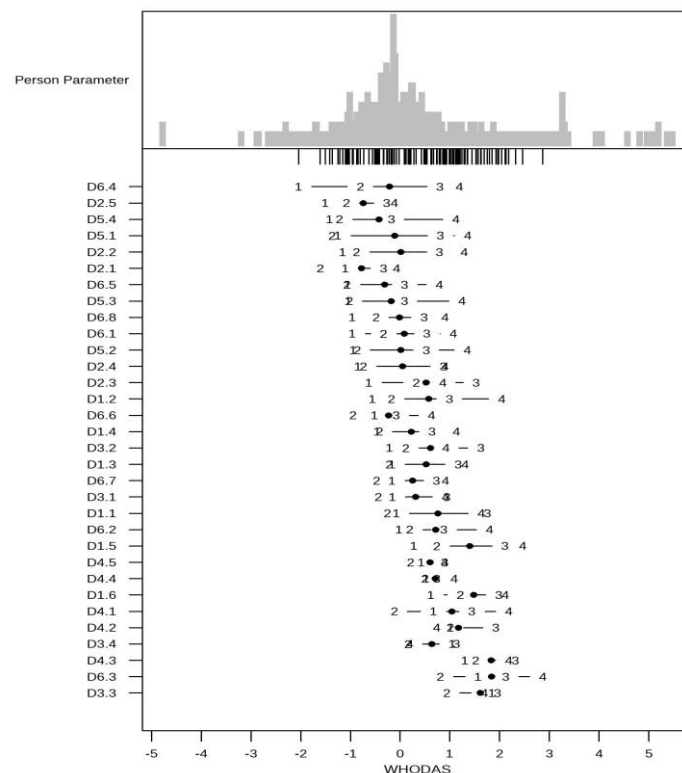
What follows reports in more details the psychometric analysis of the WHODAS 2.0 with 32 items at the start and the final approach (without imputation), deemed the most efficient and resulting in best metric properties. The final model retained the 32 items, with domains aggregated based on a conceptual, domain-based approach, preserving the assessment tool's structure. Table 5, Table 6, and Table 7 (which includes the targeting and reliability with imputed data) present the detailed Rasch statistics.

- (1) The targeting of the scale (Table 7) improved with adjustments, i.e., *item difficulties becoming more centered on the mean functioning level*.
- (2) Item dependencies in the analysis with the 32 WHODAS items inflated the reliability estimates (PSI = 0.95, Cronbach  $\alpha$  = 0.95). After adjustment for the local item dependencies, the reliability dropped to PSI = 0.89, which is still a good level of reliability, *indicating that the metric can discriminate well among levels of functioning*.
- (3) *Six items only presented perfectly ordered difficulty threshold*. Disordering consisted of a rather systematic reversing of the 2 lower categories (0 = None & 1 = Mild) and the 2 middle (2 = Moderate & 3 = Severe) categories (Figure 3). A perfect ordering could be obtained on the entire 32 items scale by reducing the number of response options through collapsing: [0 & 1] = 0, [2 & 3] = 1, and [4 = Extreme] = 2. However, the collapsing of the response options alone would not solve the multidimensionality and item dependencies so that other adjustment measures would still be required. Appendix 1 shows the person item map for the approach with collapsed response options. Ordering of thresholds is not expected anymore with testlets (Figure 5).
- (4) The residual dependencies analysis indicated strong local dependencies among the 32 items of the WHODAS 2.0 (see Figure 3), with a tendency of questionnaire items from the same domain to associate. The cut-off for LID using the mean LID + 0.2 as cut-off (Christensen, Makransky, and Horton 2017) was  $r = 0.17$ . To solve these dependencies, the items were aggregated, taking

into account the chapter structure, i.e., D1 *Understanding and Communicating*, D2 *Getting around*, D3 *Self-Care*, D4 *Getting along with people*, D5 *Life Activities*, D6 *Participation in Society*. A significant residual correlation above  $r > 0.05$  was still found between D1 *Understanding and Communicating*) and D4 *Getting along with people* ( $r = 0.19$ ) and D2 *Getting around* and D3 *Self-care* ( $r = 0.37$ ), which were also aggregated.

- (5) With Infit and Outfit expected below 1.2, the item fit was found acceptable already at the start. Only two items from the D6 *Participation in Society* domain, namely items D6.3 *How much of a problem did you have living with dignity because of the attitudes and actions of others?* ( $MSQ_{Infit} = 1.98$ ,  $MSQ_{Outfit} = 1.83$ ) and D6.4 *How much time did you spend on your health condition or its consequences?* ( $MSQ_{Infit} = 1.27$ ,  $MSQ_{Outfit} = 1.17$ ) showed poor fit on the Outfit or Infit statistic. After aggregation of the dependent items, the testlets did not show any underfit based on the Infit statistic, and with exception to one testlet, also good fit based on the Outfit statistic. The higher Outfit statistic for D5 *Life activities* ( $MSQ_{Outfit} = 1.29$ ) indicated some outlying values in this testlet, when adjusted for those the fit was good with  $MSQ_{Infit} < 1.2$ .
- (6) DIF was tested for gender and age. In the final model, the testlets aggregating D2 - *Getting around* and D3 *Self-care*, and D5 *Life activities* showed some DIF by age groups. It is to be expected that the levels of disability are impacted by the age. Only the testlet aggregating D1 *Understanding and Communicating* and D4 *Getting along with people* was not sensitive to the age of the respondents. The residuals did not present any pattern indicating DIF for gender.
- (7) Finally, the principal component analysis indicated clustering of the items by domains resulting in multidimensionality, with a 1<sup>st</sup> eigenvalue of 5.05 and a 2<sup>nd</sup> eigenvalue of 3.05. After adjustments, i.e., aggregation of items by WHODAS domains, the 1<sup>st</sup> eigenvalue dropped to 1.53 and the 2<sup>nd</sup> eigenvalue to 1.21 and supported unidimensionality.

Figure 3: Person item map after the creation of testlet



\*Indicates disordered thresholds

Figure 4: Local item dependencies before the creation of testlets

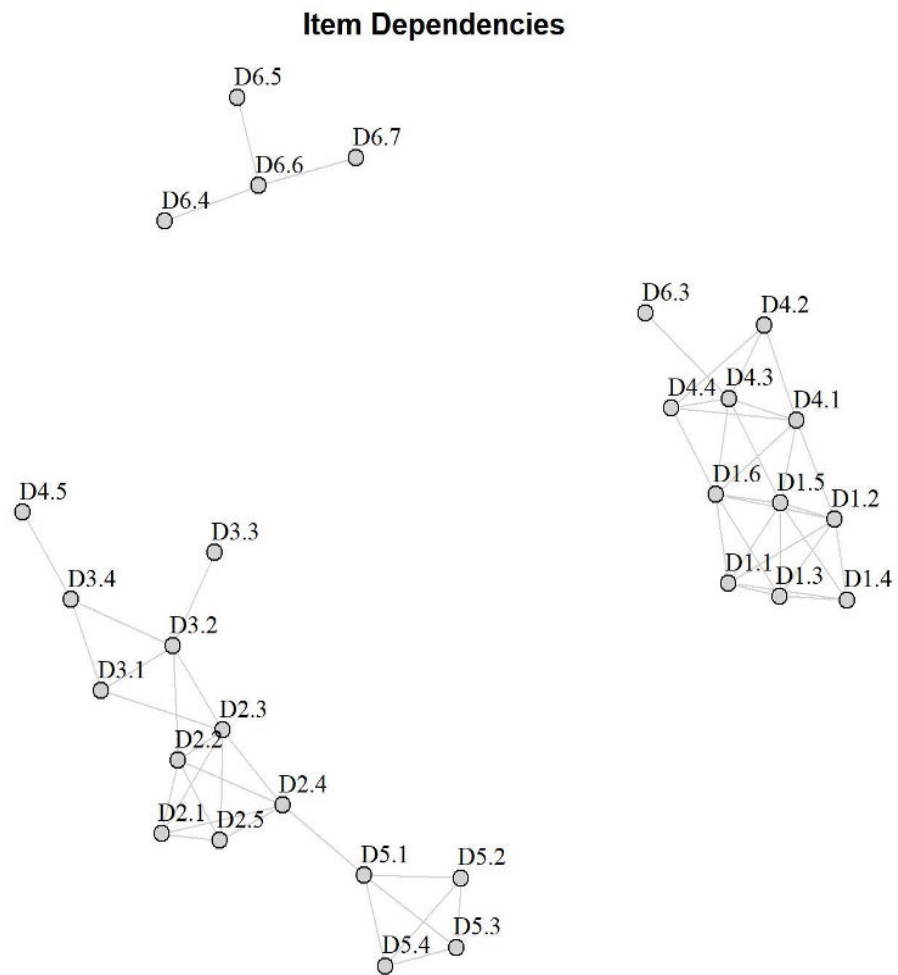


Figure 5: Person item map after the creation of testlets

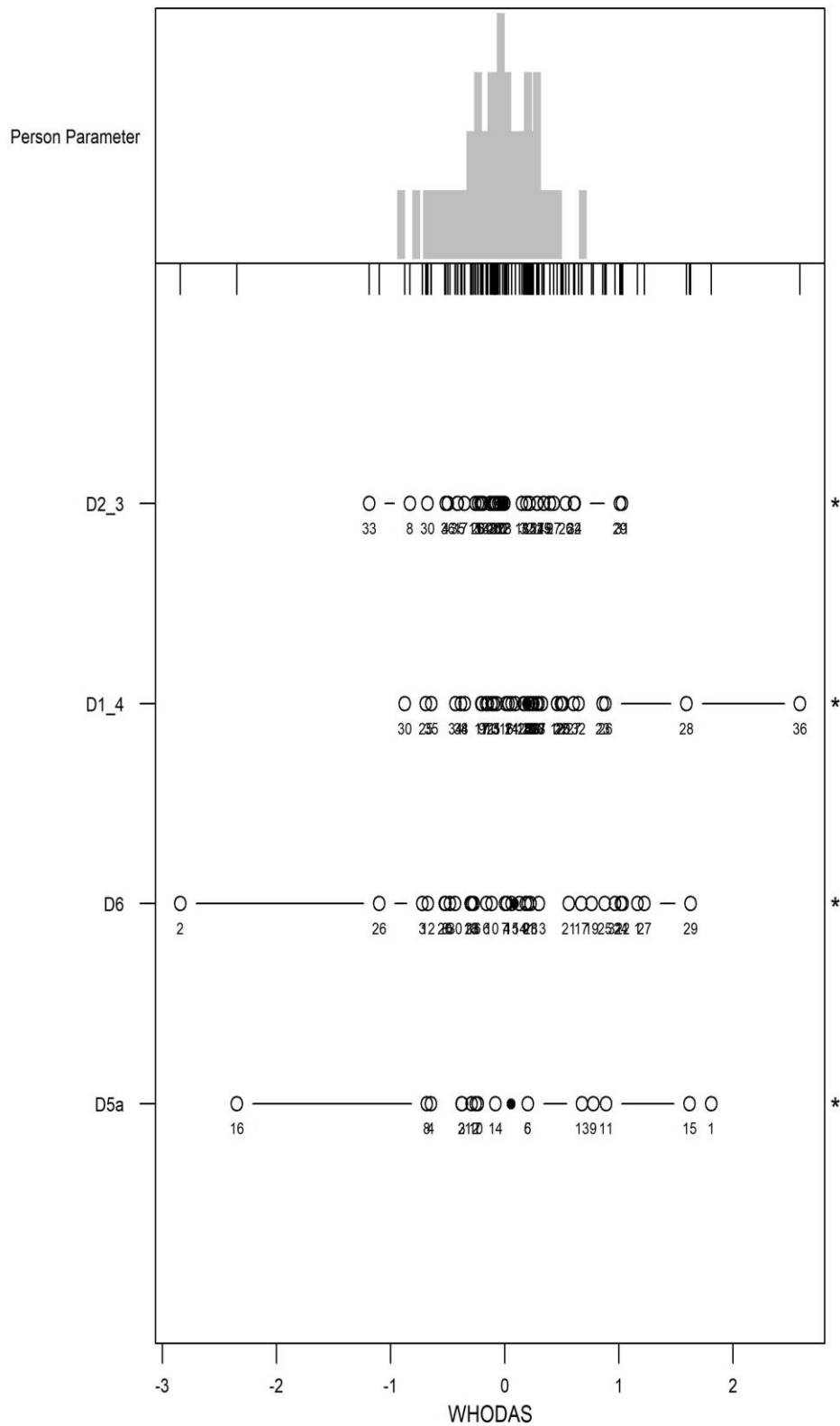


Table 5: WHODAS fit, difficulties, threshold ordering, local item dependencies, and differential item functioning without any adjustments

