



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



WB Report No: AUS0002502

Options for Including Functioning into Disability and Work Capacity Assessment **In Lithuania**

Aleksandra Posarac, Carolina Fellinghauer, Jerome Bickenbach
Brussels, Washington, DC and Vilnius

© 2017 The World Bank
1818 H Street NW, Washington DC 20433
Telephone: 202-473-1000; Internet: 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: “Aleksandra Posarac, Carolina Fellinghauer, Jerome Bickenbach. 2021. Disability Assessment in Lithuania: Options for Including Functioning into Disability and Work Capacity Assessment © 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

Report No: AUS0002502

Options for Including Functioning into Disability and Work Capacity Assessment in Lithuania

Aleksandra Posarac, Carolina Fellinghauer, Jerome Bickenbach

SPL



Table of Contents

Acknowledgements.....	1
EXECUTIVE SUMMARY	2
INTRODUCTION.....	8
PART ONE: THE INSTRUMENTS.....	9
WHODAS 2.0: Technical Details	9
DWCAO’s Activity and Ability Questionnaire: Technical Details.....	11
Comparison between A&AQ and WHODAS.....	13
ICF content comparison	13
A&AQ coefficients and the impact of functioning assessed by A&AQ on final work capacity scores	16
Conclusions about the Activity and Ability Questionnaire based on the above analysis	18
PART TWO: THE PILOT	19
Descriptive Statistics of the Pilot Sample	19
Analysis Methodology.....	20
Psychometric Analysis: Rationale and Tests	20
Results.....	23
Metric properties of WHODAS.....	23
Summary: the psychometric properties of WHODAS.....	31
Metric properties of the Activity and Ability Questionnaire	32
The suitability of A&AQ as an instrument for disability assessment.....	37
PART THREE: OPTIONS FOR INCLUDING FUNCTIONING	38
Introduction: Approaches and strategies for using WHODAS scores.....	38
Assessment options for using WHODAS to include functioning into disability determination process.....	41
Option A: Discretionary combination of medical and functioning components	41
Options B, C and D	42
The underlying problem with medical assessment and options in this Report.....	42
Option B: Using an averaging algorithm	43
Option C: Using the flagging algorithm.....	44
Option D: Using the augmenting algorithm.....	45
Examples of the inclusion strategies in practice	46
Integration strategies - Examples of individual cases.....	47
Graphical representation of the overall impact of the averaging strategy	47
CONCLUSION AND RECOMMENDATIONS.....	54
Instruments to assess functioning	54
Conclusions about the WHODAS as a functioning assessment instrument	55
Recommendations	55
Looking Ahead.....	57
APPENDICES	60
Appendix 1: Lithuania Disability and work capacity assessment and disability needs assessment .	60
Appendix 2: DWCAO’s Activity and Ability to Participation Questionnaire.....	63
Appendix 3	79
Systematic Overview to the adjustment strategies of the WHODAS items applied at mid-term	79
Appendix 4	80
References	81

List of Figures

Figure 1: Relationship between the basic work capacity (medical) and the A&AQ scores	17
Figure 2: Person item map for the WHODAS items before collapsing the response options	26
Figure 3: Local Item Dependencies before the creation of testlets.....	27
Figure 4: Score frequency distribution of WHODAS and the A&AQ scores.....	33
Figure 5: Person item map of the Activity and Ability Questionnaire	34
Figure 6: Local Item Dependencies.....	35
Figure 7: Strategy #2: (Basic work capacity 100% and WHODAS 0%).....	48
Figure 8: STRATEGY #3 (Basic work capacity 75% and WHODAS 25%) with WHODAS cut-off at the median score.....	48
Figure 9: STRATEGY #4: (Basic working capacity 50% and WHODAS 50%) with WHODAS cut-off at median score.....	49
Figure 10: STRATEGY #5: (Basic working capacity 25% and WHODAS 75%) with WHODAS cut-off at median score.....	50
Figure 11: STRATEGY #6: (Basic working capacity 0% and WHODAS 100%) with WHODAS cut-off at median score.....	50
Figure 12: STRATEGY #7 (Basic working capacity 75% and WHODAS 25%)with WHODAS cut-off at 40 points score.....	51
Figure 13: STRATEGY #8 (Basic working capacity 50% and WHODAS 50%) with WHODAS cut-off at 40 pts score.....	52
Figure 14: STRATEGY #9 (Basic working capacity 25% and WHODAS 75%) with WHODAS cut-off at 40 pts score.....	52
Figure 15: STRATEGY #10 (Basic working capacity 0% and WHODAS 100%) with WHODAS cut-off at 40 pts score).....	53

List of Tables

Table 1: 36-item WHODAS 2.0, by domain	10
Table 2: Comparing WHODAS and the A&AQ in terms of ICF categories and domains.....	14
Table 3: Distribution of the Basic Work Capacity, the Work Capacity, Activity and Ability, and WHODAS-based Score	16
Table 4: Description of pilot sample	19
Table 5: Prevalence of Health conditions in the pilot study population by ICD-10 Health Condition Category.....	20
Table 6: Frequencies and Percentages of WHODAS Responses	24
Table 7: Targeting and Reliability of WHODAS items	25
Table 8: WHODAS Item Difficulties, fit, local item dependencies, and differential item functioning at the start.....	28
Table 9: WHODAS Item Difficulties, fit, local item dependencies, and differential item functioning after adjustment.....	29
Table 10: Transformation Table for WHODAS	30
Table 11: Targeting and Reliability of Activity and Ability items	33
Table 12: Item Difficulties, fit, Local item dependencies, and differential item functioning of the Activity and Ability Questionnaire	36
Table 13: Overview of WHODAS inclusion strategies.....	40
Table 14: Work capacity and WHODAS scores and their integration strategies - Examples of individual cases.....	47

Abbreviations

AQ&A	Questionnaire about the Person's Activity and Ability to Participate
DG REFORM	European Commission's Directorate-General for Structural Reform Support
DWCAO	Disability and Work Capacity Assessment Office
DPD	Detailed project description
ICD	International Classification of Diseases
IFC	International Classification of Functioning, Disability and Health (WHO)
MSSL	Ministry of Social Security and Labor
WB	World Bank
WHO	World Health Organization
WHODAS	World Health Organization's Disability Assessment Schedule

Acknowledgements

This Report was written by Aleksandra Posarac, World Bank Lead Economist and the Lithuania Strengthening Disability Assessment Project Manager, Professor Jerome Bickenbach, Swiss Paraplegic Institute and University of Luzern, and Carolina Fellinghauer, University of Zurich, Department of Psychology, Chair for Psychological Methods, Evaluation and Statistics. This Report would not have been possible without collaboration and help from many colleagues and in particular Claudia Piferi and Marc Vothknecht (DG REFORM), Marijana Jasarevic, Social Protection Specialist (WB), Eglė Čaplikienė, Chief Advisor (for People with Disability Issues) (MSSL), and Viktorija Vasiljeva-Gringienė and Jolanta Vyšniauskienė (DWCAO). Marijana Jasarevic and Viktorija Vasiljeva-Gringienė have ably managed the WHODAS pilot. The authors have benefited from comments and advice provided by Alvydas Juocevičius and Genovaite Paliusiene, project advisors. 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 Social Security and Labor, and the Disability and Work Capacity Assessment Office without whose commitment and enormous engagement, this study would not have been possible.

EXECUTIVE SUMMARY

Highlights

This report summarizes the findings from piloting the World Health Organization's Disability Assessment Schedule (WHODAS) in Lithuania. The results from the pilot allow making three important contributions to including functioning into disability/work capacity status assessment in adults in Lithuania:

One: The pilot has assessed the psychometric properties of the *Questionnaire of the Individual's Activity and Ability to Participate (A&AQ)*, which is currently used by the Disability and Work Capacity Assessment Office (DWCAO). The comparison of A&AQ with WHODAS, which is fully based on the WHO's International Classification of Functioning, Disability and Health, shows empirically that WHODAS performs better for disability assessment and should replace the QA&A.

Two: The report proposes an empirically based strategy for including functioning into disability assessment (the so-called averaging), and

Three: This (averaging) strategy gives Lithuania the flexibility to either immediately or gradually (which we would advise) move to 50% and then 75% of functioning weight in the disability / work capacity status assessment for adult population. In this way, functioning would become critically important in the assessment of disability/ work capacity in adults.

Scope of the report

This Report was prepared as part of Output III: Recommendations on the design, implementation and assessment of a pilot at the municipal level of the World Bank (WB) led project "Improving Disability Assessment System in Lithuania" (Project).¹ The Project is implemented in cooperation with the European Commission's Directorate-General for Structural Reform Support (DG REFORM) and provides support to the Ministry of Social Security and Labor (MSSL) of the Republic of Lithuania in enhancing disability assessment. Output III, specifically, proposed a piloting exercise with two primary aims:

- (i) to assess the performance of the World Health Organization's Disability Assessment Schedule (WHODAS 2.0), in its 36-question, interviewer-conducted format;² and
- (ii) to derive recommendations concerning how functioning information and population-based metrics can best be used to augment or refine the current medical determination of disability status, with a view to contribute to the overall outcome of this project: improving the assessment of disability in Lithuania.

During the design phase of the pilot, the World Bank team in collaboration with MSSL, decided that data collected during the pilot through the current disability/work capacity assessment tool – the

¹ From the Detailed Project Description: "This Project aims at supporting the MSSL in enhancing disability assessment, through strengthening of the assessment of functioning and through related improvements in the administrative processes. More precisely, technical support and advice to the MSSL will focus on: (i) a complete situational analysis of the current approaches, including evaluation of the assessment methods and instruments currently used; (ii) recommendations for the improvements in business processes, including IT systems; (iii) the design, implementation and assessment of a pilot to strengthen the assessment of functioning and the inclusion of its results into disability assessment algorithm."

"The expected overall outcome of the project is an improved assessment of disability including functioning. Achievement of the outcome depends to a large extent on the degree of endorsement and implementation of the outputs by the Government of Lithuania and subsequent enforcement, as well as wider policy conditions, which remain outside the responsibility of the European Commission and the World Bank. Such approval and implementation remain the exclusive responsibility of the Government of Lithuania. "

² Ustun et al. 2010. Measuring health and disability: manual for WHO Disability Assessment Schedule (WHODAS 2.0). World Health Organization: Geneva. [https://www.who.int/publications/i/item/measuring-health-and-disability-manual-for-who-disability-assessment-schedule-\(whodas-2.0\)](https://www.who.int/publications/i/item/measuring-health-and-disability-manual-for-who-disability-assessment-schedule-(whodas-2.0))//.

Questionnaire of the Individual's Activity and Ability to Participate (A&AQ) – would be compared to the data collected through WHODAS. In this way, the content and structure of both instruments could be directly compared by an analysis of the pilot data, and thus the performance of these two questionnaires in the context of Lithuanian disability and work capacity assessment of adults could be compared and evaluated. This led to a third aim of the pilot:

- (iii) to compare the results of the current A&AQ questionnaire, which is used to determine weighting coefficients for modifying the medical assessment of disability and work capacity, with the results of the WHODAS tool collected during the pilot.

The World Bank team further proposed at the design stage to also pilot the Clinical Functioning Information Tool (ClinFIT)³⁴ as a potential alternative to medical reports used in disability/work capacity assessment. However, this proposal was declined by MSSL at the time.

This Report summarizes the outcomes of the pilot and presents the resulting policy recommendations. Importantly, this Report does not address potential adjustments to business and administrative procedures beyond the recommendations made by the WB team in the Report on *Disability Policy and Disability Assessment System in Lithuania* (May 2020),⁵ as these would only follow after the political decisions on the changes in the assessment methodology have been made. This Report consists of three main parts.

Part One of the Report present technical information about WHODAS and A&AQ in order to compare the structure and content of the two assessment instruments, and in light of this technical information, to make conclusions about how A&AQ performs within Lithuania's disability and work capacity assessment process.

Part Two provides descriptive statistics from the piloting of WHODAS 2.0 in its 36-question version, clinically (face to face interview) administered format. It presents the analyses of the data collected from both questionnaires during the pilot data and based on this, compares the performance of the two instruments, in terms of the agreed objective of this project, namely, to propose changes to the current disability/work capacity assessment process in Lithuania to more fully incorporate functioning information. This Part concludes with an assessment of the suitability of A&AQ for work capacity and/or disability assessment in the Lithuanian context.

Part Three describes in detail, based on the pilot outcomes, a range of options for using the WHODAS instrument and scoring metrics to integrate functioning into disability and work capacity assessment. Recommendations are presented with respect to how functioning information collected by WHODAS can be integrated into the current medical determination of disability status for a final disability assessment.

Recommendations on the inclusion of functioning into disability and work capacity assessment in Lithuania

The focus of this project is on the disability and work capacity assessment in adults, in line with the Detailed Project Description (DPD). The results of the successful piloting of WHODAS and of the comparisons with the currently used instrument A&AQ provided ample data for an evaluation of the

³ ClinFIT is the International Society for Physical Rehabilitation Medicine's (ISPRM) Universal Functioning Information Tool based on the WHO's ICF. See: www.isprm.org/.

⁵ Posarac, Aleksandra and Bickenbach, Jerome. May 2020. Disability Policy and Disability Assessment System in Lithuania. World Bank.

scientific performance of both instruments. Based on this evidence, three main recommendations are put forward. The first two recommendations are based on the demonstrated scientific soundness of WHODAS as compared to the current functioning assessment instrument (A&AQ) and the Barrême grid percentages currently used to generate the medical assessment of disability. The third recommendation concerns the proposal for a scientifically grounded algorithm for incorporating functioning information into the current medical determination of disability status.

Recommendation 1: Replace the currently used A&AQ with WHODAS-36:

The WHODAS questionnaire, in its 36-item, clinically administered format should replace the currently used Questionnaire of the Individual's Activity and Ability to Participate (A&AQ) for disability/work capacity assessment in adults in Lithuania.

Recommendation 2: Review and update the medical instrument and the Barrême table:

The medical instrument used to determine disability and the basic work capacity score should be reviewed and updated on the basis of the best medical knowledge and experience of other countries, ensuring full alignment with WHO's International Classification of Diseases, ICD-11. This would require a close collaboration with the Ministry of Health.

Alternatively, MSSL in collaboration with the Ministry of Health, may consider piloting ClinFIT, as initially proposed by the World Bank team, with a view using this information to replace medical information and scoring based on the Barrême table.

Recommendation 3: Adopt “averaging” method (Option B) for integrating functioning into disability assessment:

Based on the substantial analysis of the pilot results, several potential approaches to define a scientifically grounded algorithm for incorporating functioning information into the current medical determination of disability status are presented. On the assumption that some form of medical assessment of disability will continue to be used, these scores can be augmented in various ways to incorporate functioning information derived from the application of WHODAS. We investigate three ways of doing so and recommend what we label the 'averaging' option (Option B) that differentially 'weighs' the impact of medical and functioning components.

“Averaging” or Option B is a weighting algorithm that has two endpoints: giving medical assessment 100% weight and WHODAS score 0% and the opposite, giving WHODAS score 100% weight and the medical assessment 0%, with all the intermediate weighting option available as well. This approach gives the government of Lithuania considerable flexibility in – possibly gradually – shaping the reform of disability assessment. The chosen weighting will determine the patterns of successful or unsuccessful disability status – examples of these patterns are presented graphically in the Report.

We recommend that, first, an executive decision is taken on the relative weights of the medical assessment and WHODAS scores. This algorithm should then be used to determine disability status over a period during which the patterns of disability status can be monitored. If the chosen algorithm produces the outcomes desired – specifically an acceptable, and financially feasible, percentage of applicants who are assessed across the three levels of disability status – then that algorithm can be continued. If the outcomes are not acceptable, the algorithm can be adjusted accordingly. We recommend that the weighting starts with 50% and in two-three years moves to 75% functioning-based score and 25% medical based score.

Alternatively, the medical assessment component of disability assessment could also be eliminated. The analysis presented in this Report shows persuasively that it might be possible to use WHODAS

exclusively and still maintain a valid and reliable disability assessment process. However, we know of no country that has taken this option, and for political and historical reasons it might be challenging to do so. It must be said that there are good reasons to continue to use health information in some manner for disability assessment. Nonetheless, if this option is considered politically, Lithuania would be on scientifically sound ground to move towards a complete functioning-based disability assessment procedure. The government of Lithuania should determine which scenario is the most appropriate given political, financial and other relevant considerations.

Going beyond the inclusion of functioning into disability and work capacity assessment: Disability needs assessment and child disability assessment

While the scope of this project is limited to the disability and work capacity assessment in adults, a comprehensive reform of the disability system in Lithuania may also address (i) the needs assessment in adults certified as having disability/ limited work capacity; and (ii) disability and needs assessment in children. These are separate technical areas that should be tackled in subsequent reform steps. The reform of both areas would require extensive technical work on the ground through face-to-face interactions and separate piloting.

Disability status assessment and disability needs assessment: While sometimes confused, disability assessment and needs assessment are different technical and administrative processes with different objectives, and are based on different assessment tools:

- (i) Disability status assessment is a global summary of the 'whole person' level of disability. The summary assessment of disability must be based both on the individual's health state and on specific assessments of specific activities, summarized into a single score. To validly assess the person's level of functioning in multiple domains, the assessment instrument must be based on the ICF model and classification (such as WHODAS).
- (ii) The needs assessment identifies specific disability-associated needs, but do not assess the overall level of disability that the person experiences.⁶ Needs assessments are, by their nature, individualized and focused on specific activities that a person has difficulties performing because of one or more underlying health conditions and/or environmental barriers that are confronted in daily life (for example, sensitivities to air pollution, obstacles to mobility, discrimination in employment). Needs assessments can pinpoint which of the available supports and services the individual can benefit from in order to more fully participate in society – for example maintain employment or live independently. Needs assessments can be conducted using a variety of medical, rehabilitative and social participation clinical instruments and tools.

Importantly, WHODAS is a disability status assessment tool and not a disability needs assessment tool, as it is not granular enough to identify specific needs. However, given its psychometric performance, the information collected through WHODAS can also provide relevant initial input into the proper disability needs assessment process.

Disability assessment in children: Assessing disability in children and related assessment of their needs for support, including special educational needs is sensitive and complex technical and policy area. Such assessments are different from the disability and work capacity assessment in adults, and it is not recommended to use WHODAS for children. For the reform of the disability assessment in children, the entire child disability policy and system, including disability and disability needs

⁶ See Appendix 1 for a more detailed explanation.

assessment should be analyzed and assessed in depth, a new proposal developed and piloted and adjustments made based on collected empirical evidence.

Proposed timeline for reform implementation

Reforming disability system and policies is sensitive and complex process that requires in depth research and piloting of options, which takes planning, time, resources, and persistent effort of policy makers, practitioners and other stakeholders. Broadly speaking, it includes two key components: (i) disability and work capacity assessment as well as disability needs assessments for adults; and (ii) disability policies and disability needs assessments for children. Below, we propose a timeline for the reform and further development of the disability system and policy in Lithuania, in line with the modern understanding of disability and commitments under the United Nation Convention on the Rights of persons with Disabilities to which Lithuania is a state party.

A. Short term (next six months): Implement the reform of the disability and work capacity assessment for adults in Lithuania.

The work under this project has resulted in two major analytical reports:

- (i) *Disability Policy and Disability Assessment System in Lithuania* (May 2020). This report provides an in-depth analysis of the disability system and policies in Lithuania as they pertain to adults. The report offers a range of recommendations related to the Lithuanian disability policy and its administration. This includes, but is not limited to, programs (benefits) to support adults with disabilities, measures to support the labor market inclusion of persons with disabilities, a review of the administration of policies and programs and of disability and work capacity assessment system as implemented by DWCAO, as well as an assessment of DWCAO's management information system and a list of priority actions to improve and bring it up to date.
- (ii) This Report: *Lithuania, options for including functioning into disability and work capacity assessment*, which provides empirically based recommendations on including functioning into disability and work capacity assessment in adults.

Recommendations in both reports are focused on improving efficiency and effectiveness of disability policy and system and further developing it, while improving the quality of services provided to adults with disabilities and their well-being. For most part, recommendations in both reports are non-disruptive and relatively straightforward to implement, without the need for major regulatory framework changes or major budget resources (except for the recommendations related to DWCAO's information system that require investment).

B. Medium-term (2-3 years): reform of (i) the needs assessment for adults with disabilities; and (ii) disability policy and system for children

In the medium term, two other important elements of the overall disability policy and system in Lithuania should be reviewed. This should be based on an assessment of the current systems and the piloting of the proposed new assessment methods to ensure a sound empirical evidence.

- (i) Needs assessment for adults with disabilities:
Disability needs assessment for adults is a different process – with a different aim and using different instruments – than disability status assessment (see Annex I for more details). Optimally, disability needs assessment is conducted as a multidisciplinary administrative process, where rehabilitation professionals (medical, occupational, vocational, etc.) and social workers and, if needed, employers and the employment office work together to assess

the needs of a person with disability and refer her or him to available services with the aim of maximizing her or his functioning and activities and participation. WHODAS, while not a disability needs assessment tool, will provide important initial information on the domains of functioning which need close attention. As described later in this report, the currently used A&AQ instrument has the potential of being used as a disability needs assessment tool, with some adjustments and pilot testing.

In general, the needs assessment process may employ different tools, depending on the situation of the person whose needs are assessed. Many well-tested tools are available; however, whether and how to use them in the Lithuanian context is a matter of a careful analysis, adjustments and test piloting. Designing and testing a new disability needs assessment system will require additional resources both during the reform design phase and for the implementation of a multidisciplinary process, separate from the disability status assessment.

(ii) Disability policy and system for children:

This is a particularly complex, sensitive, and technically and human resources demanding area of disability system and policies. It plays a significant role in determining the course of life of children born with or developing intellectual and physical disabilities, congenital impairments, learning disabilities, and developmental delays. The assessment of disability in children includes the assessment of health conditions, disabilities, as well as the assessment of support needs, including an assessment of special educational needs. It requires a concerted engagement of a range of professionals, from pediatricians to nurses, development experts, social workers, and teachers, to parents and communities.

The further development of disability policy and system for children with disabilities should include the following steps: (i) an in-depth, comprehensive assessment of the current system and policies, including health, education and social protection; (ii) development of tools that need replacement or need to be introduced (example: a new tool for the assessment of special education needs based on ICF⁷) and their piloting; (iii) empirically (pilot) based recommendations. These activities require significant resources, in the same way as the implementation of recommendations is likely to require increased budget allocation to disability policy and system for children.

⁷ A good example of such a tool was developed and is used in Switzerland.

INTRODUCTION

The Minister of Social Security and Labor of the Republic of Lithuania on 3rd of March 2020, issued order No. (20.GE-31) SD-1134 requiring the Disability and Work Capacity Assessment Office (DWCAO) under MSSL to complete approximately 2,000 World Health Organization Disability Assessment Schedule 2.0 (WHODAS 2.0) questionnaires (a 36-question version) by interviewing persons who applied for the assessment of disability or work capacity for the first time. Two trainings with follow-up sessions were organized, and all assessors were provided with methodological information and instructions. A module was added to the IT system to enable pilot responses to be collected online. The pilot implementation commenced on July 1, 2020. The pilot was implemented by 43 assessors with medical education in 6 cities and 16 divisions of DWCAO, integrated into the standard assessment interview after signing a consent form. When due to the COVID lockdown restrictions, face-to-face interviews had to be discontinued; it was agreed in October 2020 to conduct them by phone. A Mid-Pilot review was conducted using information collected from $N = 1,024$ persons in November 2020. The pilot was completed with $N = 2,234$ persons in January 2021.

This Report provides descriptive statistics from the piloting of WHODAS 2.0 in its 36-question version, clinically administered format. The Report also presents the psychometric characteristics of the WHODAS scale and makes recommendations with respect to how functioning information collected by WHODAS can be integrated into the current medical determination of disability status for a final disability assessment. Finally, as mentioned, data collected during the pilot by the A&AQ questionnaire made it possible in this Report to compare the structure, content, and performance of the two questionnaires. (This Report does not address potential adjustments to business and administrative procedures.)

Part One of the Report present technical information about WHODAS and A&AQ in order to compare the structure and content of the two assessment instruments, and in light of this technical information, to make conclusions about how A&AQ performs within Lithuania's work capacity assessment process. *Part Two* presents the analyses of the data collected from both questionnaires during the pilot and based on this, compares the performance of the two instruments, in terms of the agreed objective of this project, namely, to propose changes to the disability assessment process in Lithuania to incorporate functioning information more fully. This Part concludes with conclusions about the suitability of A&AQ for work capacity and or disability assessment in the Lithuanian context. *Part Three* describes in detail, based on the pilot, a range of options for using the WHODAS instrument and scoring metrics to integrate functioning into disability and work capacity assessment.

PART ONE: THE INSTRUMENTS

WHODAS 2.0: Technical Details

In the 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 in light of 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 actually 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 is 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, in light of 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 recommended, generic, performance-based disability assessment tool. It is structured around six basic functioning domains:

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

The clinical version of the WHODAS questionnaire collects information about functioning and problems in functioning – i.e., disability – by means of a face-to-face interview conducted by a trained interviewer who asks standardized questions – and if necessary, follow-up probe questions – and in light of the responses uses WHODAS's 5-level responses scale (None, Mild, Moderate, Severe, 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-report questionnaire; it is rather a questionnaire administered in 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, clinically administered version was chosen for the pilot in order to collect information about a substantial range of functioning domains so as to create a full picture of the disability actually experienced by the respondent in their everyday life. The 36 items are shown in Table 1 by functioning domain.

