

BMJ Open Intrarater reliability of the Abilitator—a self-report questionnaire on work ability and functioning aimed at the population in a weak labour market position: a test-retest study

Miia Wikström ^{1,2}, Anne Kouvonen ^{2,3}, Matti Joensuu ¹

To cite: Wikström M, Kouvonen A, Joensuu M. Intrarater reliability of the Abilitator—a self-report questionnaire on work ability and functioning aimed at the population in a weak labour market position: a test-retest study. *BMJ Open* 2021;**11**:e057295. doi:10.1136/bmjopen-2021-057295

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2021-057295>).

Received 13 September 2021
Accepted 10 November 2021



© Author(s) (or their employer(s)) 2021. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Finnish Institute of Occupational Health, Helsinki, Finland

²Faculty of Social Sciences, University of Helsinki, Helsinki, Uusimaa, Finland

³Centre for Public Health, Queen's University Belfast, Belfast, UK

Correspondence to

Miia Wikström;
miia.wikstrom@ttl.fi

ABSTRACT

Objectives The Abilitator is a patient-reported outcome measure (PROM) of work ability and functioning of those in a weak labour market position. It covers items for work ability and self-rated health, for example, and summary scales for social, psychological, cognitive and physical functioning, as well as everyday skills. The aim of this study was to evaluate the intrarater test-retest reliability, internal consistency and basic psychometric properties of the Finnish version of the Abilitator.

Design, setting and outcome The test-retest study was conducted in European Social Fund projects in 2018–2019. The participants completed two Abilitator questionnaires over 7–14 days. The internal consistency analysis was based on data collected in 2017–2019 in services for the long-term unemployed. The reliability was assessed using correlations (r , r_s), intraclass correlation coefficient (ICC), agreement with Bland-Altman analysis and internal consistency with Cronbach's alpha.

Participants The test-retest study had 67 participants (52% men, mean age 43.9 years) and the internal consistency study 10 923 (48% men, mean age 38.58 years), respectively. Of all the participants, 80% had been unemployed for over a year.

Results The test-retest r or r_s ranged from 0.71 to 0.93 and ICC from 0.74 to 0.93 for the items and summary scales. An exception was the life satisfaction item, with an r_s of 0.60 and ICC of 0.45. A statistically significant difference was observed in the summary scale for social functioning ($t=-2.01$, $p=0.049$). Agreement was observed for all variables except social functioning. Alphas for summary scales ranged from 0.74 to 0.91.

Conclusions The Finnish version of the Abilitator is a reliable PROM for the target group and has acceptable to excellent intrarater test-retest reliability and internal consistency, apart from the life satisfaction item. Further testing is needed for the social functioning summary scale.

INTRODUCTION

In Western European welfare states, a broad selection of services promotes the health, rehabilitation, social well-being, education and employment of the working-age population. These services often require an

Strengths and limitations of this study

- This was the first test-retest reliability and internal consistency study of the Abilitator.
- It was a strength that the study was conducted, and the data collected in a realistic service setting.
- It was a strength that the study covered a wide range of participants in a weak labour market position.
- It was a limitation that some real-life service activities took place during the test-retest period.
- It was a limitation that the sample size in the test-retest study was too small to conduct further subgroup analyses, such as with different language versions of the Abilitator and participants using mobility aids.

individual assessment of the clients' work ability and functioning, to identify any needs for support and to suggest participation in appropriate service programmes. Organisations providing employment and other welfare services should also assess the effectiveness of their services.¹ Reliable data on the clients' needs and changes during the services are essential for planning the allocation of resources. The use of patient-reported outcomes (PROs) has begun to increase, and their use is now widely recommended.^{2–5}

PROs can be defined as measurements of the patients' or service clients' self-reported health, functioning, well-being and health-related quality of life.^{6–9} Patient-reported outcome measurements (PROMs) are standardised, reliable and validated instruments that include and elicit PROs.^{6,7,10,11} PROMs can be either generic or specific, depending on whether they measure the clients' perceived health status in general or in relation to specific conditions or diseases.^{7,11} PROMs can be either unidimensional, when they measure a single characteristic (construct),

or multidimensional, when multiple different constructs are measured by the same instrument.¹²

Currently, only a few generic, feasible and validated PROMs exist for multidimensionally assessing work ability and functioning of the unemployed.^{13–15} The need for PROMs with a broad view of health and well-being, especially suitable for those with comorbidities, is apparent.¹⁰ A PROM with rapid administration, good feasibility and informational output could be a starting point to support decisions and actions in practice.⁹

The concept of work ability is a combination of health, functioning, basic standard competence and the relevant occupational traits required for managing reasonable work tasks in an acceptable environment.^{16,17} Functioning is closely related to health and comprises a psychological, social, physical and cognitive dimension.^{18–24}

Unemployed people generally have poorer health and work ability than those who are employed.^{25–28} The pressures of modern working life accumulate among those who are in a weak labour market position. This is a heterogeneous group of working-age people who persistently have challenges attaining employment. This can be due to a lack of employment history, a low level of education, disabilities, chronic or multiple health problems, prolonged unemployment or a migrant background.^{29–32}

The Abilitator is a generic, multidimensional PROM for measuring the self-reported work ability and functioning of the population in a weak labour market position.^{33,34} It consists of a digital questionnaire³⁴ (online supplemental file 1), analyses the responses and produces two kinds of reports: an individual written feedback report for the respondent (online supplemental file 2) and a group-level report (online supplemental file 3). The group-level report can be constructed by the user organisation from their accumulated data in the Abilitator's digital service, in either a cross-sectional or longitudinal setting.³⁴

As an unemployed person has no job, the content related to work ability and functioning in the Abilitator corresponds to the general demands of working life, including the constructs of employability and inclusion.³³ The Abilitator's purpose is to help the service clients to identify their strengths and challenges in terms of their work ability and functioning. It is designed to provide the client with a basis for individual goal setting in a dialogue with service professionals. Another purpose is to help the professionals to implement the most suitable services for each client. It also provides the professionals with a means to follow the client's progress in a positive and empowering way.³³

A high-quality PROM is valid, reliable, responsive and interpretable.⁶ A previous study on the Abilitator's content validity found that it sufficiently comprehensively covered the aspects necessary for enabling the assessment of the overall work ability and functioning of the population in a weak labour market position.³³ The instrument also had acceptable concurrent validity for assessing different aspects of the functioning of working-age people.³⁵ However, further evidence of the Abilitator's

psychometric properties is needed, because the demand to implement the instrument in different services is growing. The reliability of the Abilitator has not been examined before.