WHODAS	Outfit <sub>1</sub>	Infit <sub>1</sub>	Item	Disordered	LID <sup>3</sup>	DIF <sup>4</sup>
Item Nbr.			Difficulty	Thresholds		
D1.1	0.73	0.7	0.72	X	D1.2, D1.3, D1.4, D1.5, D1.6	
D1.2	0.85	0.85	0.85		D1.1, D1.3, D1.4, D1.5, D1.6, D4.1	
D1.3	0.75	0.76	0.76	X	D1.1, D1.2, D1.4, D1.5, D1.6, D4.1	
D1.4	0.79	0.79	0.79		D1.1, D1.2, D1.3, D1.5, D1.6	Age
D1.5	0.81	0.84	0.82		D1.1, D1.2, D1.3, D1.4, D1.6, D4.1, D4.3	Age
D1.6	0.78	0.81	0.8		D1.1, D1.2, D1.3, D1.4, D1.5, D4.1, D4.3	Age
D2.1	0.98	0.97	0.98	X	D2.2, D2.3, D2.4, D2.5	Age
D2.2	0.72	0.73	0.73		D2.1, D2.3, D2.4, D2.5, D3.2	Age
D2.3	0.52	0.52	0.52	X	D2.1, D2.2, D2.4, D2.5, D3.2	Age
D2.4	0.53	0.53	0.53		D2.1, D2.2, D2.3, D2.5	Age
D2.5	0.8	0.82	0.81		D2.1, D2.2, D2.3, D2.4	Age, Gender
D3.1	0.68	0.69	0.68	X	D3.2, D3.4	Age
D3.2	0.67	0.68	0.68	X	D2.2, D2.3, D3.1, D3.3, D3.4	Age
D3.3	0.42	0.52	0.47	X	D3.2, D3.4	
D3.4	0.7	0.71	0.7	X	D3.1, D3.2, D3.3, D6.2	Age
D4.1	1.06	1.03	1.04	X	D1.2, D1.3, D1.5, D1.6, D4.3, D4.4	Age, Gender
D4.2	0.62	0.59	0.61	X	D4.3, D4.4	
D4.3	0.68	0.7	0.69	X	D1.5, D1.6, D4.1, D4.2, D4.4	Age
D4.4	0.95	0.84	0.89	X	D4.1, D4.2, D4.3, D4.5	
D4.5	0.7	0.73	0.71	X	D4.4	
D5.1	0.51	0.51	0.51	X	D5.2, D5.3, D5.4	
D5.2	0.57	0.57	0.57		D5.1, D5.3, D5.4	
D5.3	0.59	0.59	0.59		D5.1, D5.2, D5.4	
D5.4	0.57	0.57	0.57		D5.1, D5.2, D5.3	
D6.1	0.59	0.59	0.59		D6.2	
D6.2	1.05	0.97	1.01		D3.4, D6.1	Age
D6.3	1.98	1.83	1.9	X	no	Age, Gender
D6.4	1.27	1.17	1.22		no	Gender
D6.5	0.88	0.76	0.82	X	D6.6	Age, Gender
D6.6	0.96	0.83	0.89	X	D6.5, D6.7, D6.8	
D6.7	0.83	0.75	0.79	X	D6.6, D6.8	Age
D6.8	0.89	0.86	0.88		D6.6, D6.7	Age

1 Infit and Outfit expected below 1.2 for the absence of underfit

2 In testlets, i.e., aggregated locally dependent items, the ordering of thresholds is not expected anymore

3 Local item dependency (LID) significant if LID > mean residual correlation + 0.2

4 Differential item functioning (DIF)

Table 6: WHODAS Item Difficulties, fit, local item dependencies, and differential item functioning after adjustments

Label	WHODAS Item Nbr.	Outfit <sup>1</sup>	Infit <sup>1</sup>	Item Difficulty	Disordered Thresholds	LID <sup>3</sup>	DIF <sup>4</sup>
Testlet 1	D1.1-D1.6 & D4.1-D4.5	0.65	0.68	0.25	n.a. <sup>2</sup>	no	
Testlet 2	D2.1-D2.5 & D3.1-D3.4	0.65	0.67	0.02	n.a. <sup>2</sup>	no	Age
Testlet 3	D5.1-D5.4	1.29	1.18	0.07	n.a. <sup>2</sup>	no	Age
Testlet 4	D6.1-D6.8	0.63	0.62	0.1	n.a. <sup>2</sup>	no	Age

1 Infit and Outfit expected below 1.2 for the absence of underfit

2 In testlets, i.e., aggregated locally dependent items, the ordering of thresholds is not expected anymore

3 Local item dependency (LID) significant if LID > mean residual correlation + 0.2

4 Differential item functioning (DIF)

Table 7: Targeting and Reliability of WHODAS items

	Start		1) Domain based item aggregation		2) Domain based item aggregation with imputed data	
	Targeting		Targeting		Targeting	
	Mean	SD	Mean	SD	Mean	SD
Difficulty	0.44	1.04	0.1	0.71	0.1	0.71
Ability	0.01	1.08	0	0.38	0	0.38
	PSI	Alpha	PSI	Alpha	PSI	Alpha
Reliability	0.95	0.95	0.89	0.86	0.89	0.86

Finally, Table 8 gives the score transformation, including logit scaled Rasch ability estimates, but mainly allows to recode scores from the 32 WHODAS items into a psychometrically sound interval-scaled metric.

Table 8: Transformation Table

WHODAS 2.0	Rasch	0-100	WHODAS 2.0*	Rasch*	0-100*
Score	Logit	Score	Score	Logit	Score
32	-1.4	0	96	0.16	51
33	-1.21	6	97	0.17	51
34	-1.03	12	98	0.17	51
35	-0.86	18	99	0.18	52
36	-0.74	21	100	0.18	52
37	-0.66	24	101	0.19	52
38	-0.6	26	102	0.2	52
39	-0.55	28	103	0.2	52
40	-0.51	29	104	0.21	52
41	-0.48	30	105	0.21	53
42	-0.45	31	106	0.22	53
43	-0.42	32	107	0.22	53
44	-0.4	33	108	0.23	53

45	-0.37	33	109	0.23	53
46	-0.35	34	110	0.24	53
47	-0.33	35	111	0.24	54
48	-0.32	35	112	0.25	54
49	-0.3	36	113	0.26	54
50	-0.28	36	114	0.26	54
51	-0.27	37	115	0.27	54
52	-0.25	37	116	0.27	55
53	-0.24	38	117	0.28	55
54	-0.22	38	118	0.28	55
55	-0.21	39	119	0.29	55
56	-0.2	39	120	0.29	55
57	-0.18	40	121	0.3	55
58	-0.17	40	122	0.3	56
59	-0.16	40	123	0.31	56
60	-0.15	41	124	0.32	56
61	-0.14	41	125	0.32	56
62	-0.13	42	126	0.33	56
63	-0.12	42	127	0.34	57
64	-0.1	42	128	0.34	57
65	-0.1	43	129	0.35	57
66	-0.09	43	130	0.36	57
67	-0.07	43	131	0.37	58
68	-0.06	44	132	0.38	58
69	-0.06	44	133	0.38	58
70	-0.04	44	134	0.39	59
71	-0.04	44	135	0.4	59
72	-0.03	45	136	0.41	59
73	-0.02	45	137	0.42	60
74	-0.01	45	138	0.44	60
75	0	46	139	0.45	60
76	0.01	46	140	0.47	61
77	0.02	46	141	0.48	62
78	0.03	47	142	0.5	62
79	0.04	47	143	0.52	63
80	0.04	47	144	0.55	64
81	0.05	47	145	0.57	64
82	0.06	48	146	0.6	65
83	0.07	48	147	0.63	66
84	0.08	48	148	0.67	67
85	0.09	48	149	0.7	69
86	0.09	49	150	0.74	70
87	0.1	49	151	0.8	72
88	0.11	49	152	0.87	74
89	0.12	49	153	0.95	77
90	0.12	50	154	1.04	80
91	0.13	50	155	1.13	83
92	0.14	50	156	1.23	86
93	0.14	50	157	1.34	89
94	0.15	51	158	1.45	93
95	0.16	51	159	1.55	96
96	0.16	51	160	1.66	100

---



## 5. Summary: Psychometric properties of WHODAS

The WHODAS 2.0 with 32 items, leaving out the work and education items, requires only a few adjustments. Aggregation of the items by chapter domains solved the multidimensionality caused by the domain-based item cluster and delivered an unbiased reliability estimate. This adjustment approach preserved the original conceptual form of the instrument. **The instrument is sound and delivers a reliable interval scaled score when items are considered by domains.** The DIF found for age groups in the testlet aggregating D2 *Getting around* and D3 *Self-Care*, and D5 *Life activities* indicate that with increasing age, the levels of functioning decrease, resulting in higher levels of disability. But not all domains are affected by the age, as shown by the absence of DIF in the testlet aggregating the domains D1 *Understanding and Communicating* and D4 *Getting along with people*.

However, the data collected in Latvia showed some peculiarities that future data collection and analysis would need to address.

**First, the number of missing values is high.** A high proportion of missing values significantly diminishes the reliability of the raw score. Data can be imputed, of course, but with a high proportion of missing values data imputation of good quality is jeopardized, with lower data variance than in a more complete non-imputed dataset. With more than 20% or even 50% of missing values, some items would benefit from being more closely observed by the interviewer to understand the reasons for the non-response and encourage applicants to respond. In any case, the interviewer may want to ensure that the applicant for disability assessment understands that the functioning level is rated, that the applicant's responses are kept safe, and will not be divulged. Note that the items with very high percentages, i.e., D3.4 *Staying by yourself for a few days*, D4.4 *Making new friends*, D4.5 *Sexual activities*, and D6.1 *How much of a problem did you have in joining community activities in the same way as anyone else can?* are sensitive items and have also been problematic in disability assessments elsewhere. Further, the high proportions of missing values in these items might be a consequence of the pandemic measures affecting items such as D6.1.

**Secondly, the disordering of the response options is very prominent.** This might be caused by the dependencies among the items and the missing values in the data. Another plausible explanation is that the pilot participants had difficulties differentiating 'none' from 'mild' or 'moderate' from 'severe' functioning problems. This reinforces the need for interviewers to support the applicants in finding the rating that best describes their performance.

The two above observations point to the need to train the interviewers to probe when the respondents are hesitant to answer a question. For the assessment of disability, it is important to get answers to all questions.

With these two caveats, taken together, the seven essential statistical tests described above show that the data collected with WHODAS, under the Rasch analysis, display robust psychometric properties of validity and reliability. With a few adjustments, the scale is unidimensional and free of item dependencies with good targeting and with good reliability. Aggregating the items by domains solves observed local item dependencies and produces a unidimensional assessment metric. The domain-based testlets fit well, and a transformation table is obtained that translates the observed sum scores into an interval-scaled metric.

It is important to keep in mind that the World Health Organization developed WHODAS explicitly to statistically capture the construct of functioning from the perspective of performance – namely the actual experience of performing activities by a person with an underlying health problem in their actual everyday life environment. There is an abundance of evidence from the scientific literature –

supported by the results of this pilot – that WHODAS is a psychometric sound instrument that reliably and validly collects information about levels of disability.

Therefore, we can confidently conclude that WHODAS information is sufficiently robust and relevant, and we recommend that it is applied in the assessment of disability in Latvia in its shift from medical to a functioning based assessment.

## 6. Exploring the metric properties of the WHODAS domains

The collaborating team at the Ministry of Welfare in Latvia expressed an interest in knowing the metric properties of each of the WHODAS domains. Table 9 and Table 10 show the metric properties of each of the WHODAS domains. The results of the separate analyses indicate that the scores of most WHODAS domains could potentially be used as a stand-alone domain-specific measure if needed.

The analysis of Domain D1 *Understanding and Communicating* had to account for a dependency between the items D1.5 and D1.6 but otherwise presented ordered thresholds, item fit, unidimensionality, and good reliability (PSI = 0.88 and Cronbach  $\alpha$  = 0.91). Domain D2 *Getting around* and Domain D5 *Life Activities*, complied to all measurement assumptions with very good reliability, unidimensionality, threshold ordering, item fit, and absence of local item dependencies and did not require adjustments. Domain D3 *Self-Care* as well as Domain D4 *Getting along with People*, only required collapsing of response options to have ordered thresholds (Table 10). Otherwise, these domains also did not show any local item dependencies, presented good item fit, and unidimensionality. After the collapsing of the disordered response options, the reliability was still good with PSI = 0.81 and Cronbach  $\alpha$  = 0.85 for Domain D3 and PSI = 0.8 and Cronbach  $\alpha$  = 0.81 for Domain D4. Domain D6 - *Participation in Society*, is problematic and requires collapsing of response options and the creation of testlets.

Nonetheless, the adjustments do not come with good fit for all items and D6.3 (*How much of a problem did you have living with dignity because of the attitudes and actions of others?*) presented an Infit > 1.2. Here again, the effect of the context of the data collection on the quality of the responses could also be questioned, as many interviews took place during the Corona pandemic, so that some items of this domain about *Participation in Society* may not be applicable for the time period in question.

In summary, with the present data collection, all domains, except the domain D6, present good measurement properties and could be used to construct domain specific scores. The specific statistics for the domain specific analyses of the WHODAS are presented in Table 9, for the model level information and Table 10, for the item level.

