Cross-Cultural Validity and Construct Validity of the Dutch-Flemish PROMIS® Upper Extremity Item Bank V2.0

Masterthesis

Physiotherapy Science

Program in Clinical Health Sciences

Utrecht University

Name student:	E.J.A. (Erik-Jan) Haan
Student number:	5647770
Date:	June 25, 2018
Internship supervisor(s):	Dr. C.B. Terwee, Dr. L.D. Roorda
Internship institute:	Department of Epidemiology and Biostatistics Amsterdam Public Health research institute VU University Medical Centre, Amsterdam, The Netherlands
Lecturer/supervisor Utrecht University:	Dr. M.F. Pisters

“ONDERGETEKENDE

Erik-Jan Adriaan Haan

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. C.B. Terwee

Dr. J. van der Net

Masterthesis, Physical Therapy Sciences, Program in Clinical Health Sciences, Utrecht University, Utrecht, 2018

ABSTRACT

Aim The Patient Reported Outcomes Measurement Information System (PROMIS) Upper Extremity (UE) item bank v2.0 has been developed in the USA but is not yet available for The Netherlands or Flanders. The aim of this study was to evaluate cross cultural validity and construct validity of the Dutch-Flemish PROMIS UE item bank (v2.0) in Dutch patients with musculoskeletal upper extremity disorders.

Methods After translation, the Dutch-Flemish PROMIS UE item bank and several legacy instruments were administered through a questionnaire to adult Dutch patients visiting an orthopedic outpatient clinic. Cross-cultural validity was evaluated with Differential Item Functioning (DIF) for language (Dutch vs. English). To evaluate construct validity, hypotheses were tested for correlations between the item bank and the legacy instruments.

Results The scores of 205 Dutch patients on the item bank were compared to those of 246 US participants. Eight items showed minimal DIF for language which resulted in sufficient cross-cultural validity. The Dutch-Flemish PROMIS UE item bank had a moderate correlation with the Dutch-Flemish PROMIS Pain Intensity item ($r = -0.43$) and strong correlations ($r \geq 0.7$) with the other instruments, all correlations were as expected.

Conclusion The Dutch-Flemish PROMIS UE item bank (v2.0) has sufficient cross-cultural validity and construct validity. The next step is to further validate the item bank by full item bank calibration.

Clinical Relevance After full item bank calibration is established, the Dutch-Flemish PROMIS UE item bank (v2.0) can serve as a basis for short forms and Computerized Adaptive Testing, resulting in high measurement precision and low patient burden. This will improve quality and efficiency in clinical outcome measurement in Dutch and Flemish patients with upper extremity conditions.

Keywords: Patient-Reported Outcomes Measurement Information System; PROMIS; Upper Extremity; Validation.

INTRODUCTION

Upper extremity musculoskeletal disorders are a common health problem, with estimated point prevalence rates ranging from 1.6 to 53%, and are a huge economic burden for society [34]. With the ageing of the global population, the burden of this condition is expected to further increase, with subsequent increasing costs for health care utilization [60]. To deal with increasing health care costs, there is a growing interest in Value Based Health Care (VBHC). In VBHC, achieving high value for patients is the main goal, where value is defined as outcomes relative to costs [45, 47]. Porter et al. emphasize the importance of patient-reported outcomes in VBHC [46].

One of the proposed patient-reported outcomes in VBHC is functional status, because improving functional status is one of the reasons why patients seek care [46]. For upper extremity musculoskeletal disorders, many different Patient Reported Outcome Measures (PROMs) for functional status are used in research and in clinical healthcare [5, 27, 30, 32, 56, 62]. This variety of different PROMs hampers comparability of scores across conditions. Also, not all PROMs demonstrated strong evidence for all measurement properties [62]. In general, traditional PROMs can lead to incomplete questionnaires, contain irrelevant questions and place a high burden on respondents [6, 53]. So, currently used PROMs do not always meet the recommended minimum standards [51].

The Patient-Reported Outcomes Measurement Information System (PROMIS) was initiated by six U.S.-based academic institutions and the National Institutes of Health (NIH), with the aim to improve measurement quality and comparability of health outcomes measures and reduce the burden for respondents. This was realized by building and validating item banks for measuring specified function and health status domains [8, 9]. An item bank is a series of questions or items, all measuring the same domain, independent from disease [52]. The items in an item bank are all calibrated on the same scale, using Item Response Theory (IRT) modelling. In this way, reliable, precise and valid measuring results can be obtained. Also, IRT based item banks enable the use of short forms (subsets of items from the item bank) and Computerized Adaptive Testing (CAT). CAT uses an algorithm that selects the most informative items from the item bank, based on the individual's response to previously administered items. In this way, high measurement precision can be obtained with low respondent burden [7, 50].

The PROMIS system includes a Physical Function (PF) item bank consisting of items covering central (i.e. spinal), and upper and lower extremity functions and activities of daily living [54, 55]. From this full PF item bank, subsets of items for measuring upper extremity and lower extremity function (PROMIS UE and PROMIS Mobility) were developed, to improve the application in specific clinical settings [29]. However, the initial PROMIS UE item bank (v1.2) consisting of 16 items, showed a ceiling effect in several clinical studies with patients with upper extremity conditions, thus warranting expansion of the UE bank with items measuring higher levels of functioning [1-3, 35, 37]. Consequently, a new UE item bank (v2.0), consisting of 42 items from the v1.2 PF item bank plus four newly developed items, has recently been developed in the USA. The PROMIS UE v2.0 item bank assesses a wider range of upper extremity functioning and has higher precision when used in impaired individuals [36].