AIMS

The first aim of this study was to evaluate the Abilitator's intrarater test–retest reliability and measurement error. The second aim was to analyse the Abilitator's internal consistency. The third aim was to evaluate the basic psychometric properties of the Abilitator.

METHODS

The Abilitator self-report questionnaire

The Abilitator was developed in 2014–2017 at the Finnish Institute of Occupational Health (FIOH) in the national coordination project Social Inclusion and the Change of One's Work Ability and Capacity (Solmu), funded by the European Social Fund (ESF) Priority 5 programme (2014–2022).³³ The Abilitator is currently used in Finland to support primary-level decision-making in different services through which large numbers of clients meet professionals with various occupational backgrounds.^{34,36,37}

The Abilitator has the following nine domains (sections): *A. Personal information*, *B. Well-being*, *C. Inclusion*, *D. Mind*, *E. Everyday life*, *F. Skills*, *G. Body*, *H. Background information* and *I. Work and the Future*.³⁴ Each section contains 4–14 items (online supplemental file 1). The measure of each reported section is a summary scale, with a score of 0%–100% of the selected items, which cover different aspects of work ability and functioning³³ (figure 1). In this study, we analysed the reliability of the five individual items in section *B* and the summary scales derived from sections *C*, *D*, *E*, *F* and *G*, covering the dimensions of social functioning and inclusion, psychological functioning, everyday skills, cognitive functioning and physical functioning, respectively. The *Overall situation*

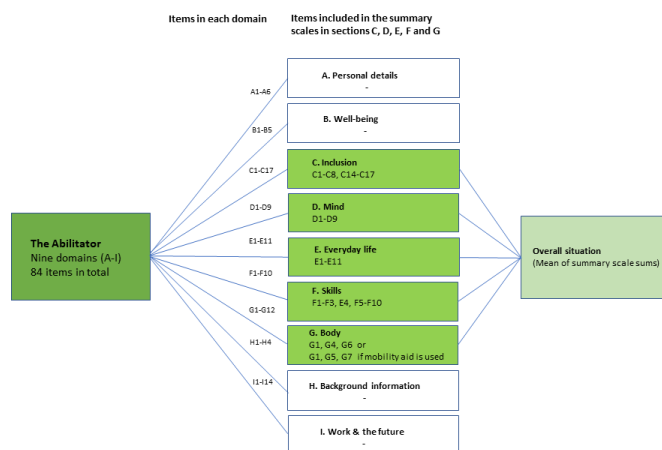


Figure 1 The Abilitator's domains, total number of items in each domain, items included in the summary scales and summary scales included in the overall situation score.

score, which is the mean of the sums derived from these five summary scales, was also included in the analysis. The Abilitator is currently available in nine languages, but we only included the Finnish version in this study.

Study populations and procedure

The data collection for the test–retest study was conducted by FIOH in 2018–2019 in cooperation with the ESF Priority 3 and 5 projects. The Priority 3 programme aims to promote employment among young unemployed adults and other groups in a weak labour market position.³⁸ The Priority 5 programme aims to improve the work ability and functioning of people outside working life to help them proceed on employment paths and to strengthen social inclusion.³⁸ First, we recruited those Priority 3 and 5 projects, that were already using the Abilitator as a part of their client assessment procedures. The recruitment was conducted via Solmu’s Facebook page and by directly contacting the potential projects’ project managers. Second, the client participants were recruited face to face at the group meetings arranged in cooperation with the ESF projects’ staff and FIOH’s researchers. All phases of the test–retest study were incorporated in the preplanned daily activities of the ESF projects. These activities included, for example, social interaction, light physical activity, gardening, cooking and rehabilitative work tasks such as sorting clothes for recycling.

The Consensus-based Standards for the selection of the Health Measurement Instruments (COSMIN) recommendation for a minimum sample size in reliability studies with two repeated measurements with intraclass correlation coefficient (ICC) 0.8 (95% CI ±0.1) is at least 50 participants.⁶ However, we aimed for an additional 30% (n=15) to account for possible non-response.³⁹ The participants completed the questionnaire at two time points (figure 2). They first voluntarily responded to the Abilitator as a part of the ESF project procedures (test 1). The second questionnaire (test 2) was preplanned in cooperation with the projects’ staff, to be completed within 7–14 days of test 1. Before completing the second questionnaire, the project participants were informed both verbally and in writing of the ongoing test–retest reliability study, data protection and the voluntary nature of participation. The participants who agreed to participate in the study completed both questionnaires independently on paper and mostly as part of group activity sessions, during which one or two project employees were available to help if any questions arose. Some participants completed both questionnaires independently but always with a project employee nearby. For the internal consistency analysis, we used a larger set of data (n=14895), which were collected between 2017 and 2019 in an online database maintained by FIOH. These data were accrued from the ESF projects and other organisations that provide services for the unemployed. FIOH obtained the organisations’ consent to use these data for research purposes.

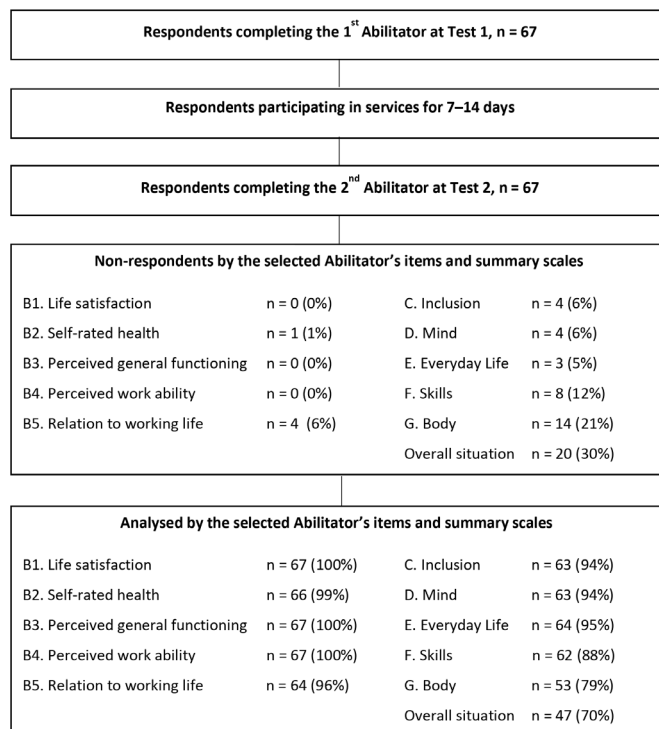


Figure 2 Flow chart of the test–retest study protocol with the number of non-respondents and analysed cases.