Table 9: Targeting, Item Parameter characteristics, Reliability, and Dimensionality of the WHODAS domains before and after adjustment

WHODAS Domain	Label	Stage	Targeting		Items			Reliability		Unidimensionality
			Difficulty Mean (SD)	Ability Mean (SD)	Threshold Ordering <sup>1</sup>	Fit <sup>2</sup>	LID <sup>3</sup>	PSI	Cronbach alpha	yes/no
Domain 1	Understanding and communicating	Start	1.43 (1.85)	0.06 (2.16)	yes	yes	yes	0.89	0.93	yes
		Final	1.17 (1.76)	0.06 (2.03)	yes	yes	no	0.88	0.91	yes
Domain 2	Getting around	Start & Final	-0.36 (2.01)	-0.02 (2.29)	yes	yes	no	0.9	0.91	yes
Domain 3	Self-care	Start	1.48 (1.47)	0.1 (2)	no	yes	no	0.83	0.9	yes
		Final	1.95 (1.65)	0.24 (2.11)	yes	yes	no	0.81	0.85	yes
Domain 4	Getting along with people	Start	1.39 (0.80)	0.24 (1.26)	no	yes	no	0.69	0.82	yes
		Final	2.33 (1.7)	0.19 (1.97)	yes	yes	no	0.8	0.81	yes
Domain 5	Life activities	Start & Final	-0.12 (2.96)	-0.05 (2.97)	yes	yes	no	0.91	0.94	yes
Domain 6	Participation in society	Start	0.25 (0.88)	0.01 (0.91)	no	no	yes	0.82	0.8	yes
		Final	0.97 (1.32)	0.01 (1.23)	yes	no	no	0.77	0.66	Yes

1 All items present ordered thresholds. For the testlets the ordering is not expected.

2 All items are fitting if fit = yes

3 Absence of local item dependencies if LID = no.

Table 10: Infit and outfit, differential item functioning, local item dependencies, recoding strategy by disordered thresholds

	Infit <sup>1</sup>		Outfit <sup>1</sup>		DIF		LID		Recoded
	Start	Final	Start	Final	Start	Final	Start	Final	Final
<b>Domain 1</b>	<b>Understanding and communicating</b>								
D1.1	0.76	0.73	0.68	0.66					
D1.2	0.77	0.75	0.72	0.69					
D1.3	0.72	0.69	0.63	0.61	Age	Age			
D1.4	0.81	0.78	0.73	0.71	Age, Gender	Age, Gender			
D1.5	0.83		0.70		Age, Gender		D1.6		
D1.6	1.03		0.91		Age		D1.5		
Testlet D1.5 & D1.6		1.04		0.87		Age			
<b>Domain 2</b>	<b>Getting around</b>								
D2.1	0.83	0.83	0.72	0.72					
D2.2	0.79	0.79	0.74	0.74	Age	Age			
D2.3	0.59	0.59	0.57	0.57					
D2.4	0.73	0.73	0.66	0.66					
D2.5	0.81	0.81	0.68	0.68	Gender	Gender			
<b>Domain 3</b>	<b>Self-care</b>								
D3.1	0.58	0.48	0.49	0.42					
D3.2	0.53	0.52	0.46	0.40					00112
D3.3	0.81	0.97	0.73	0.89					
D3.4	0.78	0.68	0.69	0.61	Age	Age			00112
<b>Domain 4</b>	<b>Getting along with people</b>								
D4.1	1.01	1.07	0.87	1.00		Age, Gender			00112
D4.2	0.46	0.61	0.38	0.35					00112
D4.3	0.56	0.78	0.47	0.48					00112
D4.4	0.58	0.49	0.49	0.44					
D4.5	0.84	0.77	0.73	0.68		Age			
<b>Domain 5</b>	<b>Life activities</b>								
D5.1	0.65	0.65	0.64	0.64					
D5.2	0.52	0.52	0.50	0.50					
D5.3	0.59	0.59	0.56	0.56					
D5.4	0.55	0.55	0.52	0.52					
<b>Domain 6</b>	<b>Participation in society</b>								
D6.1	0.74		0.73		Age		D6.2		
D6.2	0.92		0.91		Age		D6.1		
D6.3	1.31	1.24	1.31	1.14	Age, Gender	Age, Gender			
D6.4	0.96	0.87	0.96	0.86	Gender	Gender			
D6.5	0.70	0.73	0.70	0.73	Gender	Gender			
D6.6	0.70		0.69		Gender		D6.7		
D6.7	0.62		0.60		Gender		D6.6, D6.8		
D6.8	0.68		0.68		Gender		D6.7		
Testlet D6.1 & D6.2		0.84		0.80		Age			
Testlet D6.6, D6.7 & D6.8		0.59		0.57		Gender			

1 Infit and Outfit expected below 1.2 for the absence of underfit

## 7. Performance of the current assessment of disability in Latvia

In Latvia, persons with disability are categorized into these **disability status groups** by an ordinal scale:

- No disability: If **no functioning restrictions** are found or carrying out an activity is easy – it does not cause significant problems to function – disability is not found. The individual could still be identified as having a loss of general ability to work for up to 24.99 percent, but this is not regarded as a disability for the purposes of the assessment.
- Group III disability: Functioning restrictions are **moderate** if functioning is substantially limited, but not so much that the restrictions would be severe (daily activities can be done independently, but at a substantially slower pace or with more effort, or worse quality compared to the normally accepted standard for the corresponding age group). Group III disability is found when the loss of ability to work is assessed at 25.0-59.99 percent.
- Group II disability: Functioning restrictions are **severe** if functioning is substantially limited, restriction is higher than moderate, but is not very severe (most of daily activities can be done independently, but at a substantially slower pace or with more effort, or worse quality compared to the normally accepted standard for the corresponding age group, with episodic need for help or supervision. Group II disability is found when the loss of ability to work is 60.0-79.99 percent.
- Group I disability: Functioning restrictions are **very severe** if functioning is very limited or practically impossible (there is need for permanent or frequent episodic help or supervision in daily activities). Group I disability is found when the loss of ability to work is assessed as 80.0-100.0 percent.

A disability assessment system that is valid and reliable must, at a minimum, consistently map disability status groups that are ordinaly arranged by severity onto a functioning severity scale. WHODAS has been empirically shown to produce such a scale, so the pilot allowed us to evaluate the current Latvia system with a statistically valid metric of functioning. Although the disability status groups in the current system are described in terms of functioning limitations, in practice the only source of information used to assign an applicant to one of the four status groups is medical. (We are aware that a self-assessment form is also submitted with the application but, to the best of our knowledge is not systematically relied on for the determination of disability status groups. We will return to the self-assessment form below in section 8 as it needs to be discussed separately.)

As mentioned above, the only reliable way in which functioning information can be collected is, first by using a psychometrically robust and scientifically validated data collection instrument that collects functioning information from the perspective of performance, and secondly by avoiding as many of the potential biases and distortions associated with self-report as possible, by administering the instrument by a trained professional in a face-to-face interview. WHODAS is statistically and psychometrically valid and robust instrument to use – the results of the analysis of this pilot confirm that it is psychometrically strong and collects functioning information from the perspective of performance. The pilot, moreover, used trained professionals as interviewers. In short, we can rely on the WHODAS scores as presenting a valid level of functioning for each of the participants in the pilot.

There are several ways to demonstrate that the current system of determining disability status groups has problems to create disability groups, ordinaly ranked by severity, when compared to WHODAS scores.

### 7.1. Comparing disability status group determination and WHODAS scores by case studies:

The following four cases illustratively represent instances where the disability status group determined by the current medically based system does not align with the level of severity of functioning limitation identified by WHODAS score:

**A** was assessed in status group No disability. But the WHODAS functioning score is 60, indicating severe disability. She is a 59-year-old married and educated woman. She is in assisted living and cannot work because of health reasons. She was diagnosed with a malignant bladder cancer. She reports having had difficulties because of her health condition on every day of the last month and been unable to perform usual activities 2/3 of the time.

**B** was assessed in status group III or moderate functioning restriction. The WHODAS functioning score, however, is 55 or severe disability. He is a 40-year-old man with 14 years of education and is currently married and living independently in the community. He has a congenital malformation with a deformation of the spinal cord. He reports also having difficulties because of his health condition on every day of the last month. He is unable to perform usual activities or must reduce usual activities or work 1/3 of the time.

**C** is a 59-year-old married man with a WHODAS functioning score of 31 suggesting mild disability. He was determined to have status group II or severe functioning restriction due to an eye disease, specifically detachments and breaks of the retina, and a phonological disorder. He is still working and living independently in the community. His health condition is not limiting him in his daily life, and he can perform his usual activities normally without having to reduce them.

**D** is a 72-year-old woman with 15 years of education and lives with a partner. She was assessed as having disability status group 1 or severe functioning restriction. She was diagnosed with a malignant melanoma of the skin. She presents a WHODAS score of 32. Her health condition is not limiting her in performing daily life activities.

### 7.2 Comparing discrimination between disability status groups and WHODAS scores

Table 11 and Figure 6 present a comparison between the discriminatory power of the medical assessment and the WHODAS score, showing that **the current system has difficulties discriminating levels of functioning.**

For reference, in the pilot sample,

- 242 persons (11.2%) were certified as No disability,
- 831 persons (38.45%) were certified as Group III disability (moderate restrictions in functioning),
- 685 persons (31.7%) were certified as Group II disability (severe restrictions in functioning), and
- 403 persons (18.65%) were certified as Group I disability (very severe restrictions in functioning).

Table 11 shows that the No disability and Group III (moderate) disability status groups have very similar mean WHODAS functioning scores: No 44.7(6.99) and Group III – 43.8(7.09). The quartiles in Table 11 show that only the values of the medical categorization of very severe disability have WHODAS-quartile scores that are higher, indicating higher functioning problems. The quartile of the other groups, including the pilot sample score quartiles, are very close. This is further confirmed in Tables 12 and 13: While an ANOVA of the WHODAS score by disability status group is significant

( $F_{(df)} = 280.4_{(3)}$ ;  $p - value < 0.001$ , Table 12) and the Tuckey Test confirms significantly different WHODAS-scores across higher disability status groups, a differentiation of the score for moderate and no functioning restriction ( $p - value = 0.403$ , Table 13) is not possible.

The lack of discrimination between disability status groups is neither surprising nor unexpected, as the current disability status assessment is, as noted, based on inference about disability from the medical information, while WHODAS captures the performance perspective on activities and participation. **The current system, in short, does not create empirically sound status groups of disability.** This can be further confirmed by two other statistical comparisons: First Figure 6 which shows the density lines by determined disability status groups and the fact that the WHODAS scores for very severe functioning restrictions (group I) stand out, while the difference between severe, moderate and no functioning restrictions (no disability and groups III and II) is subtle, and the difference between moderate (group III) and no functioning restrictions is insignificant.

Figure 6: WHODAS-score density line by determined disability status groups

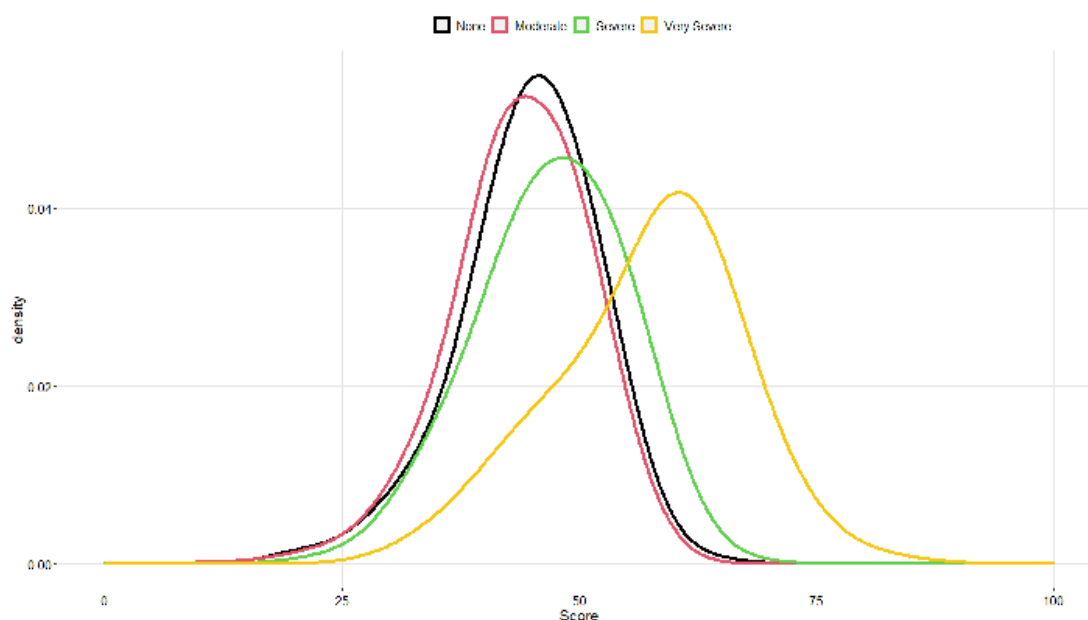


Table 11: WHODAS scores distribution for the pilot sample and by disability status group: mean, standard deviation and quartiles

	Mean	SD	25%	50%	75%
<b>Pilot</b>	47.4	9.17	42	47	53
<b>Disability (None)</b>	44.7	6.99	41	46	49
<b>Group III (Moderate)</b>	43.8	7.09	40	44	49
<b>Group II (Severe)</b>	46.9	7.85	42	48	53
<b>Group I (Very severe)</b>	57.1	9.36	51	59	63

Table 12: Analysis of variance of the WHODAS-scores by disability status groups

	DF	Sum of squares	Mean squares	F-value	P-value
Disability status	3	50995	16998	280.4	< 0.001
Residuals	2157	130781	61		

Table 13: Tukey Test for the WHODAS-scores by disability status groups

	Difference	Lower CI	Upper CI	P-values
Group III vs. none	-0.887	-2.349	0.576	0.403
Group II vs. none	2.233	0.736	3.73	0.001
Group I vs. none	12.46	10.832	14.089	< 0.001
Group II vs. Group I	3.12	2.086	4.153	< 0.001
Group III vs. Group I	13.347	12.132	14.562	< 0.001
Group III vs. Group II	10.227	8.971	11.484	< 0.001

Secondly, Table 14 presents socio-demographic characteristics of the pilot sample, disaggregated by determined disability status groups including p-value for a significance test comparing each variable by disability status groups. The groups differ significantly in terms of age (p-value < 0.001), especially the higher age for the very severe disabilities – 70+ years of age (SD = 13.66), the average age in the other disability status groups being < 60 years. Years of education also differs significantly across disability status groups (p-value < 0.001), the marital status (p-value < 0.001), living condition (p-value < 0.001), and the work status (p-value < 0.001). The percentage of divorced and widowed persons is somewhat higher in the group of persons with the severe disability status. Retired persons also feature high percentage of severe disability (61.5%), as well as those in assisted living (61.5%). Pilot participants with determined disability group III and II have higher unemployment rates (above 20%).