In the Netherlands, the Dutch-Flemish PROMIS Group translated, among others, the PROMIS PF item bank (v1.2) [54, 55, 61]. The item bank v1.2 has been validated in Dutch patients with chronic pain [18], Dutch patients with rheumatoid arthritis [43, 44] and in Dutch patients receiving physical therapy [19].

The PROMIS UE v2.0 item bank is not yet available for Dutch-Flemish populations but is desirable for clinical research and clinical application in specific populations with upper extremity impairments. Four newly developed PROMIS v2.0 items addressing upper extremity function have not been translated into Dutch-Flemish. After translation of the v2.0 items, the psychometric properties of the Dutch-Flemish PROMIS UE item bank need to be established. Therefore, the aim of the current study was to translate four new v2.0 items addressing upper extremity function and to examine the cross-cultural validity and construct validity of the Dutch-Flemish PROMIS UE item bank (v2.0) in a Dutch population of patients with musculoskeletal upper extremity disorders.

METHODS

Translation

This study was part of a larger study in which 45 newly developed PF items were translated into Dutch-Flemish, to update the v1.2 PF item bank to v2.0. The translation included the four items for the Dutch-Flemish PROMIS UE item bank (v2.0) (see Table 1).

Table 1. Description of PROMIS v2.0 items for the PROMIS UE item bank

Item name	Description
PFM2	Are you able to lift a heavy painting or picture to hang on your wall above eye-level?
PFM16	Are you able to pass a 20-pound (10 kg) turkey or ham to other people at the table?
PFM18	Are you able to continuously swing a baseball bat or tennis racket back and forth for 5 minutes?
PFC8	Does your health now limit you in opening a previously opened jar?

The translation process was performed similarly as in a previous translation of Dutch-Flemish PROMIS item banks, using state of the art methodology [4, 26, 61]. In short, the process involved 2 forward translations (1 Dutch and 1 Flemish), 1 reconciled version, 1 back translation by a native English speaker, comparison of original with back translation, and reviews by 3 bilingual experts (2 Dutch and 1 Flemish). To evaluate the comprehensibility and relevance of the items, cognitive interviews were conducted with native Dutch and Flemish patients and people from the general population.

Design and study participants for the validation study

For the validation study a cross-sectional design was used. Patients who visited the outpatient clinic of the orthopedic department of OLVG, a general hospital in Amsterdam, the Netherlands, were invited to participate. To be eligible, patients had to have a musculoskeletal disorder of the upper

extremity, had to be 18 years or older, and had to provide informed consent. Patients who were not able to read and/or write in Dutch language were excluded from this study. Also, patients who participated in another study at the orthopedic department at the same time, were excluded.

To evaluate the cross-cultural validity of the Dutch-Flemish PROMIS UE item bank with the US PROMIS UE item bank, data from a US subsample were used [36]. The subsample consists of 246 patients from an online panel, who endorsed having some difficulty due to upper extremity pain or function, aged 18 years or older.

Procedures

Patients visiting the outpatient clinic of the orthopedic department between February and May 2018, were invited to fill in a web-based (digital) or paper-and-pencil (paper) questionnaire that included, among other measures, the Dutch-Flemish PROMIS UE item bank. The study was approved by the local institutional review boards of Slotervaart/Reade (reference number P1749) and of OLVG.

Measures

First, the questionnaire included questions addressing demographic (i.e. age, gender, country of birth, educational level) and clinical characteristics (i.e. disease duration, type of disorder, location of pain).

Second, the questionnaire included the Dutch-Flemish PROMIS UE item bank (v2.0). The item bank contains 46 items addressing upper extremity function. There are two different 5-point Likert scale response scales: 1) Unable to do/With much difficulty/With some difficulty//With a little difficulty/Without any difficulty; 2) Cannot do/Quite a lot/Somewhat/Very little/Not at all. There is no timeframe for the items, but current status is inferred. Higher scores indicate better function. The total score of the Dutch-Flemish PROMIS UE item bank is expressed as a T-score which is a standardized score, with 50 representing the average score of the USA general population, with a standard deviation (SD) of 10.

Third, the questionnaire included the Dutch-Flemish PROMIS Global Health Questionnaire (v1.2). This questionnaire measures the overall evaluation of one's physical and mental health and contains 10 items. There are two subscales; global physical health (GPH; 4 items) and global mental health (GMH; 4 items)[28]. The scores of the Dutch-Flemish PROMIS UE Global Health subscales are also expressed as T-scores (average 50, SD 10). The Dutch-Flemish PROMIS Pain Intensity item (Global07r) from the Dutch-Flemish PROMIS Global Health item bank is a generic legacy instrument [28, 58]. It assesses pain intensity and consists of an 11-point numeric rating scale (NRS) with anchors 0 = "no pain" and 10 = "worst pain imaginable".