Patient and public involvement

The Abilitator was codeveloped and assessed by members of academic expert panels, practical expert panels of service professionals and target group clients.³³ In this study, neither the participants nor the service professionals were involved in the design phase or in the dissemination of the study results. However, the latter group was involved in participant recruitment, in scheduling the test–retest period, and in conducting the data collection in cooperation with research scientists from FIOH.

Data analysis

The data analysis included the items and summary scales described in the Abilitator’s group-level report³⁴ (online supplemental file 3). The study followed the guidelines of the COSMIN panel.⁶ First, we described the study population and analysed the number of missing cases. We also used the independent samples t-test to determine: (1) whether the two different ESF groups in the test–retest sample were different from each other, and (2) whether the larger sample used for the internal consistency analysis differed from the test–retest sample in terms of the reported variables. The significance level was set at $p \leq 0.05$ for all the analyses. In accordance with de Vet *et al*,⁶ the data were analysed for possible floor and ceiling effects. These effects may occur if more than 15% of all responses score at the lower or upper end of the scales. Second, we analysed the relative reliability of the test–retest data using the paired samples t-test (t) and Wilcoxon signed ranks test (z). Depending on whether the data obtained were normally distributed, the correlations between the two testing points were analysed using the Pearson’s r (r) or Spearman’s Rho (r_s) tests, and a value of ≥ 0.70 was considered a strong

correlation.^{40 41} We also calculated the ICC_{consistency} to assess the consistency of the test–retest measurements. The ICC values were interpreted as follows: ≥ 0.70 acceptable, ≥ 0.80 acceptable for research use, and ≥ 0.90 acceptable for clinical use.⁶ We further analysed the test–retest data for effect sizes for gender and ESF priority axes using two-way analysis of variance (ANOVA), where Partial Eta-squared $\eta^2 < 0.06$ was considered a small effect, $\eta^2 > 0.06–0.13$ a medium effect and $\eta^2 > 0.14$ a large effect.⁴² Third, we quantified absolute reliability using the Bland-Altman method by plotting the differences (d) between test 1 and test 2 against the means of the two measurements with 95% CI and 95% limits of agreement (LOA).⁴³ In this study, an acceptable result was that the line of equality, meaning zero, fell within the 95% CI of d.^{44 45} The measurement error reflects the intraindividual variation in the scores and was estimated as the SE of measurement (SEM_{pooled}).⁶ The SEM_{pooled} values were further converted into the smallest detectable change (SDC₉₅), which indicates the smallest within-person change with 95% CI that can be interpreted as a true change above the measurement error.⁴⁶ Fourth, an internal consistency analysis was conducted on the summary scales (C-G, Overall situation), and the acceptable level for Cronbach's alpha (α) was set at between 0.70 and 0.90.⁶ Participants who responded in a language other than Finnish or had no language choice information available were excluded from the analysis (n=3972). All analyses were conducted using IBM SPSS Statistics V.27.⁴⁷

RESULTS

Participant characteristics

Table 1 presents the participant characteristics in the test–retest study. In all, 67 participants responded to the Abilitator during test 1, 46% of whom were participants of the ESF Priority 3 projects, and 54% of the ESF Priority 5 projects. The two priority groups were

statistically significantly different in all reported variables other than item *B1* and scale *D. Mind*. On average, the Priority 3 group had higher Abilitator scores even though they were significantly older. Most participants had good perceived general functioning (*B3*) and perceived work ability (*B4*), but in the Priority 3 group, more participants were distributed in the higher point categories. The duration of unemployment (*I2*) was also statistically significantly different, as the participants in the Priority 5 group had been unemployed for a longer time and more participants had never been in employment. The test 2 response rate was 70%–100% across different sections (figure 2). The median response time from test 1 to test 2 was 14 days, with 86% (n=56) responding within 7–14 days. No floor or ceiling effects were found, as less than 15% of the responses per analysed item or summary scale scored at the lowest or highest score of the scale. The data set including the participants in the internal consistency analysis (n=10 923) was statistically significantly different ($p < 0.05$) from the test–retest participants in all reported variables other than items *B1* and *B2* and scales *C. Inclusion*, *D. Mind*, and *G. Body*. On average they were younger, with a mean age of 38.90 (95% CI 38.68 to 39.12, SD 13.32), and their scores were lower in both *B3*, with a mean of 6.86 (95% CI 6.82 to 6.89, SD 2.13), and *B4*, with a mean of 6.15 (95% CI 6.11 to 6.19, SD 2.53). In terms of duration of unemployment (*I2*), the participants in the internal consistency analysis were statistically significantly different from the Priority 3 group but not the Priority 5 group.

Relative intrarater test–retest reliability of the Abilitator

Table 2 presents the results of the relative test–retest reliability analysis of the Abilitator. The values obtained from the Pearson's r (r) and Spearman's Rho (r_s) tests showed

Table 1 Participant characteristics by the European Social Fund (ESF) Priority groups in the test–retest study.