Table 14: Pilot sample descriptive statistics by disability status groups

	None	Group III Moderate	Group II Severe	Group I Very severe	p-value
N (%)	242 (11.2)	831 (38.5)	685 (31.7)	403 (18.6)	
Gender = Male N/ (%)	91 (37.6)	340 (40.9)	305 (44.5)	161 (40.0)	0.203
Age – mean (SD)	52.74 (10.79)	53.54 (10.85)	59.40 (12.78)	70.02 (13.66)	<0.001
Years of Education – mean (SD)	12.58 (3.06)	12.72 (2.71)	12.78 (3.19)	11.77 (3.31)	<0.001
Marital Status N/ (%)					<0.001
Never married	20 (8.3)	83 (10.0)	71 (10.4)	28 (6.9)	
Currently married	118 (48.8)	396 (47.7)	312 (45.5)	158 (39.2)	
Separated	10 (4.1)	48 (5.8)	43 (6.3)	19 (4.7)	
Divorced	27 (11.2)	106 (12.8)	74 (10.8)	68 (16.9)	
Widowed	23 (9.5)	70 (8.4)	86 (12.6)	102 (25.3)	
Cohabiting	44 (18.2)	128 (15.4)	99 (14.5)	28 (6.9)	
Living Condition N/ (%)					<0.001
Independent in the community	228 (94.2)	798 (96.0)	602 (87.9)	134 (33.3)	
Assisted living	13 (5.4)	32 (3.9)	82 (12.0)	268 (66.5)	
Hospitalized	1 (0.4)	1 (0.1)	1 (0.1)	1 (0.2)	
Work Status N/ (%)					<0.001
Paid work	101 (41.7)	391 (47.1)	168 (24.5)	21 (5.2)	
Self-employed	4 (1.7)	30 (3.6)	16 (2.3)	5 (1.2)	
Non-paid work	0 (0.0)	2 (0.2)	0 (0.0)	1 (0.2)	
Student	1 (0.4)	11 (1.3)	3 (0.4)	0 (0.0)	



Keeping house	0 (0.0)	9 (1.1)	4 (0.6)	2 (0.5)
Retired	10 (4.1)	75 (9.0)	204 (29.8)	248 (61.5)
Unemployed (health reasons)	27 (11.2)	194 (23.3)	198 (28.9)	51 (12.7)
Unemployed (other reasons)	7 (2.9)	32 (3.9)	15 (2.2)	7 (1.7)
Other	92 (38.0)	87 (10.5)	77 (11.2)	68 (16.9)

### 7.3 Comparing disability status groups with WHODAS scores in terms of correlations with ICD chapters

Finally, it is important to correlate the disability status groups and WHODAS scores in terms of how they are related to the main health condition of the individual applicant. Table 15 presents WHODAS-scores by main International Classification of Diseases (ICD-10) chapters by determined disability status groups of the pilot participants. The table only shows a sharper increase of WHODAS scores for the group with very severe functioning restrictions, compared to groups without, moderate or severe functioning restrictions, corroborating our conclusion that the current method of disability assessment does not differentiate well degrees of difficulties in functioning.

With increasing disability, the reported burden of the health condition on functioning increases also, as seen, especially in the high WHODAS-scores in the group with very severe functioning restrictions, with most mean WHODAS-scores > 50, even above 60. Mean WHODAS-scores of above 50 are only exceptional in the other disability status groups: one person without disability status and 10 persons from the severe functioning restrictions reporting, respectively, a score of 58 and mean score of 51.2 (SD = 6.2) for *I Certain infectious and parasitic disease*. In the group with severe functioning restrictions, 8 persons reported an average WHODAS-score of 51.25 (SD = 2.87) for *X Diseases of the respiratory system*.

Table 15 also shows, for the sample and disaggregated by determined disability status group of the pilot participants, the number and percentage of reported main ICD condition chapters as well as the corresponding mean and standard deviation of the WHODAS score. We learn two things from this, first that the most frequently reported conditions are *II Neoplasms*, *IX Diseases of the circulatory system*, *XII Diseases of the musculoskeletal system, and connective tissue*, and *XIX Injury, poisoning and certain other consequences of external causes*. In the groups of severe and very severe functioning restrictions, *VII Diseases of the eye and adnexa* are also reported as main condition for > 10% of cases in the group. But secondly, it is important to notice that **only 2.4% of the assessed population has an ICD-code from the chapter V Mental and behavioral disorders, and 50% of these group has some form of dementia**. In our experience this is a common phenomenon, probably caused by the fact that mental and behavioral disorders, other than dementias associated with ageing are under-diagnosed.

It is clear from these separate analyses, exploring the structure and properties of the disability status group as it currently is conducted with the metric standard that the WHODAS pilot data constructed, that the status groups do not, in various ways, consistently represent a meaningful ordinal ranking of severity of disability across the participants in the pilot. Medical information is obviously relevant to the determination of disability. But these results shows that it is not suitable to rely on subjective judgement of medical professionals, based on diagnostic and other medical information alone to decide about the ordinal ranking of the severity of functioning problem of each applicant, represented by disability status groups. WHODAS offers direct, empirically robust means for collection of information about functioning and, on the basis of the kind of statistical analysis presented here, to create a linear metric of severity of functioning problem, or disability, that can validly and reliably be the basis for disability assessment.

Table 15: Frequency and Percentage of ICD chapters for the pilot sample and by disability status group as well as the mean and standard deviation (SD) of the corresponding WHODAS-scores

ICD Chapters	N (%)	Mean (SD)	N None	Mean (SD) None	N Group III	Mean (SD) Group III	N Group II	Mean (SD) Group II	N Group I	Mean (SD) Group I
I Certain infectious and parasitic diseases	28(1.3%)	47.29 (9.3)	1(0.41%)	58	14(1.68%)	41.43 (7.6)	10(1.46%)	51.2 (6.2)	3(0.74%)	58 (8.54)
II Neoplasms	389(18%)	46.33 (9.46)	34(14.05%)	44.5 (7.46)	84(10.11%)	42.81 (6.81)	163(23.8%)	44.25 (9.28)	108(26.8%)	52.79 (9.02)
III Diseases of the blood	3(0.14%)	45.33 (16.44)			2(0.24%)	36 (4.24)			1(0.25%)	64
IV Endocrine nutritional and metabolic diseases	47(2.17%)	46.32 (8.54)	2(0.83%)	43 (4.24)	17(2.05%)	41.82 (8.57)	23(3.36%)	49 (7.02)	5(1.24%)	50.6 (10.71)
IX Diseases of the circulatory system	372(17.21%)	52.22 (11)	30(12.4%)	45 (7.65)	75(9.03%)	43.85 (7.9)	125(18.25%)	47.58 (7.32)	142(35.24%)	62.25 (7.58)
V Mental and behavioral disorders	44(2.04%)	51.59 (9.37)	2(0.83%)	47.5 (7.78)	3(0.36%)	41.67 (6.66)	26(3.8%)	49.42 (6.69)	13(3.23%)	58.85 (10.69)
VI Diseases of the nervous system	124(5.74%)	49.56 (9.84)	16(6.61%)	44.38 (6.66)	35(4.21%)	42.94 (7.96)	43(6.28%)	49.63 (7.47)	30(7.44%)	59.97 (7.3)
VII Diseases of the eye and adnexa	158(7.31%)	45.44 (9.07)			31(3.73%)	38.35 (10.97)	77(11.24%)	44.32 (6.51)	50(12.41%)	51.56 (7.22)
VIII Disease of the ear and mastoid process	5(0.23%)	37.8 (9.26)			5(0.6%)	37.8 (9.26)				
X Diseases of the respiratory system	23(1.06%)	48.87 (5.59)	3(1.24%)	45.33 (3.06)	10(1.2%)	46 (5.16)	8(1.17%)	51.25 (2.87)	2(0.5%)	59 (1.41)
XI Diseases of the digestive system	27(1.25%)	46.74 (9.51)	2(0.83%)	42.5 (2.12)	11(1.32%)	44.55 (7.53)	8(1.17%)	42.62 (8.78)	6(1.49%)	57.67 (7.71)
XII Diseases of the skin and the subcutaneous tissue	11(0.51%)	46.45 (5.94)	1(0.41%)	49	7(0.84%)	46.29 (6.73)	3(0.44%)	46 (6.08)		
XIII Diseases of the musculoskeletal system and connective tissue	497(23%)	46.62 (6.71)	89(36.78%)	44.53 (6.81)	260(31.29%)	45.37 (5.74)	136(19.85%)	49.12 (6.34)	12(2.98%)	60.83 (4.22)
XIV Diseases of the genitourinary system	19(0.88%)	46.47 (9.31)	1(0.41%)	50	3(0.36%)	46.33 (9.5)	6(0.88%)	41.67 (10.69)	9(2.23%)	49.33 (8.53)
XIX Injury poisoning and certain other consequences of external causes	385(17.82%)	44.95 (7.62)	60(24.79%)	44.6 (7.28)	262(31.53%)	43.6 (6.71)	45(6.57%)	48.29 (6.98)	18(4.47%)	57.39 (9.34)
XVII Congenital malformations deformations and chromosomal abnormalities	16(0.74%)	41.19 (11.34)			4(0.48%)	36.25 (14.36)	9(1.31%)	41.44 (8.4)	3(0.74%)	47 (16.52)
XVIII Symptoms signs and abnormal clinical and laboratory findings	1(0.05%)	33			1(0.12%)	33				
XXI Factors influencing health status and contact with health services	12(0.56%)	44.75 (5.74)	1(0.41%)	40	7(0.84%)	46.29 (6.85)	3(0.44%)	44.67 (2.08)	1(0.25%)	39 (NA)

## 8. The self-assessment form and collection and use of data on functioning for the assessment

As mentioned above, a self-assessment form is submitted with the application for the assessment, and this includes functioning questions. As far as we have been able to determine, this information is not systematically used in the assessment. The question arises whether this self-reported assessment could be used, either on its own or in conjunction with WHODAS, as an option to augment the medical assessment. Self-assessment opens up a range of concerns about bias or outright fraud that are sufficiently concerning that it is unlikely that any country would be comfortable relying on this data alone to make disability determination decisions. However, it is important to evaluate the self-assessment form separately and to compare it with WHODAS to be able to advise whether the self-assessment form, might be an additional option for Latvia.

### 8.1 Descriptive statistics

The self-assessment form contains, *inter alia*, 21 questions on functioning (Table 16). It combines ICF items from the categories of body functions (b-codes) and activities and participation (d-codes). The items use the response options of ICF 0 = no, 1 = mild, 2 = moderate, 3 = severe, 4 = extreme problems in functioning. The self-assessment form was developed locally and, to the best of our knowledge, has not been tested psychometrically. This is very important as any functioning assessment instrument must meet the psychometric requirements to be valid and reliable (see sections above on psychometric properties of WHODAS) and to allow that a score of functioning along the continuum 0-100 is derived. Moreover, based on the interviews conducted for the project, our impression is that the self-assessment is not commonly considered in the assessment.

It should be noted that information on functioning can be collected and used to inform the assessment in a qualitative manner. But this way calls for judgment on how to include it in the assessment, which may vary a lot from assessor to assessor. While ordinal scales and qualitative information are more suitable for the needs assessment where excellent psychometric properties of the instrument may not be *sine qua non* for an accurate assessment of needs, a disability assessment criteria and procedure, must minimize room for discretionary decision making (see box 1 on the credibility of disability assessment).

#### Box 2: The credibility of disability assessment

The credibility and perceived legitimacy of a country's disability assessment procedure depends on a few fundamental considerations. First of all, the assessment must be **valid** to minimize 'false positives' (people assessed as disabled and receiving benefits but are not disabled) or 'false negatives' (people who should be assessed as having a disability and receiving benefits, but they are not) – see four examples above. Second, the procedure must be **reliable**, in the sense that two assessors following the same rules and criteria should be able to come to the same assessment of the same person (often called 'inter-rater reliability'). And lastly, the **decisions must be transparent and standardized**, so that the grounds for the decision-making are publicly known and their application in particular cases, when needed and applicable, independently evaluated. In short, the legitimacy of the disability assessment process depends on it being, and be seen to be, impartial, fair, and based on objective evidence.

Disability is complex and difficult to measure, and these credibility criteria are not easy to achieve in practice. Even in the most sophisticated and well-resourced countries time and other

limitations mean that mistakes can be made. Assessors rely on the supporting evidence they are provided, which may contain errors, and there are invariably differences between assessors in how the evidence is evaluated and weighed. Yet the overall accuracy of disability assessment is crucial for the political sustainability, and perceived fairness of social security and other policies that rely on disability assessment. If expert disability assessors, following the rules they have been set down, often came to different judgments about the same applicant, then the process might be viewed as arbitrary and unjust.