Fourth, the questionnaire contained three disease specific legacy instruments. The Disabilities of the Arm, Shoulder and Hand (DASH) questionnaire, subscale Disability/Symptoms measures physical function and symptoms and is applicable to patients with any or several musculoskeletal disorders of the upper limbs [33]. The DASH subscale Disability/Symptoms consists of 30 items. The time frame

for the items is the past week. The total score ranges from 0 to 100, with 100 indicating most severe disability. The DASH has satisfactory psychometric properties [5, 10, 30, 56, 57]. An official Dutch translation showed good psychometric properties [20, 65]. The Functional Index for Hand OsteoArthritis (FIHOA) is a disease-specific instrument that assesses hand OA-related functional impairment. The FIHOA consists of 10 items. There is no time frame for the items, but current status is inferred. Total scores range from 0-30, with higher scores indicating more functional impairment. The psychometric properties of the FIHOA are good [23, 24, 38]. An official Dutch translation showed good psychometric properties [66]. The Michigan Hand Outcomes Questionnaire (MHQ) is a hand-specific instrument that measures several domains and is applicable to patients with conditions of, or injury to, the hand or wrist [16]. The MHQ contains six distinct subscales. In this study, we used the MHQ subscale activities of daily living (ADLs), which assesses difficulty in performing daily activities for the right hand (5 items), the left hand (5 items) and both hands (7 items). The time frame for the items is the past week. The total score per scale is converted to a score from 0-100, with higher scores indicating less disability. The psychometric properties of the MHQ ADLs scale are good [13-16, 21, 39-41, 59, 63]. An official Dutch translation of the MHQ showed good responsiveness [64].

Statistical analysis

Demographic and clinical characteristics of the Dutch and US sample were described by descriptive statistics. Differences between the Dutch sample and the US sample were evaluated by chi-square-tests for categorical variables and independent sample-t-tests for continuous variables.

Cross-cultural validity of the Dutch-Flemish PROMIS UE item bank was assessed with Differential Item Functioning (DIF) analyses. DIF analyses examine whether people from different groups (here: language) with the same level of the trait (theta (θ), in this study the level of upper extremity function) have different probabilities of giving a certain response to an item [25, 31, 50]. There are two kinds of DIF: uniform and non-uniform. Uniform DIF exists when the magnitude of conditional dependency is relatively invariant across the trait. Non-uniform DIF exists when the magnitude or direction of DIF is variant across the trait. DIF for language (Dutch vs English) was evaluated by ordinal logistic regression models, using the R-package Lordif (version 0.3-3). An intercept model (Model 0) and three nested models were formed; Model 1 with theta as the explanatory variable, Model 2 with both theta and language as explanatory variables and Model 3 with theta, language and an interaction term for languague and theta. Uniform DIF was tested by comparing the loglikelihood values for model Model 1 and 2, and non-uniform DIF by Model 2 and 3. A McFadden's pseudo R^2 change of 2% was used as the critical value to flag for possible DIF [11, 12, 17, 50]. When items were flagged as potential DIF for language items, the impact of DIF was examined by plotting item characteristic curves (ICCs) and test characteristic curves (TCCs). The TCC plots showed the total item scores for all 46 PROMIS UE items (ignoring DIF) and the scores for only the items having DIF [11, 12].

The T-scores of the Dutch-Flemish PROMIS UE item bank were correlated to the scores on the legacy instruments. Pearson's r correlations were used for normally distributed data and Spearman's rho-coefficients for non-normally distributed data. For assessing convergent validity, we hypothesized that the Dutch-Flemish PROMIS UE item bank would have a: 1) moderate negative correlation (-0.50

< r ≤ -0.30) with the Dutch-Flemish PROMIS Pain Intensity item [28, 58], given the fact that these instruments are intended to measure different constructs; 2) strong negative correlation (r ≤ -0.50) with the DASH, Subscale Disability/Symptoms [33]; 3) strong negative correlation (r ≤ -0.50) with the FIHOA score [23, 24, 38], and 4) strong positive correlation (r ≥ 0.50) with the MHQ, Subscale ADLs score [16], given the fact that these instruments are intended to measure the same construct (upper extremity physical function).

Analyses other than DIF analyses, were done with IBM SPSS Statistics version 25 (Armork, New York, USA).

RESULTS

Translation and testing of translations

In total, 28 native Dutch or Flemish speaking participants were interviewed. The average age was 46 years (range 17 – 75), and 68 % were female. Of all participants, 68 % were patients with upper extremity disorders and 32% were participants without complaints. For the four v2.0 items from the Dutch-Flemish PROMIS UE item bank, a sufficient Dutch-Flemish translation was obtained and no separate translation for Dutch and Flemish was required. However, three out of four items (PFM2, PFM16 and PFM18) were considered not relevant or not realistic by some participants (both patients and people from the general population). Despite these issues, we decided to maintain the items without adaptation of the translation in the preliminary Dutch-Flemish PROMIS UE item bank, to investigate whether DIF for language for these items would occur.

Study participants and measurement instruments

Of 371 patients that were screened for eligibility, 67 patients did not meet the in- or exclusion criteria. Of the remaining 304 invited patients, 219 (72%) gave informed consent. Of these, 14 (5%) patients partially completed the questionnaire, which were not included in the analyses. The remaining 205 (67%) patients fully completed the questionnaire. In these 205 questionnaires, there were no missing items and all questionnaires were included in the analyses.

The demographic and clinical characteristics and the scores on the measurement instruments of the Dutch sample and the US sample are summarized in Table 2. Of the Dutch patients, 52% were female, the average age (standard deviation (SD)) was 53 (15). Of the Dutch sample, 78% were born in the Netherlands and 91% had at least a high school degree. Patients reported having pain in one or both shoulder(s), arm(s), hand(s) or fingers or in other pain locations. Reported injuries were physical (e.g. muscle) injury, overuse injury, trauma, surgery, disease (e.g. arthritis), congenital disorder, and other or unknown disorders. The results of the t-test and chi-square test showed that the Dutch patients were on average older, a larger proportion was male and a larger proportion had less than high school degree compared to the US sample.