Table 1: 36-item WHODAS 2.0, by domain

Item In the past 30 days, how much difficulty did you have in:	
Understanding and communicating	
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?
Getting around	
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]?
Self-care	
D3.1	Washing your whole body?
D3.2	Getting dressed?
D3.3	Eating?
D3.4	Staying by yourself for a few days?
Getting along with people	
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?
Life activities	
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?
Participation in society in the past 30 days:	
D6.1	How much of a problem did you have in joining in community activities in the
D6.2	How much of a problem did you have because of barriers or hindrances in the world around
D6.3	How much of a problem did you have living with dignity because of the attitudes and actions
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?

Source: WHODAS

DWCAO's Activity and Ability Questionnaire: Technical Details

In Lithuania, all persons who have been assessed 0-55 percent of working capacity are designated "persons with disabilities" and are guaranteed legislatively determined benefits according to this status. The lower the percent of working capacity scores, the more severe the disability. Work capacity is evaluated in 5 percentage point intervals, ranging from 0 to 100 (where 0 – 25 percent indicates *total incapacity for work*; 30 – 55 percent indicates a *partial capacity for work*; and 60 – 100 percent means a person is *capable of work*). The assessment consists of (i) medical criteria (hereafter **basic work capacity**) that is adjusted by a coefficient created from (ii) person's activity and ability to participate as assessed by the **Questionnaire of the Individual's Activity and Ability to Participate (A&AQ)**.

A&AQ has two parts:

Part I. Professional, work activities, and environmental accessibility – consists of questions on age, professional qualification, work experience, and work skills that the individual may use at the workplace, and adaptation of physical, work, and information environment. This part is scored by a point system in which each response category for each question is given pre-assigned points.

Part II. Activities and ability to participate – consists of 26 questions grouped under five domain headings:

1. **Mobility** (Sit-up, sitting, moving to another position; Standing up and standing; walking; Use of public and private transport; Picking up and moving things; Climbing stairs)
2. **Application of knowledge** (Concentration; Memory; Orientation in the environment and time; Understanding visual information; Understanding auditory information; Writing and counting)
3. **Interaction** (Interaction with strangers; Interaction with relatives and friends; Speaking and/or language perception)
4. **Independence** (Bathing and washing; Putting clothes on and off; Eating; Using the toilet; Taking care of own health)
5. **Daily activities** (Food preparation; Housework)

At the end of each domain, a series of dichotomous (yes/no) questions are asked about the need for assistance relevant to the domain. For example, for Mobility:

- Would technical assistance measures increase the mobility opportunities? YES NO
- Would help by another individual increase the mobility opportunities? YES NO
- Would adaptation of living environment increase the mobility opportunities? YES NO
- Would social rehabilitation services increase the mobility opportunities? YES NO

Part II is scored on the basis of a nominal scale – i.e., each item is described in terms of what the individual can or cannot do relevant to the nature of the item, and these descriptions are scored by 0,1,2,3, and 4. (See complete A&AQ in Appendix 1).

For example:

2.4.3 Eating	The individual eats independently , performs the actions safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	The individual eats independently , performs the actions safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions. Performs all actions more slowly than usually.	The individual eats independently, a minimum or average verbal help by another individual may be required (encouragement, advise) and/or preparation (e.g., put food on a plate, spread butter on bread, pour a drink) and/or a minimum contact help (e.g., to hand a cutlery, to place a piece of food in a spoon or to spear food with a fork, etc.)	When the individual is eating, a greater than average verbal and contact help by another individual is required in performing the action and/or continuous supervision of actions when the individual independently performs the action but does not understand its essence (e.g., may start eating stuff other than food products thereby endangering his/her health)	A continuous help by other individuals is required because the individual does not perform the action independently
Scoring	0	1	2	3	4

Points from Parts I and II are added to scores, and these are mapped onto coefficients:

- a score of 93-101 points: coefficient 0.7
- a score of 84-92 points: coefficient 0.8
- a score of 68-83 points: coefficient 0.9
- a score of 23-67 points: coefficient 1.0
- a score of 10–22 points: coefficient 1.1
- a score of 0-9 points: coefficient 1.2.

These coefficients are automatically applied to the medical assessment score for the final work capacity percentage.

Two initial comments about A&AQ in comparison to WHODAS should be made:

First, A&AQ uses nominal response options (i.e., descriptions of expected levels of behavior) that are then mapped onto an ordinal scale (0-4). Typically, when nominal response options are used, the relationship to ordinal scale is extremely unreliable and controversial since the link is not based on any evidence. Also, within-response multidimensionality cannot be excluded as options present more than one options regarding what to measure in a domain at each level of response. This feature is not important in purely clinical contexts when a patient's progress is being monitored to determine whether interventions are making a difference and the same nominal options are used pre- and post-intervention. But in the case of assessing levels of disability or work capacity, this arbitrariness is highly problematic as there is no empirical justification for these ordinal rankings.

Secondly, the link between the summary scores and coefficient scores is also arbitrary and without empirical basis. Even more problematic is that, given the values of the coefficients, the A&AQ

assessment of functioning has only minimal impact on the resulting assessment of work capacity based on medical criteria alone. This is borne out by the fact that in 2018, only in 1.74% of cases did the A&AQ score change the medical score meaningfully. The consequences of this feature of A&AQ – when compared to WHODAS – are demonstrated statistically below.

Comparison between A&AQ and WHODAS

Before looking more closely at the psychometric differences between A&AQ and WHODAS – based on an analysis of the full pilot dataset – we compare the two in terms of their suitability as instruments to incorporate the element of functioning into disability and work capacity assessment. As in most countries, these assessments are carried out primarily in terms of a medical determination of the applicant's perceived or documented health problems. But in Lithuania, work capacity is assessed in terms of a medical expertise that determines basic work capacity, and that score is modified in terms of coefficient derived from the A&AQ score, producing a final work capacity score. This raises the question of whether and how successfully A&AQ captures the impact of functioning on the final work capacity score?

To answer this question, two issues need to be clarified. The first relates to the functioning content of the A&AQ, i.e., how closely it is aligned with the ICF, as compared to WHODAS. (As WHODAS was expressly developed to be aligned exactly with the ICF, we use it as the benchmark.) This can be done by using a linking methodology familiar in the literature (Cieza et al. 2016) to identify ICF terms in each questionnaire. The second issue is what difference the functioning score produced by A&AQ has on the final work capacity score: does functioning as assessed by the A&AQ make a difference? Since the data from the pilot consists of results of both A&AQ and WHODAS, this can be determined empirically.

ICF content comparison

As mentioned, WHODAS was originally constructed in terms of ICF concepts and specific classification items. Although the A&AQ was not similarly constructed, it nonetheless purports to be a questionnaire for assessing functioning. Unfortunately, as Table 2 below shows, A&AQ items are either too unspecific to be linked to the specific ICF category, or else are ambiguous as they can be linked to more than one ICF category (e.g., item 2.3.3) or even more than one ICF chapter (e.g., 2.2.6). WHODAS items can unambiguously be linked to specific ICF items. As well, the items in Part I of A&AQ cannot be used to build a scale to assess ICF Activity and Participation items, although this was explicitly what A&AQ was designed to do.

Table 2: Comparing WHODAS and the A&AQ in terms of ICF categories and domains

ICF Domain	ICF Chapter	2 nd level Code	Title	WHODAS	Activity and Ability
Body Functions	b1 Mental functions	b114	Orientation functions		2.2.3
		b140	Attention functions	D1.1	2.2.1
		b144	Memory functions	D1.2	2.2.2
		b152	Emotional functions	D6.5	
Activity and participation	d1 Learning and applying knowledge	d159	Activity and participation	D6.2	
		d170	Basic learning, other specified and	D1.4	2.2.6
		d175	Writing		
		d179	Solving problems	D1.3	
	d2 General tasks and	d230	Applying knowledge, other specified and		2.2.6
		d230	Carrying out daily routine	D3.4	
	d3 Communication	d310	Communicating with - receiving - spoken	D1.5	2.2.5, 2.3.3
		d315	Communicating with - receiving -		2.2.4, 2.3.3
		d320	Communicating with - receiving - formal		2.2.4
		d325	Communicating with - receiving - written		2.2.4
		d330	Speaking		2.3.3
		d345	Writing messages		2.2.6
		d350	Conversation	D1.6	2.3.3
	d4 Mobility	d410	Changing basic body position	D2.2	2.1.1, 2.1.2
		d415	Maintaining a body position	D2.1	2.1.1, 2.1.2
		d450	Walking	D2.5	2.1.3
		d455	Moving around		2.1.6
		d460	Moving around in different locations	D2.3, D2.4	
		d470	Using transportation		2.1.4
	d5 Self-Care	d510	Washing oneself	D3.1	2.4.1
d530		Toileting		2.4.4	
d540		Dressing	D3.2	2.4.2	
d550		Eating	D3.3	2.4.3	
d570		Looking after one's health		2.4.5	

d6 Domestic-Life	d630	Preparing meals	2.5.1
	d640	Doing housework	2.5.2
	d649	Household tasks, other specified and	D5.2, D5.3, D5.4
	d699	Domestic life, unspecified	D5.1

Source: WHODAS and A&AQ.

Table 2 (Continued): Comparing WHODAS and the A&AQ in terms of ICF categories and domains

ICF Domain	ICF Chapter	2 nd level Code	Title	WHODAS	Activity and Ability
Activity and participation	d7 Interpersonal interactions and relationships	d730	Particular interpersonal relationships	D4.1	2.3.1
		d750	Informal social relationships	D4.2, D4.4	2.3.2
		d760	Family relationships		2.3.2
		d770	Intimate relationships	D4.5	
		d779	Interpersonal interactions and relationships, unspecified	D4.3	
	d8 Major life areas	d859	Work and employment, other specified and unspecified	D5.5, D5.6, D5.7, D5.8	
	d9 Community, social and civic life	d9	Community, social and civic life	D6.1	
		d940	Human rights	D6.3	
Environmental Factor		e125	Products and technology for communication		1.4
		e135	Products and technology for employment		1.4
Other		gh	General health	D6.6, D6.7	
		nc	Non-classified	D6.4, D6.8	
		pf	Personal factor		1.1, 1.2, 1.3

Source: WHODAS and A&AQ.

A&AQ coefficients and the impact of functioning assessed by A&AQ on final work capacity scores

As described above, the output of the A&AQ questionnaire is a score that is the sum of the points from questions in Parts I (Professional, work activities, and environmental accessibility) and II (Activities and ability to participate). Depending on the range of the score, the score is mapped onto a coefficient that is automatically applied to the medical assessment score to produce the final work capacity percentage. Thus, the coefficient may reduce the medical assessment score (.7) or increase it (1.2). It is likely that the entire methodology of assessing work capacity was designed so that the coefficients would be so closed to 1 that they would have a minimal impact on the final work capacity percentage.

Based on A&AQ score data from DWCAO Information System, Table 3 below shows the mean, standard deviation and percentage quantiles of the basic work capacity percentage (derived from medical diagnosis alone), and the final work capacity percentage adjusted by the A&AQ-derived coefficients. The Table also gives the A&AQ score that produced the coefficient that adjusts the basic work capacity score and, from the pilot dataset, the relevant WHODAS score.⁸

Table 3: Distribution of the Basic Work Capacity, the Work Capacity, Activity and Ability, and WHODAS-based Score

	Mean	SD	25%	50%	75%
Basic Work Capacity	46.0	14.2	35.0	45.0	55.0
Work Capacity	47.8	15.3	40.0	50.0	55.0
Activity & Ability Score	23.1	9.01	17.0	22.0	24.0
WHODAS Score	55.1	8.49	50.0	55.0	60.0

Source: WHODAS pilot data set and DWCAO Information System.

The Table shows, first of all, that the difference between the mean basic work capacity percentage and the final work capacity score that takes into account the activity and ability score is minimal (correlation is $R = 0.98$), suggesting a very low impact the A&AQ score has in the current disability assessment. Notice also that 75.0 percent of the sample has a final work capacity of 55.0 percent, which is the legislated upper cut-off to obtain disability status.

Secondly, the mean scores of the two instruments differ significantly. The sample's average A&AQ score was 23.1 ($SD = 9.01$) while for the WHODAS it was 55.1 ($SD = 8.49$). Comparison of the means with a paired t-test is highly significant (T-value = -179.89, $df = 2233$, P-value < 0.001) and suggests that for the same underlying functioning level in the assessment population, the instruments indicate significantly different scores. The lower mean of A&AQ scores suggests that the questionnaire targets a population with higher levels of disability than the WHODAS. The correlation of total scores of $R = 0.54$ indicates further that the sum scores of the two functioning measures are only moderately correlated. This suggests that the two measures are not aligned with regard to the functioning aspects that they measure.

The low sample average on the A&AQ score has to be understood in light of the definition of the response options of the questionnaire (Appendix 2). The response options range from 0 = 'No need for assistance' to 4 = 'Needs complete assistance'. Options 2 to 4 describe graduations of higher levels

⁸ For better comparability, the Activity and Ability Score has been rescaled from 0 to 100, so that A&AQ and WHODAS have the same range. Lower work capacity percentages indicate more functioning problems (or lower performance), while lower scores of the WHODAS and the Activity and Ability scores indicate better functioning (or lower level of disability).

of disability, in which individuals cannot function (totally) independently anymore. Appendix 4 shows the frequencies and percentages of responses to the A&AQ. A&AQ response options linked to scores of 3 or 4 correspond to a very high level of dependence and are rarely used in this assessment population (Appendix 4).

WHODAS, by contrast, since it has a normal score distribution curve (see Figure 4), means that the metric ranges over a broader spectrum and successfully capture the range from low to high levels of disability. The absence of ceiling effects further supports that the items, understood in a performance perspective, allow even high need individuals to report having moderate levels of disability when the individual has substantial supports available in their daily life. In other words, WHODAS more realistically captures the lived experience of disability: people with good supports will experience less disability than one might predict based on their underlying health condition alone. Someone who is blind, for example, has a high level of disability in many areas of life; yet with sufficient supports, that level of disability may be greatly reduced because the individual, though blind, can do all of the activities he or she needs, or wants to.

Figure 1: Relationship between the basic work capacity (medical) and the A&AQ scores

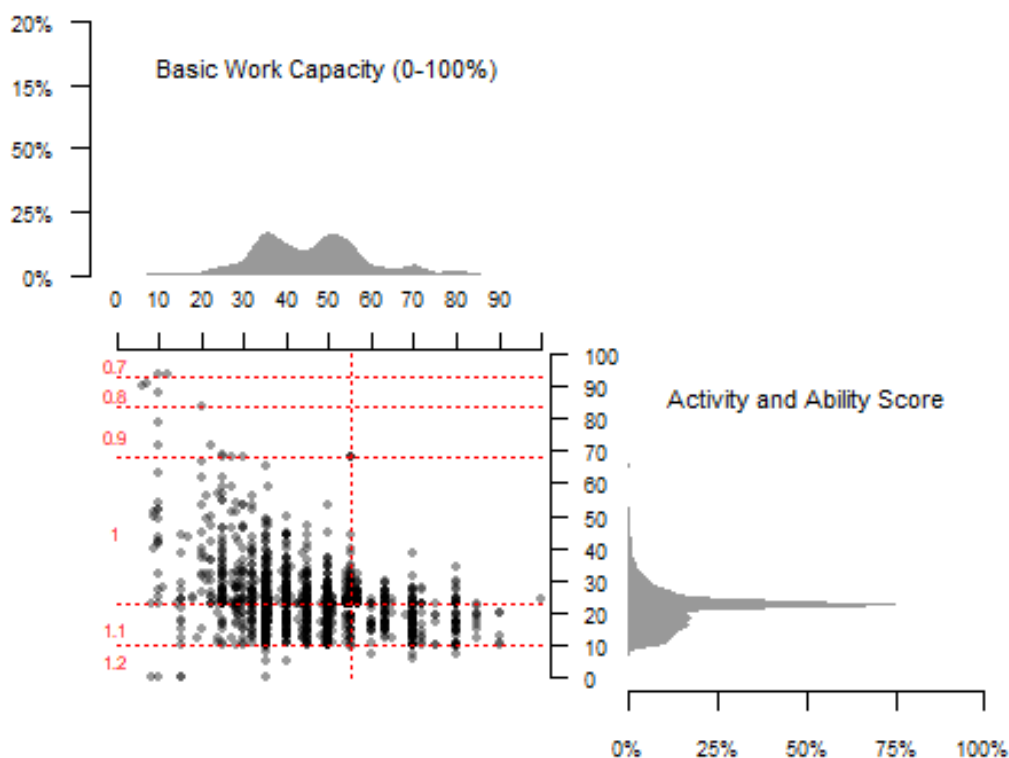


Figure 1 shows statistical details of the relationship between the basic work capacity score and the A&AQ score that explain precisely why A&AQ scores have minimal impact on the final assessment of work capacity. The figure is a scatterplot of the A&AQ scores plotted against the basic working capacity values. The dots represent individual scores from the pilot population. The red dotted horizontal lines delineate sections where a specific coefficient, the red number on the left, will be applied to adjust the basic work capacity. Coefficients <1 will decrease the basic working capacity while coefficients > 1 increase the basic work capacity.

The vertical line at 55.0 percent is the critical percentage point for determining eligibility after weighting of the basic work capacity score. A large part of the assessed population (84.0 percent) has a basic work capacity score <55.0 percent. We should expect that once A&AQ scores are applied to this basic work capacity score that at least some of them would be changed – i.e., in the figure, we would expect the dots in the scatterplot should move somewhat after the application of the coefficient, which adjusts based on functioning information. However, 87.0 percent of the population with an A&AQ score between 23 and 67 are in the area of no change, i.e., have a coefficient of 1. So, A&AQ makes very little difference for most of the pilot population. (The figure's margins provide the density distribution of the A&AQ on the right and the basic working capacity on the top. The score distribution of the A&AQ shows a sharp peak of observations in the range from 20 to 25.)

Conclusions about the Activity and Ability Questionnaire based on the above analysis

A&AQ collects information relevant to work capacity and supports a qualitative judgment about the respondent's work potential. However, both in terms of alignment with the ICF, and as a quantitative instrument that is objective, valid and reliable, A&AQ is fundamentally inadequate for several reasons:

- A&AQ is not entirely compatible with the ICF classification as there are several items that cannot be linked to the ICF.
- A&AQ contains items that are used in the scoring that are not part of the notion of functioning at all (e.g., general health), so it does not assess functioning but some other construct.
- Because of its nominal scaling, A&AQ cannot be relied on to provide non-arbitrary assessments of the extent of disability experienced by the respondent in any particular domain, and therefore, as an overall score.
- The link between summary scores, based on points from Part I and Part II, and the coefficient score is arbitrary and without any empirical basis.
- The values of the coefficients for a significant part of the population are 1 or close to 1 so that the A&AQ assessment of functioning would never have more than a minimal impact on the resulting assessment of work capacity based on medical criteria alone. As mentioned above, this was likely not accidental but a result of how the entire methodology was constructed.

PART TWO: THE PILOT

Descriptive Statistics of the Pilot Sample

As noted, during the WHODAS pilot, data was collected from 2,234 first-time applicants for disability assessment. The interview was conducted by trained professionals prior to the formal disability assessment process. This data collection flow has enabled statistical analysis and comparisons between data collected by WHODAS and information that resulted from the disability assessment process for all individuals who participated in the pilot.

Descriptive statistics for the population participating in the pilot are shown in Table 4. Participants were of age between 18 and 64 years old and capable of understanding and responding to the interviewer's questions. Information was collected from $N = 2,234$ persons. The proportion of male participants was higher (55.0 percent and 45.0 percent, respectively). The average age was 50.5 years ($SD = 11.6$), which is relatively young, almost 15 years younger than the mandatory retirement age. Most of the participants had a professional or vocational education ($N = 737$, 33.3 percent). About twenty-five percent ($N = 516$, 23.3 percent) had secondary education. Many participants had higher education, either in academia ($N = 331$, 14.9 percent) or at a higher professional education institution ($N = 393$, 17.7 percent). In total, $N = 843$ applicants (37.3 percent) were unemployed at the time of the assessment.

Most of the participants had a single primary ICD-10 health condition and one additional comorbidity ($N = 1,516$, 67.86 percent) while $N = 718$ (32.14 percent) had a single health condition without comorbidities.

Table 4: Description of pilot sample

N	2,234
Gender = Male (%)	1,229 (55)
Age - mean (SD)	50.5 (11.6)
Education Code - N (%)	
Basic	196 (8.8)
Primary	32 (1.4)
Secondary	516 (23.3)
Professional/Vocational	737 (33.3)
Higher (academia)	331 (14.9)
Higher (professional)	393 (17.7)
Special education	11 (0.5)
Employed Status = Unemployed - N (%)	843 (37.7)

Source: WHODAS pilot data set.

Table 5 presents the most frequently observed ICD-10 diagnostic chapters for the participant's primary health condition. Neoplasms are the most frequently reported ICD-10 chapter with $N = 541$ (24.22 percent) participants. Diseases of the nervous system ($N = 401$, 17.95 percent), diseases of the musculoskeletal systems ($N = 364$, 16.24 percent), and diseases of the circulatory system ($N = 314$, 14.06 percent) were experienced by more than 10.0 percent of the participants in the pilot.

Table 5: Prevalence of Health conditions in the pilot study population by ICD-10 Health Condition Category

ICD-Chapter	N	%
I Certain Infectious and Parasitic Diseases	30	1.34 %
II Neoplasm	541	24.22 %
III Diseases of the Blood	15	0.67 %
IV Endocrine Diseases	210	9.4 %
V Mental Disorders	161	7.21 %
VI Diseases of the Nervous System	401	17.95 %
VII Disease of the Ear	16	0.72 %
VII Diseases of the Eye	29	1.3 %
IX Disease of the Circulatory System	314	14.06 %
X Disease of the Respiratory System	21	0.94 %
XI Disease of the Digestive System	28	1.25 %
XII Disease of the Skin	8	0.36 %
XIII Disease of the Musculoskeletal System	364	16.29 %
XIV Disease of the Genitourinary System	15	0.67 %
XVII Congenital Malformations	5	0.22 %
XIX Injuries External Causes	62	2.78 %
XXI Factors Influencing Health Status and Contact with Health Services	6	0.27 %
Missing	8	0.36 %

Source: Lithuania WHODAS pilot data set.

Analysis Methodology

Psychometric Analysis: Rationale and Tests

Lithuania, like many European countries, has invested resources and political capital in reforming disability assessment for eligibility to benefits available to persons with disabilities from the social protection, health, and other government sectors. Traditionally, disability assessment has been a matter of using the Baremic approach⁹ to connect percentages of 'whole-person disability' directly to diagnostic categories of diseases and injuries by severity and associated impairments. The major difficulty with this approach, and the motivation for reform in European countries, is that a purely medical determination of the degree of disability that a person experiences in their lives fails to capture the essence of disability, namely functioning from the perspective of performance as defined by WHO's ICF.

But the fundamental scientific problem with the Baremic approach is at the heart of this pilot: Baremic instruments lack the basic psychometric properties that every assessment instrument must display – namely, validity and reliability. Roughly, an assessment instrument is valid when we have good reason to believe it represents what the instrument is intended to assess. An assessment instrument is reliable when it can be shown statistically that different assessments by different assessors of the same individual will yield similar results. The Baremic approach lacks these essential psychometric traits because the linkages between disability percentages and diagnostic categories are not based on empirical evidence but are at best established by the methodologically weak technique of unstructured professional consensus – several professionals coming to an agreement without

⁹ Named after François Barrême, a French mathematician from the 17th Century who invented the method.

empirical support. At worst, these links are purely speculative. Any reform of disability assessment instrumentation, therefore, must not only assess the relevant phenomenon – functioning from the perspective of performance – it must do so in a psychometrically sound manner to ensure that assessment is valid and has inter-assessor reliability.

In statistical terms, the comprehensive assessment of functioning as a component of a disability assessment process requires a very different methodology than arbitrary associations. First, sets of selected functioning items be identified that both best represent the most relevant functioning domains fit for the purposes of the assessment and can generate a summary score of disability. **Disability assessment is a summary assessment of the individual's lived experience of a health condition. What is being assessed is the entire experience, not some fragment of it.**

This entails that a very different approach is needed. Since it is neither realistic nor feasible to have an assessment tool that assesses *every* domain of functioning in a person (the ICF has more than a 1000 such domains), or to submit an applicant to a full rehabilitation diagnostic assessment, which might take several hours. We need to identify a representative set of functioning domains that can be shown, statistically, to capture as much of the entire experience of a person's functioning as possible in a summary score.

The set of ordinal-scaled functioning items assessed by a questionnaire and applied to a large number of real cases can be transformed into an interval scale by means of calibration with a psychometric model from the Rasch family.¹⁰ In this way, it is scientifically possible to identify exact numerical degrees or percentages of disability. In short, in order to truly be able to measure degrees of disability in a valid and reliable manner, we must have evidence that the resulting summary score has basic interval scale properties. Doing so is a precondition for both the validity and reliability of an assessment tool for functioning and therefore disability.