See: *Bickenbach, Jerome; Posarac, Aleksandra; Cieza, Alarcos; Kostanjsek, Nenad. 2015. Assessing Disability in Working Age Population: A Paradigm Shift from Impairment and Functional Limitation to the Disability Approach. World Bank, Washington, DC. © World Bank. <https://openknowledge.worldbank.org/handle/10986/22353> License: CC BY 3.0 IGO."*

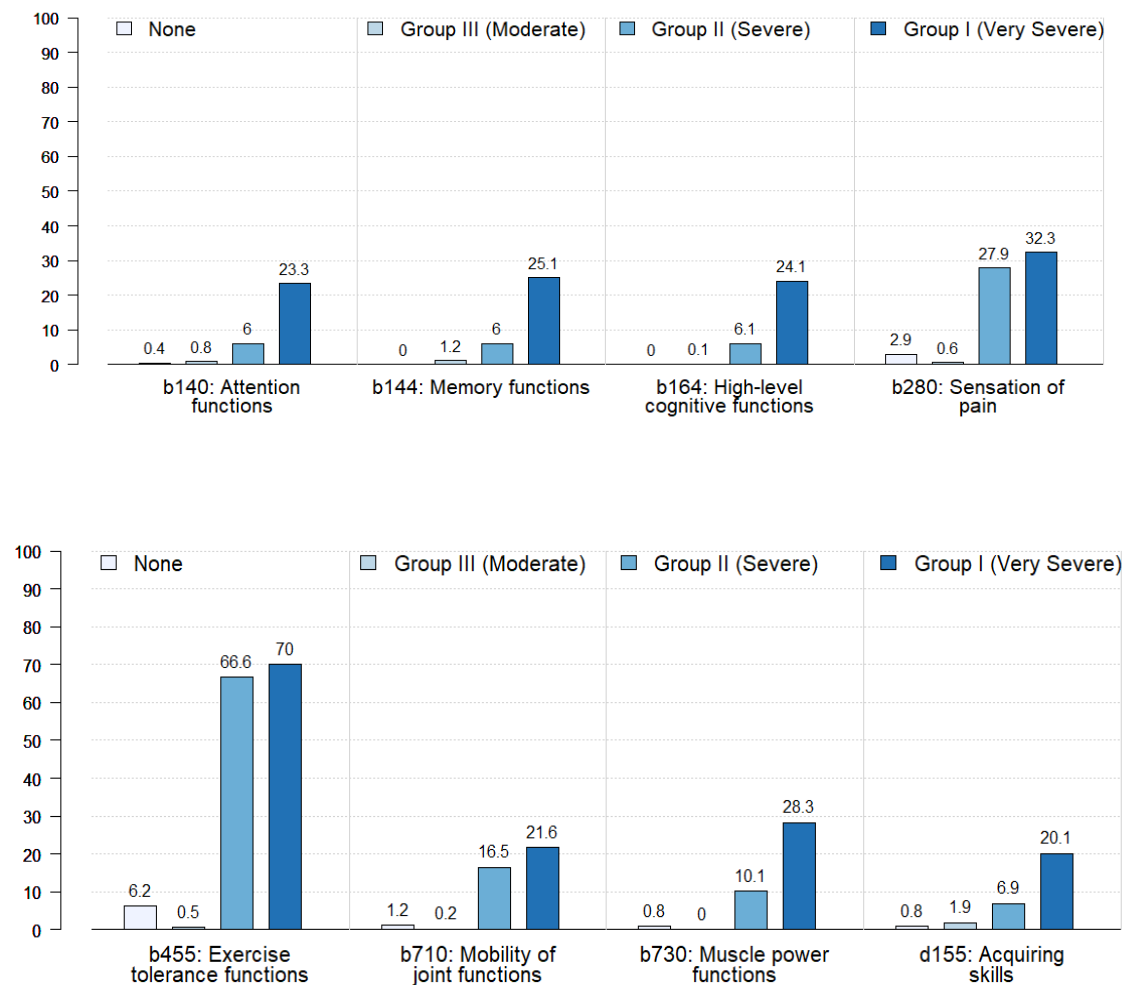
Table 16 shows the pilot sample frequencies and percentages by the chosen items of functioning and by the degree of functioning difficulty. Overall, the composition of self-reported difficulties in functioning across the sample and 21 ICF categories, is somewhat different than what one would expect based on the composition of the degree of disability among the pilot participants (18% of the sample was certified as having a severe disability and 31 as having a very severe disability). According to the self-reported problems, it can be noted that a high percentage for severe and extreme problems of functioning are found for b455 - *Exercise tolerance functions* (23.8% and 11.85%). Extreme problems in d450 - *Walking* are reported by 11.85% of the pilot population. More than 20% of the pilot reported moderate problems with b280 - *Sensation of pain* (32.76%), b455 - *Exercise tolerance functions* (31.28%), d410 - *Changing basic body position* (25.68%), d415 - *Maintaining a body position* (22.81%), and d450 - *Walking* (23.51%).

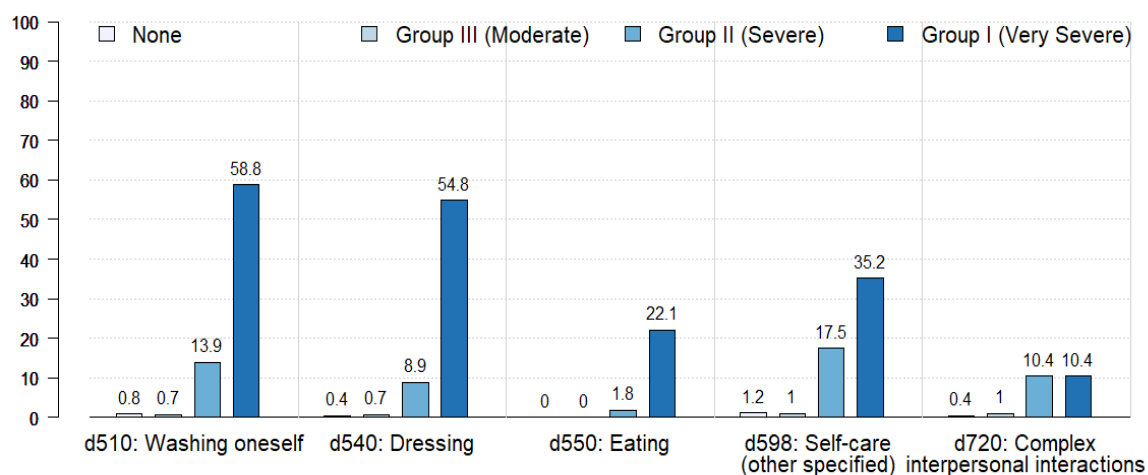
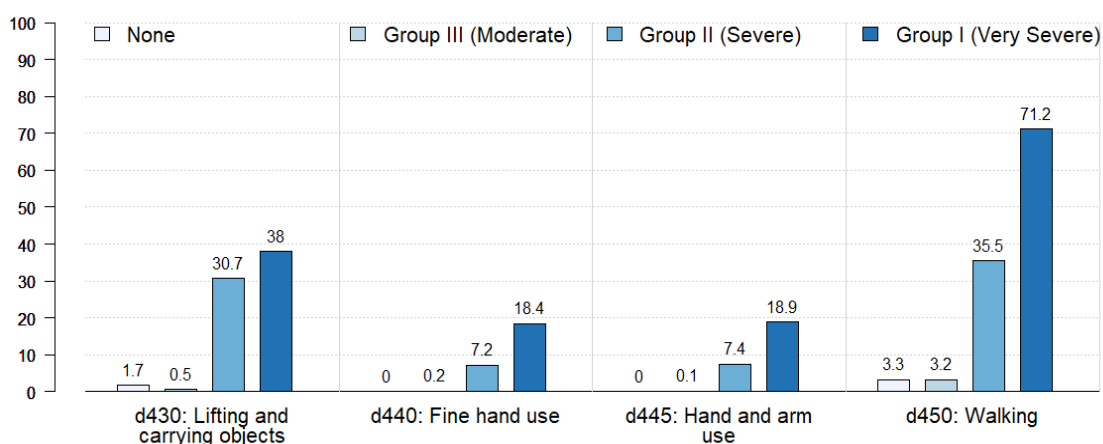
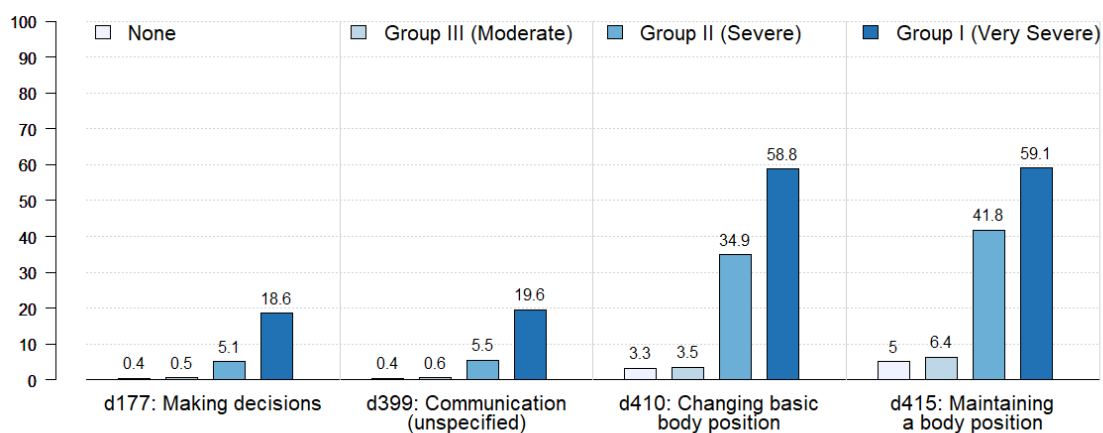
Table 16: ICF-based rating of levels of functioning problems

ICF-category	0 = No	1 = Mild	2 = Moderate	3 = Severe	4 = Extreme
b140: Attention functions	1866 (86.35%)	65 (3.01%)	87 (4.03%)	67 (3.1%)	76 (3.52%)
b144: Memory functions	1827 (84.54%)	61 (2.82%)	121 (5.6%)	69 (3.19%)	83 (3.84%)
b164: High-level cognitive functions	1922 (88.94%)	37 (1.71%)	62 (2.87%)	65 (3.01%)	75 (3.47%)
b280: Sensation of pain	1001 (46.32%)	119 (5.51%)	708 (32.76%)	260 (12.03%)	73 (3.38%)
b455: Exercise tolerance functions	641 (29.66%)	87 (4.03%)	676 (31.28%)	501 (23.18%)	256 (11.85%)
b710: Mobility of joint functions	1439 (66.59%)	120 (5.55%)	397 (18.37%)	139 (6.43%)	66 (3.05%)
b730: Muscle power functions	1699 (78.62%)	113 (5.23%)	164 (7.59%)	89 (4.12%)	96 (4.44%)
d155: Acquiring skills	1911 (88.43%)	51 (2.36%)	53 (2.45%)	64 (2.96%)	82 (3.79%)
d177: Making decisions	1935 (89.54%)	48 (2.22%)	63 (2.92%)	43 (1.99%)	72 (3.33%)
d399: Communication (unspecified)	1905 (88.15%)	53 (2.45%)	80 (3.7%)	66 (3.05%)	57 (2.64%)
d410: Changing basic body position	979 (45.3%)	114 (5.28%)	555 (25.68%)	311 (14.39%)	202 (9.35%)
d415: Maintaining a body position	993 (45.95%)	86 (3.98%)	493 (22.81%)	380 (17.58%)	209 (9.67%)
d430: Lifting and carrying objects	1300 (60.16%)	82 (3.79%)	408 (18.88%)	237 (10.97%)	134 (6.2%)
d440: Fine hand use	1714 (79.32%)	94 (4.35%)	228 (10.55%)	71 (3.29%)	54 (2.5%)
d445: Hand and arm use	1670 (77.28%)	97 (4.49%)	266 (12.31%)	78 (3.61%)	50 (2.31%)
d450: Walking	949 (43.91%)	139 (6.43%)	508 (23.51%)	307 (14.21%)	258 (11.94%)
d510: Washing oneself	1246 (57.66%)	217 (10.04%)	358 (16.57%)	160 (7.4%)	180 (8.33%)
d540: Dressing	1310 (60.62%)	232 (10.74%)	330 (15.27%)	128 (5.92%)	161 (7.45%)
d550: Eating	1860 (86.07%)	123 (5.69%)	77 (3.56%)	46 (2.13%)	55 (2.55%)
d598: Self-care (other specified)	1701 (78.71%)	47 (2.17%)	140 (6.48%)	141 (6.52%)	132 (6.11%)
d720: Complex interpersonal interactions	1876 (86.81%)	44 (2.04%)	119 (5.51%)	83 (3.84%)	39 (1.8%)

Figure 7 presents percentage of self-reported severe or extreme problems in functioning by determined disability status groups. For example, in b280 (Sensation and pain), 2.9% assessed as no disability, and 0.6% assessed as having a moderate (Group III) disability self-reported severe and/or extreme problem. The percentages were 27.9% and 32.3% In Group II and Group I, respectively.

Figure 7: Percentage of self-reported problems in functioning by the CMS determined disability groups





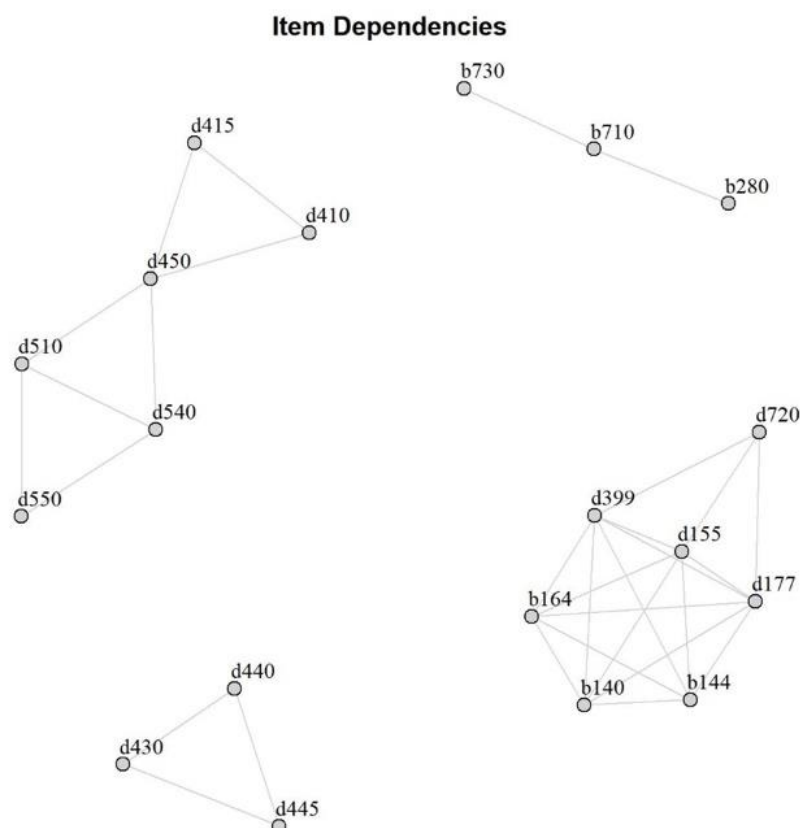
For all ICF-categories taken together, >10% of the pilot population in disability status group I, indicated functioning difficulties or problems. Highest ratings of severe and extreme functioning problems are found for d450 *Walking* (71.2%) and b455 *Exercise and Tolerance functions* (70%).