	All participants (n=67) n (%)	ESF Priority 3 (n=31) n (%)	ESF Priority 5 (n=36) n (%)
Gender	Male 35 (52) Female 32 (48)	Male 22 (71) Female 9 (29)	Male 13 (36) Female 23 (64)
Age in years mean, (min–max)	43.9 (20–62)	48.0 (20–62)	40.3 (22–59)
B3. Perceived general functioning	7.62	8.23	7.06
Mean (95% CI of mean, SD)	(7.17 to 8.06, 1.80)	(7.52 to 8.93, 1.93)	(6.54 to 7.58, 1.50)
Poor (0–5 points)	11 (16)	3 (10)	8 (22)
Fairly poor (6–7 points)	15 (22)	4 (13)	11 (31)
Good (8–9 points)	34 (51)	17 (55)	17 (47)
Excellent (10 points)	7 (10)	7 (22)	0 (0)
B4. Perceived work ability	7.02	8.03	6.09
Mean (95% CI of mean, SD)	(6.42 to 7.61, 2.41)	(7.35 to 8.72, 1.90)	(5.21 to 6.96, 1.80)
Poor (0–5 points)	14 (21)	2 (7)	12 (33)
Fairly poor (6–7 points)	20 (30)	5 (16)	15 (42)
Good (8–9 points)	24 (36)	17 (55)	7 (19)
Excellent (10 points)	9 (13)	7 (22)	2 (6)

Table 2 The relative test–retest reliability of the Abilitator

Variable	N	Paired samples t-test t, p (95% CI)	Wilcoxon signed ranks test Z, p (95% CI)	Pearson's r correlation r (95% CI)	Spearman's Rho correlation r_s (95% CI)	ICC _{consistency} (95% CI)
B1* Life satisfaction	67	−1.98, ns. (−0.42 to 0.00)	−1.89, ns. (−3.89 to 0.07)		0.60 (0.40 to 0.74)	0.45 (0.24 to 0.62)
B2* Self-rated health	67	−0.54, ns. (−0.22 to 0.12)	−0.39, ns. (−2.35 to 1.57)		0.78 (0.64 to 0.87)	0.77 (0.66 to 0.86)
B3*† Perceived general functioning	67	0.19, ns. (−0.29 to 0.35)	−0.31, ns. (−2.27 to 1.65)		0.71 (0.55 to 0.82)	0.74 (0.61 to 0.83)
B4*† Perceived work ability	67	1.13, ns. (−0.14 to 0.50)	−1.21, ns. (−3.17 to 0.75)		0.84 (0.73 to 0.90)	0.85 (0.76 to 0.90)
B5*† Relation to working life	65	0.24, ns. (−0.34 to 0.43)	−0.34, ns. (−2.31 to 1.62)		0.80 (0.68 to 0.88)	0.83 (0.74 to 0.89)
C. Inclusion*‡	63	−2.01, 0.049 (−5.60 to −0.02)	−1.80, ns. (−3.75 to 0.17)	0.80 (0.74 to 0.89)		0.80 (0.68 to 0.87)
D. Mind*‡	63	−1.44, ns. (−4.06 to 0.66)	−1.35, ns. (−3.31 to 0.61)	0.83 (0.74 to 0.89)		0.83 (0.73 to 0.89)
E. Everyday life*‡	64	0.14, ns. (−1.68 to 1.93)	−0.12, ns. (−2.08 to 1.84)		0.81 (0.69 to 0.89)	0.80 (0.69 to 0.87)
F. Skills*‡	62	0.31, ns. (−1.646 to 2.259)	−0.59, ns. (−2.55 to 1.37)		0.87 (0.77 to 0.93)	0.87 (0.79 to 0.92)
G. Body*‡§	53	−0.83, ns. (−3.87 to 1.61)	−1.10, ns. (−3.06 to 0.86)		0.91 (0.87 to 0.96)	0.90 (0.83 to 0.94)
Overall situation*‡	47	−0.72, ns. (−1.77 to 0.83)	−0.80, ns. (−2.76 to 1.16)	0.93 (0.88 to 0.96)		0.93 (0.87 to 0.96)

*The mean value is reported in the group-level report.

†Is reported in the individuals' written feedback.

‡Is reported in both the individuals' written and numeric feedback.

§Those using a mobility aid are excluded.

ICC, intraclass correlation coefficient.

strong positive (≥ 0.70) to very strong positive correlations (≥ 0.90) between test 1 and test 2. An exception was item *BI*, which showed a moderate positive correlation of 0.60. The paired samples t-test showed that only *C. Inclusion* differed statistically significantly in the repeated measurements, with a higher mean in test 2. The Wilcoxon signed ranks test revealed that even though the median of *C. Inclusion* did not change statistically significantly ($p \leq 0.05$), test 2 had more positive ranks. This difference was statistically significant when the paired values of variable *C. Inclusion* were analysed using the Abilitator's feedback categories, as the Wilcoxon signed ranks test ($z = -2.00$, $p = 0.046$) showed that there was a statistically significant positive shift as 10 participants reached one category higher in test 2. The effect sizes were small in both the paired samples t-test ($D < 0.50$) and the Wilcoxon signed ranks test ($R < 0.30$). The effect sizes analysed using two-way ANOVA revealed that neither gender nor participation in the ESF Priority 3 or 5 programmes influenced the Abilitator's test–retest results ($\eta^2 < 0.06$). The ICC values

were above the acceptable level of 0.70 in all items other than *BI*.

Absolute intrarater test–retest reliability and the internal consistency of the Abilitator

Table 3 presents the results of the absolute test–retest reliability analysis and the measurement error. In the Bland–Altman analysis, zero was included within the 95% CI of d in all measurements other than *C. Inclusion*. This implies that changes occurred in summary scale *C* within the 14-day period. The SDC_{95} for the Abilitator's summary scales varied from 5.10% to 7.71%. The internal consistency values (α , 95% CI) of the summary scales were *C. Inclusion* 0.91 (0.909 to 0.914), *D. Mind* 0.88 (0.876 to 0.883), *E. Everyday life* 0.86 (0.859 to 0.865), *F. Skills* 0.87 (0.863 to 0.870), *G. Body* 0.74 (0.733 to 0.748) or 0.75 (0.685 to 0.796) if mobility aid was used, and *Overall situation* 0.86 (0.856 to 0.867). The α level was within the acceptable range in all summary scales other than *C. Inclusion*, in which it was slightly higher than recommended.