The consensus in the scientific literature is that Rasch analysis is the most appropriate and effective statistical method for determining whether interval scale properties are evident in a summary score derived from a questionnaire. Rasch 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). (The classic Rasch model works only with dichotomous data – e.g., responses of yes/no. But WHODAS and A&AQ used polytomous scoring – e.g., responses of 0-4. Because of this, the data was calibrated with the Partial Credit Model (Masters 1982), an extension of the Rasch model suitable for polytomous responses.)

The power of Rasch analysis, and the reason it is used here to evaluate the data from the WHODAS pilot, is that it establishes the essential measurement properties required for a well-performing questionnaire suitable for assessment purposes (Bond & Fox 2001; Tennant & Conaghan 2007). Specifically, the required measurement properties involve:

- (1) **The targeting of the scale:** Intuitively, a well-performing questionnaire matches the level of 'difficulty' of its items (i.e., the chances that some proportion of the population will be assessed at a particular response level) to the population being assessed. Statistically, good targeting is achieved if the mean item difficulty and mean person ability are approximating 0. (Here, 'difficulty' means the degree of functioning, and 'ability' means the individual ability to achieve a degree of functioning. In the case of WHODAS scores, high level of ability means high level of disability)

¹⁰ Roughly, scales can either be *nominal* (where numbers serve as labels to describe or classify a phenomenon or object), *ordinal* (where numbers represent a ranking or order such as first, second, third, or mild, moderate, severe) and *interval* or *ratio* scaled (in which quantitative measurement with equidistant units is possible). The difference between the interval and the ratio scale is that only the ratio scale contains a true zero that it cannot fall below (e.g., temperature is not a ratio-scale, but height measures are).

- (2) **The reliability of the scale:** A scale is reliable when it can discriminate between levels of, in this case, functioning in the population. This is important for a disability and work capacity assessment that needs to be granular enough to differentiate people with different levels of functioning. In Rasch analysis, the reliability is given by the Person Separation Index (PSI), also sometimes called Person Separation Reliability which ranges from 0 to 1, or perfect reliability and indicates how well a scale score differentiates between levels of functioning. A PSI score above 0.8 is the standard statistical test for good reliability of the scale; values above 0.9 indicate very good reliability. The classical measure of the internal consistency of the data – the Cronbach α score – is also used to test reliability (Nunnally and Bernstein 1994).
- (3) **The ordering of the response options of the items in the questionnaire:** It is crucial that, for example, the response score 4 represents a step on the scale 'higher' than score 3, and so forth, otherwise there is no consistency to the ranking, and the questionnaire is both invalid and unreliable. An analysis of response probability curves allows us to determine whether there are response options that have this problem and decide on strategies to resolving the problem by, say, aggregating disordered response options. For example, if for an item, the response options 2 and 1 appear reversed, suggesting that an increase of difficulty cannot be discriminated, then the item responses can be recoded so that these options represent only one level of response.
- (4) **Local Item Dependencies:** Items that are correlated (i.e., 'dependent') in a questionnaire are redundant and assess approximately the same aspect of the construct of interest – here functioning. Redundancy inflates reliability, distorting this important property of questionnaires. The most widely reported statistic for the item dependencies is the Q3 matrix or correlation matrix of the Rasch residuals (Yen 1984). Residuals correlations above 0.2 are considered not acceptable. Local item dependencies are typically solved by aggregating the correlated items into testlets. In testlets, the ordering is not expected anymore.
- (5) **Fit of the items to the Rasch model:** Rasch analysis depends on being able to construct a model of the data collected by the questionnaire that shows that it is actually assessing what we want to assess, namely functioning. To succeed, data about each item of the questionnaire must 'fit' the proposed model. Items that 'overfit' tend to sharply discriminate levels of functioning, while 'underfittings' are items that cannot discriminate levels of functioning sufficiently. The fit of items is given with the 'outfit' and 'infit' statistics. The infit is less sensitive to outliers. Statistically, for good item fit, the infit and outfit values should be below 1.2 (Smith, Schumacker, and Bush 1998).
- (6) **Differential Item Functioning (DIF):** We need to be aware of the impact of factors such as gender and age on responses to items. This is important for both disability and work capacity assessment because it allows us to 'spot' items and 'flag' subgroups where at a level of functioning, the response difficulty differ significantly. This effect is called DIF. The statistical test used to determine DIF is the ANOVA that allows us to identify exogenous variables that create a lack of invariance in the item difficulty (Holland and Wainger 1993). It is common to use ANOVA on gender and age groups. [For the analysis below, age groups were defined as < 40 years ($N = 361$, 16.16 percent), 40-50 ($N = 410$, 18.35 percent), 50-60 ($N = 960$, 42.9 percent), and above 60 ($N = 503$, 22.52 percent).] It is worthwhile to note that a DIF analysis does not always indicate a metric bias but can also simply identify subgroups with higher or lower functioning (Boone, Staver et al.).
- (7) **Unidimensionality of the questionnaire:** Finally, a questionnaire should measure only one construct, in this case, functioning, as this is the assessment criterion of interest in disability and work capacity. If a questionnaire has more than one dimension, it is assessing more than one construct, which means that there is no validity to the summary total score the

questionnaire produces. Unidimensionality is assessed with a principal component of the Rasch residuals (Smith 2002). Typically, a first eigenvalue <1.8 is deemed indicative of unidimensionality. (Based on simulation analyses, Smith and Miao (1994) suggested to rather consider the size of the second eigenvalue, with values below 1.4 as more appropriately used to identify unidimensionality.)

If these measurement properties and assumptions can be met, a questionnaire can confidently be said psychometrically sound (valid and reliable); we can also be confident that the summary scores derived from the questionnaire are interval-scaled and can be used for precise measurement purposes. Each of these assumptions in the case of WHODAS and the Activity and Ability to Participate Questionnaire (A&AQ) are discussed below. (All the metric analyses were performed with the software R (Team 2016) and, more specifically, the package *mirt* for the Rasch analysis¹¹).

Results

Metric properties of WHODAS

The analysis of the dataset collected in the piloting of the 36-item WHODAS showed that the work/school items (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; and D5.8 – Getting your work done as quickly as needed) were only responded to by persons who at the time of the assessment were working or were in some form of education. These items constituted more than 60.0 percent of the missing values across the pilot. It was decided to exclude these items for the metric analysis and to construct one WHODAS-based functioning score with the remaining 32 items (see Table 6). A certain number of other items were kept, even though they had a proportion of missing values above 10.0 percent (D2 .5 – Walking long distances (13.79 percent); D3.4 – Staying by yourself (12.58 percent)) and above 20.0 percent (D4.4 – Making new friends (24.89 percent); D4.5 – Sexual activities (35.32 percent); and D6.1 – Community activities (24.44 percent), see Table 6.

¹¹R. Philip Chalmers (2012). *mirt*: A Multidimensional Item Response Theory Package for the R Environment. *Journal of Statistical Software*, 48(6), 1-29. doi:10.18637/jss.v048.i06. Available at: <https://www.jstatsoft.org/article/view/v048i06>.

Table 6: Frequencies and Percentages of WHODAS Responses

Item	No	Mild	Moderate	Severe	Extreme, cannot	Missing
D1.1	850 (38%)	789 (35.3%)	429 (19.2%)	143 (6.4%)	23 (1%)	0 (0%)
D1.2	722 (32.3%)	911 (40.8%)	441 (19.7%)	125 (5.6%)	29 (1.3%)	6 (0.3%)
D1.3	901 (40.3%)	764 (34.2%)	420 (18.8%)	117 (5.2%)	28 (1.3%)	4 (0.2%)
D1.4	839 (37.6%)	628 (28.1%)	380 (17%)	177 (7.9%)	50 (2.2%)	160 (7.2%)
D1.5	1678 (75.1%)	365 (16.3%)	142 (6.4%)	38 (1.7%)	10 (0.4%)	1 (0%)
D1.6	1449 (64.9%)	452 (20.2%)	222 (9.9%)	85 (3.8%)	26 (1.2%)	0 (0%)
D2.1	260 (11.6%)	579 (25.9%)	853 (38.2%)	418 (18.7%)	111 (5%)	13 (0.6%)
D2.2	508 (22.7%)	811 (36.3%)	620 (27.8%)	244 (10.9%)	49 (2.2%)	2 (0.1%)
D2.3	634 (28.4%)	883 (39.5%)	537 (24%)	149 (6.7%)	31 (1.4%)	0 (0%)
D2.4	543 (24.3%)	658 (29.5%)	688 (30.8%)	271 (12.1%)	66 (3%)	8 (0.4%)
D2.5	248 (11.1%)	344 (15.4%)	634 (28.4%)	467 (20.9%)	233 (10.4%)	308 (13.8%)
D3.1	574 (25.7%)	872 (39%)	557 (24.9%)	180 (8.1%)	48 (2.1%)	3 (0.1%)
D3.2	687 (30.8%)	952 (42.6%)	458 (20.5%)	111 (5%)	26 (1.2%)	0 (0%)
D3.3	1757 (78.6%)	312 (14%)	125 (5.6%)	29 (1.3%)	10 (0.4%)	1 (0%)
D3.4	655 (29.3%)	586 (26.2%)	485 (21.7%)	155 (6.9%)	72 (3.2%)	281 (12.6%)
D4.1	1153 (51.6%)	581 (26%)	279 (12.5%)	119 (5.3%)	44 (2%)	58 (2.6%)
D4.2	1463 (65.5%)	460 (20.6%)	203 (9.1%)	74 (3.3%)	20 (0.9%)	14 (0.6%)
D4.3	1548 (69.3%)	455 (20.4%)	175 (7.8%)	40 (1.8%)	8 (0.4%)	8 (0.4%)
D4.4	693 (31%)	376 (16.8%)	336 (15%)	171 (7.7%)	102 (4.6%)	556 (24.9%)
D4.5	357 (16%)	297 (13.3%)	402 (18%)	272 (12.2%)	117 (5.2%)	789 (35.3%)
D5.1	179 (8%)	645 (28.9%)	954 (42.7%)	341 (15.3%)	93 (4.2%)	22 (1%)
D5.2	206 (9.2%)	665 (29.8%)	915 (41%)	328 (14.7%)	96 (4.3%)	24 (1.1%)
D5.3	207 (9.3%)	649 (29.1%)	926 (41.5%)	329 (14.7%)	97 (4.3%)	26 (1.2%)
D5.4	103 (4.6%)	567 (25.4%)	995 (44.5%)	421 (18.8%)	122 (5.5%)	26 (1.2%)
D5.5	117 (5.2%)	200 (9%)	290 (13%)	123 (5.5%)	47 (2.1%)	1457 (65.2%)
D5.6	135 (6%)	188 (8.4%)	285 (12.8%)	125 (5.6%)	44 (2%)	1457 (65.2%)
D5.7	138 (6.2%)	204 (9.1%)	292 (13.1%)	121 (5.4%)	46 (2.1%)	1433 (64.1%)
D5.8	100 (4.5%)	209 (9.4%)	306 (13.7%)	146 (6.5%)	54 (2.4%)	1419 (63.5%)
D6.1	306 (13.7%)	543 (24.3%)	400 (17.9%)	297 (13.3%)	142 (6.4%)	546 (24.4%)
D6.2	645 (28.9%)	836 (37.4%)	483 (21.6%)	194 (8.7%)	62 (2.8%)	14 (0.6%)
D6.3	1042 (46.6%)	670 (30%)	330 (14.8%)	135 (6%)	34 (1.5%)	23 (1%)
D6.4	54 (2.4%)	593 (26.5%)	718 (32.1%)	723 (32.4%)	139 (6.2%)	7 (0.3%)
D6.5	96 (4.3%)	525 (23.5%)	762 (34.1%)	728 (32.6%)	115 (5.1%)	8 (0.4%)
D6.6	186 (8.3%)	502 (22.5%)	792 (35.5%)	653 (29.2%)	68 (3%)	33 (1.5%)
D6.7	147 (6.6%)	616 (27.6%)	810 (36.3%)	557 (24.9%)	56 (2.5%)	48 (2.1%)
D6.8	567 (25.4%)	560 (25.1%)	618 (27.7%)	337 (15.1%)	81 (3.6%)	71 (3.2%)

Source: WHODAS pilot data set.

For the WHODAS pilot, two datasets were analyzed: the first set of pilot data that was collected at the mid-point of the overall pilot and a final, complete dataset that was assembled at the end of the pilot. For the mid-pilot dataset, a first series of psychometric analyses were conducted, and several strategies were tested to decide which approach would best accommodate items with missing values but also allow for analysis of local item dependencies and issues of multidimensionality. The strategy that worked best at this stage (see Appendix 3) involved not imputing the data, aggregation into testlets, and no item recoding. The same strategy was adopted for the analysis of the final, complete dataset. The mid-point observations on the pilot data ($N = 1,024$) were confirmed with the final complete sample ($N = 2,234$).

The whole scale showed multidimensionality with a strong tendency of the items to load by WHODAS domains. Only a few items cross-loaded, and only a few items were free of dependencies. To solve the multidimensionality and local item dependencies, the correlating items were aggregated, considering the domain structure of the WHODAS. The detailed statistics are shown in Table 7 for the reliability and quality of targeting, Table 8 presents data for the fit statistics at the start of the analysis, and Table 9 shows the fit statistics after adjustments.

What follows are descriptions of the characteristics of the data and results of the psychometric analyses performed on the final sample using the best adjustment approach decided on at the mid-pilot data statistical analysis:

- (1) **The targeting of the scale:** The targeting of the scale improved with adjustments, i.e., item difficulties becoming more centered on the general difficulty estimate. However, aggregation of the scale items somewhat narrowed the measurement scope (Table 7).
- (2) **The reliability of the scale:** The reliability, inflated at the beginning of the analysis because of the item dependencies ($PSI = 0.94$, Cronbach $\alpha = 0.95$) was found to be good after adjustments were undertaken ($PSI = 0.87$, Cronbach $\alpha = 0.83$); Table 7.

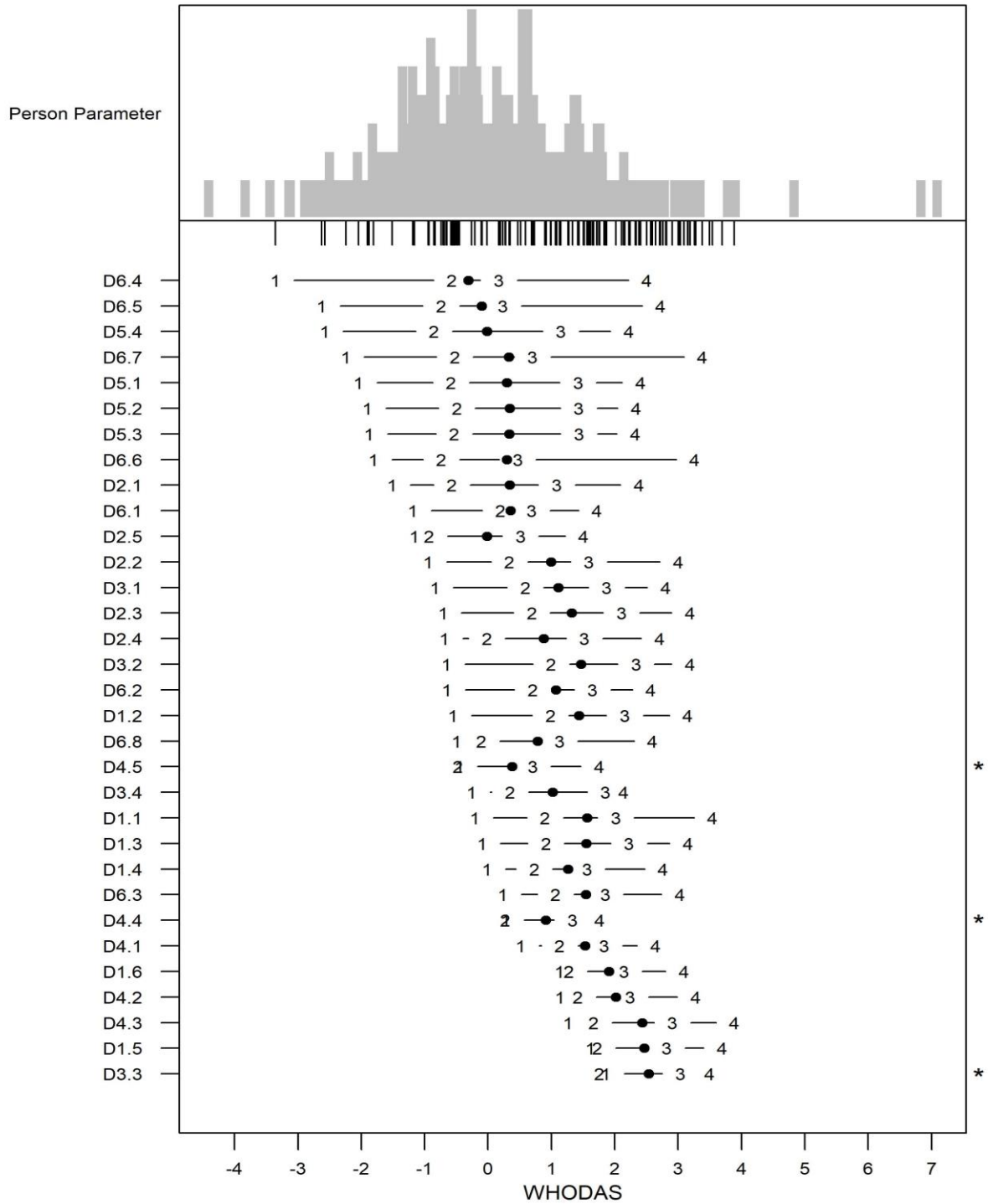
Table 7: Targeting and Reliability of WHODAS items

	Targeting			
	Start Mean	SD	Final Mean	SD
Difficulty	1.00	1.57	0.31	0.83
Ability	0.00	1.03	0.00	0.42
	PSI	Alpha	PSI	Alpha
Reliability	0.94	0.95	0.87	0.83

Source: WHODAS pilot data set.

- (3) **The ordering of the response options:** Threshold ordering was rather good at the start, with only 3 items (D3.3 – Eating, D4.4 – Making new friends, and D4.5 – Sexual activities) showing disordered thresholds (Figure 2).
- (4) **Local Item Dependencies:** The analysis of the residual dependencies showed strong local dependencies among the 32 items of the WHODAS 2.0 (see Figure 3), with a tendency for questionnaire items from the same domain to associate. To address these dependencies, the items were aggregated, taking into account the chapter structure. The domain 6, *Participation in society*, was kept as two subsets, items D6.1 to D6.4 and items D6.5 to D6.8, as the correlational structure indicated independence of these two subsets (Figure 3). A residual correlation above $r = 0.2$ was found between domain 1 (*Understanding and Communicating*) and domain 4 (*Getting along with people*), which were aggregated accordingly. The thresholds of the testlets are not expected to be ordered.

Figure 2: Person item map for the WHODAS items before collapsing the response options



*indicate disordered thresholds

Figure 3: Local Item Dependencies before the creation of testlets

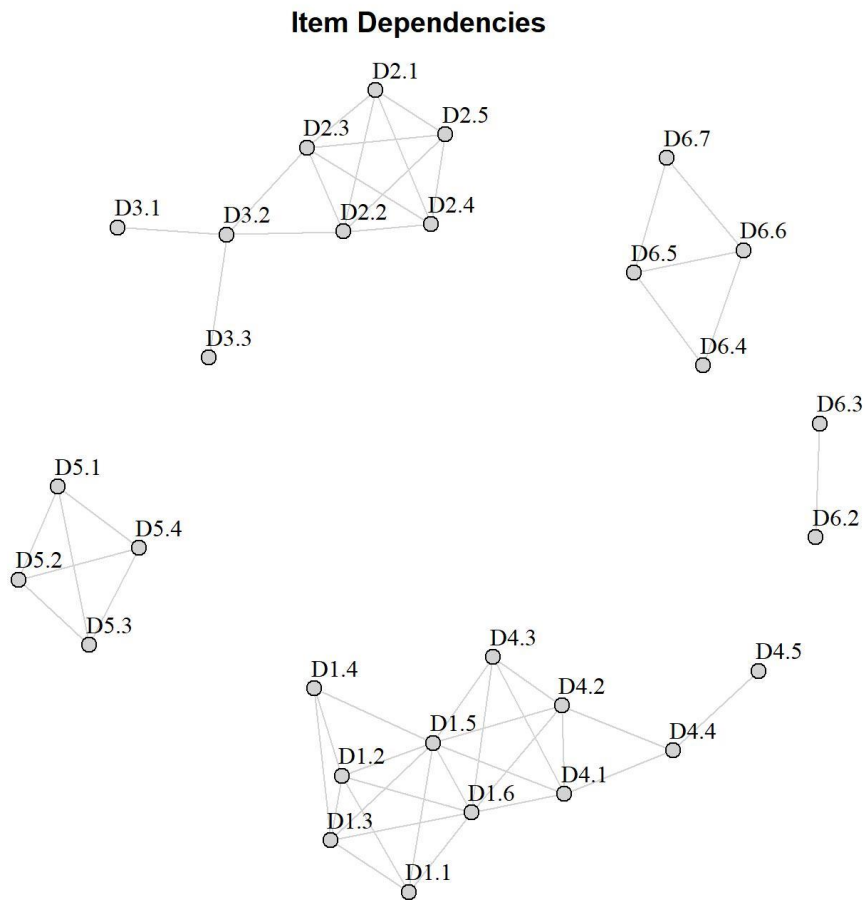


Table 8: WHODAS Item Difficulties, fit, local item dependencies, and differential item functioning at the start

Item Nbr.	Outfit ¹	Infit ¹	Item	Disordered Thresholds	LID ²	DIF ³
D1.1	1.13	1.12	1.56		D1.2, D1.3, D1.5, D1.6	Age, Gender
D1.2	1.13	1.11	1.44		D1.1, D1.3, D1.4, D1.5, D1.6	Age, Gender
D1.3	1.12	1.12	1.55		D1.1, D1.2, D1.4, D1.5, D1.6	Age
D1.4	1.19	1.12	1.26		D1.1, D1.2, D1.3, D1.5	Age, Gender
D1.5	1.07	1.02	2.46		D1.1, D1.2, D1.3, D1.4, D1.6, D4.1, Age	
D1.6	1.16	1.09	1.91		D1.1, D1.2, D1.3, D1.5, D4.1, D4.2, Age	
D2.1	0.96	0.96	0.35		D2.2, D2.3, D2.4, D2.5	Age
D2.2	1.04	1.05	1.00		D2.1, D2.3, D2.4, D2.5, D3.2	Age
D2.3	0.85	0.89	1.32		D2.1, D2.2, D2.4, D2.5, D3.2	Age
D2.4	0.84	0.87	0.88		D2.1, D2.2, D2.3, D2.5	Age
D2.5	1.04	1.05	-0.02		D2.1, D2.2, D2.3, D2.4	Age, Gender
D3.1	0.85	0.89	1.11		D3.2	Age, Gender
D3.2	0.95	0.97	1.47		D2.2, D2.3, D3.1, D3.3	Age, Gender
D3.3	1.06	1.00	2.54	x	D3.2	Gender
D3.4	0.90	0.96	1.02			
D4.1	1.27	1.20	1.53		D1.5, D1.6, D4.2, D4.3, D4.4	Age
D4.2	1.00	1.07	2.02		D1.5, D1.6, D4.1, D4.3, D4.4	Age
D4.3	1.30	1.23	2.43		D1.5, D1.6, D4.1, D4.2	Age
D4.4	1.50	1.24	0.91	x	D4.1, D4.2, D4.5	Age
D4.5	1.16	1.19	0.38	x	D4.4	Age
D5.1	0.73	0.73	0.30		D5.2, D5.3, D5.4	Age
D5.2	0.64	0.64	0.34		D5.1, D5.3, D5.4	Age
D5.3	0.70	0.70	0.34		D5.1, D5.2, D5.4	Age
D5.4	0.74	0.74	-0.01		D5.1, D5.2, D5.3	Age
D6.1	0.98	1.01	0.36			
D6.2	0.89	0.93	1.07		D6.3	
D6.3	1.19	1.07	1.55		D6.2	Age, Gender
D6.4	1.30	1.28	-0.31		D6.5, D6.6	
D6.5	1.08	1.08	-0.10		D6.4, D6.6, D6.7	Gender
D6.6	1.23	1.19	0.30		D6.4, D6.5, D6.7	Age, Gender
D6.7	1.17	1.18	0.33		D6.5, D6.6	Age
D6.8	0.92	0.92	0.78			Gender

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

² Local item dependency (LID) significant with $r > 0.2$

³ Differential item functioning (DIF)

- (5) **Fit of the items to the Rasch model:** The item fit, with infit and outfit expected below 1.2, was found very acceptable already at the start, with only 5 out of the 32 items showing infit or outfit above the cut-off (D4.1 – *Dealing with strangers*; D4.3 – *Getting along with close people*; D4.4 – *Making new friends*; D6.4 – *Time on health condition*; and D6.6 – *Health as drain on financial*

resources) (Table 8). After aggregation, all testlets showed acceptable infit and outfit values, below 1.2 (Table 9).

