Translation, validation and reliability of the Dutch Late-Life Function and Disability Instrument Computer Adaptive Test

Master thesis

Physiotherapy Science

Program in Clinical Health Sciences

Utrecht University

Name student:	Remco Arensman
Student number:	3857298
Date:	July 4th, 2014
Internship supervisor(s):	Dr. M.F. Pisters, Prof. Dr. R.A. de Bie
Internship institute:	Center for Physical Therapy Research and Innovation in Primary Care, Julius Health Care Centers & Department of Rehabilitation, Nursing Science and Sport, Brain Center Rudolf Magnus, University Medical Center Utrecht, Utrecht, The Netherlands
Lecturer/supervisor Utrecht University:	Dr. J. van der Net
Word count:	3780

"ONDERGETEKENDE

Remco Martijn Arensman,

bevestigt hierbij dat de onderhavige verhandeling mag worden geraadpleegd en vrij mag worden gefotokopieerd. Bij het citeren moet steeds de titel en de auteur van de verhandeling worden vermeld."

Examiner

Dr. M.F. Pisters

Assessors:

Dr. M.F. Pisters

Dr. C. Veenhof

Masterthesis, Physiotherapy Science, Program in Clinical Health Sciences, Utrecht University, Utrecht, 2014

SAMENVATTING

Introductie: Het adequaat kunnen meten van fysieke functie en beperkingen in participatie bij ouderen is van groot belang voor het inschatten en voorspellen van de zorgbehoeften van ouderen in de klinische praktijk. De Late-Life Function and Disablity Instrument Computer Adapted Test (LLFDICAT) is ontwikkeld voor het meten van fysieke functie en beperkingen in participatie in onderzoek naar ouderenzorg en de klinische praktijk en lijkt een veelbelovend instrument te zijn. Het doel van dit onderzoek was om de LLFDICAT te vertalen naar het Nederlands en het onderzoeken van de validiteit en betrouwbaarheid bij Nederlands sprekende thuiswonende ouderen.
Methode: De item bank van de LLFDICAT is vertaald door middel van een voorwaarts-terugwaarts vertaal methode. Een steekproef van 40 thuiswonende ouderen heeft deelgenomen aan het onderzoek. De deelnemers hebben de LLFDICAT, WHO-Disability Assessment Schedule 2.0 (WHODAS 2.0), PF-10 en de 10-meter loop test (10MLT) uitgevoerd. Na twee dagen, maar binnen twee weken vond een hertest van de LLFDICAT plaats. Pearson's r was berekend voor de validiteit en de ICC 2,1, SEM en SDC voor de betrouwbaarheid.
Resultaten: De resultaten lieten een correlatie zien tussen de LLFDICAT Functie schaal en de PF-10 van r = 0,804. Verder waren er negatieve correlaties met de WHODAS 2.0 (r = -0,657) en de 10MLT (r = -0,614). De ICC 2,1 voor de LLFDICAT Functie schaal was 0,904 (95% betrouwbaarheids interval 0,827-0,948), met een SEM van 2,51 en een SDC van 6,97. Voor de LLFDICAT Participatie schaal was de ICC 0,775 (95% betrouwbaarheids interval 0,613-0,874), met een SEM van 4,28 en SDC 11,86 points.
Conclusie: De Nederlandse LLFDICAT heeft een sterke validiteit en laat een goede betrouwbaarheid zien bij het meten van fysieke functie en beperkingen in participatie bij thuiswonende ouderen. Het is een kort, valide en betrouwbaar meetinstrument, geschikt voor gebruik in onderzoek en de klinische praktijk.

ABSTRACT

Background: Being able to adequately assess physical function and disability in the elderly population is vital for estimating and predicting healthcare needs in clinical practice. The Late-Life Function and Disability Instrument Computer Adapted Test (LLFDICAT) was developed to assess physical functioning and disability in gerontology research and clinical practice. The LLFDICAT appears to be a promising instrument for the assessment of physical function and disability in the elderly. The aim of this study was to translate the LLFDICAT to the Dutch language and to investigate the validity and reliability in a sample of Dutch-speaking community dwelling older persons.
Methods: The item bank of the LLFDICAT was translated using a forward-backward method. A sample of 40 older adults was recruited and completed the LLFDICAT, WHO-Disability Assessment Schedule 2.0 (WHODAS 2.0), PF-10 and 10-meter walk test (10MWT). The LLFDICAT was retested in two to ten days. Pearson's *r* was calculated to assess the concurrent validity of the LLFDICAT and the ICC 2,1, SEM and SDC were calculated to assess reliability.
Results: A correlation of *r* = 0,804 was found between the LLFDICAT Function scale and the PF-10. Additionally, the LLFDICAT Disability scale showed a negative correlation with both the WHODAS 2.0 (*r* = -0,657) and the 10MWT (*r* = -0,614).
The LLFDICAT Function scale showed an ICC of 0,904 (95% Confidence Interval 0,827-0,948), with a low SEM (2,51 points) and SDC (6,97 points). The LLFDICAT Disability scale showed slightly higher values for the ICC (0,775; 95% Confidence Interval 0,613-0,874) the SEM (4,28 points) and the SDC (11,86 points).
Conclusions: The Dutch LLFDICAT shows strong validity and high reliability when used to assess physical function and disability in community dwelling elderly.
The LLFDICAT is a short, valid and reliable instrument for use in both research and clinical practice settings.