Table 2. Demographic and clinical characteristics of the Dutch sample (n = 205) and the US sample (n = 246)

	Dutch sample (n=205)		US sample (n = 246)	
Age mean (SD), range	53	(15), 18-87	48	(14), 18-85 ***
Gender, n (%)				
Male	99	(48)	76	(31) ***
Female	106	(52)	170	(69) ***
Country of birth, n (%)				
Netherlands	160	(78)	-	
Other	45	(22)	-	
Social status, n (%)				
Single	64	(31)	-	
Married or living together	122	(60)	-	
Living apart together	3	(1)	-	
Living with parents	6	(3)	-	
Other	10	(5)	-	
Educational level, n (%)				
Less than high school degree	18	(9)	6	(2) ***
High school degree	20	(10)	53	(22)
Some college	70	(34)	81	(33)
College degree	14	(7)	80	(32)
Advanced degree	83	(40)	26	(11)
Employment status, n (%)				
Full time	76	(37)	-	
Part time	38	(19)	-	
Student	5	(2)	-	
Unpaid, volunteer, household	18	(9)	-	
Retired	39	(19)	-	
Unemployed	8	(4)	-	
Other	21	(10)	-	
Social benefits, n (%)				
Sick listed	38	(20)	-	
Disability benefit	21	(10)	-	
Unemployment benefit	5	(2)	-	
Other	17	(8)	-	
No social benefit	124	(60)	-	
Duration of complaints, n (%)				
< 1 month	20	(10)	-	
1-3 months	21	(10)	-	
3-6 months	27	(13)	-	
6-12 months	35	(17)	-	
1-2 years	43	(21)	-	
2-5 years	30	(15)	-	
>5 years	29	(14)	-	
Location of pain,[a] n (%)				
Shoulder(s)	157	(76)	-	
Arm(s)	116	(57)	-	
Hand(s)	44	(21)	-	
Fingers	41	(20)	-	

Other	41	(20)	-
No pain	15	(7)	-
Type of disorder,[a] n (%)			
Physical injury (e.g. muscle)	40	(20)	-
Work-related overuse injury	22	(11)	-
Non-work-related overuse injury	9	(4)	-
Trauma	66	(32)	-
Surgery	20	(10)	-
Disease (e.g. arthritis)	6	(3)	-
Congenital	5	(2)	-
Other	38	(19)	-
Unknown	34	(17)	-

[a]Multiple answers were allowed.
*p < 0.05
***p < 0.001

The scores on the Dutch-Flemish PROMIS UE item bank, the PROMIS Global Health Questionnaire and the legacy instruments are shown in Table 3. The mean T-score of the Dutch-Flemish PROMIS UE item bank in the Dutch sample was 34.6 (SD = 8.6) and the corresponding mean for the US sample was 36.5 (SD = 7.0). The results of the t-test showed that the level of upper extremity function was lower in the Dutch sample.

Table 3. Scores on the measurement instruments.

	Dutch sample (n=205)		US sample (n = 246)	
T-scores of the PROMIS Upper Extremity item bank				
Mean (SD)	34.6	(8.6)	36.5	(7.0) *
Range	14.1 – 61.2		20.1 – 61.2	
T-scores of the PROMIS Global Health Questionnaire				
Subscale Global Physical Health			-	
Mean (SD)	43.0	(7.1)	-	
Range	23.7 – 60.3		-	
Subscale Global Mental Health				
Mean (SD)	47.0	(9.9)	-	
Range	24.1 – 67.6		-	
Legacy instruments mean (SD)				
PROMIS Global Health Pain Intensity	4.9	(2.7)	-	
DASH Subscale Disability/Symptoms	36.4	(20.9)	-	
FIHOA	7.0	(7.2)	-	
MHQ Subscale Activities of Daily Living	81.3	(19.9)	-	

PROMIS Global Health Pain Intensity (0-10); DASH, Disabilities of the Arm, Shoulder and Hand Subscale Disability/Symptoms (0-100); FIHOA, Functional Index for Hand OsteoArthritis (0-30); MHQ, Michigan Hand Outcomes Questionnaire Subscale Activities of Daily Living (0-100).
*$p < 0.05$

Cross-cultural validity

Eight items were flagged for DIF for language. Four items had some level of uniform DIF and four items had some level of non-uniform DIF (see Table 4.) For all items with uniform DIF (PFB19r1, PFB20r1, PFC21r1 and PFC43), the Dutch patients were more likely to endorse higher response categories, compared with the US participants, indicating that the items were easier for Dutch patients. For the items with non-uniform DIF, the item scores for item PFB16r1 were higher for Dutch participants compared with the US participants at lower levels of theta, indicating that this item was easier for Dutch patients compared with US participants, with lower levels of upper extremity functioning. For the items PFB28r1, PFM2 and PFM16, the item scores were lower for Dutch participants compared with the US participants, at lower levels of theta, indicating that these items were more difficult for Dutch patients with lower levels of upper extremity functioning (see the ICC plots in Appendix A).