- (6) **Differential Item Functioning:** DIF was tested for gender and age. Most items and testlets appeared to be affected by the age of the participants, which ranged up to 64 years. In the final model with testlets, gender effects are seen in the testlets aggregating the items from domain 3 *Self-care* and domain 6 *Participation in society*. To keep the sum scores comparable across the entire population without facilitating subgroups with higher difficulties, items were not adjusted for the observed DIF (Table 8-9).
- (7) **Unidimensionality of the questionnaire:** The principal component analysis indicated multidimensionality with items clustering by domains and with a 1st eigenvalue of 5.40 and a 2nd eigenvalue of 2.81. After adjustments, i.e., aggregation of items by WHODAS domains, the 1st eigenvalue dropped to 1.82 and the 2nd eigenvalue of 1.33, supporting unidimensionality according to the defined criterion.

Table 9: WHODAS Item Difficulties, fit, local item dependencies, and differential item functioning after adjustment

WHODAS Label Item No.	Outfit ¹	Infit ¹	Item Difficulty	Disordered Thresholds	LID ³	DIF ⁴
Testlet 1 D1.1-D1.6 & D4.1-1.18		1.18	0.54	n.a. ²	no	Age
Testlet 2 D2.1-D2.5	0.90	0.91	0.26	n.a. ²	no	Age
Testlet 3 D3.1-D3.4	0.71	0.72	0.59	n.a. ²	no	Age, Gender
Testlet 4 D5.1-D5.4	0.78	0.78	0.07	n.a. ²	no	Age
Testlet 5 D6.1-D6.4	0.71	0.71	0.27	n.a. ²	no	Age, Gender
Testlet 6 D6.5-D6.8	0.82	0.81	0.10	n.a. ²	no	Gender

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

³ Local item dependency (LID) significant with $r > 0.2$

⁴ Differential item functioning (DIF)

Finally, Table 10 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 10: Transformation Table for WHODAS

WHODAS Score	Rasch Logit	0-100 Score	WHODAS* Score	Rasch* Logit	0-100* Score
0	-2.71	0	64	0.49	64
1	-2.25	9	65	0.5	64
2	-1.83	18	66	0.52	65
3	-1.58	22	67	0.53	65
4	-1.42	26	68	0.55	65
5	-1.29	28	69	0.56	66
6	-1.18	31	70	0.58	66
7	-1.1	32	71	0.59	66
8	-1.02	34	72	0.6	66
9	-0.95	35	73	0.62	67
10	-0.89	36	74	0.63	67
11	-0.83	38	75	0.65	67
12	-0.78	39	76	0.66	68
13	-0.73	40	77	0.68	68
14	-0.68	41	78	0.69	68
15	-0.64	41	79	0.71	68
16	-0.6	42	80	0.72	69
17	-0.56	43	81	0.74	69
18	-0.52	44	82	0.75	69
19	-0.48	45	83	0.77	70
20	-0.44	45	84	0.78	70
21	-0.41	46	85	0.8	70
22	-0.38	47	86	0.81	71
23	-0.34	47	87	0.83	71
24	-0.31	48	88	0.84	71
25	-0.28	49	89	0.86	71
26	-0.25	49	90	0.87	72
27	-0.22	50	91	0.89	72
28	-0.2	50	92	0.9	72
29	-0.17	51	93	0.92	73
30	-0.14	51	94	0.93	73
31	-0.12	52	95	0.95	73
32	-0.09	52	96	0.96	73
33	-0.06	53	97	0.98	74
34	-0.04	53	98	0.99	74
35	-0.02	54	99	1.01	74
36	0.01	54	100	1.02	75
37	0.03	55	101	1.04	75

Source: Lithuania WHODAS pilot data set.

Table 10 (Continued): Transformation Table for WHODAS

WHODAS Score	Rasch Logit	0-100 Score	WHODAS*	Rasch*	0-100*
			Score	Logit	Score
38	0.05	55	102	1.05	75
39	0.07	56	103	1.07	76
40	0.09	56	104	1.09	76
41	0.11	56	105	1.11	76
42	0.13	57	106	1.12	77
43	0.15	57	107	1.15	77
44	0.17	58	108	1.17	78
45	0.19	58	109	1.19	78
46	0.21	58	110	1.21	78
47	0.22	59	111	1.23	79
48	0.24	59	112	1.26	79
49	0.26	59	113	1.28	80
50	0.28	60	114	1.32	81
51	0.29	60	115	1.36	81
52	0.31	60	116	1.4	82
53	0.32	61	117	1.45	83
54	0.34	61	118	1.52	85
55	0.36	61	119	1.59	86
56	0.37	62	120	1.67	88
57	0.39	62	121	1.74	89
58	0.4	62	122	1.82	91
59	0.42	63	123	1.9	92
60	0.43	63	124	1.97	94
61	0.45	63	125	2.05	95
62	0.46	63	126	2.13	97
63	0.48	64	127	2.21	98
64	0.49	64	128	2.28	100

Source: Lithuania WHODAS pilot data set.

Summary: the psychometric properties of WHODAS

Taking together the seven essential statistical tests described above show that the data collected with WHODAS, under the Rasch analysis, displays very robust psychometric properties of validity and reliability. With a few adjustments, the scale is well targeted 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 to translate observed sum scores into an interval scaled metric.

It is important to keep in mind that WHO 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. 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 to augment the disability percentage score by health condition assigned by medical

assessment in order to enhance the accuracy and validity of the disability and work capacity assessment process in Lithuania.

Metric properties of the Activity and Ability Questionnaire

The *Questionnaire of the Individual's Activity and Ability to Participate* (A&AQ) was filled out by a DWCAO assessor during the interview with applicants for work capacity assessment. The A&AQ was created by the Lithuanian Ministry of Social Security and Labor (MSSL) in order to generate data that can be used to create weighting coefficients for the work capacity assessment. Based on the A&AQ, a coefficient ranging from 0.7 to 1.2 is derived that adjusts the score from the basic work capacity assessment – derived from the purely medical assessment – with 'activity and ability to participate' information. Based on the ICF content comparison reported above, it is reasonable to say most of the items assessed functioning in the ICF sense (i.e., the 'activity and ability' construct is similar to the 'functioning' construct). Therefore, the adjusted work capacity score that results from the application of the derived coefficient, and then used to determine the eligibility of the person for benefits, can tentatively be viewed as a functioning-augmented assessment.

Although the A&AQ functions analogously to the WHODAS, a relevant 'head-to-head' comparison between the two required us to perform the same kind of metric analysis on A&AQ as was done for WHODAS. In this way, the A&AQ score's measurement properties and the statistical quality of the resulting coefficients could be evaluated.

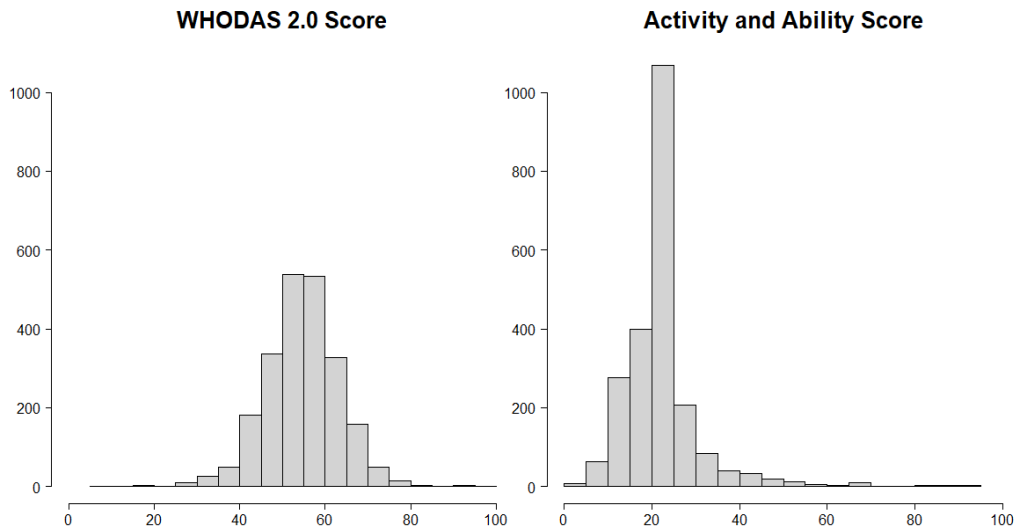
The metric analysis of the A&AQ was conducted with the 26 items that build the sum score of interest. Six individuals showed a high number of missing values (> 20 items) and were excluded from the analysis. A total of $N = 2,228$ individuals represented the study population for the psychometric analysis with the Rasch model. The frequencies and proportions of ratings for each item are shown in Appendix 3.

It is significant that the kurtosis¹² of the distribution of the A&AQ score is extremely high (*kurtosis* = 12.74). In principle, a kurtosis between -2 and +2 is considered acceptable and supports the claim that the data is normally distributed. Here, about 35.0 percent of the respondents of the scale achieved a score of 22 or 23. By comparison, the kurtosis of WHODAS is fully within the acceptable range and as a result shows a relatively normal distribution of values. Since, as in the Rasch analysis, each score translates into one ability estimate, it can be expected that the distribution of the 'activity and ability' estimates will be equally poor. While the Rasch model can still be computed, as it does not presuppose normally distributed population scores, it can be expected that the general reliability of the questionnaire will be affected.

As Figure 4 below shows, the difference in kurtosis values between WHODAS and A&AQ makes a substantial difference in the face validity of the two instruments. While WHODAS displays a normal distribution of severity of disability – intuitively representing the 'natural' distribution of health conditions and functioning limitations across a population – A&AQ radically 'peaks' at a mild level of disability so that nearly a third of the assessed population experiences levels of disability to that degree. WHODAS, in short, discriminates more levels of functioning across the assessed population, which makes it the basis for a more effective and arguably equitable functioning metric.

¹² In statistics, kurtosis is a form of distortion of a probability distribution, compared to the 'normal distribution', graphed as a so-called 'bell curve' in which the peak is in the center and the two sides ('tails') gently slope downward. The normal distribution is said to have kurtosis value of 0. A positive kurtosis is characterized by peaked curve and fewer outlier to the norm, whereas a negative kurtosis is characterized by a flatter curve and more outliers to the norm.

Figure 4: Score frequency distribution of WHODAS and the A&AQ scores



What follows shows the results of the metric analysis of the A&AQ in terms of the same seven measurement assumptions and statistical tests used to analyze WHODAS. As for WHODAS, the scale is first calibrated with all the items and then with the adjustments necessary to achieve metrical soundness.

- (1) **The targeting of the scale:** The targeting is shown in Table 11, and as already mentioned above, the population is very peaked. After the Rasch analysis, the mean difficulty of the questionnaire is 1.42 logits and the standard deviation 1.89 logits. The person ability parameter has an SD = 0.71 around the mean set to zero by the Rasch model (Figure 5). A mean item difficulty of zero would be expected for very good targeting of the instrument to the population. The mean difficulty of 1.42 logits by contrast, means that high scores, i.e., higher disability, are less likely than what the scales aim to measure.
- (2) **The reliability of the scale:** scale reliability is relatively good with a $PSI = 0.84$ and a Cronbach $\alpha = 0.86$. Yet this score is inflated by item dependencies and multidimensionality (see below) (Table 11).

Table 11: Targeting and Reliability of Activity and Ability items

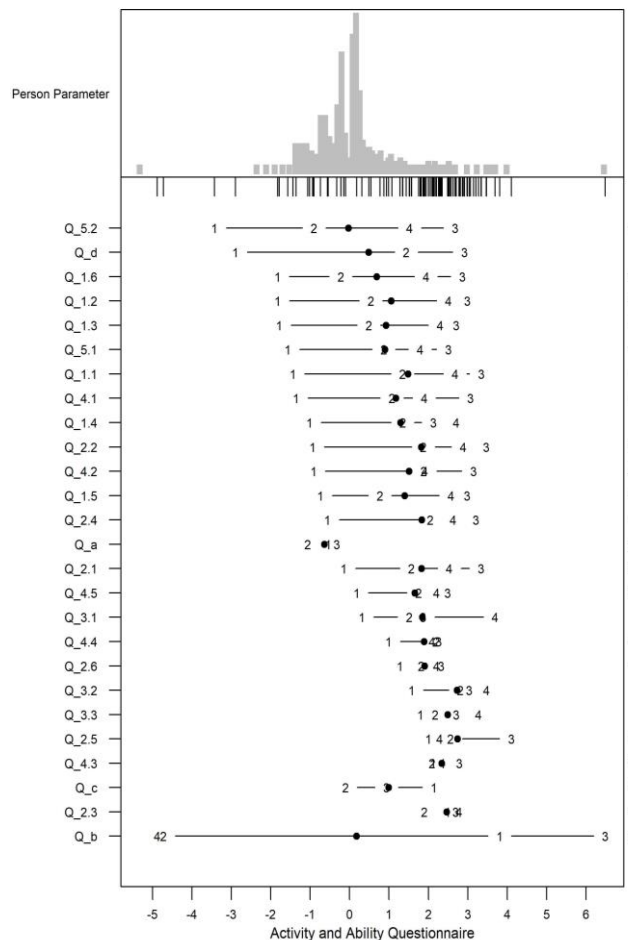
Targeting						
	Start		1) Statistically based item aggregation		2) Domain-based item aggregation	
	Mean	SD	Mean	SD	Mean	SD
Difficulty	1.42	1.89	0.64	0.96	0.39	1.38
Ability	0.00	0.71	0.01	0.30	0.00	0.38
Reliability						
	PSI	Alpha	PSI	Alpha	PSI	Alpha
Reliability	0.84	0.86	0.67	0.49	0.72	0.69

- (3) **The ordering of the response options:** Threshold ordering is problematic, with most items showing disordered thresholds (Figure 5). This indicates that the response options do not work as intended.
- (4) **Local Item Dependencies:** The analysis shows that there are many residual dependencies between items above the cut-off of $r = 0.2$ (see Figure 6). Items of Domains 2. *Application of Knowledge* and Domain 3. *Interaction* were associated, as well as Domains 1. *Mobility*, 4.

Independence, and 5. Daily activities are affected by this. Further, the items Q_b *Professional qualification*, and Q_c *Work experience and work skills* are correlated highly. This means that A&AQ has multiple redundancies that undermine the reliability of the total score.

- (5) **Fit of the items to the Rasch model:** The item fit, with infit and outfit ideally below 1.2, was found to be good for most items of the A&AQ scale. Specifically, problematic items, with infit or outfit above 1.2, are Q_a *Age groups*, Q_b *Professional qualification*, and Q_c *Work experience and work skills* (Table 11). Although these items are also part of the A&AQ score, they do not represent what the instruments is supposed to be assessing, namely, functioning from the perspective of performance (Table 12).
- (6) **Differential Item Functioning:** DIF was tested for both gender and age. Most items are sensitive to the age of the participants. Lack of invariance in the difficulty of items for the gender's respondent is seen in items Q_1.5 *Picking up and moving of things*, Q_2.1 *Concentration*, Q_2.2 *Memory*, Q_2.3 *Understanding visual information*, and Q_4.5 *Taking care of own health* (Table 12).
- (7) **Unidimensionality of the questionnaire:** The principal component analysis indicated that the items cluster by domains which results in multidimensionality, with a 1st eigenvalue of 5.1 and a 2nd eigenvalue of 2.34. Multidimensionality means that the A&AQ does not assess one coherent construct, namely functioning, but in fact, assesses several constructs that are not conceptually linked. A&AQ is therefore not an appropriate instrument for assessing functioning in a consistent and valid manner.

Figure 5: Person item map of the Activity and Ability Questionnaire



*indicate disordered thresholds

Figure 6: Local Item Dependencies

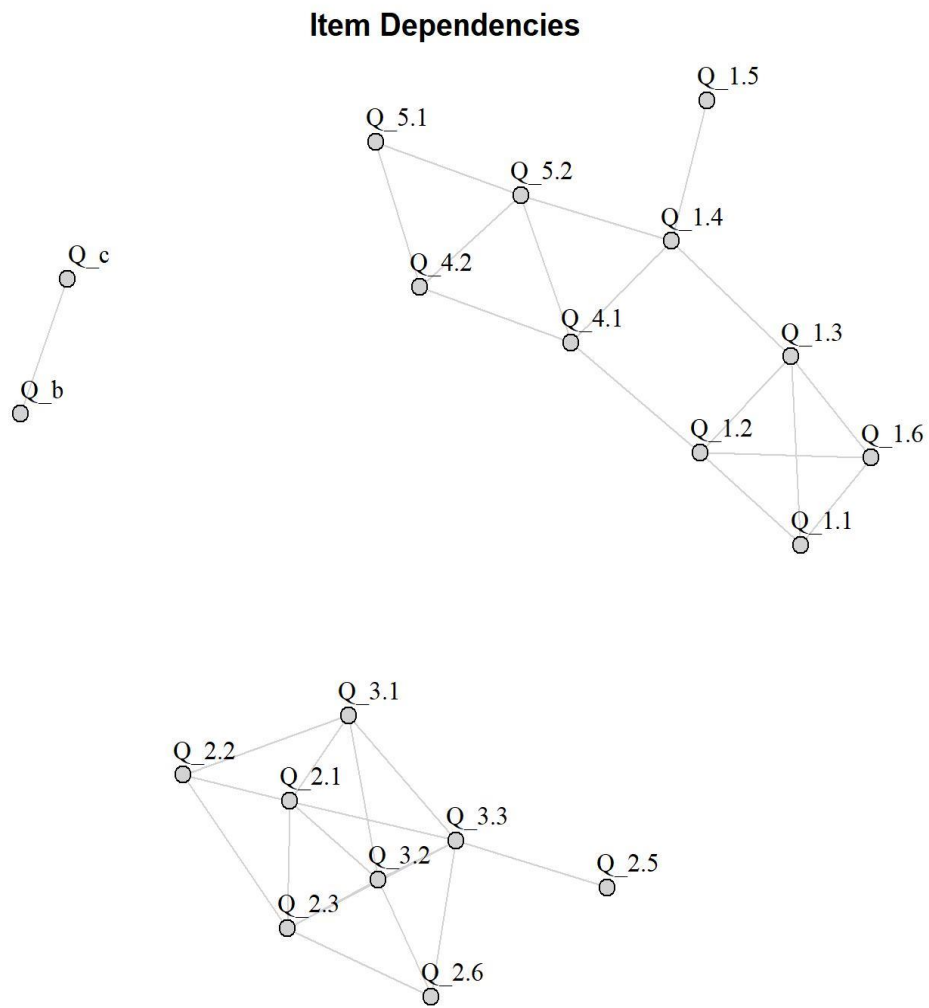


Table 12: Item Difficulties, fit, Local item dependencies, and differential item functioning of the Activity and Ability Questionnaire

Activity and Ability	Outfit ¹	Infit ¹	Item Difficulty	Disordered Thresholds	LID ²	DIF ³
Q_a	2.07	1.3	-0.64	x		
Q_b	1.77	1.32	1.77	x	Q_c	
Q_c	1.48	1.16	1.48	x	Q_b	Age
Q_d	0.87	0.91	0.87			Age
Q_1.1	0.91	0.92	0.91	x	Q_1.2, Q_1.3, Q_1.6	Age
Q_1.2	0.88	0.89	0.88	x	Q_1.1, Q_1.3, Q_1.6, Q_4.1	Age
Q_1.3	0.86	0.88	0.86	x	Q_1.1, Q_1.2, Q_1.4, Q_1.6,	Age
Q_1.4	0.74	0.76	0.74		Q_1.3, Q_1.5, Q_4.1, Q_5.2	Age
Q_1.5	0.92	0.93	0.92	x	Q_1.4	Age, Gender
Q_1.6	0.84	0.85	0.84	x	Q_1.1, Q_1.2, Q_1.3	Age
Q_2.1	1.01	1.02	1.01	x	Q_2.2, Q_2.3, Q_3.1, Q_3.2, Q_3.3	Age, Gender
Q_2.2	0.97	0.99	0.97	x	Q_2.1, Q_2.3, Q_3.1	Age, Gender
Q_2.3	0.83	1	0.83	x	Q_2.1, Q_2.2, Q_2.6, Q_3.2, Q_3.3	Age
Q_2.4	1.09	1.09	1.09	x		Age, Gender
Q_2.5	1.2	1.07	1.2	x	Q_3.3	Age
Q_2.6	0.87	0.93	0.87	x	Q_2.3, Q_3.2, Q_3.3	Age
Q_3.1	1.1	1.08	1.1		Q_2.1, Q_2.2, Q_3.2, Q_3.3	Age
Q_3.2	0.94	0.97	0.94		Q_2.1, Q_2.3, Q_2.6, Q_3.1, Q_3.3	Age
Q_3.3	0.99	1.01	0.99		Q_2.1, Q_2.3, Q_2.5, Q_2.6, Q_3.1,	Age
Q_4.1	0.82	0.83	0.82	x	Q_1.2, Q_1.4, Q_4.2, Q_5.2	Age
Q_4.2	0.81	0.81	0.81	x	Q_4.1, Q_5.1, Q_5.2	Age
Q_4.3	0.81	0.82	0.81	x		Age
Q_4.4	0.89	0.94	0.89	x		
Q_4.5	0.81	0.84	0.81	x		Age, Gender
Q_5.1	0.77	0.78	0.77	x	Q_4.2, Q_5.2	
Q_5.2	0.73	0.76	0.73	x	Q_1.4, Q_4.1, Q_4.2, Q_5.1	

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

² Local item dependency (LID) significant with $r > 0.2$

³ Differential item functioning (DIF)

As for WHODAS, the A&AQ showed multidimensionality and locally dependent items; however, given the lower reliability of the scale with poorly distributed scores, a solution that would fully satisfy the assumptions of the Rasch model was not possible. The items of the first part, i.e. the person factors, showed poor fit, which statistically supports that they do not work to assess functioning. The nominal responses could not be calibrated to ordered response difficulty thresholds. The scale in general, showed poor targeting. The levels of functional dependence that the scale is able to measure are far above the level observed in the assessed population in general.

The suitability of A&AQ as an instrument for disability assessment

A&AQ is used by DWCAO to augment the basic work capacity assessment that is derived from a purely medical assessment by using a score of overall 'activity and ability' to generate weighting coefficients ranging from 0.7 to 1.2 adjusts the score from the basic work capacity assessment. Analyzing the basic features of A&AQ, using an ICF content comparison, it could be concluded that:

1. On its face, the 'activity and ability' construct in A&AQ is analogous to the 'functioning' construct that is the basis for WHODAS, so that there is a prima facie reason to believe that A&AQ adds the functioning dimension to work capacity assessment.
2. The ICF content comparison, however, shows that some of the A&AQ items are either too vague to be clearly linked to ICF or simply are not functioning-relevant items at all, so that though analogous, the 'activity and ability' construct is not identical to functioning.
3. The A&AQ relies on nominal response options that are then mapped onto an ordinal scale (0-4). For assessing levels of work capacity or disability more broadly, this arbitrariness is problematic as there is no empirical justification for these ordinal rankings.
4. The link between the summary scores and coefficient scores that A&AQ generates is completely arbitrary and without any empirical basis. The result, as confirmed by empirical evidence, is that the A&AQ assessment of functioning has only minimal impact on the resulting assessment of work capacity based on medical criteria alone.

These issues strongly suggest that A&AQ is not an adequate instrument for the use to which it is being put in the Lithuanian disability and work capacity assessment process. Moreover, in light of the comparison between A&AQ and WHODAS in terms of the metric analysis, further points must be added to this list:

5. A&AQ does not target a range of levels of functioning that is appropriate for disability or work capacity assessment: most of the scores collected from the applicants who went through the pilot were in a five-unit range, from 20 to 25 points.
6. The reliability of the A&AQ, if adjusting for all local item dependencies that inflate the reliability estimate is not sufficient to consider this assessment tool fit for measurement (see Table 12).
7. The Rasch analysis shows that A&AQ is unrepairably multidimensional and of low reliability, which means that it does not assess one coherent construct – functioning or even 'activity and ability' – but several. This means that the summary score does not validly capture a single construct, in this case, functioning, that can be used to generate weighting coefficients.

Taking these seven points together, the conclusion is that A&AQ is not a suitable instrument for validly and reliably generating scores and related weighting coefficients¹³ for work capacity or disability assessment. We recommend that A&AQ be replaced by WHODAS.

¹³ Here we mean “statistically generating coefficients”, not coefficients generated by experts.

PART THREE: OPTIONS FOR INCLUDING FUNCTIONING

Introduction: Approaches and strategies for using WHODAS scores

As was shown above, the Rasch analysis of WHODAS based on pilot data shows that this instrument has strong measurement properties. Although the items in WHODAS tend to cluster by ICF domains, which results in some item dependencies, multidimensionality, and biased reliability estimates, this is not a problem since aggregating items by domains creates a perfectly sound metric. In short, and as the literature on the use of WHODAS in various contexts has repeatedly shown, WHODAS is a superior tool for measuring functioning and disability with high reliability and discrimination. For this reason, Rasch-transformed total scores will have interval scale properties and a reliable WHODAS score can be derived that not only is a valid assessment of the degree of disability but can be easily used for additional statistical analyses of individual or population-level disability data.

In this section of the Report, we analyze and discuss options for how WHODAS can be utilized in the Lithuanian context to replace A&AQ and more validly and reliably integrate functioning information into disability assessment and work capacity assessment. This Report has shown that WHODAS successfully collects functioning information, and based on the pilot data, it does so with strong psychometric properties of validity and reliability. But how can WHODAS scores be used in the Lithuanian context to improve disability and work capacity assessment?