Keywords:

Physical function, Disability, Late Life Function and Disability Instrument, Elderly, Validity, Reliability

INTRODUCTION

It is expected that before 2020, for the first time since the beginning of recorded history, the number of adults aged 65 and older will outnumber children under the age of 5.[1] This trend of global aging imposes new challenges on global health care in order to meet the needs of the aging population.[2] As people age, they often develop multiple diseases and require more health care services.[3] Being able to adequately assess physical function (PF) and disability in the elderly population is vital for estimating and predicting health care needs in clinical practice.[4, 5] As a result, PF and disability have become part of the comprehensive geriatric assessment used in geriatric clinical care and standard outcome measures in gerontology research.[6, 7]

Not surprisingly, a large number of measurement instruments have been developed to assess PF or disability of older persons in gerontology research and clinical practice.[8] Patient reported outcome measures are preferred because of their low cost and convenience.[9] However, patient reported outcome measures often suffer from limitations, such as measuring only a single construct, being multidimensional with no apparent conceptual structure, lacking sensitivity to detect important changes or having floor or ceiling effects when used for evaluative purposes.[10-12]

To overcome these limitations, the Late-Life Function and Disability Instrument (LLFDI) was developed.[13, 14] The LLFDI is a self-report questionnaire designed to assess PF and disability in older adults living in the community.[13, 14] It consists of 2 subscales, the 32-item function scale and the 16-item disability scale. The LLFDI showed excellent test-retest reliability for the function component (Intraclass Correlation Coefficient (ICC)=0.87-0.98) and moderate to good reliability for the disability component (ICC=0.68-0.91).[13-15] Both components showed expected differences in summary scores of known-functional groups supporting validity.[13, 14] Additionally, responsiveness to meaningful change, moderate to strong construct and predictive validity have been demonstrated.[16] However, the LLFDI suffers from two major limitations. Similar to comparable patient reported outcome measures for PF and disability, it takes a long time to complete the LLFDI (>20 minutes on average for the combined function and disability scales) and all questions are administered to all patients regardless of their applicability, making it difficult to use in clinical care.[17]

To alleviate respondent burden without sacrificing precision and sensitivity, a Computer Adaptive Test (CAT) version of the LLFDI was developed using item response theory methods.[7] Additionally, item response theory based CAT instruments have several other advantages over conventional instruments.[18] Firstly, using high precision CAT instruments reduces the number of subjects needed in research settings or smaller effect sizes can be found with the same number of subjects. Secondly, CAT instruments are suitable for monitoring individual patients in day to day clinical care due to the low respondent burden. The LLFDI-CAT appears to be a promising patient reported outcome measure for the assessment of PF and disability in the elderly. However, in order to use the LLFDI-CAT in research and clinical practice in the Netherlands, it has to be translated and the psychometric properties for the Dutch population of the elderly have to be investigated. Therefore, the aim

of this study is to translate the LLFDI-CAT to the Dutch language and to investigate the validity and reliability in Dutch-speaking community dwelling older persons.

METHODS

Study design

In the current study the LLFDI-CAT was translated to the Dutch language. To investigate the validity and reliability of the Dutch LLFDI-CAT a cross-sectional design with a test-retest moment was used. The current study was approved by the Medical Ethics Committee of the University Medical Centre Utrecht, Utrecht, the Netherlands.

Phase 1: Translation

The translation protocol was based on the guidelines proposed by Beaton et al.[19] However, some changes were made to accommodate the differences between translating a fixed item instrument and a CAT item bank. The most important changes were a change to stage II and the omission of the pre-test in a small sample from the target population. The protocol as used in this study can be seen table 1. Figure 1 shows the flow diagram of the translation procedure. After completing stage IV, the software for the Dutch translation of the LLFDI-CAT was produced by the original developer of the instrument.

Table 1 Translation procedure for the Late-Life Function and Disability Instrument Computer Adapted Test item bank

Stage of the translation	Procedure
Stage I	All the items from the LLFDI-CAT version database and all texts were translated into Dutch by two independent translators. Both translators were bilingual with Dutch as their mother tongue. The first translator was aware of the concept examined by the questionnaire. The second translator was unaware of the concept being investigated and had no clinical background. The goal was to create two translations reflecting clinical equivalence and the language used by the general population.
Stage II	The first translator created a first draft of the combined translation. The first draft was thoroughly checked by the second translator, who listed any inconsistencies or translations he did not agree with. The first draft and the list from the second translator were discussed during a meeting with both translators and an independent observer until consensus on the combined translation was reached. All changes made were registered by the observer. This stage was completed when the combined translation was finished.
Stage III	Two independent bilingual translators with English as their mother tongue and Dutch as their second language translated the common translation back to English. Both translators were unaware of the concept being explored and had no medical background. When both back-translations were finished, content agreement with the original version was checked by two independent reviewers to ensure consistent translation. Any inconsistencies or conceptual errors in the translation were documented and the corresponding items were changed.
Stage IV	An expert committee consisting of a methodologist, a medical professional, one of the forward and one of the backward translators and a language professional consolidated the final version of the translated item pool. The expert committee made sure the translation and adaptation were idiomatically, semantically, experientially and conceptually equivalent. Any issues encountered by the expert committee and the changes made to the items were documented. In a case of persisting disagreement on the translation, the original developer of the instrument was contacted in order to clarify the item and its underlying construct.

Phase 2: Validity and reliability study

Study population

Community dwelling independently ambulatory older adults were recruited for the validity and reliability study. Participants were recruited through convenience sampling in the region of Leiden and Utrecht in the Netherlands. A local physiotherapy practice was asked to hand out invitations and information letters to potential participants. Additionally, invitations and information letters were mailed to residents of senior citizen apartment buildings. The inclusion criteria were: (a) aged 65 years or older, (b) independently ambulatory with or

without an assistive device, (c) community dwelling, (d) must have provided informed consent prior to participation. Potential participants were excluded when they: (1) underwent joint replacement surgery in the lower extremities within the previous six months, (2) were hospitalized within the previous three months, (3) were unable to walk 10 meters without assistance from another person, (4) were living in a nursing home or a similar facility at the time of inclusion. All participants received written information about the aim and procedures of the study and informed consent was obtained from all participants prior to participation. Demographic data was collected and is listed in table 2.