Table 3 The absolute test–retest reliability of the Abilitator with SE of measurement and smallest detectable change

Variable	N	Pooled SE of measurement (SEM _{pooled})	Smallest detectable change (SDC ₉₅)	Bland-Altman analysis with limits of agreement (LOA)				
				<i>d</i>	SD _{diff}	SE of <i>d</i>	95% CI of <i>d</i>	LOA
B1* Life satisfaction	67	0.61	2.16	−0.21	0.86	0.11	−0.42 to 0.01	−1.89 to 1.47
B2* Self-rated health	67	0.49	1.94	−0.05	0.68	0.85	−0.21 to 1.23	−1.39 to 1.30
B3* [†] Perceived general functioning	67	0.93	2.70	0.03	1.31	0.16	−0.29 to 0.35	−2.55 to 2.61
B4* [†] Perceived work ability	67	0.92	2.70	0.18	1.30	0.16	−0.14 to 0.50	−2.40 to 2.73
B5* [†] Relation to working life	65	1.10	2.91	0.05	1.55	0.19	−0.34 to 0.43	−2.98 to 3.10
C. Inclusion* [‡]	63	7.73	7.71	−2.80	11.08	1.40	−5.60 to −0.02	−11.50 to 5.88
D. Mind* [‡]	63	6.58	7.11	−1.70	9.40	1.20	−4.10 to 0.70	−20.10 to 16.70
E. Everyday life* [‡]	64	5.75	6.64	0.13	7.24	0.91	−1.68 to 1.93	−14.10 to 14.32
F. Skills* [‡]	62	5.44	6.47	0.31	7.70	0.98	−1.65 to 2.30	−14.80 to 15.40
G. Body* ^{‡,§}	53	7.33	7.50	−1.13	9.93	1.40	−3.90 to 1.61	−20.60 to 18.34
Overall situation* [‡]	47	3.37	5.10	−0.47	4.43	0.65	−1.80 to 0.83	−11.50 to 5.90

*The mean value is reported in the group level report.

†Is considered in the individuals' written feedback.

‡Is considered in both the individuals' written and numeric feedback.

§Those using a mobility aid are excluded, *d*=mean difference between Tests 1 and 2, SD_{diff}=SD deviation of mean difference, SE of *d*=SE error of the mean difference, 95% CI of *d*=95% CI of *d*.

DISCUSSION

In this study, we assessed the intrarater test–retest reliability, measurement error, internal consistency and basic psychometric properties of the Finnish version of the Abilitator, in line with the COSMIN guidelines. The study covered the items and the summary scales presented in the Abilitator's group-level report. The results showed that the relative reliability of the Abilitator varied from acceptable to excellent. The test–retest correlations (r , r_s) were high or very high and the intraclass correlations (ICC) fulfilled the acceptable, research or clinical use criteria. In addition, the means of the two tested time points were not different. An exception was item *B1. Life satisfaction*, with unacceptable r_s and ICC. It thus seems to describe the respondent's situation only at the moment of responding. In addition, the mean of summary scale *C. Inclusion* showed a small but statistically significant positive change during the test–retest period. This change was also statistically significant when analysed using the Abilitator's feedback categories. However, the effect sizes indicate that the impact of these changes was small. In terms of absolute reliability, a small degree of variation in the measurements among individuals was observed. The analyses showed acceptable agreement (LOA)

between test 1 and test 2, with no systematic pattern, except in summary scale *C. Inclusion*, in which a small but true change occurred beyond SEM and SDC. The Abilitator had good to excellent internal consistency, which shows that the items in each summary scale are closely related. However, in summary scale *C. Inclusion*, α was only borderline acceptable. In addition, neither floor nor ceiling effects were observed for any of the analysed items or summary scales, which means that neither the items nor the scales need to be reduced or extended when this instrument is used in the target population.

There are four possible explanations for the significant differences in summary scale *C. Inclusion*. First, this scale might not be a reliable PROM. The analyses conducted do not support this explanation, as the scale's test–retest correlation was high and intraclass correlation reached acceptable criteria. However, the internal consistency of summary scale *C. Inclusion* was slightly too high, and therefore item reduction could be considered in the future. Second, this scale might be very responsive and sensitive to small changes in social functioning and social inclusion. This aspect will be analysed in more detail in a future study of the Abilitator's responsiveness. Third, between test 1 and test 2, the participants took part in ESF project

activities. These activities focused on strengthening social inclusion and social interaction. Examination of the test–retest results of the individual items revealed that the significant positive change occurred in items *C7*. ‘I feel part of society’ and *C16*. ‘I find it easy to get to know new people’. These aspects of inclusion and social functioning might have been strengthened by the services in which the study population participated. Another test–retest study should thus be conducted without service activities during the test–retest period. Fourth, the dimensions of social functioning and social inclusion may be very sensitive to change in general, and they may be even more challenging to reliably measure in the target population. It has been suggested that another person, such as a family member, could assess social functioning objectively alongside the respondent’s self-report.⁴⁸ However, it is known that prolonged unemployment may lead to social exclusion, marginalisation, reduced self-esteem and feelings of shame.^{49–54} Therefore, it might be more encouraging for the target population for the assessment procedures to be based on self-reports.

As there are no gold standard measures of self-reported work ability and functioning for the population in a weak labour market position, we compared the results of the Abilitator with two validated, commonly used self-report instruments: the Work Ability Index (WAI)¹⁴ and the WHO Disability Assessment Schedule (WHODAS 2.0).¹⁵ The WAI is used by occupational health services and research to assess employee work ability.¹⁴ It has also been administered to the unemployed.⁵⁴ Overall, the WAI has been found to be a reliable instrument for self-reported work ability.^{55–60} The mean ICC values ranged from 0.72 to 0.84,^{55 56 60} and α levels from 0.65 to 0.82.^{57–59} WHODAS 2.0 is a generic PROM instrument for measuring health and disability.¹⁵ The test–retest reliability of WHODAS 2.0 was good, with ICC at the domain level ranging from 0.93 to 0.96, and at the item level from 0.69 to 0.89. The internal consistency of WHODAS 2.0 was $\alpha=0.96$ overall and ranged from 0.79 to 0.96 at the domain level.^{61 62} The WAI and WHODAS 2.0 studies had larger sample sizes and longer response periods than the current study of the Abilitator. Those studies with a test–retest period over 14 days may have tested the stability of the phenomenon rather than reliability of the method.⁵ Moreover, it was not clearly stated whether any intervention was implemented during the test–retest period, which was not the case in our study. The Abilitator reached similar test–retest reliability and internal consistency to that of the WAI and WHODAS 2.0. However, WHODAS 2.0 reached ICC levels high enough for clinical use more often than the WAI and the Abilitator. The main difference in the test–retest study results was that the WAI and WHODAS 2.0 had no items or sections with significant differences within the response period, but the Abilitator did, in summary scale *C*. *Inclusion*.