## 8.2 Psychometric characteristics of the 21 ICF categories from the self-assessment form

We tested the psychometric properties of the functioning categories in the self-assessment form. The objective was to investigate whether these ICF categories could be validly and reliably summarized in one summary score as in the case of WHODAS?

A Rasch analysis showed that using these ICF categories to build a functioning scale would be problematic. Only two items are not dependent on others, i.e., *b455 Exercise tolerance functions* and *d598 Self-care (other specified)*. Figure 8 presents significant dependencies between ICF-categories, they only correspond to some extent to the structure of the ICF (e.g., d430, d440, d445 as part of *Carrying, moving and handling objects*). These dependencies also affect the dimensionality and show that they define a multidimensional construct. The reliability is at an acceptable level (PSI = 0.88), but only until adjusting for the item dependencies (item dependencies are known to inflate reliability estimates). After adjustment, the reliability dropped to PSI = 0.66, which is insufficient. To some extent this low reliability can also be explained by the poor targeting of the ICF-categories to the sample's functioning level as shown in Figure 12. The mean difficulty (in logits) of the ICF-categories was 1.02 (SD.1.36) while the samples mean level of functioning (in logits) was 0.01 (SD = 0.73). This indicates, that taken together, the selection of ICF-categories would overestimate the levels of functioning problems. Regarding the item fit, only two items showed some misfit: *b280 Sensation of Pain* and *d720 Complex Interpersonal Interactions*. Detailed statistics regarding the fit of the ICF-categories to the Rasch model are presented in Table 17.

Figure 8: Local item dependencies between the ICF-categories used for the self-assessment.



The person item map in Figure 9 shows first that a high proportion of the applicants do not show any problems or difficulty in the selected ICF categories. The first bar in the upper part of the Figure represents 7.96% of the sample. The lower part indicates the positions of the difficulty thresholds for each item. The rating scale with 5 options (1= no problem, 2 = mild, 3 = moderate, 4 = severe, 5 = extreme) is not working properly. All items present disordered thresholds. Based on their location (black dots), the items where most applicants would indicate more functioning problems, are b455 *Exercise tolerance functions* and d450 *Walking*. The items where fewest applicants would indicate problems are d720 *Complex interpersonal interactions* and d550 *Eating*.

Figure 9: Person item map for the ICF-categories, without any adjustments of the data

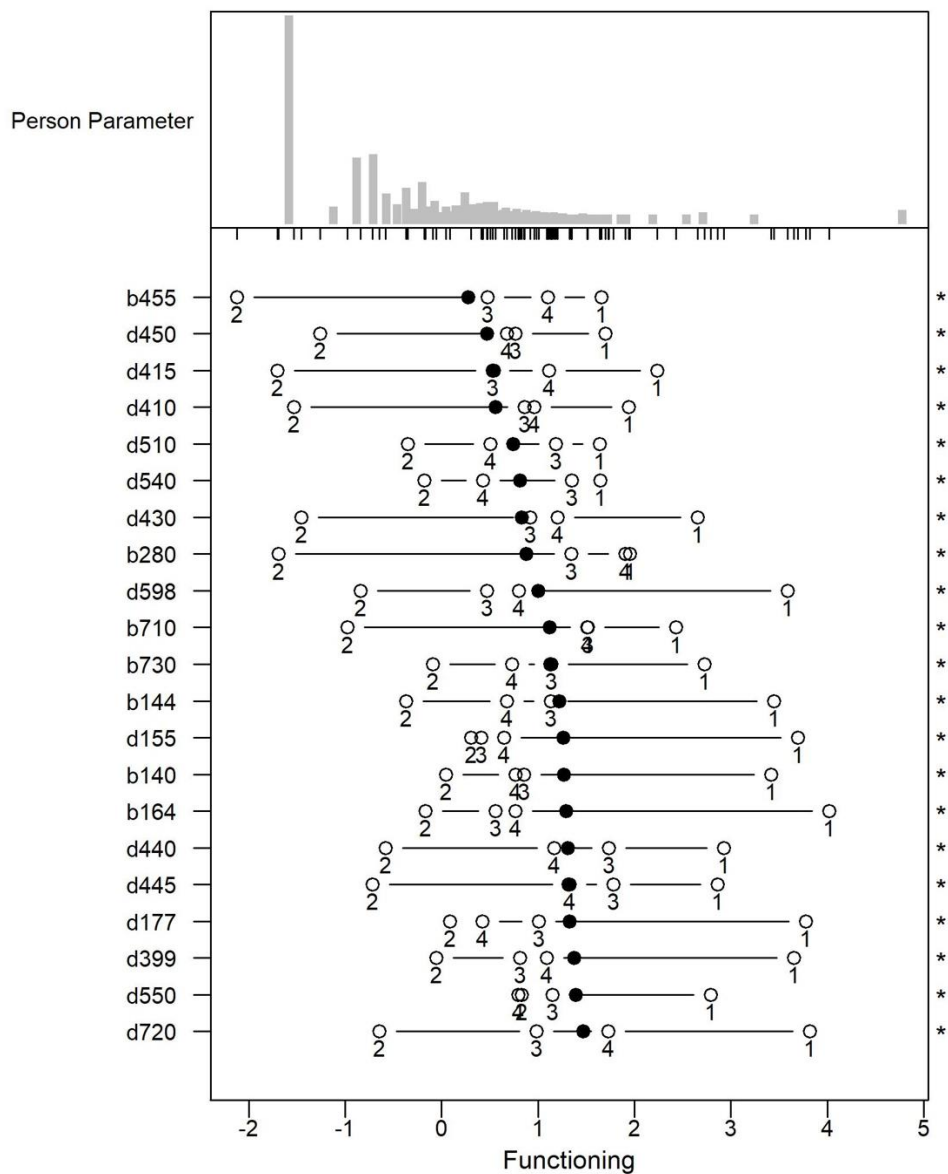




Table 17: ICF categories fit, difficulties, threshold disordering, local item dependencies, and differential item functioning without any adjustments

ICF-code	Label	Outfit <sup>1</sup>	Infit <sup>1</sup>	Difficulty	Disrd. Thresh.	LID <sup>3</sup>	DIF <sup>4</sup>
<b>b140</b>	Attention functions	0.56	0.92	1.27	X	b144, b164, d155, d177, d399	
<b>b144</b>	Memory functions	0.76	0.93	1.22	X	b164, d155, d177, d399	
<b>b164</b>	High-level cognitive functions	0.96	1.05	1.29	X	b144, d155, d177, d399	
<b>b280</b>	Sensation of pain	1.19	1.22	0.88	X	b710	Gender, Age
<b>b455</b>	Exercise tolerance functions	1.08	1.05	0.28	X		
<b>b710</b>	Mobility of joint functions	1	1.15	1.12	X	b280, b730	Age
<b>b730</b>	Muscle power functions	1.07	1.15	1.12	X	b710	
<b>d155</b>	Acquiring skills	0.58	0.96	1.26	X	b144, b164, d177, d399, d720	
<b>d177</b>	Making decisions	0.63	0.94	1.32	X	b144, b164, d155, d399, d720	Age
<b>d399</b>	Communication (unspecified)	0.91	1.03	1.38	X	b144, b164, d155, d177, d720	Age
<b>d410</b>	Changing basic body position	0.64	0.7	0.56	X	d415, d450	
<b>d415</b>	Maintaining a body position	0.64	0.7	0.54	X	d410, d450	
<b>d430</b>	Lifting and carrying objects	0.85	1.02	0.83	X	d440, d445	
<b>d440</b>	Fine hand use	0.81	1.03	1.31	X	d430, d445	Age
<b>d445</b>	Hand and arm use	0.87	1.07	1.31	X	d430, d440	Age
<b>d450</b>	Walking	0.64	0.71	0.47	X	d410, d415, d510, d540	Age
<b>d510</b>	Washing oneself	0.72	0.78	0.74	X	d450, d540, d550	Age
<b>d540</b>	Dressing	0.72	0.78	0.81	X	d450, d510, d550	Age
<b>d550</b>	Eating	1.19	1	1.39	X	d510, d540	Age
<b>d598</b>	Self-care (other specified)	1.19	1.18	1	X		
<b>d720</b>	Complex interpersonal interactions	1.02	1.32	1.47	X	d155, d177, d399	Age

1 Infit and Outfit expected below 1.2 for the absence of underfit

3 Local item dependency (LID) significant if LID > mean residual correlation + 0.2

4 Differential item functioning (DIF)

In summary, the selection of the ICF categories from the self-assessment form fails to achieve the essential statistical properties required to measure functioning. As it stands, therefore, the self-assessment form cannot be used to create a summary score of functioning. There may, nonetheless, be reasons to keep the self-assessment form, as we will discuss in the next section.

## 9. Options for including functioning in disability assessment

This section presents options for including functioning into disability assessment in Latvia in a systematic, formalized, and transparent mode. For the source of functioning information, we use WHODAS because its good psychometric characteristics observed in other studies were confirmed in the Latvian pilot.

The challenge is how best to combine medical and functioning information. Although functioning information is directly relevant to disability, purely medical information is also important in order to support a valid and fair assessment of disability in individuals applying for benefits. Medical information, in the ICF terms, is information about the “intrinsic” capacity of the body and mind. In many instances, the biomedical problems people have can make all the difference to what they experience in their lives. A person in chronic pain, a person missing a limb, or experiencing severe depression is experiencing disability, and it may not matter much what his or her environment is. The body makes a difference in disability, and ignoring the body, or downplaying the importance of the

body distorts the concept of disability. Moreover, medical information gives us a longitudinal perspective: we know what to expect as the disease progresses, what complications or secondary conditions may arise in the future. This too is relevant to the overall determination of the degree of disability that a person experiences.

Unlike other countries, Latvia has chosen to define disability in terms of disability status groups rather than in terms of a percentage of disability that an individual experiences. The disability status groups are ordinal in nature and represent four levels of severity: no disability, moderate (group III), severe (group II) and very severe (group I). Somewhat arbitrarily, these ordinal groups are made artificially linear by assigning each group to a range, or band, of percentage: <24.99% - no disability; 25-59.99% - moderate; 60-79.99% severe and 80-100% very severe).

We have no information where these percentage numbers came from and whether they are based on some scientific evidence. We assume that they were decided on based on expert deliberations. As they are part of the current system, however, we need to take them into account as we develop and propose options for moving forward with our recommendations.

That being said, our first task is to try to define bands using WHODAS Rasch transformed data that can parallel the disability status group percentage bands. Only by doing so is it possible to suggest ways in which the medically based disability status group procedure can be integrated with the WHODAS disability linear metric.

### 9.1 Proposals for WHODAS disability severity bands

We are not aware of any previous attempt to create ranges or bands of percentages of disability severity using WHODAS data. There are suggestions, however, of cut-off points in terms of which these bands could be constructed. There are studies that report the 90<sup>th</sup> even 95<sup>th</sup> percentile of WHODAS-scores as the best cut-off to diagnose severe disability in some specific groups, such as post-partum women (Mayrink et al. 2018) or elderly (Ferrer et al. 2019). Minimally clinically important difference score for the WHODAS have not been established yet (Federici et al. 2017). A score of 40, after rescaling, aligns with the results of a large survey conducted on Australian households using WHODAS (Andrews et al. 2009). In addition, Yen et al. (2017) have shown that data from WHODAS-scores in the Taiwanese population of applicants for disability benefits obtained scores around this same cut-off (median at 40.57). The WHODAS Manual provides norm values for Item Response Theory (IRT) derived WHODAS-scores ranging from 0-100 aggregated across general populations from 10 countries from a WHO multi-country survey conducted in the early 2000s. Finally, the World Report on Disability suggests that 15% of the population have some form of disability. Based on the norm values for the WHODAS, the 85% (100% - 15%) percentile of the WHODAS score range are found for scores > 25. A WHODAS score of 25 can represent a cut-point to discriminate no from mild levels of disability.

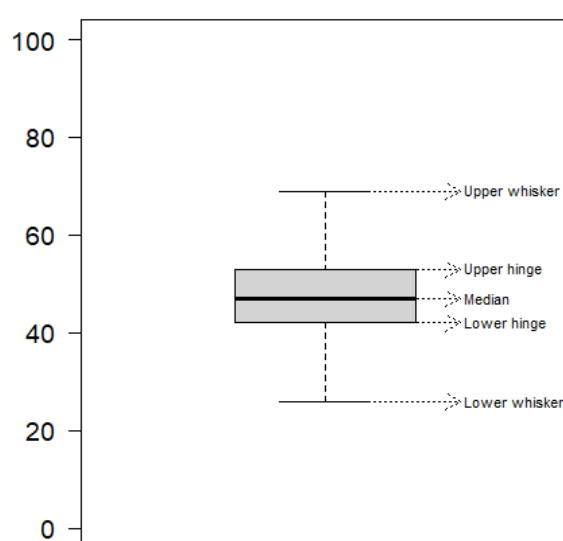
Fortunately, the World Bank also has experience in implementing WHODAS pilots in Latvia, Lithuania, and Greece and these data are potentially useful for constructing Rasch-based WHODAS score distributions from disability assessment applicants from these pilot studies. Table 18 shows the WHODAS score distribution in the three countries and suggests cut points that could be used to describe the level of disability that an individual experiences. These cut points aggregate information from the literature, from WHODAS officially published norm-values and information on WHODAS score distributions collected in Latvia, Lithuania, and Greece. The suggested cut-offs were not computed based on a sound statistical methodology and will need revision, once the WHODAS is implemented and more data points are collected over time on the continuum from low to high levels of disability.