Assessment of validity

To assess the measurement domain validity, construct validity of the Dutch LLFDI-CAT was assessed. Construct validity was defined as "the degree to which scores of an patient reported outcome instrument are consistent with hypotheses (for instance with regard to internal relationships, relationships to scores of other instruments or differences between relevant groups) based on the assumption that the patient reported outcome instrument validly measures the construct to be measured".[20] Measurements were completed by a researcher in the participant's own home environment or in a local physiotherapy practice. The participants were asked to complete the Dutch LLFDI-CAT on a laptop computer. Additionally, paper forms of the World Health Organization Disability Assessment Schedule (WHODAS 2.0) and the RAND-36 Physical Function scale (PF-10) were completed.[21, 22] Lastly, the 10-meter walk test (10MWT) was completed by all participants.[23] The selection of the comparators was based on the recommendation by Kayes et al. to use objective measurement instruments in conjunction with subjective measurement scales to measure complicated constructs and on the validation studies performed on the original instrument.[24, 25]

Assessment of reliability

The measurement domain reliability consists of internal consistency, reliability and measurement error. For this study, reliability and measurement error were assessed using a retest moment for the LLFDI-CAT. Participants were contacted by phone within two to 14 days after completion of the testing procedure. During this contact a researcher administered the LLFDI-CAT to obtain the retest data. Additionally, the time to complete the interview was recorded to assess the time it takes to complete the LLFDI-CAT.

The Late-Life Function and Disability Instrument Computer Adapted Test

To develop the LLFDI-CAT, a large item bank containing items for both the functioning and the disability scales of the LLFDI was constructed. Using item response theory methods, the items in the item bank were calibrated on a scale ranging from 0 to 100 with a mean of 50 and a standard deviation of 10.[7] All items received their own location on the scale and were ordered from "easy" to "difficult". The distances between the item locations on the scale are known. After all items were calibrated, a software program was developed for the LLFDI-CAT. The program selects an item from the middle of the scale and consequent items are selected

based on the responses to previous items. After an item is completed, the software calculates a patient score and standard error. When reaching a pre-set level of precision or when a pre-set number of items have been administered, the final patient score and standard error are calculated. To achieve precision and sensitivity levels similar to the original fixed item instrument, the software was programmed to stop when either 10 items per scale were completed or the patient score standard error was < 3,0.[7] The LLFDI-CAT showed high ICCs (>0.87) with the original fixed item instrument, while at the same time greatly reducing patient burden.[7]

Comparator instruments

The WHODAS 2.0 was designed to assess the functioning of an individual in six activity domains: understanding and communicating, getting around, self-care, getting along with people, life activities and participation in society.[21] The WHODAS 2.0 36-item self-report version was used in this study. The 36 items are scored on a 5-point Likert scale (none; mild; moderate; severe; extreme, cannot do). After scoring, an algorithm is used to convert the item scores to a scale score ranging from 0 (no disability) to 100 (full disability). Cronbach's α for the different subscales of the WHODAS 2.0 ranged from 0.7 to 0.97 in rehabilitation patients and 0.77 to 0.98 in patients with chronic diseases.[26, 27] Correlations between the WHODAS 2.0 score and the different domains of the SF-36 ranged from -0.29 to -0.65 in patients with chronic diseases.[27]

The PF-10 consists of 10 items designed to sample PF through three attributes: self-care, mobility and other physical activities and body movements.[22] Items are scored on a 3-point Likert scale (yes, limited a lot; yes, limited a little; no, not limited at all). The raw scores are converted to a scale score ranging from 0 to 100 with higher scores representing better PF. Internal consistency of the PF-10 has been reported as high (Cronbach's α=0.82) for older adults.[28] The PF-10 has been found to be a unidimensional index of PF in patients from seven different countries.[10, 29] The PF-10 has previously been used as a comparator in a validation study of the LLFDI Ffunction scale and showed high correlation (r=0.88).[30]

The 10MWT aims to measure walking speed.[23] The participants were asked to walk a distance of 10 meters at comfortable walking speed. Test-retest reliability of the 10MWT has been reported as excellent (ICC 0.96-0.98) with small standard error of measurement (0.004-0.008 m/s).[31] An argument can be made for the construct validity of gait speed in measuring disability as a recent review found that low gait speed is one of the most powerful predictors of disability in activities of daily living in community-dwelling elderly.[32]

Statistical analysis

Data was entered into SPSS 20.0 (SPSS Inc., Chicago, IL, USA) by a researcher, verified by a different researcher and consequently analysed.

The required sample size for the validity study was calculated using $\alpha = 0.05$; $\beta = 0.20$ and effect size $r = 0.50$ based on data from a validation study on the original instrument.[25] Additionally, for the reliability study a lowest acceptable ICC value of 0.60 ($\rho 0=0.60$) was

chosen based on the upper boundary of 'moderate' from the benchmarks proposed by Landis and Koch.[33] A value slightly lower than the lowest value for the ICC found in a recent reliability study of the LLFDI was chosen as the expected ICC (ρ1=0.80).[15] A required minimum sample size of 40 subjects was found using tabulated values provided by Walter et al. with ρ0=0.60; ρ1=0.80; α = 0.05; β = 0.20; and 2 measurements per subject.[34]

In order to assess concurrent validity, correlations using Pearson's *r* were calculated. Based on the results from previous validation studies and the minimum effect size required for a strong correlation according to Cohen's convention, the following hypotheses were set a priori[15, 25, 35]:

- The scores of the subjects on the Disability subscale of the LLFDI-CAT will have a negative correlation of at least 0.5 with their scores on the WHODAS 2.0
- The scores of the subjects on the Function subscale of the LLFDI-CAT will have a positive correlation of at least 0.5 with their scores on the PF-10
- The scores of the subjects on the Disability subscale of the LLFDI-CAT will have a negative correlation of at least 0.5 with their scores on the 10MWT