What follows describes strategies for including a WHODAS-derived summary score for disability and work capacity assessments. Following similar analyses done in other countries, several methods were tested on the final pilot dataset. These can be grouped into three principal strategies (1) averaging the medical assessment score with the WHODAS score to arrive at a final work capacity or disability assessment score; (2) flagging persons above a certain WHODAS cut-off for additional assessment or other administrative response; and (3) as in the current approach with A&AQ, augmenting the medical assessment score by means of coefficients generated from WHODAS data:¹⁴

- (1) **Averaging** – averaging the basic medical assessment score and WHODAS score. Below we show the results of eight strategies (#3 - #10) that were tested using different weighting combinations. This approach is based on the theory that, together, medical and functioning scores contribute, to different degrees, to a realistic and valid assessment of disability or work capacity.
- (2) **Flagging** – identifying persons above a WHODAS cut-off and flagging these individuals to request from them additional information or reassessment, or otherwise altering the overall disability percentage to account for the reported level of functioning. Strategies #11 to #15 represent different flagging scenarios. The flagging approach is based on the assumption that medical information on its own distorts or otherwise misrepresents the true extent of the disability the individual experiences so that when an individual has a WHODAS score that is over some cut-off, this suggests that the medical score does not adequately capture the experience of disability and more information, or reassessment, is required.
- (3) **Augmenting** – As in the current use of A&AQ, the basic medical score can be altered (i.e. raised, lowered or kept the same) in terms of the WHODAS score by means of a score-based coefficient. (In this Report, it was decided only to lower this value.) Strategies #16 to #17 represent three potential coefficients that can be used for augmenting. This approach relies on the insight that at the core of disability and work capacity assessment is the medical

¹⁴ It is important to add that as WHODAS is used more and more data will be collected, and this data can be further analyzed using the techniques in this Report to continually uptake and recalibrate the various proposals that are suggested here. Moreover, these data have other potential policy applications, in identifying disability trends and planning for the future.

problem the individual experiences, but at the same time that experience is modified (to some extent) by environmental factors that need to be taken into account to augment or adjust the medical score.

Averaging, Flagging and Augmenting are three of a number of potential approaches to bringing together two scores that measure different phenomena but which, together, constitute our best assessment of disability or work capacity. These three are, arguably, the most intuitively obvious approaches to merging diverse assessments into a single overall assessment. Each is grounded in the ICF understanding of disability as the outcome of an interaction between the underlying health condition and impairments of a person and the physical, human-built, interpersonal, attitudinal, social, economic, and political environment in which the person lives and acts. They differ, however, in how they weigh the impact of the medical and environmental determinants of disability.

Table 13 gives an overview of the testing strategies that were considered. For comparison purposes Strategy #1 was included as the current situation in which the basic medical score is altered by coefficients based on the A&AQ scores and Strategy #2 is the case in which functioning is ignored and only the medical score is used.

The averaging strategies #3 to #10 aggregate medical score and the WHODAS score by giving WHODAS increasingly higher weight (25%, 50%, 75%, 100%) either by setting critical level of WHODAS at the median or at 40 (the justification for 40-cut-off is provided in the next section).

The flagging strategies are of two types. Strategies #11 to #13 include those who, in addition to receiving a positive disability assessment based on the medical assessment, add those with a WHODAS score above a cut-off, again the baseline score of 40, who have scored in the 3rd and 4th quantile of the WHODAS score generated by the WHODAS pilot survey data. Strategies #14 and #15 consider the distribution of the WHODAS score within an ICD disease category and flag additional persons based on their position within that disease category's specific score distribution, i.e., above the 3rd or the 4th quantile.

The augmenting strategies #16 and #17 are two strategies that diminish the medically assessed work capacity percentage by a coefficient < 1 if the WHODAS score is above a certain cut-off and indicates higher disability. (Intuitively, we are recognizing in this way that when WHODAS scores indicate high levels of disability that this score should readjust the medical score by a lowering coefficient.)

Table 13: Overview of WHODAS inclusion strategies

General Approach	Nbr.	Description of eligibility formula	Cut-off	Comment	Total eligible persons	Newly eligible	Potential exclusion
Actual approach	#1	Adjusted Basic Working Capacity	55% as cut-off	Actual strategy	1,889		
No approach	#2	Baseline: Work Capacity (100%)	55% as cut-off		1,873		
Averaging:	#3	Weighted mean of Work Capacity (75%) and WHODAS (25%)	Bivariate cut-off-line through 55% Work Capacity and the median of WHODAS		1,701	15	187
	#4	Weighted mean of Work Capacity (50%) and WHODAS (50%)			1,625	31	279
	#5	Weighted mean of Work Capacity (25%) and WHODAS (75%)			1,449	58	482
	#6	Weighted mean of Work Capacity (0%) and WHODAS (100%)			1,187	152	838
Averaging:	#7	Weighted mean of Work Capacity (75%) and WHODAS (25%)	Bivariate cut-off-line through 55% Work Capacity and the approximate normative cut-off (40 pts)		1,934	72	11
	#8	Weighted mean of Work Capacity (50%) and WHODAS (50%)			2,031	179	21
	#9	Weighted mean of Work Capacity (25%) and WHODAS (75%)			2,133	287	27
	#10	Weighted mean of Work Capacity (0%) and WHODAS (100%)			2,160	340	53
Flagging:	#11	Work Capacity as #1 or WHODAS above a cut-off	WHODAS > 40	Normative cut-off	2,213	340	0
	#12	Work Capacity score as #1 or WHODAS above a cut-off	WHODAS above 3 rd Q (>55)		2,025	152	0
	#13	Work Capacity score as #1 or WHODAS above a cut-off	WHODAS above 4 th Q (>60)		1,951	78	0
	#14	Work Capacity score as #1 or WHODAS above a cut-off	WHODAS above 3 rd Q by HC		1,990	125	0
	#15	Work Capacity score as #1 or WHODAS above a cut-off	WHODAS above 4 th Q by HC		1,921	56	0

Augmenting	#16	if Working Capacity > 55 AND if WHODAS from 40 to 4th Q THEN Working Capacity x 0.8 if WHODAS > 4th Q THEN Working Capacity x 0.6	55% as cut-off	2,055	182	0
	#17	if Working Capacity > 55 AND if WHODAS from 40 to 4th Q THEN Working Capacity x 0.6 if WHODAS > 4th Q THEN Working Capacity x 0.5	55% as cut-off	2,212	339	0

Source: WB team simulations.

Assessment options for using WHODAS to include functioning into disability determination process

Four options to include functioning into disability assessment in Lithuania were modeled and statistically tested using a variety of Averaging, Flagging, and Augmenting approaches and statistical strategies. Each option follows the ICF theory in as much as it combines the medical component of assessment¹⁵ with a functioning component, assessed by WHODAS. Option A is the situation in which WHODAS scores are taken into account in a purely discretionary manner. Options B, C, and D are based on statistically derived algorithms.

Each of these assessment options is described below, with advantages and disadvantages of each. Our framework for evaluating these options – based on the scientific literature – are key scientific principles that determine the credibility of any disability or work capacity assessment process: **validity** (the extent to which the option relies on a true assessment of disability); **reliability** (the ability of the option to arrive at the same assessment of the same case by different assessors); **transparency** (the degree to which the assessment process and outcomes can be described and understood by all stakeholders); and **standardization** (the extent to which the process resists distortion or alteration over time and across locations).

Option A: Discretionary combination of medical and functioning components

This is the option in which an individual or committee reviews medical scores and the WHODAS scores and makes a judgment about the extent of disability as the individual or committee sees fit. This is a purely discretionary option, and it is surprisingly common in practice. As an option for disability, it has the (minimal) advantage of simplicity, administrative convenience, and low cost. On the disadvantage side, however, this approach is subject to manipulation, or whim, totally lacks validity and reliability, and is utterly non-transparent. The option is given here in part as a contrast to the remaining options B, C, and D, but also, in fairness, because some countries continue to rely on this option for disability assessment. We do not recommend this option.¹⁶

¹⁵ As explained above, we have not reviewed the medical assessment tables used by DWCAO. The review would require a different testing approach, including a review of scientific research and evidence, and in particular an investigation into the methodology used to generate the percentage scores. We suggest that MOLSS and DWCAO could compare their medical assessment tables with similar tables used in other EU countries to see whether the percentage scores for health conditions are roughly similar across countries.

¹⁶ Anecdotal evidence suggests that medical professionals involved in the assessment of disability are convinced that they “know best” and are capable of taking into account functioning and the experience of disability as part of the medical description of the applicant's situation. One often hears medical assessors claim that they take functioning fully into account when examining medical records. One implicit result from the pilot is that this assumption is false.

Options B, C and D

The three remaining options all depend on statistically derived algorithms, which makes them very different from Option A. In different ways and for different reasons, each of the remaining options satisfy not only the basic psychometric properties of validity and reliability but each, to different degrees, strive to achieve transparency and standardization. The three options are based on extensive statistical testing performed, using the pilot data, of the Averaging, Flagging, and Augmenting approaches described above. Three preliminary technical points should be kept in mind:

1) The baseline used for all strategies tested – namely Strategy #2 – does not lead to the number of successful applicants under the current system but uses only the ICD health condition information to determine the work capacity percentage, with a cut-off at 55% for determining eligibility for benefits. Strategy #1 is the approach that was actually applied, which corrected for the A&AQ score. The baseline number of successful applicants of the 2,234 analyzed for the pilot was 1,873. It should be noted that this is a very high rate of success inasmuch 83.0 percent of applicants were assigned a work capacity percentage of 55.0 percent or lower.

2) In order to interpret the results of the statistically tested strategies, it is important to notice that the work capacity percentages and WHODAS scores are radically different: the work capacity score distribution is heavily skewed toward the lower end of the scale (this is reflected in the fact that 1,873 of 2,234 applicants were assigned a percentage of disability of 55.0 percent or less), with an average of 46.0 percent and the 4th quantile at 55.0 percent. By contrast, the score distribution of WHODAS is statistically normal, with a mean of 55.0 percent and a 3rd quantile at 60.0 percent (see Appendix 3). What this means is that it is reasonable to expect that the more reliance on WHODAS scores the final assessment is, the fewer applicants will be found eligible. (It should also be kept in mind that WHODAS will not only change the overall number of successful applicants but also will change who is successful and who is not: in some instances, WHODAS scores will raise the overall percentage based on the Work Capacity percentage, in other instances it will lower it.)

3) As noted above, we posit the standardized WHODAS score of 40 as the cut-off for 'significant disability' – that is a level of disability that warrants state intervention to support an individual. Scientifically speaking, it is essential to create a cut-off since there is no 'gold standard' for when disability is significant. Ultimately, the cut-off is a socio-political decision that should be transparent and evidence-based in the sense that it represents a plausible threshold based on an analysis of disability prevalence in a population. The score of 40 used in these analyses (and standardized by means of the Rasch Transformation Table 10) 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 underlying problem with medical assessment and options in this Report

There is another important issue that needs to be appreciated. In our view, the medically determined score used in Lithuanian is based on a Baremic system with all of the inherent problems associated with Baremic systems mentioned above: The essential psychometric properties of validity and reliability are either unknowable or demonstrably absent for all Baremic systems. This is because the asserted linkages or associations between whole person, disability percentages, and diagnostic categories found in these systems are not based on empirical evidence but are almost invariably established by the methodologically weak technique of unstructured professional consensus – several professionals coming to an agreement without empirical support. In some instances, even this minimal evidence-based is missing, and the linkages are purely speculative.

The upshot of this is that no modification of the current Lithuanian disability and work capacity assessment system will produce a thoroughly valid and reliable assessment, given the problems with the medical component. On the assumption that it is very unlikely that Lithuania will be in a position to change its medical assessment strategy into one that is scientifically more robust, the best tactic available for reform is to try to minimize the impact of, or partially correct for, the difficulties of the medical assessment. As we mention below, the averaging algorithm is, on this point, the most likely to be successful in this regard.

Option B: Using an averaging algorithm

Once again, since the basic work capacity percentages are heavily skewed in favor of 40-50.0 percent, it is inevitable that by directly averaging this score with the WHODAS score the number of applicants who are found disabled and eligible for benefits decreases, as WHODAS may show higher levels of functioning in the applicant that may compensate the work capacity percentage below the cut-off (e.g., up to 838 could be found work able with Strategy #6). At the same time, the composition of those assessed as disabled will change as well when WHODAS indicates that an individual who would have been assessed >55.0 percent in the current system but does not in fact experience incapacitating difficulties, and vice versa.

To get a full sense of the range of possible approaches under Option B, four weighting schemes were tested when creating the 8 strategies:

- 75.0 percent basic work capacity percentage & 25.0 WHODAS score
- 50.0 percent basic work capacity percentage & 50.0 percent WHODAS score
- 25.0 percent basic work capacity percentage & 75.0 percent WHODAS score
- 0.0 percent basic work capacity percentage & 100.0 percent WHODAS score

There are, of course, many approaches to weighting that might be adopted (and any other arrangement can be constructed, and its consequences determined using the same methods as used for these four), but these four are perhaps the most intuitive. As there is little scientific literature or international consensus on where the cut-off point in WHODAS scores lies for 'significant disability', each of the four strategies (with the baseline strategy excluded) use these two cut-offs:

- work capacity cutoff (55.0 percent) and median of WHODAS score (55 points)
- work capacity cutoff (55.0 percent) and WHODAS score of 40 (as recommended in Andrews et al, 2009 & Yen & al. 2017)

Advantages of Option B:

- An assessment of the level of functioning plays a significant role in the determination of eligibility for disability benefits so that the eligibility for benefits is not solely based on purely medical criteria, and in particular on the crude basic work percentages that are not based on empirical evidence; this option avoids this.
- The averaging approach minimalizes the impact of the inherent psychometric problems with the basic work capacity percentage based on the Baremic medical assessment instrument used.
- The assessment of the level of functioning is empirically and statistically verified.
- This option yields high levels of validity and reliability.
- Merging the results of two assessments scaled by means of 'weighted averaging' is fully objective, transparent, and non-discretionary.
- The method is not sample-dependent.

Disadvantages of Option B:

- There are, potentially, an infinite number of combinations of weighting schemes (i.e., ‘strategies’), each of which generates a different set of eligible applicants and has different budgetary and political consequences. This is an unavoidable fact about the nature of disability as a continuum and the fact that there is no scientifically verified or objectively determined cut-off of severity for eligibility.
- Any strategy selected will be objectionable to individuals who, under that strategy, will not be eligible. This signals the need for clear and transparent information dissemination and a solid grievance redress system that may include using tools for clinical testing and determination of functioning, such as ClinFIT20,¹⁷ or other tools used or recommended by rehabilitation specialists. It should also be noted that any new method adopted by DWCAO will apply to new applicants only. To smooth the transition, disability recertification may be staged over several years and/ or be conducted for the new cohort only.

Option C: Using the flagging algorithm

There are two types of flagging strategies: Strategies #11 to #13 use the basic work capacity score but identify or flag those individuals with WHODAS scores above 40.0 (#11) or those with WHODAS scores above the 3rd and the 4th quantile (#12, #13). (As mentioned, the cut-off of 40 is the only level suggested in the literature for ‘significant’ disability.) Strategies #14 and #15 similarly use the basic work capacity threshold for eligibility and flag, by health condition, the additional individuals who, for that health condition, have a WHODAS score above the 3rd or 4th quantile. In effect, this approach uses health condition-specific cut-offs rather than the single cut-off of 40. The rationale for this second, more complex algorithm is that the impact of health conditions on peoples’ day-to-day life (i.e., the actual disability they experience) intuitively varies, and it is important to contextualize the WHODAS score to capture this fact.

Table 13 shows that setting the cut-off for the WHODAS score at 40 and flagging those individuals with a Work Capacity < 55.0 percent (i.e., strategy #11) would result in a large number of newly disabled individuals (N = 340). On the other hand, increasing the cut-off to the 3rd quantile (WHODAS score 55) as in strategy #12 or even further to the 4th quantile (WHODAS score of 60) as in strategy #13 would reduce the number of newly eligible individuals drastically, first to 152 and then to 78. Further refining the flagging approach by using the quantile found within health condition types would include slightly less individuals – 125 and 56 for strategies #14 and #15, respectfully. As with the averaging approach, the correlation among the groups of individuals who become eligible in the different strategies is high.

Advantages of Option C:

- Scientifically robust and based on actual data.
- Matches the basic intuition that the purely medical approach may miss individuals who, as reported in the WHODAS score, are experiencing more functioning problems in their lives than the health condition they have suggests they do.
- High levels of validity and reliability.

¹⁷ ClinFIT20 is the official disability assessment tool of the International Society of Physical and Rehabilitation Medicine (ISPRM) that is currently being used in China and Japan, with other countries expressing interest in adopting it. See ClinFIT: ISPRM’s Universal Functioning Information Tool based on the WHO’s ICF, available at: <http://www.jisprm.org> on Friday, July 12, 2019, IP: 62.98.194.95.

Disadvantages of Option C:

- This option assumed that the WHODAS score should never lower the score of an individual who, based on the basic work capacity score alone, was assigned a percentage of disability < 55.0 and thus qualified for disability benefits. It was inevitable that the only possible impact of the WHODAS score was to increase the number of successful applicants. **Hence, this option defeats the objective of integrating functioning into disability assessment.**
- There is no scientific or statistical way to determine which approach is better, but inevitably (and unavoidably) depends on a socio-political decision informed by economic considerations.
- The flagging approach is vulnerable to political manipulation as the criteria for determining which individuals to 'flag' is discretionary.
- For the purpose of integrating functioning into disability assessment, the option is not appropriate as it defeats the objective of integrating functioning into disability assessment across the board (to all applicants). For this reason, we do not recommend this option.

Option D: Using the augmenting algorithm

The augmenting approach (represented by strategies #16 and #17) reproduces an approach that is used not only in Lithuania currently but in many European countries (Germany, France, England, Switzerland, and others), namely modifying the score assigned by a disability assessment committee by means of a coefficient (here < 1) that represents the additional functioning information captured by the WHODAS score. The underlying intuition behind this approach, and presumably the motivation in Lithuania, is to avoid relying entirely on a medical determination of disability, especially when such an approach undervalues the actual impact of health conditions on a person's life and functioning performance.

Two strategies for using the augmenting approach are presented here (there are, in theory, many other possibilities). The strategies use the normative score of 40 and the 4th quantile score of 60 as cut-off values to multiply the basic work capacity by coefficients of 0.8 and 0.6, respectively, when the WHODAS score exceeded these values. As can be seen in Table 13 above, the outcome of strategy #16 is very close to the outcome of strategy #12, which also uses the 3rd WHODAS pilot sample quantile as cut-off, and similarly the outcome of strategy #17, with stronger coefficients, correlates highly with strategy #11 which flagged persons with a WHODAS score above 40. Strategy #16 adds 182 individuals, while strategy #17 adds 339.

Advantages of Option D:

- Using a coefficient value generated statistically is a common tactic used widely and is familiar in the Lithuanian context as well, so the change will not be overly disruptive.
- A coefficient approach (increasing the medically-determined disability percentage in light of functioning scores) is the most intuitive way to combine the scores of very different assessments – medical and functioning – into a single score.
- This option incorporates the insight that a medical determination alone can often miss instances where people actually have moderate to high disability needs.
- This option, because of the psychometric properties of WHODAS, has high levels of validity and reliability, but only for a relatively small number of applicants.

Disadvantages of Option D:

- As with Option C, D assumes that the WHODAS score will never improve the score of an individual who, based on the basic work capacity alone, would be assigned a percentage of disability < 55.0 percent and so qualify for disability benefits. (Since the pilot shows that currently, more than 80.0

percent of applicants have a basic work capacity that is < 55.0 percent, it was inevitable that the only possible impact of the WHODAS score would be to increase the number of successful applicants. ***So, arguably, this option defeats the objective of fully integrating functioning into disability assessment, whatever the resulting consequences.***

- This approach does not sufficiently lessen the impact of the Baremic approach to determine basic work capacity percentages.
- As with Option C, there are many possible variations of this approach with different outcomes – in this Report only two were tested. Although the coefficient approach itself is intuitively understandable and can be made transparent to the public, the scientific and statistical justification for Option D is somewhat technical and may not be easily understandable by the lay public.

Examples of the inclusion strategies in practice

The options presented above may seem too abstract. To make them more concrete, four individual cases based on data are described below, with WHODAS and (for comparison with the current baseline) A&AQ scores, so that the outcome in terms of eligibility of these options can be shown for these individuals.

A: is a 56-year-old married woman with moderate bipolar disorder and an underlying heart condition. She reports 12 years of education, no professional qualifications and used to work as an employee of a printing company. She is unemployed for health reasons at the time of the assessment. Her basic work capacity was 63.0 percent; however, the disability assessment with WHODAS showed a score of 73, which corresponds to a very high level of disability, largely above the average. Her A&AQ score of 22 supports a level of disability at the population average.

B: is a married 38.5-year-old man with a severe eye disease that causes reduced visual functions in both eyes which can, however, be corrected for higher acuity. His basic work capacity is estimated at 30.0 percent. He went through higher education and is employed at the time of the assessment, working as a computer specialist. The A&AQ score of 14 and the WHODAS score of 18 would support very low levels of disability.

C: is a 60-year-old divorced man with disabling back problems. He reports 11 years of education up to the secondary level and worked as a driver. Presently, he is unemployed. His basic work capacity is 70.0 percent due to moderate movement restrictions. However, both the A&AQ score of 37 as well as the WHODAS score of 73 indicate high functioning problems.

D: is an 18-year-old man with a disease of the nervous system in the form of a benign epileptic syndrome without cognitive or personality disorders. He reports low levels of disability with an A&AQ score of 12 and a WHODAS score of 28. He has secondary level education and no profession. His basic work capacity is estimated to be 50.0 percent.

How would these four individuals be assessed with the combined based work capacity percentage and the (Rasch-adjusted) WHODAS score in the seventeen strategies (including the baseline strategies of pure basic working capacity #1, and baseline adjusted with A&AQ score #2)? The results for each of the averaging strategies (green = eligible for benefits; red = not eligible for benefits) are shown in Table 14.

Integration strategies - Examples of individual cases

Table 14: Work capacity and WHODAS scores and their integration strategies -
Examples of individual cases

	Work	WHODAS	Actual Approach										Flagging					Augmenting	
			#1	#2	#3	#4	#5	#6	#7	#8	#9	#10	#11	#12	#13	#14	#15	#16	#17
A	63%	73	Red	Red	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
B	25%	18	Green	Green	Green	Red	Red	Red	Green	Green	Red	Red	Green	Green	Green	Green	Green	Green	Green
C	70%	73	Red	Red	Red	Green	Green	Green	Red	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
D	50%	28	Green	Green	Red	Red	Red	Red	Green	Red	Red	Red	Green	Green	Green	Green	Green	Green	Green

The four cases that have been selected are extreme examples but illustrate well the impact of these Options. In principle, one would expect that low work capacity goes along with high disability, i.e., high WHODAS scores, and contrarily that high work capacity goes along with low functioning scores. The cases all have incongruent scores on the medically based work capacity and the functioning score assessed with WHODAS. The data also presented a negative correlation between the two measurements but with a small coefficient ($r = -0.23$).

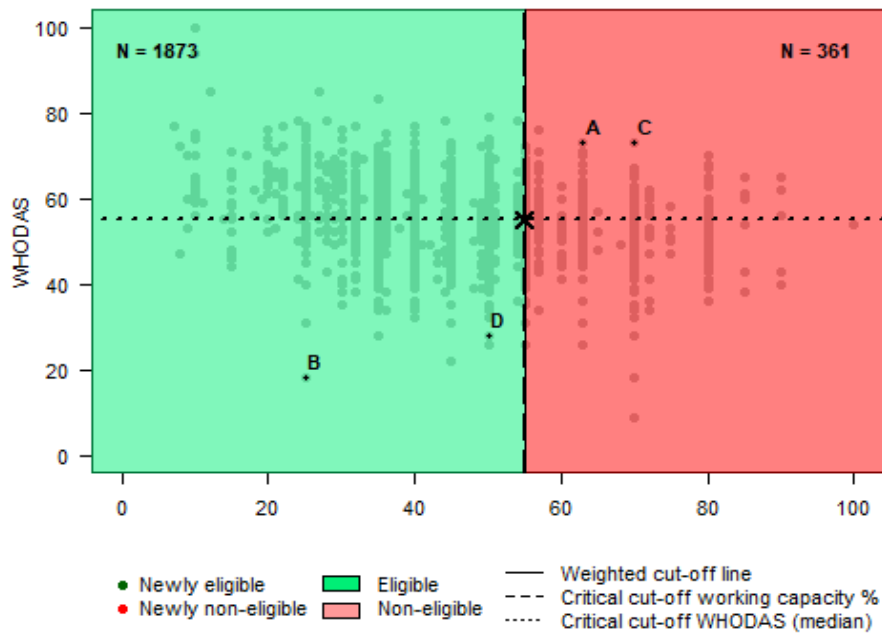
Graphical representation of the overall impact of the averaging strategy

What follows will illustrate graphically how the averaging options function with five relative weightings of the work capacity score and the WHODAS score – weighing basic work capacity at 100%, WHODAS at 0%; weighing basic work capacity at 75%, WHODAS at 20%; and so on.