This study had several strengths. First, it was conducted in a realistic service setting with a target population that used real-life services. Second, the study covered a wide

range of participants within the target group, as the population in a weak labour market position is very heterogeneous. As anticipated, the ESF Priority 5 participants were in a weaker labour market position than the ESF Priority 3 participants. The larger population in the internal consistency analysis had poorer work ability and general functioning than the test–retest population. Despite these strengths, the study also had limitations. The first was that the population in the test–retest study used services within the response time period. Therefore, the results of the test–retest analyses might have been influenced by the activities conducted in the ESF projects. The second weakness of this study was the sample size. Even though it was sufficient for the test–retest analysis, a larger sample may have enabled subgroup analyses. For example, in terms of summary scale *G. Body*, we had to exclude those who used a mobility aid, because the sample size was too small for a reliable analysis in this subgroup.

There are also practical implications of this study. First, those who use the Abilitator can now assume that the instrument is reliable. However, the results are only generalisable to the target group and to those responding in Finnish. Second, the new information about the Abilitator’s SEM and SDC allows for easier interpretation of its results. Third, this study showed that *life satisfaction* is not a reliable aspect to measure in the target group. It also showed that rapid positive changes in social functioning and inclusion may occur when the target group is participating in low-threshold service activities such as those that the ESF projects provide.

CONCLUSION

This study showed that the Finnish version of the Abilitator is a reliable tool to use with individuals in a weak labour market position, and that it has acceptable to excellent intrarater test–retest reliability and internal consistency. Overall, the Abilitator meets the reliability requirements for systematic use in different services aimed at the target group. In most parts, the Abilitator also meets the reliability requirements for research and clinical use. However, in item *B1*, conclusions about the respondents’ life satisfaction should be drawn with caution, and further testing needs to be conducted to determine the reliability of summary scale *C*. *Inclusion*. Future research on the Abilitator’s psychometric properties should cover its structural and predictive validity and responsiveness. Moreover, item reduction and the definition of a clinically important change could promote the Abilitator’s usability and strengthen its position as a systematically implemented PROM in a wide range of primary level services.

Twitter Miia Wikström @WikstromMiia and Anne Kouvonon @AKouvonon

Acknowledgements The staff of the ESF Priority 5 Project Solmu participated in recruitment of ESF Priority 3 and 5 projects that took part in this study. The staff and participants of one ESF Priority 3 project, and three ESF Priority 5 projects volunteered to cooperate in data collection. In addition, Adjunct Professor Vesa A.

Niskanen, PhD from the University of Helsinki generously consulted with us in the data analysis.

Contributors MW was responsible for the study design, data collection, data analyses, and writing of the manuscript. MW is acting as the guarantor of this study. AK and MJ contributed to the study design, data analyses and critically revising the manuscript. MJ also participated in the data collection. MJ is the principal investigator of the Abilitator studies at FIOH. All the authors have read and approved the manuscript.

Funding This work was supported by the European Social Fund (ESF) Priority 5 programme via Solmu-project, grant number S20237. The funders played no role in the study design, data collection, data analysis, interpretation or writing of the report.

Competing interests All authors have completed the ICMJE uniform disclosure form at <http://www.icmje.org/disclosure-of-interest/> and declare: no support from any organisation for the submitted work; no financial relationships with any organisations that might have an interest in submitted work in the previous three years; no other relationships or activities that could appear to have influenced the submitted work.

Patient consent for publication The target group clients signed their informed consent to participate in the co-development workshops or cognitive interviews. The academic expert panels gave verbal consent to voluntarily participate as specialists to improve or co-develop the Abilitator's content.

Ethics approval All parts of this study were approved by the ethics board of the Finnish Institute of Occupational Health in June 2017, reference number ETR Joensuu 6/2017. The three external expert panels involved in the co-development process were informed both verbally and in writing of the research and voluntary participation was emphasised before co-development took place.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Access to the data can be requested from the corresponding author in line with the FIOH data sharing policy.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Miia Wikström <http://orcid.org/0000-0001-7737-8392>