Table 4. Items with Language DIF (4 – uniform DIF and 4 non-uniform DIF)

Items with uniform DIF			
Item name	**English**	**Dutch-Flemish**	**McFaddens pseudo R^2 change***
PFB19r1	Are you able to squeeze a new tube of toothpaste?	Kunt u tandpasta uit een nieuwe tube knijpen?	0.027
PFB20r1	Are you able to cut a piece of paper with scissors?	Kunt u met een schaar een stuk papier knippen?	0.026
PFB21r1	Are you able to pick up coins from a table top?	Kunt u munten van een tafel oppakken?	0.039
PFC43	Are you able to use your hands, such as for turning faucets, using kitchen gadgets, or sewing?	Kunt u uw handen gebruiken, bijvoorbeeld om kranen open en dicht te draaien, keukengerei te gebruiken of te naaien?	0.033
Items with non-uniform DIF			
Item name	**English**	**Dutch-Flemish**	**McFaddens pseudo R^2 change****
PFB16r1	Are you able to press with your index finger (for example ringing a doorbell)?	Kunt u met uw wijsvinger ergens op drukken (bijvoorbeeld een deurbel)?	0.020
PFB28r1	Are you able to lift 10 pounds (5 kg) above your shoulder?	Kunt u 5 kilo boven uw schouder tillen?	0.027
PFM2	Are you able to lift a heavy painting or picture to hang on your wall above eye-level?	Kunt u een zwaar schilderij of fotolijst optillen om boven ooghoogte aan de muur te hangen?	0.024
PFM16	Are you able to pass a 20-pound (10 kg) turkey or ham to other people at the table?	Kunt u aan tafel een grote en zware schaal met eten (10 kilo) doorgeven aan een ander?	0.025

*R^2 change for comparing Model 1 and 2
**R^2 change for comparing Model 2 and 3

The total impact of DIF (for language) on the TCC is shown in Fig. 1. The left graph shows the TCC for all 46 UE items (ignoring DIF), and the right graph shows the TCC for the eight items having DIF. These curves show that the UE total score is only slightly different for Dutch participants then for US participants, indicating minimal impact of DIF by language.

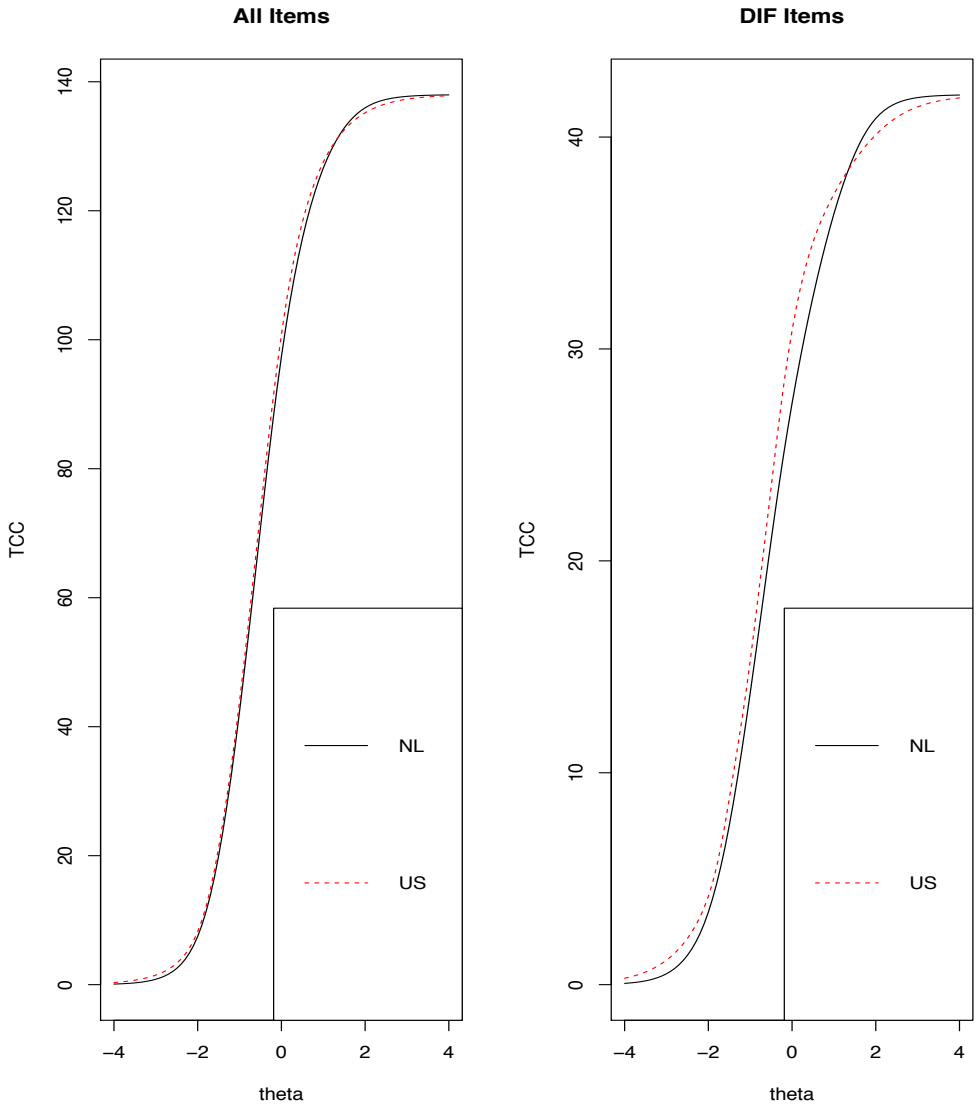

Fig. 1. The total impact of DIF for language on the test characteristic curves (TCC). The graph shows the relation between the total item scores on the y-axis and the level of upper extremity function (theta) on the x-axis. The left graph shows the TCC for all 46 Dutch-Flemish (NL) and United States (US) PROMIS Upper Extremity items (ignoring DIF). The right graph shows the TCC for just the eight items having DIF.

Construct validity

All correlations between the Dutch-Flemish PROMIS UE T-score and the legacy instruments were according to the a-priori formulated hypotheses (see Table 5).