The averaging approach can be easily depicted by the mean of a cartesian coordinate system with the work capacity score on the x-axis and the WHODAS score on the y-axis. The weighted cut-off-line separates between eligible and non-eligible individuals. Like a clock hand, the separation line moves with increasing weight of the WHODAS for individuals who are either newly included for or newly excluded individuals from disability benefits. The coordinate system approach can be easily implemented in practice to actually 'locate' specific individuals on the graph, based on their working capacity assessment and WHODAS scores. This makes it possible, at a glance, to see if an individual is on the line and not clearly in any group, so that the person can be allocated to one or the other group. For concreteness as well, the four described individuals, A, B, C, and D, are located in each graph.

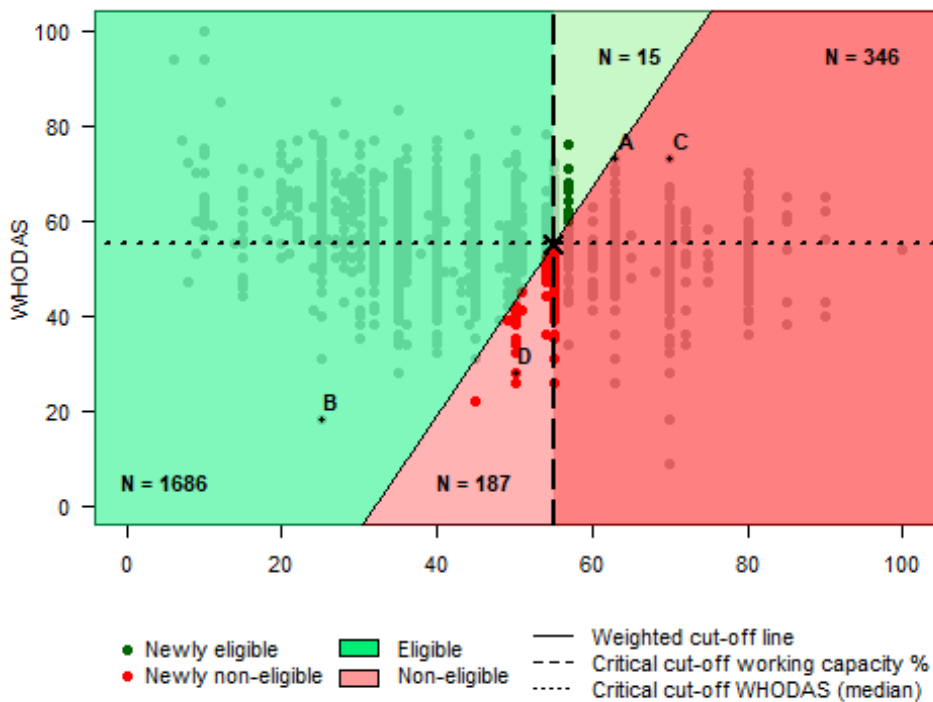
Starting with Strategy #2 (Figure 7), in which only the basic work capacity score is considered, the cartesian field is divided vertically at a cut-off of 55%, with eligible individuals on the left side and non-eligible individuals, with a higher work capacity percentage, on the right side:

Figure 7: Strategy #2: (Basic work capacity 100% and WHODAS 0%)



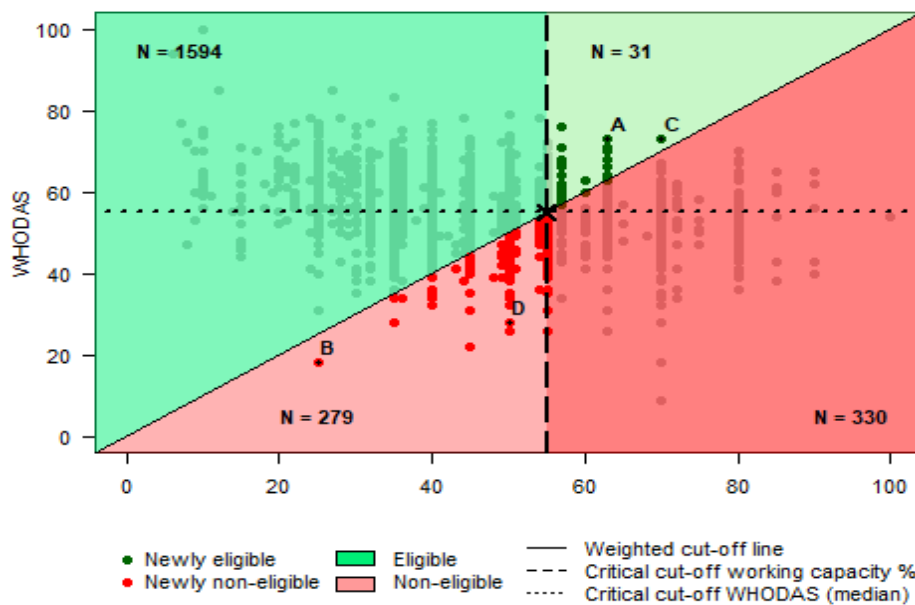
Without any adjustment to the baseline assessment and with the actual approach that aims to adjust the basic work capacity by means of the A&AQ information, the cases A and C would not be eligible. A has an estimated basic work capacity of 63.0 percent and C of 70.0 percent, hence neither of them was granted disability status. Cases B and D on the other hand, have a basic work capacity below 55.0 percent, with 25.0 percent and 50.0 percent respectively and were granted disability status.

Figure 8: STRATEGY #3 (Basic work capacity 75% and WHODAS 25%) with WHODAS cut-off at the median score



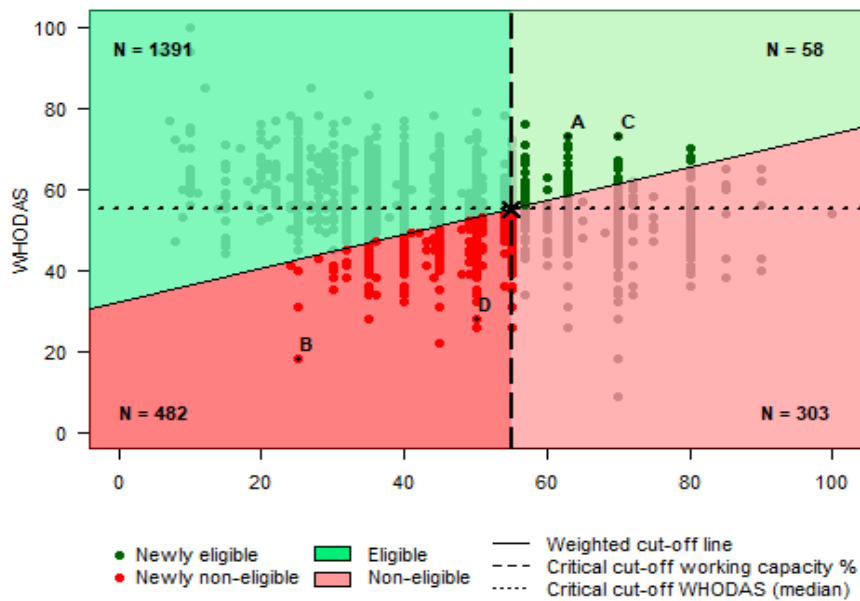
In the averaging strategies #3 to #6, the cut-off of the WHODAS is set at the median (55 points). In strategy #3, WHODAS contributes 25.0 percent to the basic work capacity percentage, this would change the disability status of case D. D is an 18-year-old man with a benign epileptic syndrome without cognitive disorders. He reports very little disability, as shown by the small WHODAS score, and his basic work capacity of 50.0 percent is just below the cut-off. With the inclusion of the WHODAS score and the described functioning level, D would not be justified for disability status. In fact, a total of $N = 187$ (8.4 percent) individuals would become not eligible if entering 25.0 percent functioning information in the process. A total of $N = 15$ individuals, on the other hand, may be now retained for disability status when taking into account their high levels of disability assessed with the WHODAS.

Figure 9: STRATEGY #4: (Basic working capacity 50% and WHODAS 50%) with WHODAS cut-off at median score



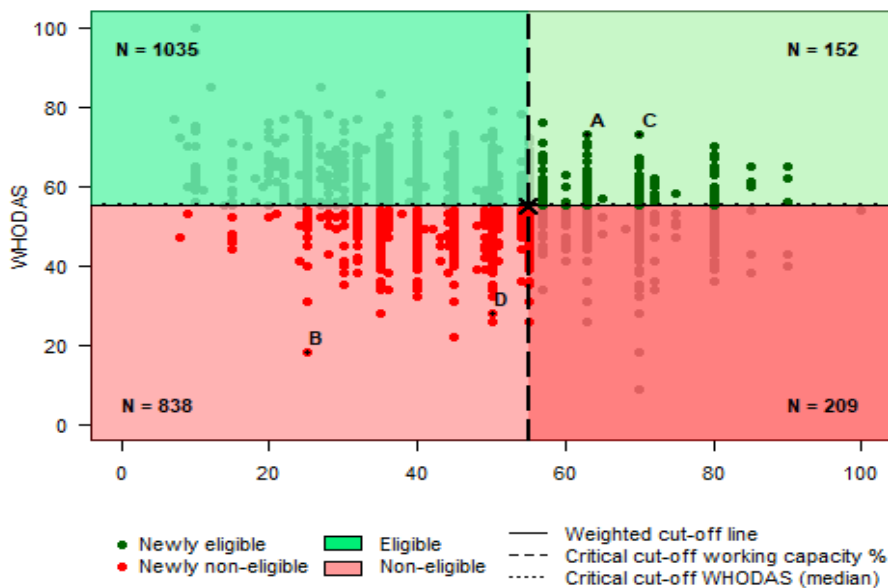
In strategy #4, WHODAS contributes 50.0 percent to the basic work capacity percentage. This would change the disability status of all four cases, which showed opposing scores on the work capacity assessment and functioning assessment based on WHODAS. A total of $N = 279$ (12.5 percent) individuals would not be considered eligible anymore, as their disability levels are too low with respect to their work percentages. A total of $N = 31$ (1.4 percent) would become eligible. All these individuals had work capacity scores above 55 but with high levels of disability based on WHODAS.

Figure 10: STRATEGY #5: (Basic working capacity 25% and WHODAS 75%) with WHODAS cut-off at median score



The Strategy #5, as well as Strategy #6, are here for illustration. This represents an approach where the functioning assessment would overweight the medical assessment. A total of $N = 482$ (21.6 percent) individuals would not be considered eligible anymore, as their disability levels are too low with respect to their work percentages. A total of $N = 58$ (2.6 percent), on the other hand, may now become eligible. WHODAS contributes 75.0 percent to the basic work capacity percentage. It should be said that in practice this is highly unlikely to happen since it suggests that a person has a high level of problems in functioning that cannot be explained in terms of his or her underlying health problems.

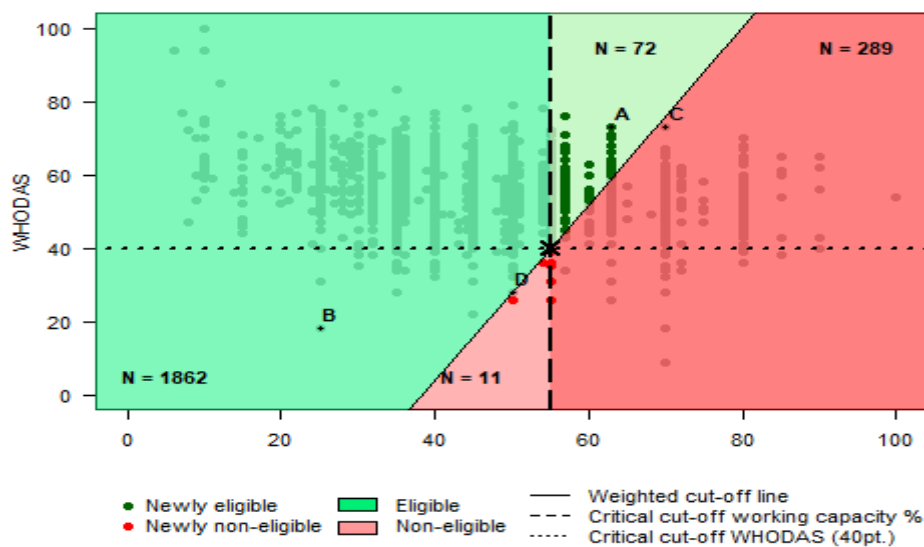
Figure 11: STRATEGY #6: (Basic working capacity 0% and WHODAS 100%) with WHODAS cut-off at median score



Looking at the three strategies of #4, #5, #6 (Figures 9, 10, and 11) together, we see that increasing the weight of the WHODAS to 50% or more would reverse the situation, so that cases A and C, despite

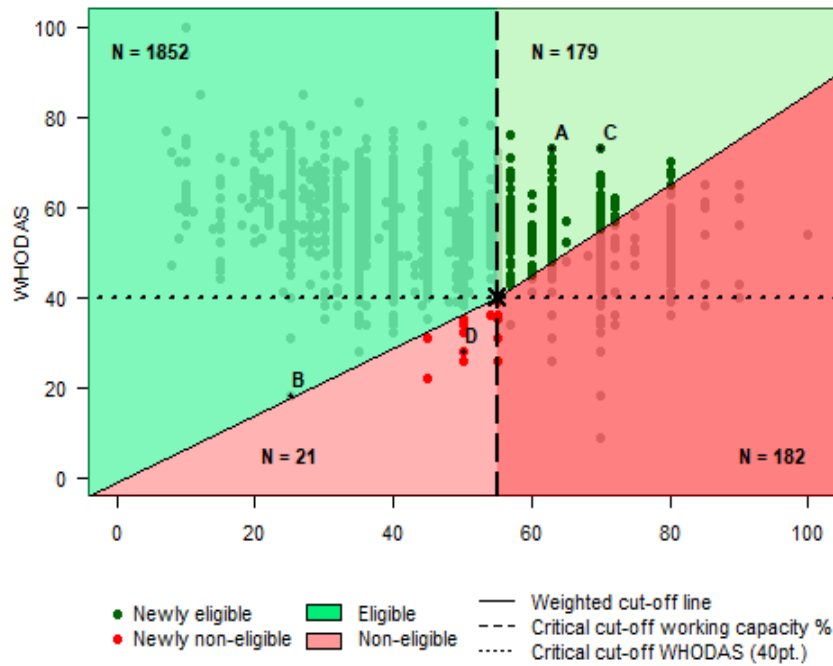
their work capacity above the 55.0 percent cut-off, would both become eligible for disability. Cases A and C present very high functioning scores, above the 4th quantile of the population. On the other hand, B and D, of younger ages and with functioning scores below the first quantile would not be eligible anymore for disability status. Case B, with a work capacity of 25.0 percent has a severe eye disease and works as a computer specialist. The WHODAS does not specifically assess visual acuity so that his loss of capacity cannot be assessed directly or only through limitations in his participation. However, he endorses only 2 WHODAS items, the time spend on the health condition and relaxation, as representing a moderate problem. A strategy like in #6 which would be based 100.0 percent on the WHODAS, would possibly not capture the full impact of his disability when working as a computer specialist.

Figure 12: STRATEGY #7 (Basic working capacity 75% and WHODAS 25%)with WHODAS cut-off at 40 points score



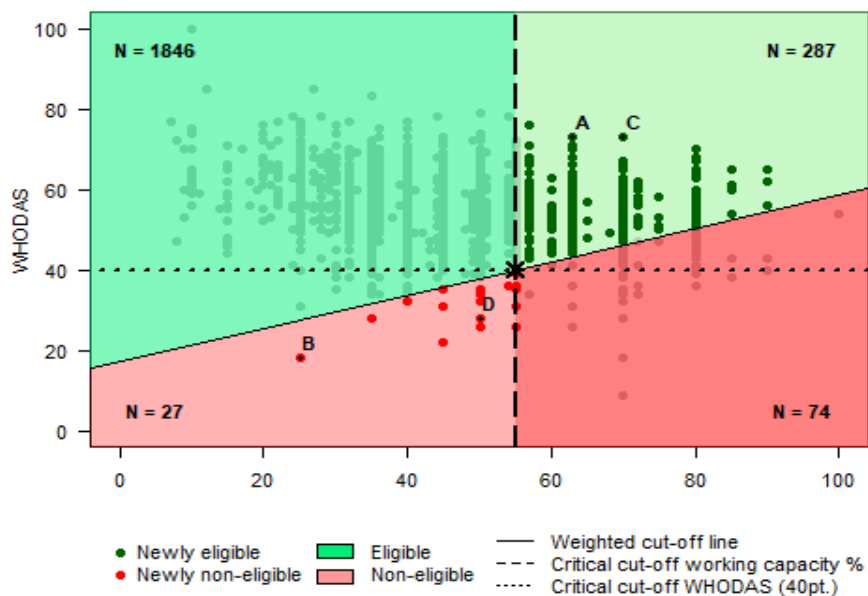
For Strategy #7, when lowering the cut-off of the WHODAS score to 40, as for averaging strategies #7 to #10, the inclusion of the 4 cases would change again. Unlike strategy #3, the lower cut-off would benefit now case A, which would be eligible with the 25.0 percent contribution of the WHODAS score, on the other hand, this would not be enough weight to include the case C, despite very high functioning problems. Case D is straddling between eligibility and non-eligibility and may require an external viewpoint to decide. By lowering the cut-off of the WHODAS to 40, the functioning will be expected lower to consider a person for non-eligible. At the same time, eligibility becomes easier than with Strategies #2 to #6. For the first time, the number of newly eligible individuals ($N = 72$, 3.2 percent) exceed the newly non-eligible individuals ($N = 11$, < 1.0 percent).

Figure 13: STRATEGY #8 (Basic working capacity 50% and WHODAS 50%) with WHODAS cut-off at 40 pts score



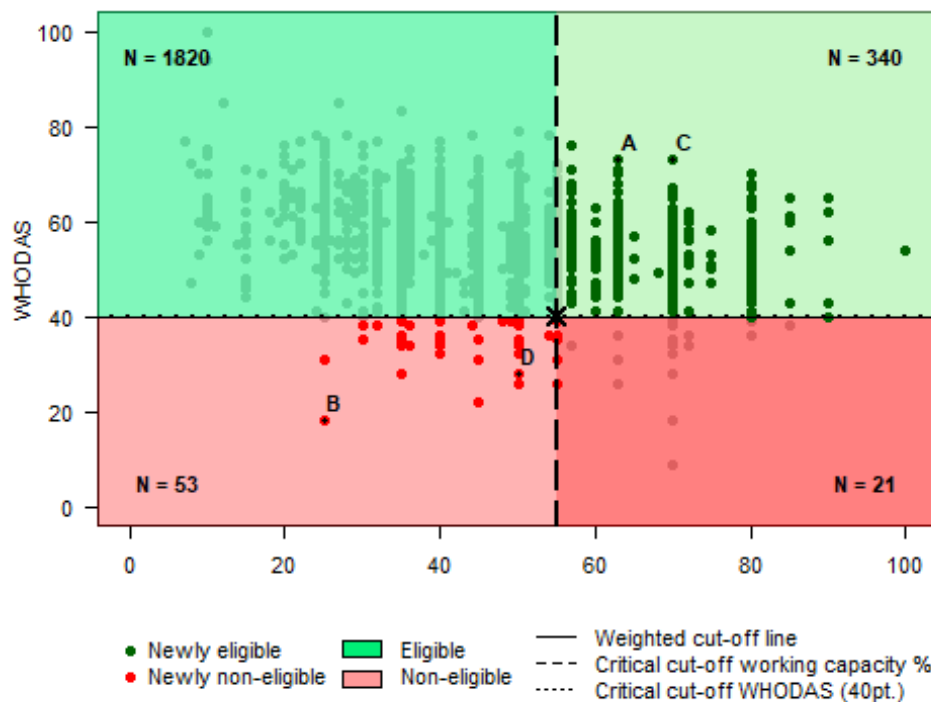
With Strategy #8 the number of newly eligible individuals represents 8% (N = 179). Only N = 21 (< 1.0 percent) would not be eligible anymore given their low functioning levels. Here as previously, only case D, the 18-year-old man the benign epileptic syndrome, may be “penalized” by the inclusion of the functioning information. Case B, still eligible in Strategy #8 would have been excluded with a critical WHODAS score level at the median and the same general weight of the functioning information.

Figure 14: STRATEGY #9 (Basic working capacity 25% and WHODAS 75%) with WHODAS cut-off at 40 pts score



With strategy #9 and 75.0 percent contribution of the WHODAS, case B, the IT specialist with severe eye problems, would not remain eligible for benefits. A total of $N = 287$ (12.8 percent) of the assessed population may now gain eligibility by adding the functioning information, and only a marginal proportion would lose their disability status ($N = 27$, 1.2 percent). Strategies #8 and #9, with a WHODAS contribution of 75.0 percent or more, i.e., when the functioning overweight the medical perspective, would again reverse the starting situation.

Figure 15: STRATEGY #10 (Basic working capacity 0% and WHODAS 100%) with WHODAS cut-off at 40 pts score)



Strategy #10, similarly to Strategy #6 will only use the functioning information assessed through WHODAS. By lowering the critical cut-off, lower disability levels would be retained for disability status. A total of $N = 340$ individuals (15.2 percent) would then be considered as having a significant disability level. This situation, of course, finalizes the weighting scheme but is not an earnest alternative, as WHODAS would possibly not capture the full impact of the disability.

We have seen that the impact of the averaging strategy on the final determination of disability can be easily visualized. Positioning individuals in a cartesian coordinate system, with the help of the cut-off lines and colored fields, makes it possible to immediately spot where a score combination lies in terms of eligibility for disability status.

Visualization of the flagging strategy #11 to #13 in a cartesian system would be possible but would show us far less as it would just further cut horizontally the right (red) part (> 55) of the coordinate system, so that values above a certain WHODAS score would be flagged and reconsidered. Flagging strategies #14 to #15 that use health condition-specific cut-off to flag outlying cases would only work in health condition-specific coordinate systems. Visualizations of the augmenting strategies visualizations would not clearly represent the changes in eligibility.

Although the visualization of the flagging strategy (#11 to #15) and augmenting strategies (#16 to #17) are not useful, it is clear that all four individuals – A, B, C, and D – would be eligible for all of as these strategies do not modify the basic work capacity percentage but instead add persons with functioning

scores above a certain level. As all four selected individuals' cases show either a very high WHODAS scores (A&C), or a very low work capacity percentage (B & D), their eligibility is guaranteed, although not on the basis of the basic work capacity score alone. (At the same time, arguably, it would be advisable to reconsider individual A under the flagging strategy: individual A has an estimated basic work capacity of 63.0 percent and a very high WHODAS score, i.e., many functioning problems. This combination raises the question whether A has a mental condition, such as bipolar disorder syndrome, or other factor, that is the cause of the functioning problems.)

CONCLUSION AND RECOMMENDATIONS

This Report presents the results of analyses, based on data from the pilot, i) to assess the performance of WHO's Disability Assessment Schedule (WHODAS 2.0), in its 36-question, interviewer-conducted format; ii) to compare WHODAS with the currently used work capacity assessment tool – the Questionnaire of the Individual's Activity and Ability to Participate (A&AQ) – and to assess its performance; and iii) to present, and evaluate, a range of options for using WHODAS instrument and its scoring metrics to augment or refine the current medical determination of disability and work capacity assessment.

In this final part of the Report, we present summary conclusions from previous sections and, on the basis of these conclusions, recommendations to achieve the aim of Output III of this project, namely how functioning information and population-based metrics can best be used to augment or refine the medical determination of disability status to satisfy the overall outcome of this project to improve the assessment of disability in Lithuania. *In other words, these recommendations aim to create an assessment system for disability/work capacity that assesses disability as 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, in light of the person's state of health.*

Instruments to assess functioning

Conclusions about the Activity and Ability Questionnaire as a functioning assessment instrument

Our conclusions about the currently used A&AQ instrument are based on a content and structure comparison with the WHODAS question, as well as the detailed Rasch-based analysis.

While we have found that the 'activity and ability' construct the A&AQ assesses is at least analogous to the ICF notion of 'functioning', it cannot be said that it fully aligns with ICF since some A&AQ items are too vague or not relevant to functioning-relevant items at all. More importantly, as a quantitative instrument that is objective, valid and reliable and therefore suitable for functioning assessment, A&AQ is fundamentally inadequate for several reasons:

- The A&AQ relies on nominal response options that are then mapped onto an ordinal scale (0-4). For assessing levels of work capacity or disability more broadly, this arbitrariness is problematic as there is no empirical justification for these ordinal rankings.
- The link between the summary scores and coefficient scores that A&AQ generates is completely arbitrary and without any empirical basis. The result, as confirmed by empirical evidence, is that the A&AQ assessment of functioning has only minimal impact on the resulting assessment of work capacity based on medical criteria alone.
- A&AQ does not target a range of levels of functioning that is appropriate for disability or work capacity assessment: most of the scores collected from the applicants who went through the pilot were in a five-unit range, from 20 to 25 points.
- The reliability of the A&AQ, if adjusting for all local item dependencies that inflate the reliability estimate is not sufficient to consider this assessment tool fit for measurement.

- The Rasch analysis shows that A&AQ is irreparably multidimensional and of low reliability, which means that it does not assess one coherent construct – functioning or even 'activity and ability' – but several. This means that the summary score does not validly capture a single construct, in this case, functioning, that can be used to generate weighting coefficients.