Anne Kouvonen <http://orcid.org/0000-0001-6997-8312>

Matti Joensuu <http://orcid.org/0000-0001-9432-2278>

REFERENCES

- Rychetnik L, Frommer M, Hawe P, *et al.* Criteria for evaluating evidence on public health interventions. *J Epidemiol Community Health* 2002;56:119–27.
- PROMIS® Home. PROMIS Health Organization. Available: <https://www.promishealth.org/> [Accessed 9 Aug 2021].
- National Health Service (NHS). Guide to patient reported outcome measures (PROMs). NHS digital. Available: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/patient-reported-outcome-measures-proms/guide-to-patient-reported-outcome-measures-proms> [Accessed 9 Aug 2021].
- Canadian Institute for Health Information. Patient-Reported outcome measures (PROMs). CIHI. Available: <https://www.cihi.ca/en/patient-reported-outcome-measures-proms> [Accessed 9 Aug 2021].
- Finnish Institute for Health and Welfare (THL). The Finnish TOIMIA network. Available: <https://thl.fi/en/web/functioning> [Accessed 9 Aug 2021].
- de Vet H, Terwee C, Mokkink L, *et al.* *Measurement in medicine. A practical guide*. Cambridge: University Press, 2011.
- Weldring T, Smith SMS. Article commentary: patient-reported outcomes (PROs) and patient-reported outcome measures (PROMs). *Health Serv Insights* 2013;6:HSI.S11093–68.
- Valderas JM, Kotzeva A, Espallargues M, *et al.* The impact of measuring patient-reported outcomes in clinical practice: a systematic review of the literature. *Qual Life Res* 2008;17:179–93.
- Lohr KN, Zebrack BJ. Using patient-reported outcomes in clinical practice: challenges and opportunities. *Qual Life Res* 2009;18:99–107.
- Coulter A. Measuring what matters to patients. OECD health Ministers commit to patient reported measures of performance. *BMJ* 2017;356:1–2.
- Dawson J, Doll H, Fitzpatrick R, *et al.* The routine use of patient reported outcome measures in healthcare settings. *BMJ* 2010;340:c186.
- Brodersen J, Doward L, Thorsen H. Writing Health-Related Items for Rasch Models – Patient-Reported Outcome Scales for Health Sciences: From Medical Paternalism to Patient Autonomy. In: Christensen KB, Kreiner S, Mesbah M, eds. *Rasch models in health*, 2012.
- Finger ME, Escorpizo R, Bostan C, *et al.* Work rehabilitation questionnaire (WORQ): development and preliminary psychometric evidence of an ICF-based questionnaire for vocational rehabilitation. *J Occup Rehabil* 2014;24:498–510.
- Tuomi K, Ilmarinen J, Jahkola M. Työkyvyndeksi [Work Ability Index]. Työterveyssholto 19, 1997. Helsinki: Finnish Institute of Occupational Health
- Üstün T, Kostanjsek N, Chatterji S, eds. *Measuring Health and Disability: Manual for WHO Disability Assessment Schedule: WHODAS 2.0*. Geneva: World Health Organization, 2010.
- Tengland P-A. The concept of work ability. *J Occup Rehabil* 2011;21:275–85.
- Ilmarinen J, Gould R, Järviskoski A. Diversity of Work Ability. In: Gould R, Ilmarinen J, Järvisalo J, *et al.*, eds. *Dimensions of work ability. Results of the health 2000 survey*. Helsinki: Finnish Centre for Pensions, The Social Insurance Institution, National Public Health Institute, Finnish Institute of Occupational Health, 2008: 13–23.
- World Health Organization (WHO). *How to use the ICF. A practical manual for using the International classification of functioning, disability and health (ICF)*. Geneva: WHO, 2013.
- World Health Organization (WHO). *International classification of functioning, disability and health: ICF*. Geneva: WHO, 2001.
- Aalto A-M. Suositus psyykkisen toimintakyvyn mittaamiseksi väestötutkimuksissa [Recommendation for measuring psychological functioning in population research], 2011. The Finnish TOIMIA network. Available: <https://thl.fi/en/web/functioning/toimia-functioning-measures-database> [Accessed 9 Aug 2021].
- Tiikkainen P, Heikkinen R-L. Sosiaalisen toimintakyvyn arviointi ja mittaaminen väestötutkimuksissa [The evaluation and measurement of social functioning in public health studies], 2011. The Finnish TOIMIA network. Available: <https://thl.fi/en/web/functioning/toimia-functioning-measures-database> [Accessed 9 Aug 2021].
- Leemann L, Kuusio H, Hämäläinen R-M. Sosiaalinen osallisuus [Social Inclusion]. Sosiaalisen osallisuuden edistämisen koordinaatiohanke (Sokra), 2015. Helsinki; Finnish Institute for Health and Welfare. Finnish. Available: <https://thl.fi/en/web/thlfi-en/research-and-development/research-and-projects/project-to-co-ordinate-the-promotion-of-social-inclusion-sokra>
- Painter T, Stewart AL, Carey S. Measurement, and expectations. *Adv Renal Replace Ther* 1999;6:110–23.
- Tuulio-Henriksson A. Kognitiivisen toimintakyvyn arviointi väestötutkimuksissa [The evaluation of cognitive functioning in public health studies], 2011. The Finnish TOIMIA network. Available: <https://thl.fi/en/web/functioning/toimia-functioning-measures-database> [Accessed 9 Aug 2021].
- Hult M, Pietilä A-M, Koponen P, *et al.* Good work ability among unemployed individuals: association of sociodemographic, work-related and well-being factors. *Scand J Public Health* 2018;46:375–81.
- Hult M, Pietilä A-M, Saarinen T. Socio-demographic and work-related factors associated with the work ability of the unemployed. NES2016 - Ergonomics in Theory and Practice. *Proceedings of 48th Annual Conference of Nordic Ergonomics and Human Factors Society*, 2016:70–4.
- Szlachta E, Gawlik-Chmiel B, Kallus KW. Do the long-term unemployed regard themselves as able to work? *J Public Health* 2012;20:505–11.
- Pensola T, Järviskoski A, Järvisalo J. Unemployment and work ability. In: Gould R, Ilmarinen J, Järvisalo J, *et al.*, eds. *Dimensions of work*