Table 5. Correlations between the Dutch-Flemish PROMIS UE T-scores and the legacy instruments.

Legacy Instrument	Hypothesized r	Observed r
Dutch-Flemish PROMIS Pain Intensity	- 0.50 < r ≤ - 0.30	- 0.43
DASH subscale Disability/Symptoms	r ≤ - 0.50	- 0.87
FIHOA	r ≤ - 0.50	- 0.86
MHQ Subscale Activities of Daily Living	r ≥ 0.50	0.81

DASH, Disabilities of the Arm, Shoulder and Hand Subscale Disability/Symptoms; FIHOA, Functional Index for Hand OsteoArthritis; MHQ, Michigan Hand Outcomes Questionnaire Subscale Activities of Daily Living.

DISCUSSION

The aim of this study was to translate four new PROMIS v2.0 items for the Dutch-Flemish PROMIS UE item bank and to investigate the cross-cultural validity and construct validity of the item bank in Dutch patients with musculoskeletal disorders of the upper extremity. A sufficient Dutch-Flemish translation was obtained for all items, some items showed relevance issues but all items were maintained in the preliminary item bank. DIF analyses showed eight items having some DIF for language, but the impact of DIF on the total score was minimal, indicating sufficient cross-cultural validity. The construct validity for the item bank was sufficient, because all four predefined hypotheses for correlations with legacy instruments were confirmed.

This is the first study on the validity of the 46 item PROMIS UE item bank v2.0 outside the US. Comparable to a previous study on the Dutch-Flemish PROMIS PF item bank, we also found good cross-cultural validity, although some items in both studies showed some DIF for language [18]. In the cognitive interviews conducted after translation in the current study, three items (PFM2, PFM18 and PFM16) were regarded as unrealistic by some participants (both patients and people from the general population). Two of these items, which reflect higher levels of upper extremity function (PFM2 and PFM16) also showed (non-uniform) DIF and responses showed that they were more difficult for Dutch participants with lower levels of upper extremity function. This might indicate that these items will be less suitable to maintain in the final Dutch-Flemish PROMIS UE item bank, but this has to be investigated in the final item bank calibration. We therefore decided to keep the items in this preliminary version of the item bank.

Studies in populations with upper extremity conditions found ceiling effects for the PROMIS V1.2 UE item bank [1-3, 35, 37]. In the current study, T-scores with normal distribution of the scores were found, indicating there were no floor- or ceiling effects for the PROMIS UE item bank v2.0 in impaired individuals. The distribution of the scores on the FIHOA and MHQ Subscale ADLs however were skewed, indicating a ceiling effect for these instruments in our mixed sample of patients with shoulder, arm or hand complaints. So, the Dutch-Flemish PROMIS UE item bank v2.0 has an improved measurement range compared to these legacy instruments and compared to the v1.2 version of the PROMIS UE item bank.

In the current study, we found a correlation for the Dutch-Flemish PROMIS UE item bank with the DASH of -0.87, which is comparable to findings from the study of Beckman et al. who found a correlation for the UE v1.2 CAT with the DASH of -0.80 in an orthopedic upper extremity (non shoulder) sample [2]. Döring et al. found a correlation of 0.81 for the PROMIS UE CAT (v1.2) and the Quick DASH in a sample with hand and upper extremity conditions [22]. A similar correlation with the Quick DASH (0.82) was found in the study of Kaat et al. in a sample with isolated upper extremity fractures [37]. So, the correlation of the Dutch-Flemish PROMIS UE item bank with the DASH is in line with previous studies on the v1.2 version of the PROMIS UE item bank, indicating that both instruments measure similar constructs.

Strenghts of the current study were that the Dutch sample consisted of patients with a wide variety of musculoskeletal upper extremity conditions with a broad range of functioning. The results on the T-score for the Dutch-Flemish PROMIS UE item bank indicate that the item bank is suitable for measuring upper extremity function across different conditions and showed no floor- or ceiling effects. Also, this study was conducted according to the COSMIN Risk of Bias checklist, ensuring adequate study design for cross-cultural validity/measurement invariance and for construct validity, with sufficient sample sizes [42]. In addition, the international PROMIS guidelines for instrument development and validation were followed [48].

This study has some limitations. First, the Dutch and US sample differed in age, gender and educational level, and the US sample was a non-clinical sample, which might influence the comparability of results. Second, due to a limited sample size, we were not able to perform DIF analyses for age, gender and educational level within the Dutch sample. However, previous studies on the Dutch-Flemish PROMIS PF item bank, showed negligible impact of DIF for these aspects on total physical function [18, 19, 44]. Third, we had a limited number of responses to the paper-and pencil questionnaires, which limits generalizability to this way of administration mode, although in a previous study no different results for the way of administration mode was found [18].

In line with previous work of the Dutch-Flemish PROMIS group, the results of this study add to the evidence for PROMIS instruments for the Dutch and Flemish population. Following the PROMIS guidelines, cross-cultural validation is the first recommended step after translation of PROMIS items banks [49]. Once cross-cultural validity has been established, further development of the item bank is warranted. We recommend to expand the current study in a larger sample, with a minimal sample size of n = 500, for the full item bank calibration. Afterwards, PROMIS short forms and CAT can be developed for use in clinical practice and research.

In conclusion, this study found sufficient cross-cultural validity and construct validity of the preliminary Dutch-Flemish Upper Extremity item bank v2.0. Further validation of the item bank is now warranted and the final item bank will have the potential of improved measurement of upper extremity functioning in the Dutch-Flemish population.