The validity of the LLFDI-CAT was interpreted using the conventions of Cohen for effect sizes of Pearson's *r* (0.10 small; 0.30 medium; 0.5 strong).[35]

To assess reliability, the ICC 2,1 absolute agreement was calculated.[36, 37] As high between-subjects variability inflates the ICC scores, classifying the ICC scores in categories such as low, medium or high provides little information.[38] However, the ICC combined with the Standard Error of Measurement (SEM) provide the information required to draw conclusions regarding the reliability of the LLFDI-CAT. Additionally, the 95% limits of agreement (LoA) were calculated as a graphic representation of agreement and the smallest detectable change (SDC) was used to provide clinicians with information regarding the change in scores required to measure 'true' change in an individual patient.[39, 40]

RESULTS

Demographics

The included sample for the validation and reliability study of the LLFDI-CAT consisted of 40 subjects and was predominantly female. More detailed information regarding the sample can be found in Table 2. The participants in the sample were predominantly female (72,5%) and chronic diseases reported were COPD (n=10), diabetes mellitus type I or II (n=5), cardiovascular disease (n=6), arthrosis (n=3), stroke (n=2) and other (n=11). The mean score was 50,89 (SD 7,95; range 36,81-76,43) on the LLFDI-CAT Function scale and 49,86 (SD 8,94; range 30,32-64,95) on the LLFDI-CAT Disability scale. The mean time required to administer the LLFDI-CAT as measured during the retest of the LLFDI-CAT was 8 minutes 59 seconds (range 5 minutes 45 seconds to 14 minutes 4 seconds). One participant had two missing scores on the PF-10. The missing values were imputed with the personal scale mean as suggested in the RAND-36 manual.[41] Additionally, for one participant body length was missing preventing calculation of the BMI.

Validity

The LLFDI-CAT Function scale and Disability scale showed absolute correlations exceeding r = 0,5 with the comparators. In Table 3 can be seen that the correlation between the LLFDI-CAT Function scale and the PF-10 (r = 0,804) confirms the hypothesis that participant scores on the Function subscale of the LLFDI-CAT have a positive correlation of at least 0,5 with their scores on the PF-10. Additionally, the LLFDI-CAT Disability scale shows a negative correlation with both the WHODAS 2.0 (r = -0,657) and the 10MWT (r = -0,614), confirming both hypotheses regarding the validity of the LLFDI-CAT Disability scale.

Reliability

The results from the reliability study are shown in Table 4. All participants were available for the retest of the LLFDI-CAT and the average number of days between the test and retest ranged from two to eight (mean 4,48 days). The LLFDI-CAT Function scale showed an ICC value of 0,904 (95% Confidence Interval 0,827-0,948), with a SEM (2,51 points) and SDC (6,97 points). The LLFDI-CAT Disability scale showed a value of 0,775 (95% Confidence Interval 0,613-0,874) for the ICC, 4,28 for the SEM and 11,86 for the SDC. Figures 2 and 3 agreement between test and retest scores for the LLFDI-CAT scales is shown in Bland-Altman plots with the LoA.

Table 2 Characteristics of the study sample

	Study sample n = 40			
	n	%	Mean (SD)	Range
Age, in years			79,20 (7,9)	65-93
Female	29	72,5		
BMI	39		27,4 (5,1)	20-45
Marital status				
Single	4	10		
Married	16	40		
Widow/widower	20	50		
Chronic Diseases				
None	15	37,5		
One	16	40		
Two	5	12,5		
Three or more	4	10		
Walking aid				
None	22	55		
Cane or wheeled walker	17	42,5		
Walking aid and wheelchair	1	2,5		
Education				
< 6 years	17	42,5		
6 to 12 years	22	55		
> 12 years	1	2,5		

SD = Standard Deviation; BMI = Body Mass Index; LLFDI CAT = Late Life Function and Disability Instrument Computer Adapted Test

Table 3 Correlations of the Late Life Function and Disability Instrument Computer Adapted Test version sub scales with the comparators using Pearson's *r and the Coefficients of Determination*

	WHODAS-2.0	PF-10	10MWT
LLFDI CAT Function scale		0,804*†	
LLFDI CAT Disability scale	-0,657*		-0,614*

*WHODAS 2.0 = World Health Organisation Disability Assessement Schedule 2.0; CoD = Coefficient of Determination; PF-10 = RAND-36 Physical Function scale; 10MWT = 10 Meter Walk Test; LLFDI CAT = Late Life Function and Disability Instrument Computer Adapted Test; * = significant correlation at the p<0,000 level; † = 39 cases included in the analysis*

Table 4 Intraclass Correlation Coefficients, Standard Error of Measurement and Smallest Detectable Change of the Late Life Function and Disability Instrument Computer Adapted Test sub scales

	Test mean (SD)	Retest mean (SD)	$ICC_{2,1}$	SEM	SDC
LLFDI CAT Function scale	50,89 (7,95)	51,47 (8,26)	0,904	2,51	6,97
LLFDI CAT Disability scale	49,86 (8,94)	49,27 (9,12)	0,775	4,28	11,86

SD = Standard Deviation; ICC = Intraclass Correlation Coefficient; SEM = Standard Error of Measurement; SDC = Smallest Detectable Change; LLFDI CAT = Late Life Function and Disability Instrument Computer Adapted Test