- ability. *Results of the health 2000 survey*. Helsinki: Finnish Centre for Pensions, The Social Insurance Institution, National Public Health Institute, Finnish Institute of Occupational Health, 2008: 123–30.
- 29 Oivo T, Kerätär R. Osatyökykyisten reitit työllisyyteen -etuudet, palvelut, tukitoimet. [The paths of the partly disabled to employment – benefits, services and support]. Report 43, 2018. Ministry of Social Affairs and Health: Helsinki. Finnish. Available: <http://urn.fi/URN:ISBN:978-952-00-3997-4>
- 30 Kerätär R. You'll see more when you look from a distance. Multidimensional assessment of work ability and needs for rehabilitation among long-term unemployed. PhD thesis, 2016. Acta Universitatis Ouluensis D Medica 1340. Available: <http://urn.fi/urn:isbn:9789526210865>
- 31 Goud S, Fenger H. *National report on the labour market position of vulnerable groups in the Netherlands*. Rotterdam: Erasmus University, 2014. https://repub.eur.nl/pub/77323/Meis_204553.pdf
- 32 Arnkil R, Spangar T, Jokinen E. Selvitys heikossa työmarkkina- asemassa olevien palveluista Pohjoismaissa sekä Alankomaissa, Iso-Britanniassa, Saksassa ja Ranskassa [Report on the available services for the population in a poor labour market situation in the Nordic countries and in the Netherlands, Great Britain, Germany and France], 2012. Ministry of employment and the economy. Finnish. Available: <https://tem.fi/documents/1410877/3342347/Selvitys+heikossa+ty%C3%B6markkina-asemassa+olevien+palveluista+14022012.pdf/8a63619b-9fec-4936-858c-dfdc40679673/Selvitys+heikossa+ty%C3%B6markkina-asemassa+olevien+palveluista+14022012.pdf?t=1477044580000>
- 33 Wikström M, Anttila H, Savinainen M, et al. Development and content validity of the Abilitator: a self-report questionnaire on work ability and functioning aimed at the population in a weak labour market position. *BMC Public Health* 2020;20:327.
- 34 Kykyviisari, ©, Kykyviisari. Abilitator® website. Finnish Institute of Occupational Health. Available: Abilitator Kykyviisari <https://sivusto.kykyviisari.fi/> [Accessed 9 Aug 2021].
- 35 Savinainen M, Jauhainen O, Heikkilä H, et al. The concurrent validity of the Abilitator: a self-assessment questionnaire on functioning and work ability. *J Public Health* 2020;17.
- 36 Ministry of Economic Affairs and Employment. Local government pilots on employment - Ministry of Economic Affairs and Employment (tem.fi). Available: <https://tem.fi/en/local-government-pilots-on-employment> [Accessed 9 Aug 2021].
- 37 Työkykyohjelma - Sosiaali- ja terveysministeriö (stm.fi) [Work Ability program - Ministry of Social Affairs and Health]. Available: <https://stm.fi/en/work-ability-programme1> [Accessed 9 Aug 2021].
- 38 European social fund (ESF) programme for sustainable growth and jobs. Available: <http://www.rakennerahastot.fi/web/en> [Accessed 9 Aug 2021].
- 39 Bujang M, Baharum N. A simplified guide to determination of sample size requirements for estimating the value of intraclass correlation coefficient: a review. *Arch Orofac Sci* 2017;12:1–11.
- 40 Akoglu H. User's guide to correlation coefficients. *Turk J Emerg Med* 2018;18:91–3.
- 41 Mukaka MM. Statistics corner: a guide to appropriate use of correlation coefficient in medical research. *Malawi Med J* 2012;24:69–71.
- 42 Maher JM, Markey JC, Ebert-May D. The other half of the story: effect size analysis in quantitative research. *CBE Life Sci Educ* 2013;12:345–51.
- 43 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.
- 44 Giavarina D. Understanding Bland Altman analysis. *Biochem Med* 2015;25:141–51.
- 45 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60.
- 46 Terwee CB, Bot SDM, de Boer MR, et al. Quality criteria were proposed for measurement properties of health status questionnaires. *J Clin Epidemiol* 2007;60:34–42.
- 47 IBM SPSS statistics 27. IBM. Available: <https://www.ibm.com/products/spss-statistics> [Accessed 9 Aug 2021].
- 48 Simonen M. Sosiaalinen toimintakyky [Social functioning]. In: Heimonen S, Karvinen E, Pohjalainen P, et al, eds. *Toimintakyvyn laaja-alainen arviointi ja tukeminen [Holistic assessment and support of functioning]*. Helsinki: Age Institute, 2009. https://www.ikainstituutti.fi/content/uploads/2016/08/Oraita_1-2009-pdf.pdf
- 49 Oksanen K. Valtioneuvoston tulevaisuusselonteon 1. OSA. Jaettu ymmärrys työn murroksesta. Government Report on the Future, part 1. [A shared understanding of the transformation of work] Helsinki: Prime Minister's Office, 2017. Finnish. Available: <https://julkaisut.valtioneuvosto.fi/handle/10024/80036>
- 50 Tøge AG, Blekesaune M. Unemployment transitions and self-rated health in Europe: a longitudinal analysis of EU-SILC from 2008 to 2011. *Soc Sci Med* 2015;143:171–8.
- 51 Björklund O, Söderlund M, Nyström L, et al. Unemployment and health: experiences narrated by young Finnish men. *Am J Mens Health* 2015;9:76–85.
- 52 Schuring M, Robroek SJW, Otten FWJ, et al. The effect of ill health and socioeconomic status on labor force exit and re-employment: a prospective study with ten years follow-up in the Netherlands. *Scand J Work Environ Health* 2013;39:134–43.
- 53 Robroek SJW, Schuring M, Croezen S, et al. Poor health, unhealthy behaviors, and unfavorable work characteristics influence pathways of exit from paid employment among older workers in Europe: a four year follow-up study. *Scand J Work Environ Health* 2013;39:125–33.
- 54 Holm P, Jalava J, Ylöstalo P. Työttömien työkyky vuonna 2005 [Working capacity of the unemployed]. Työpoliittinen tutkimus 308., 2006. Helsinki: Pellervo economic research PTT. Finnish. Available: <https://www.ptt.fi/media/liitteet/rap195.pdf>
- 55 Silva Junior S da, Vasconcelos A, Griep R, Rotenberg L. Test-Retest reliability of the work ability index (WAI) in nursing workers. *Rev Bras Epidemiol* 2013;16.
- 56 Renosto A, Biz P, Hennigton E. Test-Retest reliability of the works ability index in metallurgical workers in southern Brazil. *Rev Bras Epidemiol* 2009;12.
- 57 Gould R, Koskinen S, Seitsamo J. Data and methods. In: Gould R, Ilmarinen J, Jarvisalo J, et al, eds. *Dimensions of work ability. Results of the health 2000 survey*. Helsinki: Finnish Centre for Pensions, The Social Insurance Institution, National Public Health Institute, Finnish Institute of Occupational Health, 2008: 25–32.
- 58 Serguskin L. Työkyvyn ja sen osa-alueiden pysyvyys ja muutos: esimiesten 10-vuotisseurantatutkimus [Permanence and change of work ability and its components: a 10-year follow-up study] A Master Thesis in psychology, 2008. University of Jyväskylä. Finnish. Available: <http://urn.fi/URN:NBN:fi:juyu-201010052935>
- 59 Radkiewicz P, Widerszal-Bazyl M. Psychometric properties of work ability index in the light of comparative survey study. *Int Congr Ser* 2005;1280:304–9.
- 60 de Zwart BCH, Frings-Dresen MHW, van Duivenbooden JC. Test-Retest reliability of the work ability index questionnaire. *Occup Med* 2002;52:177–81.
- 61 Ustün TB, Chatterji S, Kostanjsek N, et al. Developing the world Health organization disability assessment schedule 2.0. *Bull World Health Organ* 2010;88:815–23.
- 62 Garin O, Ayuso-Mateos JL, Almansa J, et al. Validation of the "World Health Organization Disability Assessment Schedule, WHODAS-2" in patients with chronic diseases. *Health Qual Life Outcomes* 2010;8:51.