REFERENCES

1. Anthony CA, Glass NA, Hancock K, Bollier M, Wolf BR, Hettrich CM. Performance of PROMIS Instruments in Patients With Shoulder Instability. *The American journal of sports medicine.* 2017;45:449-453.
2. Beckmann JT, Hung M, Voss MW, Crum AB, Bounsanga J, Tyser AR. Evaluation of the Patient-Reported Outcomes Measurement Information System Upper Extremity Computer Adaptive Test. *The Journal of hand surgery.* 2016;41:739-744.e734.
3. Beleckas CM, Padovano A, Guattery J, Chamberlain AM, Keener JD, Calfee RP. Performance of Patient-Reported Outcomes Measurement Information System (PROMIS) Upper Extremity (UE) Versus Physical Function (PF) Computer Adaptive Tests (CATs) in Upper Extremity Clinics. *The Journal of hand surgery.* 2017.
4. Bonomi AE, Cella DF, Hahn EA, Bjordal K, Sperner-Unterweger B, Gangeri L, Bergman B, Willems-Groot J, Hanquet P, Zittoun R. Multilingual translation of the Functional Assessment of Cancer Therapy (FACT) quality of life measurement system. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 1996;5:309-320.
5. Bot SD, Terwee CB, van der Windt DA, Bouter LM, Dekker J, de Vet HC. Clinimetric evaluation of shoulder disability questionnaires: a systematic review of the literature. *Annals of the rheumatic diseases.* 2004;63:335-341.
6. Boyce MB, Browne JP, Greenhalgh J. The experiences of professionals with using information from patient-reported outcome measures to improve the quality of healthcare: a systematic review of qualitative research. *BMJ Qual Saf.* 2014;23:508-518.
7. Cella D, Gershon R, Lai JS, Choi S. The future of outcomes measurement: item banking, tailored short-forms, and computerized adaptive assessment. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2007;16 Suppl 1:133-141.
8. Cella D, Riley W, Stone A, Rothrock N, Reeve B, Yount S, Amtmann D, Bode R, Buysse D, Choi S, Cook K, Devellis R, DeWalt D, Fries JF, Gershon R, Hahn EA, Lai JS, Pilkonis P, Revicki D, Rose M, Weinfurt K, Hays R, Group PC. The Patient-Reported Outcomes Measurement Information System (PROMIS) developed and tested its first wave of adult self-reported health outcome item banks: 2005-2008. *Journal of clinical epidemiology.* 2010;63:1179-1194.
9. Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, Rose M. The Patient-Reported Outcomes Measurement Information System (PROMIS): progress of an NIH Roadmap cooperative group during its first two years. *Medical care.* 2007;45:S3-s11.
10. Changulani M, Okonkwo U, Keswani T, Kalairajah Y. Outcome evaluation measures for wrist and hand: which one to choose? *Int Orthop.* 2008;32:1-6.
11. Choi S, Gibbons LE, Crane PK. Logistic Ordinal Regression Differential Item Functioning using IRT, version 0.3-3. Available at: https://cran.r-project.org. Accessed May 30, 2018.
12. Choi SW, Gibbons LE, Crane PK. lordif: An R Package for Detecting Differential Item Functioning Using Iterative Hybrid Ordinal Logistic Regression/Item Response Theory and Monte Carlo Simulations. *J Stat Softw.* 2011;39:1-30.
13. Chung BT, Morris SF. Reliability and internal validity of the michigan hand questionnaire. *Ann Plast Surg.* 2014;73:385-389.
14. Chung BT, Morris SF. Confirmatory factor analysis of the Michigan Hand Questionnaire. *Ann Plast Surg.* 2015;74:176-181.
15. Chung KC, Hamill JB, Walters MR, Hayward RA. The Michigan Hand Outcomes Questionnaire (MHQ): assessment of responsiveness to clinical change. *Ann Plast Surg.* 1999;42:619-622.