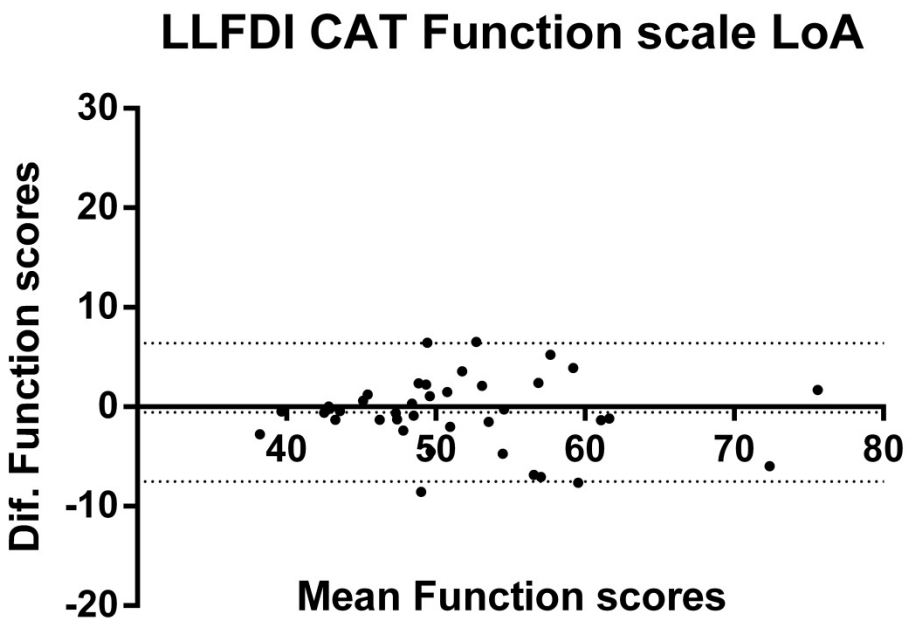

Figure 1 Bland-Altman plot of the test and retest scores of the Late Life Function and Disability Computer Adapted Test Function scale with 95% limits of agreement

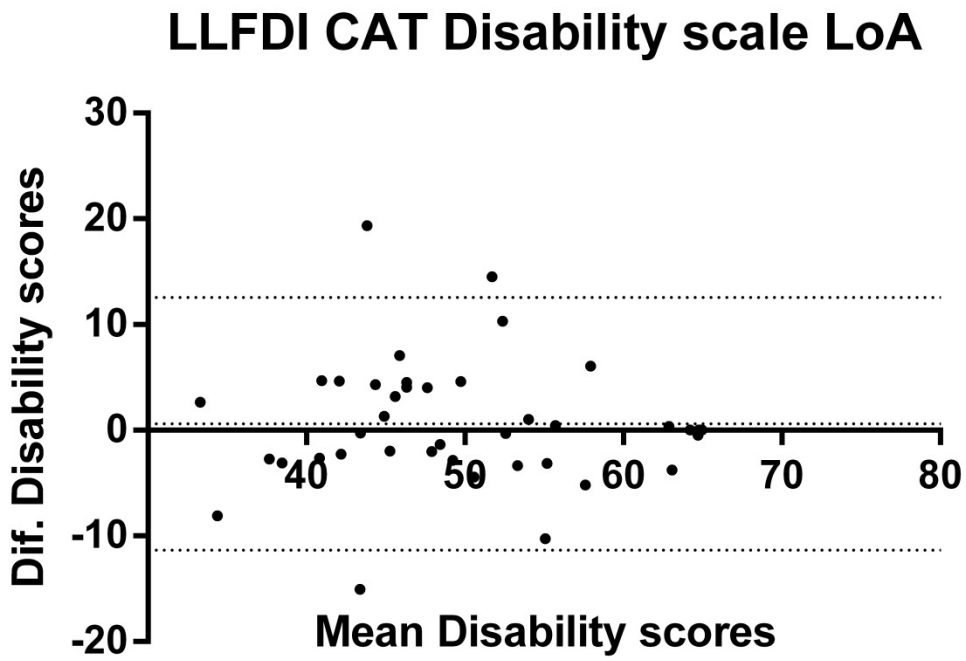

Figure 2 Bland-Altman plot of the test and retest scores of the Late Life Function and Disability Computer Adapted Test Disability scale with 95% limits of agreement

DISCUSSION

The aim of this study was to translate the LLFDI-CAT to the Dutch language and to investigate the concurrent validity and reliability for Dutch-speaking community dwelling older persons. The results found in the validity study confirm the construct validity of both the Function and the Disability scales of the LLFDI-CAT. All hypotheses set were confirmed by the strong correlations found. The correlation found between the LLFDI-CAT Function scale and the PF-10 is of the magnitude and direction as was expected. The strong correlation of 0,804 confirms the construct validity of the LLFDI-CAT Function scale. This is not surprising as both instruments aim to measure PF through the use of items reflecting physical abilities related to the use of the upper or lower extremities. Currently, there have been no other studies reporting on the construct validity of the LLFDI-CAT Function scale. Instead, the results from the present study were compared with the results from studies investigating the construct validity of the LLFDI. The correlation between the LLFDI-CAT Function scale and the PF-10 found in this study are in accordance with values found by Dubuc et al. (r = 0,85) and Lapier et al. (r = 0,83) for the original LLFDI.[42, 43] However, Roaldsen et al. reported a lower correlation of r = 0,52.[15] As the correlations found in this study are similar or higher than those of the LLFDI, the construct validity of the LLFDI-CAT Function scale is further supported.[15]

The correlations found between the LLFDI-CAT Disability scale and the WHODAS 2.0 and 10MWT were strong according to Cohen's convention, confirming its construct validity.[35] A possible explanation for the lower values when compared with the LLFDI-CAT Function scale is that disability is a broad construct. As a result, fixed item instruments lack the high number of items required to measure the entirety of the construct.[44] Similarly, a performance test like the 10MWT does not capture the parts of disability caused by mental health, social or environmental factors. Therefore, it was expected that the construct "disability" was only partly measured by the comparators, resulting in lower correlations. This is further supported by the theoretical basis on which the respective instruments were based. The LLFDI-CAT Disability scale aims to measure disability in the elderly, where both the PF-10 and the WHODAS 2.0 were designed for the general population.[10, 13, 14, 21] There are currently no known studies using the WHODAS 2.0 or the 10MWT as comparators for the construct validity of the LLFDI or the LLFDI-CAT Disability scale. However, the LLFDI has been compared with other self-report questionnaires, such as the London Handicap Scale (r = 0,47-0,66), the Western Ontario and McMasters Universities Osteoarthritis Index (r = -0,47) and a performance test, the 20-meter walk test (r = 0,37).[42, 43, 45] The correlations found in these studies are similar or lower than the correlations found in the current study further confirming the construct validity of the LLFDI-CAT Disability scale.