Conclusions about the WHODAS as a functioning assessment instrument

With respect to alignment with the ICF, as has been mentioned, WHO developed WHODAS explicitly to statistically capture the construct of disability 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.

Moreover, on the basis of evidence from the scientific literature on multiple applications and use-cases for WHODAS, as well as the in-depth analysis of the measurement properties of the WHODAS carried out in the pilot and reported here, we are confident that WHODAS information is sufficiently robust and relevant to augment the disability percentage score by health condition assigned by medical assessment in order to enhance the accuracy and validity of the disability and work capacity assessment process in Lithuania.

Recommendations

The objective of the WHODAS pilot has been to recommend empirically based options to strengthen the inclusion of functioning into disability assessment in Lithuania. In the light of the empirical analysis presented above, the following is recommended concerning the instruments used to assess disability in Lithuania:

Recommendation 1: Replace the currently used A&AQ with WHODAS-36:

The WHODAS questionnaire, in its 36-item, clinically administered format should replace the currently used Questionnaire of the Individual's Activity and Ability to Participate (A&AQ) for disability/work capacity assessment in adults in Lithuania.

Recommendation 2: Review and update the medical instrument and the Barrême table:

The assessment of disability combines medical information and functioning information. While our project did not include a review of the medical instrument/ Baremic table with health conditions/ impairments and assigned percentages of disability/ work (in)capacity used currently in disability assessment in Lithuania, given advances in medical science, practice and technology, a periodic review and adjustment in the Baremic table is highly advisable. We thus recommend:

Efforts should be made that the medical instrument used to determine disability and the basic work capacity score is reviewed and updated on the basis of the best medical knowledge and experience of other countries and is fully aligned with WHO's International Classification of Diseases, ICD-11. This would require a close collaboration with the Ministry of Health.

Alternatively, MSSL in collaboration with the Ministry of Health, may consider piloting ClinFIT, as initially proposed by the World Bank team, with a view using this information to replace medical information and scoring based on the Barrême table.)

Integrating functioning information into disability assessment

In the above sections of this Report, we have analyzed and presented four options for the combined application of the current medical assessment and WHODAS for disability assessment and work capacity assessment in Lithuania:

- Option A: Discretionary combination of medical and functioning components
- Option B: Using an averaging algorithm
- Option C: Using the flagging algorithm
- Option D: Using the augmenting algorithm

For each of these four options we have discussed advantages and disadvantages and have presented several integration scenarios for each one of them, as well as each scenarios' impact on the number of persons assessed as having a disability (relative to the baseline).

Overall, we have concluded that Option B: Using an averaging algorithm, performs the best with regards to the objective of fully integrating functioning into disability assessment.

We thus recommend

Recommendation 3:

Adopt the “averaging” method (Option B below) for integrating functioning into disability assessment.

The government of Lithuania should determine which scenario is the most appropriate given political, financial and other relevant considerations. “Averaging” or Option B below gives the government of Lithuania considerable flexibility in how it wants to shape its reform of the disability assessment. Option B is a weighting algorithm that has two endpoints: giving medical assessment 100% weight and WHODAS score 0% and the opposite, giving WHODAS score 100% weight and the medical assessment 0%. There are, of course, many intermediate weighting options. Which is chosen will generate different patterns of successful or unsuccessful disability status – and examples of these patterns are presented graphically below in the Report.

The analysis that is presented in this Report shows persuasively that it might be possible to eliminate the medical assessment component of disability assessment entirely and use WHODAS exclusively and still maintain a valid and reliable disability assessment process. We know of no country that has taken this option, and for political and historical reasons it might be challenging to do so. It must be said that are good reasons to continue to use health information in some manner for disability assessment. Nonetheless, Lithuania would be on scientifically sound ground to take the step to move towards a complete functioning-based disability assessment procedure

In any event, we recommend that, first, an executive decision is made about what the relative weights of the medical assessment and WHODAS scores will be instituted, and then, secondly use this algorithm to determine disability status over a period of time, during which the patterns of disability status can be monitored. If the chosen algorithm produces the outcomes desired – specifically an acceptable, and financially feasible, percentage of applicants who are assessed across the three levels of disability status – then that algorithm can be continued; if the outcomes are not acceptable, the algorithm can be adjusted accordingly. To be practical, we recommend that the weighting starts with 50% and in two-three years moves to 75% functioning-based score and 25% medical based score.

It is also important to note that continued use of WHODAS in the disability assessment process will produce a stream of data that can be used to update the analysis provided in the Report, giving the Ministry vital evidence of trends and disability patterns. For example, as the Covid-19 pandemic is

likely that countries will experience so-called 'long Covid' as a chronic health problem with potential disabling consequences. In order to predict future health and social support requirements for this population, the accumulated data from the application of the WHODAS can be used.

Related to the Recommendation 3, there are three more recommendations:

3.1 DWCAO should establish Statistics and Research (S&R) Unit that would conduct the analyses mentioned above.

3.2 Develop Capacity: Should this unit be established prior to the Project closing; the World Bank Team will train the staff in the relevant statistical analysis techniques.

3.3 If the decision is made to switch to WHODAS and adopt the averaging method for weighting the functioning and medical scores, needed adjustments in the DWCAO IT system should be made, including the development of the statistical algorithm for averaging, training of staff and the deployment of the new method.

The following considerations are relevant as well:

- While empirically based and more objective, the averaging approach is not very different from the coefficient approach currently in use: changing/ adjusting the medically determined disability percentage in light of WHODAS functioning scores. Hence, the transition to a new assessment algorithm will not be overly disruptive.
- This option incorporates the insight that a medical determination alone can (i) miss instances where people actually have moderate to high disability needs; (ii) overestimate the impact of the health condition when people actually have mild to low disability needs.
- This option, because of the psychometric properties of WHODAS, has high levels of validity and reliability.
- The option gives the government of Lithuania considerable flexibility by offering a range of scenarios (see above) with predictable eligibility outcomes, given existing applicant trends. Moreover, the chosen scenario can be altered in light of the collection of statistical information as the new assessment algorithm is implemented reflecting changing trends.
- It should also be noted that any new method adopted by DWCAO should apply only to new applicants for disability assessment. To smooth the transition, disability recertification may be staged over several years.
- The transition to a new questionnaire should technically be relatively easy in terms of the software adjustments. Other reflections and recommendations on the adjustments to the administrative processes will follow once the decision on the choice of the scenario is made.

Looking Ahead

Reforming disability system and policies is sensitive and complex process that requires in depth research and piloting of options, which takes planning, time, resources, and persistent effort of policy makers and practitioners and other stakeholders. Broadly speaking, it includes two key components: (i) disability policies and system, including disability and work capacity and disability needs assessments for adults; and (ii) disability policies and system, including disability and disability needs assessments for children. Below, we propose a roadmap for the reform and further development of the disability system and policy in Lithuania, in line with the modern understanding of disability and commitments under the United Nation Convention on the Rights of persons with Disabilities to which Lithuania is a state party.

Phase One, short term (next six months): Disability policy and system, including disability and work capacity assessment for adults in Lithuania. Under the DG Reform project implemented by the World Bank, two major analytical reports were prepared:

- (1) *Disability Policy and Disability Assessment System in Lithuania* (May 2020). This report provides results of an in-depth analysis and assessment of the disability system and policies in Lithuania as they pertain to adults. The report offers a range of recommendations pertaining to the Lithuanian disability policy and system, including, but not limited to, programs (benefits) to support adults with disabilities, labor market inclusion of persons with disabilities, policy and programs implementation arrangements and disability and work capacity assessment system as implemented by DWCAO, including DWCAO's management information system and a list of priority actions to improve and bring it up to date, and
- (2) This Report: *Lithuania, options for including functioning into disability and work capacity assessment*, which provides empirically based recommendations on including functioning into disability and work capacity assessment in adults.

Recommendations in both reports are focused on improving efficiency and effectiveness of disability policy and system and further developing it, while improving the quality of services provided to adults with disabilities and their well-being. For most part, recommendations in both reports are relatively easy to implement, without major regulatory framework changes, they are non-disruptive and do not require major budget resources (except for the recommendations related to DWCAO's information system that require investment).

Phase two, medium-term (2-3 years)

This phase would comprise two other important elements of the overall disability policy and system in Lithuania:

One: Disability policy and system for children.

This is a particularly complex, sensitive, and technically and human resources demanding area of disability system and policies, including an assessment of health conditions, disabilities and a range of assessment of needs for support, including an assessment of special educational needs. It plays a significant role in determining the course of life of children born with or developing intellectual and physical disabilities, congenital impairments, learning disabilities, and developmental delays. It requires a concerted engagement of a range of professionals, from pediatricians, to nurses, to development experts, social workers, and teachers, to parents and communities.

As noted, the current project with DG Reform does not include children with disabilities. To further develop its disability policy and system for children with disabilities, the following steps are a must: (i) an in-depth, comprehensive assessment of the current system and policies, including health, education and social protection; (ii) development of tools that need replacement or need to be introduced (example: a new tool for the assessment of special education needs based on ICF¹⁸) and their piloting; (iii) empirically (pilot) based recommendations. These activities require significant resources. Furthermore, the implementation of recommendations is likely to require increased budget allocation to disability policy and system for children and the Government should start considering this ahead of time.

One very beneficial step in planning and pursuing the reform of the child disability policy and system in Lithuania will be to twin with one of the EU countries that have achieved significant success in creating an inclusive system for children with disabilities. One of the best such examples is Portugal.

¹⁸ A good example of such a tool is developed by

Two: Needs assessment for adults with disabilities:

As explained (see the Overview and Appendix 1), disability needs assessment for adults is a different process – with a different aim and using different instruments – than needs assessment. As explained in this report, the currently used A&AQ instrument, although not scientifically acceptable for a summary score of disability status assessment, with some adjustments and pilot testing has the potential of being used as a disability needs assessment tool.

What is important to keep in mind is that optimally, disability needs assessment is conducted as a multidisciplinary administrative process, where rehab professionals (medical, occupational, vocational, etc.) and social workers and if needed employers, employment office, etc. specialists work together to assess the needs of a person with disability and refer her or him to available services with the aim of maximizing her or his functioning and activities and participation. WHODAS, while not a disability needs assessment tool, will provide important initial information on the domains of functioning which need close attention.

Moreover, the needs assessment process may employ different tools, depending on the situation of the person whose needs are assessed. Many well tested tools are available; however, whether and how to use them in the Lithuanian context is a matter of a careful analysis, adjustments and test piloting.¹⁹

For example, evidence suggests that it is very important to make the transition from sick leave to some variation of work as seamless as possible, otherwise once an individual leaves the workplace and takes up some form of income replacement, then it is extremely unlikely that he or she will ever return to the labor market. If, however a robust needs assessment process is adopted and the individual, while on sick leave, can work together with his or her employer, with DWCAO, and a rehabilitation specialist, then, on the basis of information from the needs assessment, a return to work plan can be developed that would ensure the transition from sick leave back to work, either in its original form or some modification to account for permanent change in functioning status as determined by the needs assessment.

Designing and testing a new disability needs assessment system will require additional resources both during the reform design phase and for the implementation of a multidisciplinary process, separate from the disability status assessment.

¹⁹ Selb M, Gimigliano F, Proding B, Stucki G, Pestilli G, Iocco M, Boldrini P. Toward an International Classification of Functioning, Disability and Health clinical data collection tool: the Italian experience of developing simple, intuitive descriptions of the Rehabilitation Set categories. *European Journal of Physical and Rehabilitation Medicine* 2017 April; 53(2):290-8. Finger M, Escorpizo R, Bostan C, De Bie R. Work Rehabilitation Questionnaire (WORQ): Development and Preliminary Psychometric Evidence of an ICF-Based Questionnaire for Vocational Rehabilitation. *Journal of Occupational Rehabilitation* (2014) 24:498–510.

APPENDICES

Appendix 1: Lithuania Disability and work capacity assessment and disability needs assessment

Disability/work capacity assessment and needs assessment are two separate processes. **Disability assessment is used to establish the whole person 'status' of disability.** Once this status is formally established and a person is issued a certificate of disability, this person is formally eligible to various social insurance and other benefits, provided that she or he meets other benefit and service specific criteria. **Needs assessment is an assessment that identifies the needs the individual has because of his or her health condition and impairments, for the purpose of providing supports and services to optimize functioning, and often specifically to return to work.** These processes are very different, in purpose, outcome and methodology:

A status disability assessment is a process for quickly dividing the applicant population into two broad groups: those not having a disability/limited work capacity and those having a disability/limited work capacity. Those assessed as having disability/limited work capacity are also assessed for the level and duration of disability/limited work capacity. Depending on the level of certified disability, persons are eligible to receive various publicly financed allowances and services. In many ways, formal disability certification is established as a formal gate through which persons with disabilities can access those benefits.

A disability needs assessment, by contrast, assumes that the person has already been determined to have a level of disability that makes them eligible for some benefit, and then investigates by means of detailed questions and other investigations precisely which of the available supports and services the person would benefit from given their disability-related needs.

It should be noted that whatever instrument is used for needs assessment, it must be based on functioning, since limitations in functioning in some physical or mental domain create needs for supports and services. In other words, although status disability assessment and needs assessment are very different, they both should be based on functioning information. From our perspective there is only one relevant model of disability, and that is disability – understood in terms of the ICF – as limitation of functioning in one or more domains in interaction with the person's environment.

About disability/work capacity status assessment:

Lithuania already has a disability status assessment that has conceptually moved from a medical model to one based on functioning. Building on the strengths of the existing system by incrementally reforming it on the basis of empirical evidence from the WHODAS pilot, is a smart strategy. It is non-disruptive and allows for a gradual shift to a methodology in which functioning plays a dominant role.

Given the options laid out in this Report, Lithuania could put WHODAS into place in the current disability status assessment process and adopt an averaging strategy that gradually moves toward an algorithm that assigns 75% weight given to functioning in, say, 3-5 years. It could be done in a shorter period, but a smart strategy is to start with a lower weighting (e.g., 50%), collect and record data systematically and then perform an analysis using the techniques from the Report on a much bigger sample, before moving to 75%. To facilitate this, an analytical/statistical unit should be established at DWCAO.

Excluding entirely medical information from disability status assessment is not a good idea. It is actually in conflict with the ICF, which defines disability as an outcome of interaction between a health condition and the person's' environment. Therefore, in ICF terms, it is essential to have medical

information about the person's health condition and impairments. Because of a significant impairment, what in the ICF is called the *intrinsic capacity* of a body will be reduced irrespective of environmental accommodation and support. A blind person will always experience some disadvantage, regardless of all support she/he may be provided. So medical information will always be relevant to disability status assessment.

About disability needs assessment:

It is important to keep in mind that the adult population of people who are identified as being disabled in terms of a disability status assessment is not homogenous. Roughly speaking there are four distinct groups of persons who may seek to be assessed for their needs related to disability:

1. Working age adults on sick leave or otherwise unemployed with impairments.
2. Retired individuals.
3. Children and children transitioning to adults.
4. Individuals with congenital impairments (intellectual, birth defects, genetic diseases) who may never have worked but, as children will have been assessed for disability, following the rules in Lithuania.

Although members of each of these groups may end up with the same level of whole person disability after a status assessment, when it comes to needs assessment procedures, assessment instruments and criteria of eligibility will be very different. Or in other words, the content and format of a personal needs plan would be very different for these four populations.

For group 1, the person likely sought a disability status assessment because he or she is moving out of sick leave because the health problem, whether work-related or not, has not resolved itself and the person believes that they cannot work. Experience in other countries confirms that it is essential to ensure that people in this situation are as soon as possible directed toward supports and services that enable them to return to work, either the same or a different job. The aim of a needs assessment is to serve the central purpose of disability policy for individuals at risk of permanent disability or experiencing the onset of disability, namely, to provide the supports and services needed to optimize functioning in order to continue work and be active and participate in all aspects of life. One important factor to consider is that we have strong empirical evidence that once persons leave employment due to disability, most never return to employment unless the transition back to work is made easy.

Other countries have instituted practices that help to achieve this result. Needs assessment is administered early in the process – perhaps even before disability status assessment is done – and a multidisciplinary team, usually led by a rehabilitation professional will meet with the individual, employment counselors, and perhaps the individual's employer to map out step-by-step return to work plan. The aim is to ensure that each person does not leave the labor market entirely, but has the supports and services, including vocational rehabilitation, required to realistically return to work. In case of an earnings differential – where for example the health problem creates an impairment that makes it impossible, even with supports, to return to the same job, the person should be referred to status disability assessment in order to receive disability pension (remember that a disability pension is an insurance-based pension that compensates for a loss of income due to disability), while continuing working in the new job. Individually and socially, it is always better to have a disabled individual active and working with support measures, including the provision of a disability pension.

For retired individuals (who, we assume, have access to old-age social security pension), it makes more sense to have a disability status assessment first to determine whether the individual has a disability that can be accommodated with supports and services (perhaps to return to another job if he or she wishes) or whether they require more substantial support in the form of personal assistance and long-

term care. This step should be followed by a need assessment – tailored to this population – that would be the basis for eligibility for relevant supports and services.

For Group 3 and 4, a different assessment processes will be required, depending on how these groups have been assessed during their childhood years. This especially pertains to persons with intellectual impairments and other congenital impairments that have lasted since birth. These individuals will already have been receiving some form of support and as they transition to adulthood it may become appropriate to use the needs assessment process to determine the need for continuous support and assistance and eligibility for regular employment or some version of specialized work, such as social enterprise or sheltered workshops.

Questions are sometimes asked about the status of WHODAS in various other contexts including needs assessment, level of health care, work incapacity, and opportunities for social integration. Related to the use of WHODAS in needs assessment: it is true that WHODAS might be used for needs assessment; but it would not be a good instrument for this because it is far too generic and does not provide sufficiently detailed information to support decisions about supports and services. However, the information from WHODAS could provide some useful insights about areas where a person may need focused support.

International experience suggests that there are many instruments and questionnaires that can be successfully used for needs assessment purposes. Some of these are generic and are commonly used in rehabilitation assessments and are commonly referred to as 'dependency measures' (SF-36, FIM, etc.), others are specifically related to major life areas, and in part work for working age adults and education for children. Needs assessment instruments do not establish grades or percentage of whole person disability (i.e., they are not status disability instruments). Instead, they look more specifically at key areas of physical and medical functioning in order to identify limitations that can be improved by means of supports and services. Because of this purpose, they tend to be more specific and focused, for example work for working age and education for children. For seriously impaired individual, needs assessment focus is on 'independent living', that is the basic functioning capacity to live on one's own. In these severe cases, the primary support is personal assistant or informal, long term care provision.

Appendix 2: DWCAO's Activity and Ability to Participation Questionnaire

(Questionnaire form)

DISABILITY AND WORKING CAPACITY ASSESSMENT OFFICE
UNDER THE MINISTRY SOCIAL SECURITY AND LABOUR

QUESTIONNAIRE OF A PERSON'S ACTIVITY AND ABILITY TO PARTICIPATE

(date)

(forename, surname of the individual)

(forename and surname of the person's (representative) parents, custodian (guardian) or of his/her authorized representative)

(forename and surname of the employee of the Disability and Working Capacity Assessment Office under the Ministry of Social Security and Labor having performed the assessment and completed the questionnaire)

I have been made familiar with the procedure of the assessment of a degree of working capacity, I am aware of the significance of the Questionnaire of a person's activity and ability to participate (hereinafter – the Questionnaire) in assessing a degree of working capacity.

A person (his/her representative) _____

(signature)

(forename and surname)

(date)

The first part of the Questionnaire shall be completed base on the documents and information provided for the purpose of establishing the working capacity.

When completing the Questionnaire, please mark the appropriate point (by circling it) and enter the total number of points score

1. Professional, work activities, and environmental accessibility		Points
1.1. Age	55 years and more	3
	45–54 years	2
	35–44 years	1
	Up to 35 years	0
1.2. Professional qualification	Does not hold professional qualification or cannot exercise the professional qualification held	4
	Vocational rehabilitation is required	3
	Does not hold professional qualification or cannot exercise the professional qualification held, but can do works that require other qualification	2
	Professional qualification restored or a new professional qualification acquired during the vocational rehabilitation programme	1
	Holds a professional qualification and can exercise it	0
1.3. Work experience and work skills that the individual may use at the workplace	Has no work experience or work skills, cannot exercise the existing ones and cannot acquire them	3
	Lost work experience and work skills because of interruption of employment of more than 3 years	2
	Has no work experience and work skills but can acquire them	1
	Has work experience and work skills, can exercise them	0

1.4. Adaptation of physical, work and information environment	Complex adaptation of both physical, work and information environment and/or help by a personal assistant at the workplace are required	3
	Complex adaptation of a work environment or help by a personal assistant at the workplace are required	2
	Non-complex adaptation of a physical or work, or an information environment is required	1
	Adaptation of a physical, work and information environment is not required	0
Assessment of professional, work activities, and of environmental accessibility		

The second part of the Questionnaire contains questions related to the daily activities of the individual. When completing the Questionnaire, please mark the appropriate option (by circling it) of help required by the individual.

2.Activities and ability to participate	Assessment criteria (in points)				
	0	1	2	3	4
2.1. Mobility (moving)					
2.1.1. Sit-up, sitting, moving to another position All columns should be aligned to the left but numbers (0, 1, 2, 3, 4) should all be centered.	Sits-up, sits, changes seating safely (without threatening himself/herself and/or those around him/her realizing the meaning of the actions	Sits-up, changes seating on his/her own, sometimes aids are required (higher chair, stick, crutches, etc.), sometimes requires help, encouragement from another individual	Sits-up, sits, changes seating on his/her own using aids (higher chair, stick, crutches, etc.). Sometimes requires a minimum contact help when performing an action, sometimes – encouragement or care by another individual in creating conditions in order for the action to be performed (e.g., putting a slippery board underneath the buttocks, raising or lowering the footrest)	The individual does not perform actions on his/her own and safely (may threaten himself/herself and/or those around him/her. However, using aids and with help by another individual may sit-up, sit, change the position	Continuous help by others is needed because the individual does not make any actions by himself/herself
Scoring for sit-up, sitting, moving to another position	0	1	2	3	4
2.1.2. Standing up and standing	Stands up and stands for more than 30 minutes (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	Stands up and stands on his/her own for more than 30 minutes, sometimes aids are required (stick, crutches, etc.), sometimes requires help, encouragement from or care by another individual	Stands up and stands on his/her own for up to 30 minutes using aids (higher chair, stick, crutches, etc.). Sometimes requires a minimum contact help when performing an action (e.g. a support) sometimes – encouragement or care	Aids (higher chair, stick, crutches, etc.) and help by other individuals are required because the individual does not make actions on his/her own and safely	Continuous help by others is needed because the individual does not make any actions by himself/herself

			by another individual in creating conditions in order for the action to be performed (e.g., putting a slippery board underneath the buttocks, raising or lowering the footrest)		
Scoring for standing up and standing	0	1	2	3	4
2.1.3. Walking	The individual is fully independent, walks at least 200 meters without having rest. Does not use aids, walks safely across various surfaces. Carries out actions safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	The individual is independent – walks at least 200 meters without having rest, may use aids when necessary (stick, crutches, walker, etc.). Action takes longer or gait is unsafe, sometimes the care by another individual, verbal correction are needed. Manages to overcome obstacles safely	Cannot walk a distance of more than 200 meters without having rest, uses aids (stick, crutches, walker, etc.). A minimum contact help (hold-up in case of loss of balance or assisting with rotating and changing the direction of movement, or stepping across the threshold)	Cannot walk a distance of more than 200 meters without having rest. Aids are always required (stick, crutches, walker, etc.), and assistance by another individual (hold-up in case of loss of balance or assisting with rotating and changing the direction of movement, or stepping across the threshold). Assistance by one individual is sufficient	Continuous help by others is needed because the individual does not make any actions by himself/herself
Scoring for walking	0	1	2	3	4