Table 18: Rasch-based 0-100 WHODAS-score ranges  
in Latvia, Lithuania, and Greece – suggested cut-points

WHODAS-score range	Latvia Rasch-based 0-100 score	Lithuania Rasch-based 0-100 score	Greece Rasch-based 0-100 score	Proposed cut points for the Rasch-based 0-100 WHODAS-score
< (Mean – 1SD)	< 38	< 46.6	< 29.5	Cut-off chosen based on 15% of disability from the WRD and looking at the cut-point in norms from the WHODAS manual. This approach delivers a cut-point that departs from the distribution of WHODAS scores at the lower end found in the determined groups of disability among the pilot participants. 0 to 25 [label: No disability]
(Mean – 1SD) to Mean	38 to 47.4	46.6 to 55.1	29.5 to 46.6	26 to 45 [label: Moderate] Note that this group includes “mild difficulty - 1” answers to WHODAS questions, hence, this category actually should be “mild to moderate”.
Mean to (Mean + 1SD)	47.4 to 56.6	55.1 to 63.6	46.6 to 63.8	46 to 60 [label: Severe]
(Mean + 1SD) to 100	> 56.6	> 63.6	> 63.8	61 to 100 [label: Very severe]

The distribution of the WHODAS scores is shown in boxplots. The hinges display a confidence interval around the median (50th percentile) and represent the 25th and 75th percentile, or first and third quartile (Q1 and Q3). The whiskers represent the reasonable extremes of the data, i.e., the minimum and maximum values that do not exceed a certain distance from the median, here  $Q3 + 1.5 * IQR / \sqrt{n}$  or  $Q1 - 1.5 * IQR / \sqrt{n}$ . IQR represents the interquartile range. Values beyond the whiskers are outliers, i.e., the most extreme upper and lower scores of a distribution. Figure 10 shows again the terminology used in this report to visualize the position of the scores. In some figures an additional segment is added to indicate the 90th percentile.

Figure 10: Boxplot - Terminology



## 9.2 Distribution of WHODAS scores using the proposed Rasch based cut offs

Below, Tables 19-21 present distribution of WHODAS Rasch scores using the proposed cut offs from Table 18 (the last column) above. Table 19 presents socio-demographic characteristics of the population stratified by disability levels based on the WHODAS cut-points. First a very low percentage of the population would have no disability (1.18%); this is much less than the number of persons having no disability status (11.2% in Table 14). this may be explained by the fact that patients are referred to SMC by medical doctors who are expected to send to SMC only cases with health state that significantly impacts “intrinsic body capacity”. The percentage of very severe disability is 8.6%, significantly lower than what the current system determines as having a very severe disability (18.65%). The major shift is thus from very severe to severe and no disability to mild to moderate group. Consequently, the mild to moderate and particularly severe are larger than in the current medical assessment of disability (41.5% vs 38.45% and 48.8% vs. 31.7%). There is an effect of gender, age, marital status, living situation and the working situation characteristics on the observed values of WHODAS-based functioning level. In comparison, for example, gender was not significant when stratifying by the SMC determined disability status groups.

Table 19: WHODAS-based disability severity – descriptive statistics

	None	Moderate	Severe	Very Severe	p-value
<b>N (%)</b>	<b>26 (1.2)</b>	<b>913 (41.5)</b>	<b>1074 (48.8)</b>	<b>189 (8.6)</b>	
Gender = Male (%)	17 (65.4)	420 (46.0)	400 (37.2)	79 (41.8)	<0.001
Age – mean (SD)	46.35 (17.74)	55.59 (12.51)	58.41 (12.32)	74.20 (13.36)	<0.001
Years of Education – mean (SD)	13.08 (3.19)	12.84 (2.91)	12.52 (3.03)	10.93 (3.15)	<0.001
Marital Status (%)					<0.001
Never married	6 (23.1)	82 (9.0)	108 (10.1)	11 (5.8)	
Currently married	15 (57.7)	434 (47.5)	476 (44.3)	78 (41.3)	
Separated	0 (0.0)	45 (4.9)	69 (6.4)	7 (3.7)	
Divorced	0 (0.0)	103 (11.3)	154 (14.3)	27 (14.3)	
Widowed	0 (0.0)	89 (9.7)	136 (12.7)	63 (33.3)	
Cohabiting	5 (19.2)	160 (17.5)	131 (12.2)	3 (1.6)	
Living Condition (%)					<0.001
Independent in the community	25 (96.2)	880 (96.4)	874 (81.4)	6 (3.2)	
Assisted living	1 (3.8)	31 (3.4)	198 (18.4)	183 (96.8)	
Hospitalized	0 (0.0)	2 (0.2)	2 (0.2)	0 (0.0)	
Work Status (%)					<0.001
Paid work	11 (42.3)	384 (42.1)	292 (27.2)	0 (0.0)	
Self-employed	0 (0.0)	27 (3.0)	29 (2.7)	0 (0.0)	
Non-paid work	0 (0.0)	2 (0.2)	1 (0.1)	0 (0.0)	
Student	3 (11.5)	11 (1.2)	2 (0.2)	0 (0.0)	
Keeping house	0 (0.0)	5 (0.5)	11 (1.0)	0 (0.0)	
Retired	6 (23.1)	156 (17.1)	262 (24.4)	131 (69.3)	
Unemployed (health reasons)	2 (7.7)	145 (15.9)	309 (28.8)	18 (9.5)	
Unemployed (other reasons)	2 (7.7)	34 (3.7)	22 (2.0)	3 (1.6)	
Other	2 (7.7)	149 (16.3)	146 (13.6)	37 (19.6)	

Table 20 presents the pilot sample groups by WHODAS scores and the SMC determined disability status groups.

Table 20: WHODAS-based disability level versus medical disability status

		<b>Disability Status Groups</b>			
		No	Group III	Group II	Group I
<b>Functioning levels based on WHODAS</b>	None	4	15	4	0
	Moderate	116	452	278	54
	Severe	121	363	390	188
	Very Severe	1	1	13	161

The next table (Table 21) Frequency and percentage of ICD chapters for the pilot sample and by WHODAS-based functioning level as well as the mean and standard deviation (SD) of the corresponding WHODAS-scores.

Table 21: Frequency and percentage of ICD chapters for the pilot sample and by WHODAS-based functioning level as well as the mean and standard deviation (SD) of the corresponding WHODAS-scores

	Pilot N (%)	Pilot Mean (SD)	None N (%)	None Mean (SD)	Moderate N (%)	Moderate Mean (SD)	Severe N (%)	Severe Mean (SD)	Very Severe N (%)	Very Severe Mean (SD)
I Certain infectious and parasitic diseases	28(1.3%)	47.3(9.3)			11(1.2%)	37.6(4.8)	16(1.5%)	52.8(4.1)	1(0.6%)	66
II Neoplasms	389(18%)	46.3(9.5)	5(21.7%)	16.6(9.4)	178(19.8%)	39.1(4.6)	181(17%)	51.8(4.2)	25(14.2%)	64.2(4.2)
III Diseases of the blood	3(0.1%)	45.3(16.4)			2(0.2%)	36(4.2)			1(0.6%)	64
IV Endocrine nutritional and metabolic diseases	47(2.2%)	46.3(8.5)			21(2.3%)	38.4(4.7)	25(2.4%)	52.4(4.3)	1(0.6%)	61
IX Diseases of the circulatory system	372(17.2%)	52.2(11)	3(13%)	22.7(2.5)	98(10.9%)	39.3(4.3)	178(16.8%)	52.5(4.4)	93(52.8%)	66.4(4.6)
V Mental and behavioral disorders	44(2%)	51.6(9.4)			11(1.2%)	39.7(4.3)	24(2.3%)	52.1(4)	9(5.1%)	64.8(3.8)
VI Diseases of the nervous system	124(5.7%)	49.6(9.8)	2(8.7%)	20.5(6.4)	41(4.6%)	40.4(4.2)	63(5.9%)	52.1(4.3)	18(10.2%)	64.9(4.2)
VII Diseases of the eye and adnexa	158(7.3%)	45.4(9.1)	5(21.7%)	20.2(5.7)	75(8.3%)	40(4.6)	71(6.7%)	51.2(3.9)	7(4%)	63.7(2.6)
VIII Disease of the ear and mastoid process	5(0.2%)	37.8(9.3)			4(0.4%)	35(7.9)	1(0.1%)	49		
X Diseases of the respiratory system	23(1.1%)	48.9(5.6)			5(0.6%)	40.8(2.6)	18(1.7%)	51.1(3.8)		
XI Diseases of the digestive system	27(1.2%)	46.7(9.5)			13(1.4%)	38.8(5)	12(1.1%)	52.5(5)	2(1.1%)	63.5(0.7)
XII Diseases of the skin and the subcutaneous tissue	11(0.5%)	46.5(5.9)			4(0.4%)	40.5(3.7)	7(0.7%)	49.9(3.8)		
XIII Diseases of the musculoskeletal system and connective tissue	497(23%)	46.6(6.7)	3(13%)	21(1.7)	215(23.9%)	40.9(3.3)	270(25.4%)	50.9(3.7)	9(5.1%)	63.3(2.4)
XIV Diseases of the genitourinary system	19(0.9%)	46.5(9.3)			6(0.7%)	35(4.8)	13(1.2%)	51.8(4.9)		
XIX Injury poisoning and certain other consequences of external causes	385(17.8%)	44.9(7.6)	3(13%)	24.3(1.2)	200(22.2%)	39.4(4.4)	173(16.3%)	50.6(3.4)	9(5.1%)	64.6(3.5)
XVII Congenital malformations deformations and chromosomal abnormalities	16(0.7%)	41.2(11.3)	2(8.7%)	25(0)	8(0.9%)	36.5(4.5)	5(0.5%)	50.2(4.1)	1(0.6%)	66
XVIII Symptoms signs and abnormal clinical and laboratory findings	1(0%)	33			1(0.1%)	33(NA)				
XXI Factors influencing health status and contact with health services	12(0.6%)	44.8(5.7)			7(0.8%)	40.7(2.3)	5(0.5%)	50.4(3.8)		

As indicated in Table 18, WHODAS scores < 25 indicate no disability, scores 26-45 indicate (mild to) moderate disability, scores of 46-60 indicate severe disability, and scores of 61 to 100 very severe disabilities. Data in Table 21 indicate some variability within the functioning boundaries, i.e., some conditions are closer to the lower boundary while some are closer to the upper boundary of the groups. For example, for *moderate functioning problems*, an average score of 35 (SD = 4.8) is observed for applicants with XIV *Diseases of the genitourinary system* (N = 6) and an average of 40.9 (SD = 3.3) with XIII *Diseases of the musculoskeletal system* (N = 215). How significant this difference is in terms of disability is hard to say now, indicating that as the data is collected over time, it would need to be analyzed and bands of groups adjusted.

### 9.3 Options for changing the current Latvian disability assessment system

Below we present three options for changing the current Latvia disability assessment system, which in effect are modes for integrating functioning information. All options require that WHODAS be used in the assessment process at some point; but the difference is how the information from WHODAS is used, and when. All options also keep the disability status groups as they are now, there is no reason to change this approach to disability determination; it serves its purposes.

#### 1. Flagging mechanism

The least disruptive change to the current system would be to keep the medical assessment and disability status groupings as is, including the self-assessment form but to use the WHODAS Rasch score derived bands in the determination decision. A systematic procedure can be devised so that any individual whose WHODAS score places them in a band that is different from the disability status group medically assessed is 'flagged'. Whether WHODAS rates the disability percentage level higher or lower than the range for the status group assigned, that individual's case needs to be reconsidered so that the divergence is explained. The explanation might point to the nature of the health condition the person has – e.g., a condition that will inevitably worsen may need to be in a higher status group than WHODAS will indicate, or WHODAS may be indicating more functioning problems than are typically experienced by a person with that health condition.

With this strategy, the WHODAS scores are used as indicators of the extent of problems in functioning, although the cut-offs that create the percentage bands may need to be refined based on a larger set of cases and more insight into the relationship between the population's lived experience and the reported scores.

For example, referring back to the four cases, A, B, C, and D described above, under this approach, the WHODAS score of the A person indicates that she has severe disability, meaning a lot of difficulties in her day-to-day life; on the other hand, persons' C and D scores indicate a mild disability (Figure 11). The determined disability status groups of these persons by SMC do not correspond to the functioning measured by the WHODAS. These four individuals are only a selection. Interestingly, this approach seems to highlight more individuals with extremely low WHODAS-scores, i.e., < 25 (no disability) with moderate and severe determined disability status where the status could be reconsidered.