The ICC 2,1 scores of 0,904 for the Function scale and 0,775 for the Disability scale show high reliability and are sufficiently high to indicate that the instrument is stable over repeated measurements when no change is expected. Moreover, the narrow LoA as reported in Figures 2 and 3 show high agreement between measurements. However, as can be seen in Figure 3, the LoA for the Disability scale are wider due to six cases. It is unclear what caused the large

difference in scores between the repeated measurements as the involved cases did not differ from the other cases in any of the other variables. A possible explanation could be that some items from the LLFDI-CAT disability scale were not consistently interpreted by these select cases. However, this could not be verified in the current study. The low SEM (2,51-4,28) further support reliability as the SEM quantifies the error of the LLFDI-CAT and the random variation in repeated measurements.[46] The SDC values (6.97-11,86) derived from the SEM show the minimal change scores required to exceed the measurement error of the instrument. However, more research is required to assist in establishing minimal clinically relevant changes. The reliability findings are consistent with reliability studies on the LLFDI (ICC range 0,44-0,98), the Hebrew translation (ICC range 0,46-0,90) and the Swedish translation of the LLFDI (ICC range 0,82-0,91).[13-15, 47] Only one study has reported on the SDC of the LLFDI and found scores almost identical to the SDC scores found in this study (2,9 for the Function scale and 4,1 for the Disability scale), further supporting the excellent reliability and precision of the LLFDI-CAT.[15] In addition to high precision, the other advantage of a CAT instrument over fixed item instruments is the low time required for completion of the instrument. Completing the LLFDI-CAT took less than 9 minutes, reducing patient burden by as much as 50% compared with the fixed item LLFDI.[7]

This study has some limitations. First, this study is the first to attempt to translate the CAT version of the LLFDI and one of first studies in which an existing CAT instrument is translated and validated. Consequently, no guidelines or protocols for the translation and validation procedure of existing CAT instruments exist. To overcome this problem, an existing protocol originally designed for use in the translation of fixed item instruments was adapted and used for the translation of the LLFDI-CAT.[19] A compromise made during the translation process was the omission of the pre-test of the translated item bank in a small sample of older adults. As a consequence, some items could have been improperly translated. To overcome this problem a language expert was asked to review the translated item bank to ensure proper translation and to make sure the items were easy to understand for older adults. As the use of CAT instruments becomes more prevalent in research and clinical practice, the need for high quality translations will rise. The practical problems encountered during the translation process in the current study highlight the need for guidelines on the translation of CAT instruments.

Secondly, the high percentage of women in the sample can make the results difficult to generalize. However, as the age of the population advances the percentage of women increases to up to 72,1% in 90-year olds.[48] The mean age of the participants in this study was 79,2 years with 72,5% being women, therefore the number of women in the sample was only slightly higher than in the general population of the same age.

Lastly, no performance tests were used as a comparator for the LLFDI-CAT Function scale. As a result, it is possible the construct "function" the LLFDI-CAT Function scale aims to measure is not adequately assessed. However, Bean et al. found that performance tests and PRO measures assess different aspects of older person's functioning.[49] The PRO measures are more complete as they also associate with psychosocial and health factors, where

performance tests do not.
The thoroughness with which the validity and reliability of the LLFDI-CAT were tested provide a clear understanding of its psychometric properties. The LLFDI-CAT appears to be a valid and reliable instrument for the assessment of PF and disability in community dwelling elderly. Additionally, the high precision and low patient burden make the instrument preferable in both research and clinical care settings as a replacement of traditional fixed item instruments for evaluative purposes. However, future research should further expand on this study by investigating the responsiveness of the LLFDI-CAT. When responsiveness of the LLFDI-CAT has been confirmed, it can be confidently used to assess the effectiveness of treatment strategies aimed at improving PF and disability in the elderly.

CONCLUSION

The Dutch LLFDI-CAT shows good construct validity and high reliability for assessment of PF and disability in community dwelling elderly and can be used for evaluative purposes in clinical practice. Furthermore, the advantages of the LLFDI-CAT over traditional instruments make it preferable over traditional fixed item instruments.

REFERENCES

1. National Institute on Aging, National Institutes of Health, U.S. Department of Health and Human Services. Global Health and Aging. 2011;11-7737:2. Available from: http://www.nia.nih.gov/sites/default/files/nia-who_report_booklet_oct-2011_a4__1-12-12_5.pdf. Accessed 10-10-2013.

2. Economist Intelligence Unit. Healthcare strategies for an ageing society. 2009. Available from: http://www.eiu.com/site_info.asp?info_name=Philips_ageing&page=noads. Accessed 10-10-2013.

3. Arai H, Ouchi Y, Yokode M, et al. Toward the realization of a better aged society: messages from gerontology and geriatrics. *Geriatr Gerontol Int*. 2012;12:16-22.

4. Guralnik JM, Simonsick EM, Ferrucci L, et al. A short physical performance battery assessing lower extremity function: association with self-reported disability and prediction of mortality and nursing home admission. *J Gerontol*. 1994;49:M85-94.

5. Guralnik JM, Ferrucci L, Simonsick EM, Salive ME, Wallace RB. Lower-extremity function in persons over the age of 70 years as a predictor of subsequent disability. *N Engl J Med*. 1995;332:556-561.