2.1.4. Use of public and private transport	Uses a public and private transport on his/her own and safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	Uses a public and private transport on his/her own and safely, sometimes aids are required (handrails, crutches, sticks, etc.), sometimes help by another individual is required (to provide with information, to encourage, etc.)	Can use a public and private transport only with help by another individual, aids are always required (handrails, crutches, sticks, etc.). Aids allow using a public and private transport adapted for disabled individuals, in the case of specially adapted transport infrastructure	Can only use a public and private transport adapted for the needs of disabled individuals, in the case of specially adapted transport infrastructure. Always uses aids (handrails, crutches, sticks, etc.)	Continuous help by others is needed. Can only use a special transport (ambulance or other vehicles specially adapted for disabled individuals)
Scoring of use of public and private transport	0	1	2	3	4
2.1.5. Picking up and moving of things	Picks up, lifts up and moves on his/her own and safely things that weight less than 3 kilograms (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	Picks up, lifts up and moves on his/her own things that weight less than 3 kilograms, sometimes aids are required (stick, crutches, etc.) or help by another individual, the action is performed more slowly by distributing the weight on both hands	Always uses aids (stick, crutches, etc.) to pick up, lift up and move things that weight less than 3 kilograms, the limitation in one hand, loss of balance are possible, sometimes help by another individual is required (giving, hold-up, encouragement, etc.)	Cannot pick up, lift up and move weights of 3 kilograms. Aids (stick, crutches, etc.) and help by another individual (giving, hold-up, encouragement, etc.) are always required for the action to be performed	Continuous help by others is needed because the individual does not make any actions by himself/herself
Scoring of picking up and moving of things	0	1	2	3	4
2.1.6. Climbing stairs	Fully independent - climbs up and down the stairs to the second floor without using any additional means, without holding upon	Is nearly independent - climbs up and down the stairs to the second floor. However, handrails, stick or	Cannot climb to the second floor without having rest, aids are required (support, handrails, stick, etc.). A minimum contact help	Cannot climb to the second floor, aids are always required, a contact help by one individual is sufficient	Continuous help by others and aids are required because the individual does not make any actions by himself/herself

	handrails. Carries out actions safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	another support are required	is required (hold-up, stabilization of balance)		
Scoring of climbing stairs	0	1	2	3	4
Assessment of the need for assistance in increasing mobility	Would technical assistance measures increase the mobility opportunities? (tick <input checkbox"="" checked="" type="checkbox>): <input type="/> YES <input type="checkbox"/> NO Would help by another individual increase the mobility opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO Would adaptation of living environment increase the mobility opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO Would social rehabilitation services increase the mobility opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO				
2.2. Application of knowledge					
2.2.1. Concentration	Finds no difficulty to concentrate on activities (lasting not less than 10 minutes)	The individual manages to concentrate on activities, to focus attention, but not longer than for 10 minutes, sometimes aids are required (notes, electronic reminders), encouragement or reminder by another individual	The individual concentrates on activities only after being reminded and/or following verbal encouragement by another individual	A continuous external motivation is required even for the short concentration (lasting up to 10 minutes), can be easily distracted from the task. Constant reminders, encouragement and similar forms are necessary.	A continuous help by other individuals is required because the individual is unable to concentrate even for a short task
Scoring of concentration	0	1	2	3	4
2.2.2. Memory	Is able to memorize information from different fields, can link it to other information	Is able to memorize information from different fields, sometimes aids are required (notes, reminders), may forget details of information	Remembers the things that are important for him/her or his/her family members only using aids (notes, reminders) or with help by another individual	Does not remember by himself/herself the things that are important for him/her or his/her family members in basic daily activities. Uses aids on continuous basis,	A continuous help by other individuals is required because the individual has completely lost memory functions

		that has not been used for a long time	(reminder, encouragement)	constant verbal reminder by another individual is required (encouragement to start, continue and end activities), control over the course of actions is required	
Scoring of memory	0	1	2	3	4
2.2.3. Orientation in the environment and time	Is well oriented in time and environment without help by others. Performs actions in a secure manner (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	Is well oriented in environment without help by others, sometimes help by another individual may be required (explanation, instruction, reminder)	Is poorly oriented in environment and time without aids (cane for blind, means of communication, talking watches, rings, etc.), sometimes help by another individual is also required (sign language interpreter, guide, etc.)	No orientation in environment and time, does not control own emotions and behaviour (from aggression to total apathy), underestimates his/her possibilities, aids are always required (cane for blind, means of communication, talking watches, rings, etc.) and help by another individual	A continuous help by other individuals is required because the individual completely does not understand the surrounding environment, is not oriented in time
Scoring of orientation in environment and time	0	1	2	3	4
2.2.4. Understanding of visual information	Understands visual information, is able to read a written text	Understands visual information, is able to read written text, sometimes aids are required (magnifying glasses, contact lenses, etc.) or help by another individual (to explain information)	Understands visual information, is able to read written text only using aids (magnifying glasses, etc.), sometimes help by another individual is required	Partially understands visual information, does not read a written text. Always uses aids (magnifying glasses, contact lenses, etc.) and help by another individual	A continuous help by other individuals is required because the individual completely does not understand usual visual information or completely does not see it

Scoring of understanding of visual information	0	1	2	3	4
2.2.5. Understanding of auditory information	Understands auditory information, is able to speak complex sentences in a comprehensible manner	Understands auditory information, is able to speak in a comprehensible manner, sometimes aids are required or help by another individual (to explain information)	Understands only commonly spoken language and responds more slowly. Aids are always required (hearing aids, etc.) using which the individual can hear and speak in short sentences, sometimes help by another individual is required (sign language interpreter)	Does not understand auditory information (although can hear it). Can read of lips only individual words, sounds, pronounces individual words in a way that makes it difficult to understand them, communicates in sign language. Always uses aids, help by another individual is required (to translate from and to sign language, contact help, plainly expressed spoken language, mimicry)	A continuous help by other individuals and aids are required because the individual completely does not understand usual auditory information or completely does not hear it
Scoring of understanding of auditory information	0	1	2	3	4
2.2.6. Writing and counting	Is able to convey information independently in writing	Is able to write text, count independently. However, this takes longer than usually. Sometimes aids are required (adapted writing instrument, information technologies, etc.)	Is able to write only very short and simple text and to count. Aids are required and sometimes help by another individual	The individual is unable to write and count individually. Aids are always required and help by another individual	A continuous help by other individuals is required because the individual is able neither to write nor to count
Scoring of writing and counting	0	1	2	3	4

Assessment of the need for assistance in applying knowledge	Would technical assistance measures increase the opportunities of knowledge application? (tick <input checkbox"="" checked="" type="checkbox>): <input type="/> YES <input type="checkbox"/> NO Would help by another individual increase the opportunities of knowledge application? <input type="checkbox"/> YES <input type="checkbox"/> NO					
2.3. Interaction						
2.3.1. Interaction with strangers	Has no difficulties in interacting with strangers	Reluctantly interacts with strangers, may have minor speech and / or perceptual impairments. Sometimes help by another individual is required (encouragement, motivation, etc.)	Limited interaction with strangers, avoids or cannot maintain social contacts. Aids are always required (information technologies, notes, communication aids, etc.), sometimes help by another individual is required	Is unable to interact (due to physical, mental or intellectual condition), without much help from others the individual is at risk of social exclusion. Aids are always required (information technologies, notes, communication aids, etc.) and help by another individual	A continuous help by other individuals is required because the individual completely does not interact. Interaction is impossible even with help of others	
Scoring of interaction with strangers	0	1	2	3	4	
2.3.2. Interaction with relatives and friends	Has no difficulties in interacting with relatives and friends	Reluctantly interacts with relatives and friends, may have minor speech and / or perceptual impairments. Sometimes help by another individual is required (encouragement, motivation, etc.)	Limited interaction with relatives and friends, avoids or cannot maintain social contacts. Aids are required (information technologies, notes, communication aids, etc.), sometimes help by another individual is required (initiative, encouragement, motivation, stimulation, etc.)	Aids are required when interacting (information technologies, notes, communication aids, etc.) and help by another individual because the individual is unable to interact (due to physical, mental or intellectual condition), without much help from others the individual is at risk of social exclusion	A continuous help by other individuals is required because the individual completely does not interact. Interaction is impossible even with help of others	

Scoring of interaction with relatives and friends	0	1	2	3	4
2.3.3. Speaking (creating of messages during interaction) and/or language perception (accepting of messages during interaction)	Smoothly expresses thoughts, realizes the situation, is able to express own needs and/or understands spoken language, and responds accordingly to the message spoken	Lacks fluency in speaking, speaks in individual words, using gestures and mimicry, or is able to express in writing his/her needs and/or understands spoken language	Does not speak. However, is able to express his/her needs with help of gestures and other signs, or in writing, and/or understands simply expressed spoken language but responds only with certain mimics or difficult to understand gestures	Does not speak and with help of certain signs that not everyone understands is able to express the basis, most essential needs and/or understands only the simplest instructions or questions, but does not react to them	A continuous help by other individuals and aids are required because the individual does not speak and is unable to express his / her needs with gestures and other signs, and/or completely does not understand even the simplest instructions or questions, gestures, mimicry messages, and does not react to them
Scoring of speaking and/or language perception	0	1	2	3	4
Assessment of the need for assistance that increases the interaction opportunities	Would technical assistance measures increase the interaction opportunities? (tick <input checkbox"="" checked="" type="checkbox>): <input type="/> YES <input type="checkbox"/> NO Would help by another individual increase the interaction opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO Would help in decision making increase the interaction opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO Would social rehabilitation services increase the interaction opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO				
2.4. Independence					
2.4.1. Bathing and washing	Can take care of personal hygiene independently and safely (wash, bathe, care for individual body parts)	The individual manages to wash, bathe, to dry the body with a towel independently, the adapted environment and/or prostheses / orthoses are required, verbal assistance may be required (to	A minimum contact help is required (e.g. to rub body parts with a sponge and to hand preparations and items (help may be required in drying the back, legs, the injured body part with a towel)	A greater than average contact help is required when the individual is washing, bathing, drying the body with a towel	A continuous help by other individuals and aids are required because the individual cannot wash and bathe independently

		encourage, describe actions) and/or to prepare a bath and washing preparations and items (to clean a bath, to fill it is water)			
Scoring of washing and bathing	0	1	2	3	4
2.4.2. Putting clothes on and off	The individual manages to put clothes and shoes on and off, chooses the right outfit and does this safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	The individual manages to put clothes and shoes on and off, chooses the right outfit, it only takes longer for him/her to do this than for a healthy individual, the individual is not safe enough or uses prostheses / orthoses, verbal assistance may be required (encouragement, advise) and/or preparation (to put on prostheses, splints or to put clothes on and off)	A minimum contact help is required (e.g. when starting to put clothes on or to deal with fine elements of outfit (such as buttons, clips, buckles, laces) or sometimes to advice about proper outfit, to describe actions of putting clothes on and off and/or encourage to put clothes on and off. Aids are always required (orthoses, prostheses, etc.)	A greater than average contact help is required when the individual is putting clothes and shoes on and off, does not choose proper outfit on his/her own. Aids are always required (orthoses, prostheses, etc.) and help by another individual	A continuous help by other individuals is required because the individual does not perform the action independently
Scoring of putting clothes on and off	0	1	2	3	4
2.4.3. Eating	The individual eats independently, performs the actions safely (without threatening himself/herself and/or those around him/her),	The individual eats independently, performs the actions safely (without threatening himself/herself and/or those around him/her),	The individual eats independently, a minimum or average verbal help by another individual may be required (encouragement,	When the individual is eating, a greater than average verbal and contact help by another individual is required in performing the action and/or	A continuous help by other individuals is required because the individual does not perform the action independently

	realizing the meaning of the actions	realizing the meaning of the actions. Performs all actions more slowly than usually	advises) and/or preparation (e.g. put food on a plate, spread butter on bread, pour a drink) and/or a minimum contact help (e.g. to hand a cutlery, to place a piece of food in a spoon or to spear food with a fork, etc.)	continuous supervision of actions when the individual independently performs the action but does not understand its essence (e.g. may start eating stuff other than food products thereby endangering his/her health)	
Scoring of eating	0	1	2	3	4
2.4.4. Using the toilet	The individual uses the toilet independently and does this safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	The individual uses the toilet independently and does this safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions. Aids are sometimes required (stick, crutches, walker, etc.) and help by another individual	The individual is able to use the toilet independently, aids are required (stick, crutches, walker, raiser for toilet seat, a special chair, etc.), verbal help may be required (encouragement, telling the actions) and/or a minimum or average contact help by another individual (e.g. to hold, to help in putting clothes on and off)	A greater than average contact help by another individual is required when the individual is using the toilet, when the individual is not self-aware of the process (does not control it individually) but can cope with the toilet related matters when another individual controls the process. Aids are always required (stick, crutches, walker, raiser for toilet seat, a special chair, etc.)	The individual requires a continuous contact help by another individual in performing the action because the individual does not understand or control urination and / or defecation actions, and is dependent on the help of another individual
Scoring of using the toilet	0	1	2	3	4
2.4.5. Taking care of own health	The individual carries out activities related to	The individual carries out activities related to	When reminded by another individual, the	Help by another individual is required	The individual requires a continuous help by

	health care (visiting doctors, following doctors' instructions, taking medications, etc.) independently and meaningfully	health care (visiting doctors, following doctors' instructions, taking medications, etc.) independently and meaningfully. The individual understands that it is necessary to take medications and takes them. The individual is able to choose the necessary medications, knows when, what medications and in what doses to take, does not forget to take them. Sometimes help by another individual is required (reminder, encouragement). Performs the actions safely (without threatening himself/herself and/or those around him/her), realizing the meaning of the actions	individual manages, without help by another individual or with the minimal help by another individual, to select medications, their quantity, what medications he/she needs to take and takes them independently, visits doctors, follows their instructions	because the individual does not realize that he/she needs to take medications (may resist to this) and/or is unable to select medications, does not understand in what doses and when to take medications. Does not understand when he/she needs to visit doctors or to follow their instructions	another individual because the individual himself/herself does not realize that he/she needs to take medications and/or is unable to take medications. Does not understand that he/she needs to visit doctors and to follow their instructions. The individual is dependent on the actions of another individual. Medications are injected and/or administered via a probe and/or must be administered orally
Scoring of taking care of own health	0	1	2	3	4
Assessment of the need for assistance that increases independence of the individual	<p>Would technical assistance measures increase the independence opportunities? (tick <input checkbox"="" checked="" type="checkbox>): <input type="/> YES <input type="checkbox"/> NO</p> <p>Would help by another individual increase the independence opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO</p> <p>Would adaptation of the living environment increase the independence opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO</p> <p>Would social rehabilitation services increase the independence opportunities? <input type="checkbox"/> YES <input type="checkbox"/> NO</p> <p>Would help in decision making increase the independence opportunities of the individual? <input type="checkbox"/> YES <input type="checkbox"/> NO</p> <p>Would social rehabilitation services increase the independence of the individual? <input type="checkbox"/> YES <input type="checkbox"/> NO</p>				

2.5. Daily activities					
2.5.1. Food preparation	Can prepare food independently and safely (without threatening himself/herself and/or those around him/her)	Can prepare food independently and safely (without threatening himself/herself and/or those around him/her). However, aids and/or help by another individual are sometimes required. Doing so takes longer than usually	Can prepare food independently if the living environment is adapted for this purposes. Always uses aids, help by another individual is sometimes required (to encourage, to hand, bring something, to cut products, to tell the course of actions, etc.). Food preparation takes longer than usually	Is unable to prepare food independently, aids, specially adapted living environment and help by another individual are always required (to encourage, to hand, bring something, to cut products, to pour food and drinks, to tell the course of actions, etc.).	Is unable to prepare food, is completely dependent on care (help) by another individual
Scoring of food preparation	0	1	2	3	4
2.5.2. Housework	Performs housework independently and safely, without threatening himself/herself and/or those around him/her, realizing the meaning of the actions	Performs housework independently and safely, without threatening himself/herself and/or those around him/her, realizing the meaning of the actions. Aids and/or help by another individual are sometimes required	Can perform housework only using aids (prostheses, walkers, wheelchair, etc.), help by another individual is sometimes required (encouragement, motivation, telling sequence of actions, etc.). Does not plan housekeeping actions, it takes longer for the individual to perform activities than for a healthy individual (verbal help is required – advises, recommendations)	Is unable to do housework independently. Help by another individual is required in performing housework, aids and specially adapted living environment are always required	Is unable to perform housework. Complete supervision (help) by another individual is required

Scoring of housework completed	0	1	2	3	4
Assessment of the need for assistance in daily activities	Would technical assistance measures facilitate daily activities? (tick <input checkbox"="" checked="" type="checkbox>): <input type="/> YES <input type="checkbox"/> NO Would help by another individual facilitate daily activities? <input type="checkbox"/> YES <input type="checkbox"/> NO Would adaptation of the living environment facilitate daily activities? <input type="checkbox"/> YES <input type="checkbox"/> NO Would social rehabilitation services facilitate daily activities? <input type="checkbox"/> YES <input type="checkbox"/> NO				
Total score:					

The assessment has been carried out and the questionnaire has been completed by

(name of the position held)

(signature)

(forename and surname)

I have made myself familiar with

Individual (his/her representative)

(signature)

(forename and surname)

Notes _____

Appendix 3

Systematic Overview to the adjustment strategies of the WHODAS items applied at mid-term

#	Stage	Name	Not Imputed			Imputed			
			LID	PSI	Misfit	LID	PSI	Misfit	
1	1	Start	All	yes	0.943	yes	yes	0.951	yes
2	1a	Start	Testlet by Domains D6 together	yes	0.867	no	yes	0.884	no
3	1a	Start	Testlet by Domains D6.1-4 D6.5-8	yes	0.876	no	yes	0.893	no
4	1a.1*	Start	<i>Testlet by Domain (+Testlet solving 3)</i>	<i>no</i>	<i>0.872</i>	<i>no</i>	yes	0.886	yes
5	1b	Start	Testlet LID	yes	0.88	yes	yes	0.895	yes
6	1b.1	Start	Testlet LID (+Testlet solving 5)	yes	0.874	no	yes	0.884	no
7	1b.2	Start	Testlet LID (+Testlet solving 6)	yes	0.866	no	yes	0.894	yes
9	2	Reduced	Without Items > 20% Missing	yes	0.942	yes	yes	0.945	yes
10	2a	Reduced	Testlet by Domains D6 together	yes	0.868	no	yes	0.868	no
11	2a	Reduced	Testlet by Domains D6.1-4 D6.5-8	yes	0.876	no	yes	0.888	no
12	2a.1	Reduced	Testlet by Domain (+Testlet solving 11)	no	0.867	yes	no	0.881	yes
13	2b	Reduced	Testlet LID	yes	0.881	yes	yes	0.895	yes
14	2b.1	Reduced	Testlet LID (+Testlet solving 13)	yes	0.865	no	no	0.89	yes
15	3	Reduced	Without Items > 10% Missing	yes	0.938	yes	yes	0.939	yes
16	3a	Reduced	Testlet by Domains D6 together	yes	0.861	no	yes	0.873	no
17	3a	Reduced	Testlet by Domains D6.1-4 D6.5-8	yes	0.869	no	yes	0.881	no
18	3a.1	Reduced	Testlet by Domain (+Testlet solving 17)	<i>no</i>	<i>0.861</i>	<i>no</i>	no	0.875	yes
19	3b	Reduced	Testlet LID	yes	0.871	yes	no	0.884	yes
20	3b.1	Reduced	Testlet LID (+Testlet solving 19)	<i>no</i>	<i>0.864</i>	<i>no</i>	no	0.884	yes
21	4	WHODAS	12-Item Version	yes	0.873	yes	yes	0.892	yes
22	4.1	WHODAS	Testlet LID (+solving 21)	yes	0.837	yes	<i>no</i>	<i>0.818</i>	<i>no</i>
23	4.2	WHODAS	Testlet LID (+solving 22)	<i>no</i>	<i>0.823</i>	<i>no</i>			

in italic, issue free outcomes

* selected adjustment approach

Appendix 4

Frequencies and Percentages of the Activity and Ability Questionnaire

Code	Domain	Item	Up to 35 years	45-54 years	35-44 years	Up to 35 years	Missing	
Q_a	Professional, Work activities, and environmental accessibility	Age	253 (11.32%)	279 (12.49%)	674 (30.17%)	1023 (45.79%)	5 (0.22%)	
			Score = 0 <i>with qualifications/ with experience / no adaptations</i>	Score = 1	Score = 2	Score = 3	Score = 4	Missing
Q_b		Professional qualification	672 (30.08%)	11 (0.49%)	1140 (51.03%)	2 (0.09%)	404 (18.08%)	5 (0.22%)
Q_c		Work experience and work skills	1611 (72.11%)	183 (8.19%)	254 (11.37%)	180 (8.06%)	0 (0%)	6 (0.27%)
Q_d		Adaptation	114 (5.1%)	1642 (73.5%)	430 (19.25%)	42 (1.88%)	0 (0%)	6 (0.27%)
			Score = 0 <i>No Problem</i>	Score = 1	Score = 2	Score = 3	Score = 4	Missing
Q_1.1	1. Mobility	Sit-up, sitting, moving to another position	395 (17.68%)	1386 (62.04%)	412 (18.44%)	27 (1.21%)	8 (0.36%)	6 (0.27%)
Q_1.2		Standing up and standing	246 (11.01%)	1186 (53.09%)	724 (32.41%)	57 (2.55%)	15 (0.67%)	6 (0.27%)
Q_1.3		Walking	250 (11.19%)	1150 (51.48%)	732 (32.77%)	74 (3.31%)	22 (0.98%)	6 (0.27%)
Q_1.4		Use of public and private transport	531 (23.77%)	1241 (55.55%)	365 (16.34%)	74 (3.31%)	17 (0.76%)	6 (0.27%)
Q_1.5		Picking up and moving of things	590 (26.41%)	1037 (46.42%)	541 (24.22%)	47 (2.1%)	13 (0.58%)	6 (0.27%)
Q_1.6		Climbing stairs	190 (8.5%)	866 (38.76%)	1057 (47.31%)	84 (3.76%)	31 (1.39%)	6 (0.27%)
Q_2.1	2. Application of knowledge	Concentration	939 (42.03%)	998 (44.67%)	263 (11.77%)	20 (0.9%)	8 (0.36%)	6 (0.27%)
Q_2.2		Memory	602 (26.95%)	1350 (60.43%)	254 (11.37%)	17 (0.76%)	5 (0.22%)	6 (0.27%)
Q_2.3		Orientation in the environment and time	1979 (88.59%)	183 (8.19%)	48 (2.15%)	12 (0.54%)	6 (0.27%)	6 (0.27%)
Q_2.4		Understanding of visual information	775 (34.69%)	1228 (54.97%)	200 (8.95%)	18 (0.81%)	7 (0.31%)	6 (0.27%)
Q_2.5		Understanding of auditory information	1897 (84.91%)	285 (12.76%)	40 (1.79%)	3 (0.13%)	3 (0.13%)	6 (0.27%)
Q_2.6		Writing and counting	1618 (72.43%)	461 (20.64%)	108 (4.83%)	26 (1.16%)	15 (0.67%)	6 (0.27%)
Q_3.1	3. Interaction	Interaction with strangers	1148 (51.39%)	786 (35.18%)	219 (9.8%)	68 (3.04%)	7 (0.31%)	6 (0.27%)
Q_3.2		Interaction with relatives and friends	1779 (79.63%)	398 (17.82%)	41 (1.84%)	8 (0.36%)	2 (0.09%)	6 (0.27%)
Q_3.3		Speaking and/or language perception	1824 (81.65%)	325 (14.55%)	61 (2.73%)	14 (0.63%)	4 (0.18%)	6 (0.27%)
Q_4.1	4. Independence	Bathing and washing	402 (17.99%)	1282 (57.39%)	486 (21.75%)	38 (1.7%)	20 (0.9%)	6 (0.27%)
Q_4.2		Putting clothes on and off	612 (27.39%)	1335 (59.76%)	246 (11.01%)	21 (0.94%)	14 (0.63%)	6 (0.27%)
Q_4.3		Eating	1906 (85.32%)	251 (11.24%)	52 (2.33%)	11 (0.49%)	8 (0.36%)	6 (0.27%)
Q_4.4		Using the toilet	1523 (68.17%)	571 (25.56%)	90 (4.03%)	23 (1.03%)	15 (0.67%)	12 (0.54%)
Q_4.5		Taking care of own health	1109 (49.64%)	870 (38.94%)	194 (8.68%)	33 (1.48%)	16 (0.72%)	12 (0.54%)
Q_5.1	5. Daily activities	Food preparation	325 (14.55%)	1241 (55.55%)	554 (24.8%)	69 (3.09%)	33 (1.48%)	12 (0.54%)
Q_5.2		Housework	30 (1.34%)	629 (28.16%)	1389 (62.18%)	119 (5.33%)	55 (2.46%)	12 (0.54%)

References

- Andrews, Gavin, Alice Kemp, Matthew Sunderland, Michael von Korff, and Tevik Bedirhan Ustun. 2009. "Normative Data for the 12 Item WHO Disability Assessment Schedule 2.0." *PLOS ONE* 4 (12): 1–6. <https://doi.org/10.1371/journal.pone.0008343>.
- Bond, Trevor G., and Christine M. Fox. 2001. *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. Mahwah, NJ: L. Erlbaum.
- Cieza A, Fayed N, Bickenbach J, Prodinger B. Refinements of the ICF Linking Rules to strengthen their potential for establishing comparability of health information. *Disabil Rehabil* 2016:1-10.
- 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*.
- Masters, Geoff N. 1982. "A Rasch Model for Partial Credit Scoring." *Psychometrika* 47 (June): 149–74.
- Nunnally, Jum C., and Ira H. Bernstein. 1994. *Psychometric Theory*. 3rd ed. New York; London: McGraw-Hill.
- Posarac, A. and Bickenbach, J. May 2020. *Disability Policy and Disability Assessment System in Lithuania*. World Bank.
- 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.
- Team, R Core. 2016. "R: A Language and Environment for Statistical Computing." Vienna, Austria: R Foundation 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., S. Chatterji, N. Kostanjsek, J. Rehm, C. Kennedy, J. Epping-Jordan, S. Saxena, M. von Korff, and C. Pull. 2010. "Developing the World Health Organization Disability Assessment Schedule 2.0." *Bull World Health Organ* 88 (11): 815–23. <https://doi.org/10.2471/BLT.09.067231>.
- Yen, C. F., T. Y. Chiu, T. H. Liou, W. C. Chi, H. F. Liao, C. C. Liang, and R. Escorpizo. 2017. "Development of Activity and Participation Norms Among General Adult Populations in Taiwan." *Int J Environ Res Public Health* 14 (6). <https://doi.org/10.3390/ijerph14060603>.
- 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>.