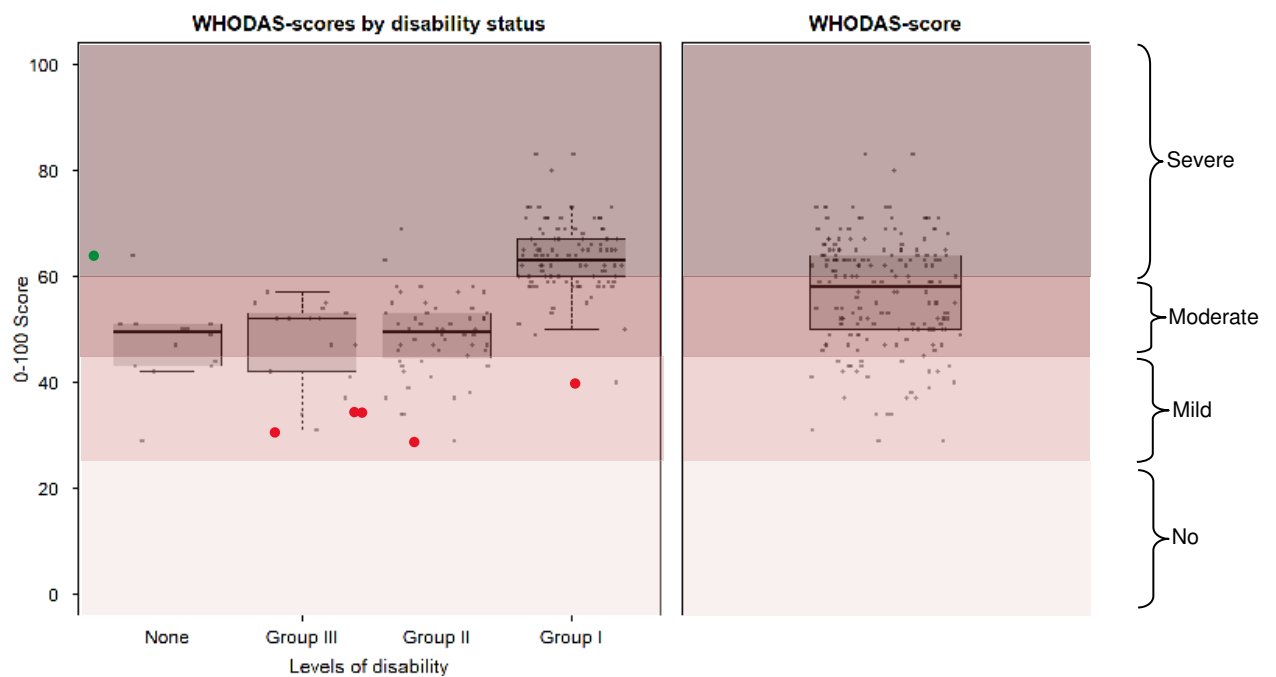
One individual has a score even higher than the example case A in the no disability determined group, but otherwise severe disability based on WHODAS seems only to occur in the severe and especially very severe disability status groups. Figure 12 provides the same illustration but only with the data points and score distributions of the persons with a cerebrovascular disease (I6). The figure indicates again a few contradicting data points that have been highlighted in red, when the score is unexpectedly low or green when it is too high for the given disability status group. Figure 12 further

shows that the distribution of the WHODAS-scores for the individuals with a cerebrovascular disease and very severe disability status are mostly found in the range of severe disability (scores 60-100).

Figure 11: WHODAS-score distribution by disability status using WHODAS cut points for the Rasch-based scores (0-100)



Figure 12: WHODAS-score distribution by disability status for ICD-codes I6 – Cerebrovascular diseases using WHODAS cut points for the Rasch-based scores (0-100)





## *II. Priority to WHODAS summary scores*

A more radical suggestion is, in effect, reverse the sequencing of disability assessment in Latvia by using the WHODAS summary score, and the percentage bands described above, to determine for each applicant the disability status group to which they belong – from No Disability to Very Severe Disability. The argument in favor of doing so has been made several times in this document: WHODAS is a psychometrically powerful and scientifically robust questionnaire that has been explicitly designed to capture precisely what disability assessment is about: namely, the overall, 'whole person' level of functioning problems that people actually experience in their daily lives, taking full account of all environmental barriers and facilitators. We have repeatedly described this as ICF's perspective of performance arguing that disability assessment is a matter of validly assessing the actual lived experience of people with health conditions. This is exactly what WHODAS does.

However, this is not to say, nor is this option defining a process that ignores the essential medical information that we have repeatedly said is essential to disability assessment. What this option suggests is that WHODAS summary score, based on Rasch-derived metric scale, can provide the first estimate of which disability status group the applicant appropriately belongs to. Medical information can then be used to adjust or refine this first estimate to reflect the nature of the applicant's underlying health condition.

A good example of how this might work is the case of cancer, a potential very serious and often incurable and progressive disease which, in its early stages, and often years after first onset, is not associated with high level of disability. The WHODAS score for an individual recently diagnosed with lung cancer or some other serious form of cancer, would be categorized as Group III or moderate, or possibly even No disability, because the disease has at that point did not disrupt the kinds of activities that are measured in the WHODAS. However, as a medical record would indicate, progressively the cancer will have worsening effects on functioning, and this fact is important information that needs to be taken into account in the disability assessment process.

## *III. Comprehensive disability assessment (our recommended option)*

This last option allows for incrementally introduced reforms without being disruptive to existing procedures and practices, but nonetheless constitutes an important revision that brings functioning information into the disability assessment system in Latvia. ***It is our recommended option***, so we describe it in more detail. A comprehensive disability assessment depends on three sources of information:

1. *Functioning information* presented as a summary score for a 'whole person' level of functioning. The WHODAS pilot has confirmed, and we have reported above, that the WHODAS 36-question version is both feasible to introduce into the Latvian system but has the desired psychometric properties of validity and reliability.
2. *Health status information* coming from the health sector. We suggest that the collection and quality of medical information can be improved. The medical referral form should be revised to explicitly require the primary health condition, as well as secondary conditions and other co-morbidities to be listed (and their ICD codes), and description of diagnostic test results, the state of health and proposed therapies. This information should be typed and electronically submitted to the SMC (this will likely require the assistance of a nurse or other assistant to transfer the physicians' handwritten notes to more legible format).
3. *Information about the applicant's environment* – family, home, school, workplace, community. The most direct way of getting this information is from the self-assessment form, which can be revised to include questions on personal and demographic data, household composition, living arrangements, housing situation, education and employment. This should be the primary function

of this form: the functioning questions are not necessary because they are already included in WHODAS. If possible, the form could also include questions about sick leave and the benefits and services the person is already receiving, including assistive technology. The form might also have questions on what benefits and services the person believes they need, as well as their wishes and plans. In short, we propose that a well-structured needs assessment section is included in the self-assessment form as a way of collecting information that the SMC could use to assess needs and propose existing benefits and services to improve disability experience.

As already explained, a significant challenge in including functioning into disability assessment in Latvia is the fact that the current system does not assign individual percentages of disability but uses an ordinal scale (with underlying percentage bands of “difficulties in functioning”). The disability assessment method proposed here avoids this issue by bringing together three sources of information for a compressive and individualized assessment that is grounded in functioning information. The assessment procedure should be the same for all applicants irrespective of whether the health condition or impairment was caused by age, accident or occupational disease or work accident.

Procedurally, we suggest the following process (many existing steps in the current disability assessment procedure would remain the same, we are only listing the changes):

1. A person submits an application along with the revised medical referral and revised self-assessment form. Application, medical referral and self-assessment should all be in electronic format and part of the applicant’s electronic file. When the file is composed, a cross-check/verification of data should be run, and inconsistencies or missing information flagged.
2. The appointment interview is scheduled electronically, and the person is informed.
3. A SMC employee prepares the file for the face-to-face interview. The assessors should not have any connection to the applicant (if there is even a remote connection, she/he should be recused from the assessment).
4. Prior to the interview, a trained assessor not participating in the face-to-face meeting, administers WHODAS. The answers should be immediately marked in an electronic file, so that an automated algorithm can generate the WHODAS Rasch score immediately.
5. During the interview – recommended is presence of two assessors, an administrative assistant and the applicant and possibly one more person close to the applicant, the assessors evaluate disability experience of the applicant. They should use medical information, self-assessment information (revised content) and information from WHODAS (but without the WHODAS Rasch score). The referring physician might be present as well and present the medical case for disability. The assessors should be trained in interview techniques and a guidance on what and how to ask should be prepared. The interview should be recorded but only with the consent of the applicant.
6. Based on the interview and the documents, the assessors prepare evaluation and propose the group of disability with a comprehensive justification using the assessment guidelines. The proposal should also include a section on proposed benefits and support measures. The evaluation form is sent to the supervisor.
7. The supervisor reviews the evaluation and compares the proposed disability group with the WHODAS Rasch score. If the two overlap, the evaluation is completed, and a certificate is issued along with the proposed interventions. The person may be automatically referred to the benefits and service administrators without the need to apply for them separately. If the proposed disability group is different from the WHODAS Rasch score-based group, the supervisor should appoint a different assessor to review all documents about the case and make her/his proposal to the supervisor. Should it be needed, this assessor may talk to the applicant, her/his physician or any other person who may provide additional information. The case should then be discussed in the case meeting chaired by the supervisor and in the presence of all assessors. Optimally, the decision on the disability group would be reached unanimously. If the consensus is not possible, the chair decides.

This proposal moves disability assessment system toward holistic, comprehensive assessment of disability. The systematic and transparent inclusion of functioning will not require dramatic changes in the organization of disability assessment. The proposed changes in the self-assessment form will enhance the assessment of needs, which will make decisions about benefits and interventions to accompany the certification easier for all parties concerned.

The new system will require adjustments in the information management system, and the assessment instructions and guidelines. It is very important to establish a statistical and analytical unit at the Ministry or SMC to analyze WHODAS and other disability related data to (i) fine tune the WHODAS Rasch cut-offs, (ii) analyze disability trends; (iii) conduct analytical and statistical research needed for the development of evidence-based disability policies, and system, including disability assessment. An alternative would be for the Ministry and SMC to establish a formalized collaboration with one of the premier universities in Latvia.

## References

- Aleksandra Posarac, Elina Celmina and Jerome Bickenbach. 2020. Disability Policy and Disability Assessment System in Latvia © World Bank.
- Bond, Trevor G., and Christine M. Fox. 2001. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: L. Erlbaum.
- Boone, W. J. 2016. "Rasch Analysis for Instrument Development: Why, When, and How?" *CBE Life Sci Educ* 15 (4). <https://doi.org/10.1187/cbe.16-04-0148>.
- Christensen, Karl Bang, Guido Makransky, and Mike Horton. 2017. "Critical Values for Yen's Q3: Identification of Local Dependence in the Rasch Model Using Residual Correlations." *Applied Psychological Measurement* 41 (3): 178–94. <https://doi.org/doi:10.1177/0146621616677520>.
- Federici, Stefano, Marco Bracalenti, Fabio Meloni, and Juan V. Luciano. 2017. "World Health Organization Disability Assessment Schedule 2.0: An International Systematic Review." *Disability and Rehabilitation* 39 (23): 2347–80. <https://doi.org/10.1080/09638288.2016.1223177>.
- Fellinghauer, C. S., B. Proding, and A. Tennant. 2018. "The Impact of Missing Values and Single Imputation Upon Rasch Analysis Outcomes: A Simulation Study." *J Appl Meas* 19 (1): 1–25.
- Ferrer, Michele Lacerda Pereira, Monica Rodrigues Perracini, Flávio Rebutini, and Cassia Maria Buchalla. 2019. "WHODAS 2.0-BO: Normative Data for the Assessment of Disability in Older Adults." *Revista de Saude Publica* 53: 19. <https://doi.org/10.11606/S1518-8787.2019053000586>.
- Holland, P. W., and H. Wainger. 1993. *Differential Item Functioning*. Edited by N. J. Hillsdale. Erlbaum.
- Mair, Patrick, Reinhold Hatzinger, and Marco Johannes Maier. 2019. *eRm: Extended Rasch Modeling*.
- Maraes, Ida. 2013. "Local Dependence." In *Rasch Models in Health*, 111–30. John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781118574454.ch7>.
- Masters, Geoff N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47 (June): 149–74.
- Mayrink, Jussara, Renato T. Souza, Carla Silveira, José P. Guida, Maria L. Costa, Mary A. Parpinelli, Rodolfo C. Pacagnella, et al. 2018. "Reference Ranges of the WHO Disability Assessment Schedule (WHODAS 2.0) Score and Diagnostic Validity of Its 12-Item Version in Identifying Altered Functioning in Healthy Postpartum Women." *International Journal of Gynecology & Obstetrics* 141 (S1): 48–54. <https://doi.org/10.1002/ijgo.12466>.
- Nunnally, Jum C., and Ira H. Bernstein. 1994. *Psychometric Theory*. 3rd ed. New York; London: McGraw-Hill.
- Rasch, G. 1960. *Probabilistic Models for Some Intelligence and Attainment Tests*. Copenhagen: [s.n.].
- Smith, E. V. 2002. "Detecting and Evaluating the Impact of Multidimensionality Using Item Fit Statistics and Principal Component Analysis of Residuals." *J Appl Meas* 3 (2): 205–31.
- Smith, R. M., and C. Y. Miao. 1994. "Assessing Unidimensionality for Rasch Measurement." In *Objective Measurement: Theory into Practice. Volume 2*. Greenwich: Ablex: M. Wilson.

- Smith, R. M., R. E. Schumacker, and M. J. Bush. 1998. "Using Item Mean Squares to Evaluate Fit to the Rasch Model." *J Outcome Meas* 2 (1): 66–78.
- Stekhoven, D. J., and P. Buhlmann. 2012. "MissForest–Non-Parametric Missing Value Imputation for Mixed-Type Data." *Bioinformatics* 28 (1): 112–18.  
<https://doi.org/10.1093/bioinformatics/btr597>.
- Team, R Core. 2016. "R: A Language and Environment for Statistical Computing."
- Tennant, A., and P. G. Conaghan. 2007. "The Rasch Measurement Model in Rheumatology: What Is It and Why Use It? When Should It Be Applied, and What Should One Look for in a Rasch Paper?" *Arthritis Rheum* 57 (8): 1358–62. <https://doi.org/10.1002/art.23108>.
- Ustun, T. B. 2009. *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule (Whodas 2.0)*. Geneva: World Health Organization.
- Yen, Wendy M. 1984. "Effects of Local Item Dependence on the Fit and Equating Performance of the Three-Parameter Logistic Model." *Applied Psychological Measurement* 8 (2): 125–45.  
<https://doi.org/10.1177/014662168400800201>.

## Appendix 1

Appendix 1: Person item map after collapsing of the response options into three categories