6. National Institutes of Health Consensus Development Conference Statement: geriatric assessment methods for clinical decision-making. *J Am Geriatr Soc*. 1988;36:342-347.

7. Jette AM, Haley SM, Ni P, Olarsch S, Moed R. Creating a computer adaptive test version of the late-life function and disability instrument. *J Gerontol A Biol Sci Med Sci*. 2008;63:1246-

1256.

8. McDowell I, Newell C. *Measuring Health: A Guide to Rating Scales
and Questionnaires. 2nd Ed. New York: Oxford University Press; 1996.* 2nd ed ed. New York: Oxford University Press; 1996.

9. Verbrugge LM, Jette AM. The disablement process. *Soc Sci Med*. 1994;38:1-14.

10. Haley SM, McHorney CA, Ware JE,Jr. Evaluation of the MOS SF-36 physical functioning scale (PF-10): I. Unidimensionality and reproducibility of the Rasch item scale. *J Clin Epidemiol*. 1994;47:671-684.

11. Hays RD, Sherbourne CD, Mazel RM. The RAND 36-Item Health Survey 1.0. *Health Econ*. 1993;2:217-227.

12. McHorney CA. Methodological and psychometric issues in health status assessment across populations and applications. In: Albrecht GL, Fitzpatrick R, eds. *Advances in Medical Sociology*. Vol 5. Greenwich, CT: JAI Press LTD; 1994:281-282-304.

13. Jette AM, Haley SM, Coster WJ, et al. Late life function and disability instrument: I. Development and evaluation of the disability component. *J Gerontol A Biol Sci Med Sci*. 2002;57:M209-16.

14. Haley SM, Jette AM, Coster WJ, et al. Late Life Function and Disability Instrument: II. Development and evaluation of the function component. *J Gerontol A Biol Sci Med Sci*. 2002;57:M217-22.

15. Roaldsen KS, Halvarsson A, Sarlija B, Franzen E, Stahle A. Self-reported function and disability in late life - cross-cultural adaptation and validation of the Swedish version of the late-life function and disability instrument. *Disabil Rehabil*. 2013.

16. Beauchamp MK, Schmidt CT, Pedersen MM, Bean JF, Jette AM. Psychometric properties of the Late-Life Function and Disability Instrument: a systematic review. *BMC Geriatr*. 2014;14:12-2318-14-12.

17. Gignac MA, Cao X, Mcalpine J, Badley EM. Measures of disability: Arthritis Impact Measurement Scales 2 (AIMS2), Arthritis Impact Measurement Scales 2-Short Form (AIMS2-SF), The Organization for Economic Cooperation and Development (OECD) Long-Term Disability (LTD) Questionnaire, EQ-5D, World Health Organization Disability Assessment Schedule II (WHODASII), Late-Life Function and Disability Instrument (LLFDI), and Late-Life Function and Disability Instrument-Abbreviated Version (LLFDI-Abbreviated). *Arthritis Care Res (Hoboken)*. 2011;63 Suppl 11:S308-24.

18. Bjorner JB, Chang CH, Thissen D, Reeve BB. Developing tailored instruments: item banking and computerized adaptive assessment. *Qual Life Res*. 2007;16 Suppl 1:95-108.

19. Beaton DE, Bombardier C, Guillemin F, Ferraz MB. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine (Phila Pa 1976)*. 2000;25:3186-3191.

20. Mokkink LB, Terwee CB, Patrick DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of clinical epidemiology JID - 8801383*.

2010;63:737-745.

21. World Health Organization. The World Health Organization Disability Assessment Schedule Phase II Field Trial Instrument. Geneva, Switzerland: The World Health Organization; 1999.

22. Stewart A, Ware J. *Measuring Functioning and Well-being: The Medical Outcomes Study Approach*. Durham, NC: Duke University Press; 1991.

23. Collen FM, Wade DT, Bradshaw CM. Mobility after stroke: reliability of measures of impairment and disability. *Int Disabil Stud*. 1990;12:6-9.

24. Kayes NM, McPherson KM. Measuring what matters: does 'objectivity' mean good science? *Disabil Rehabil*. 2010;32:1011-1019.

25. Sayers SP, Jette AM, Haley SM, Heeren TC, Guralnik JM, Fielding RA. Validation of the Late-Life Function and Disability Instrument. *J Am Geriatr Soc*. 2004;52:1554-1559.

26. Pösl M, Cieza A, Stucki G. Psychometric properties of the WHODASII in rehabilitation patients. *Qual Life Res*. 2007;16:1521-1531.

27. 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;8:51-7525-8-51.

28. Bohannon RW, DePasquale L. Physical Functioning Scale of the Short-Form (SF) 36: internal consistency and validity with older adults. *J Geriatr Phys Ther*. 2010;33:16-18.

29. Raczek AE, Ware JE, Bjorner JB, et al. Comparison of Rasch and summated rating scales constructed from SF-36 physical functioning items in seven countries: results from the IQOLA Project. International Quality of Life Assessment. *J Clin Epidemiol*. 1998;51:1203-1214.

30. Hand C, Richardson J, Letts L, Stratford P. Construct validity of the late life function and disability instrument for adults with chronic conditions. *Disabil Rehabil*. 2010;32:50-56.

31. Peters DM, Fritz SL, Krotish DE. Assessing the reliability and validity of a shorter walk test compared with the 10-Meter Walk Test for measurements of gait speed in healthy, older adults. *J Geriatr Phys Ther*. 2013;36:24-30.

32. Vermeulen J, Neyens JC, van Rossum E, Spreeuwenberg MD, de Witte LP. Predicting ADL disability in community-dwelling elderly people using physical frailty indicators: a systematic review. *BMC Geriatr*. 2011;11:33-2318-11-33.

33. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33:159-174.

34. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med*. 1998;17:101-110.

35. Cohen J. A power primer. *Psychol Bull*. 1992;112:155-159.

36. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull*. 1979;86:420-428.

37. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients.

Psychol Methods. Mar 1996;1:30-46.

38. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. *J Strength Cond Res*. 2005;19:231-240.

39. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*. 1986;1:307-310.

40. de Vet HC, Terwee CB, Knol DL, Bouter LM. When to use agreement versus reliability measures. *J Clin Epidemiol*. 2006;59:1033-1039.

41. van der Zee K, Sanderman R. *Het Meten Van De Algemene Gezondheidstoestand
met De RAND-36 : Een Handleiding.* second ed. Groningen: Noordelijk Centrum voor Gezondheidsvraagstukken, NCG.; 2012.

42. Dubuc N, Haley S, Ni P, Kooyoomjian J, Jette A. Function and disability in late life: comparison of the Late-Life Function and Disability Instrument to the Short-Form-36 and the London Handicap Scale. *Disabil Rehabil*. 2004;26:362-370.

43. Lapier TK. Utility of the late life function and disability instrument as an outcome measure in patients participating in outpatient cardiac rehabilitation: a preliminary study. *Physiother Can*. 2012;64:53-62.

44. Jette AM. Assessing disability in studies on physical activity. *Am J Prev Med*. 2003;25:122-128.

45. Riddle DL, Jensen MP. Construct and criterion-based validity of brief pain coping scales in

persons with chronic knee osteoarthritis pain. *Pain Med*. 2013;14:265-275.

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

47. Melzer I, Kurz I, Sarid O, Jette AM. Relationship between self-reported function and disability and balance performance measures in the elderly. *J Rehabil Res Dev*. 2007;44:685-691.

48. Giesbers H, Verweij A, Beer Jd. Bevolking: Wat is de huidige situatie? *Volksgezondheid Toekomst Verkenning, Nationaal Kompas Volksgezondheid*. 2014.

49. Bean JF, Olveczky DD, Kiely DK, LaRose SI, Jette AM. Performance-based versus patient-reported physical function: what are the underlying predictors? *Phys Ther*. 2011;91:1804-1811.

Figures and tables

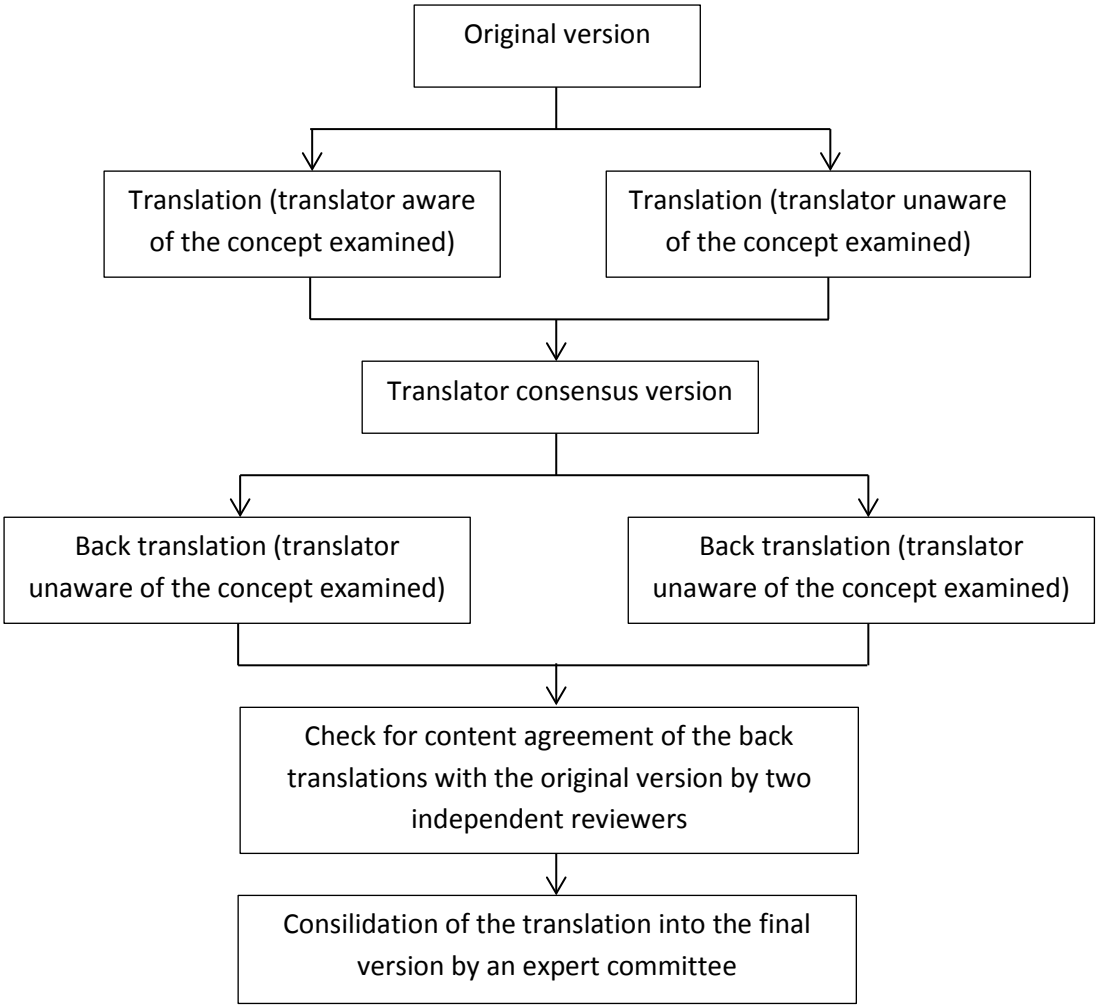

Figure 3. The translation and cross-cultural adaptation procedure of the Dutch Late-Life Function and Disability Instrument Computer Adapted Test version