16. Chung KC, Pillsbury MS, Walters MR, Hayward RA. Reliability and validity testing of the Michigan Hand Outcomes Questionnaire. *The Journal of hand surgery.* 1998;23:575-587.
17. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. *Medical care.* 2006;44:S115-123.
18. Crins MHP, Terwee CB, Klausch T, Smits N, de Vet HCW, Westhovens R, Cella D, Cook KF, Revicki DA, van Leeuwen J, Boers M, Dekker J, Roorda LD. The Dutch-Flemish PROMIS Physical Function item bank exhibited strong psychometric properties in patients with chronic pain. *Journal of clinical epidemiology.* 2017;87:47-58.
19. Crins MHP, van der Wees PJ, Klausch T, van Dulmen SA, Roorda LD, Terwee CB. Psychometric properties of the PROMIS Physical Function item bank in patients receiving physical therapy. *PloS one.* 2018;13:e0192187.
20. De Smet L, De Kesel R, Degreef I, Debeer P. Responsiveness of the Dutch version of the DASH as an outcome measure for carpal tunnel syndrome. *J Hand Surg Eur Vol.* 2007;32:74-76.
21. Dias JJ, Rajan RA, Thompson JR. Which questionnaire is best? The reliability, validity and ease of use of the Patient Evaluation Measure, the Disabilities of the Arm, Shoulder and Hand and the Michigan Hand Outcome Measure. *J Hand Surg Eur Vol.* 2008;33:9-17.
22. Doring AC, Nota SP, Hageman MG, Ring DC. Measurement of upper extremity disability using the Patient-Reported Outcomes Measurement Information System. *The Journal of hand surgery.* 2014;39:1160-1165.
23. Dreiser RL, Maheu E, Guillou GB. Sensitivity to change of the functional index for hand osteoarthritis. *Osteoarthritis Cartilage.* 2000;8 Suppl A:S25-28.
24. Dreiser RL, Maheu E, Guillou GB, Caspard H, Grouin JM. Validation of an algofunctional index for osteoarthritis of the hand. *Rev Rhum Engl Ed.* 1995;62:43S-53S.
25. Embretson SE, Reise SP. *Item response theory for psychologists.* Mahwah, NJ: Lawrence Erlbaum Associates; 2000.
26. Eremenco SL, Cella D, Arnold BJ. A comprehensive method for the translation and cross-cultural validation of health status questionnaires. *Eval Health Prof.* 2005;28:212-232.
27. Forget NJ, Higgins J. Comparison of generic patient-reported outcome measures used with upper extremity musculoskeletal disorders: linking process using the International Classification of Functioning, Disability, and Health (ICF). *J Rehabil Med.* 2014;46:327-334.
28. Hays RD, Bjorner JB, Revicki DA, Spritzer KL, Cella D. Development of physical and mental health summary scores from the patient-reported outcomes measurement information system (PROMIS) global items. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2009;18:873-880.
29. Hays RD, Spritzer KL, Amtmann D, Lai JS, Dewitt EM, Rothrock N, Dewalt DA, Riley WT, Fries JF, Krishnan E. Upper-extremity and mobility subdomains from the Patient-Reported Outcomes Measurement Information System (PROMIS) adult physical functioning item bank. *Archives of physical medicine and rehabilitation.* 2013;94:2291-2296.
30. Hoang-Kim A, Pegreffi F, Moroni A, Ladd A. Measuring wrist and hand function: common scales and checklists. *Injury.* 2011;42:253-258.
31. Holland P, Wainer H. *Differential item functioning.* Hillsdale, NJ: Lawrence Erlbaum Associates; 1993.
32. Huang H, Grant JA, Miller BS, Mirza FM, Gagnier JJ. A Systematic Review of the Psychometric Properties of Patient-Reported Outcome Instruments for Use in Patients With Rotator Cuff Disease. *The American journal of sports medicine.* 2015;43:2572-2582.
33. Hudak PL, Amadio PC, Bombardier C. Development of an upper extremity outcome measure: the DASH (disabilities of the arm, shoulder and hand) [corrected]. The Upper Extremity Collaborative Group (UECG). *Am J Ind Med.* 1996;29:602-608.
34. Huisstede BM, Bierma-Zeinstra SM, Koes BW, Verhaar JA. Incidence and prevalence of upper-extremity musculoskeletal disorders. A systematic appraisal of the literature. *BMC musculoskeletal disorders.* 2006;7:7.

35. Hung M, Voss MW, Bounsanga J, Crum AB, Tyser AR. Examination of the PROMIS upper extremity item bank. *Journal of hand therapy : official journal of the American Society of Hand Therapists.* 2016.
36. Kaat AJ, Buckenmaier CC, Cook KF, Rothrock NW, Schalet BD, Gershon RC, Vrahas MS. The Expansion and Validation of a New Upper Etremity Item Bank for the Patient Reported Measurement Information System (PROMIS). *Manuscript in preparation.* 2018.
37. Kaat AJ, Rothrock NE, Vrahas MS, O'Toole RV, Buono SK, Zerhusen T, Jr., Gershon RC. Longitudinal Validation of the PROMIS Physical Function Item Bank in Upper Extremity Trauma. *Journal of orthopaedic trauma.* 2017;31:e321-e326.
38. Klokker L, Terwee CB, Waehrens EE, Henriksen M, Nolte S, Liegl G, Kloppenburg M, Westhoven R, Wittoek R, Kjeken I, Haugen IK, Schalet B, Gershon R, Bliddal H, Christensen R. Hand-related physical function in rheumatic hand conditions: a protocol for developing a patient-reported outcome measurement instrument. *BMJ Open.* 2016;6:e011174.
39. London DA, Stepan JG, Calfee RP. Determining the Michigan Hand Outcomes Questionnaire minimal clinically important difference by means of three methods. *Plast Reconstr Surg.* 2014;133:616-625.
40. Maia MV, de Moraes VY, Dos Santos JB, Faloppa F, Belloti JC. Minimal important difference after hand surgery: a prospective assessment for DASH, MHQ, and SF-12. *SICOT J.* 2016;2:32.
41. McMillan CR, Binhammer PA. Which outcome measure is the best? Evaluating responsiveness of the Disabilities of the Arm, Shoulder, and Hand Questionnaire, the Michigan Hand Questionnaire and the Patient-Specific Functional Scale following hand and wrist surgery. *Hand (N Y).* 2009;4:311-318.
42. Mokkink LB, de Vet HCW, Prinsen CAC, Patrick DL, Alonso J, Bouter LM, Terwee CB. COSMIN Risk of Bias checklist for systematic reviews of Patient-Reported Outcome Measures. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2018;27:1171-1179.
43. Oude Voshaar MA, Ten Klooster PM, Glas CA, Vonkeman HE, Krishnan E, van de Laar MA. Relative performance of commonly used physical function questionnaires in rheumatoid arthritis and a patient-reported outcomes measurement information system computerized adaptive test. *Arthritis & rheumatology (Hoboken, N.J.).* 2014;66:2900-2908.
44. Oude Voshaar MA, ten Klooster PM, Glas CA, Vonkeman HE, Taal E, Krishnan E, Moens HJ, Boers M, Terwee CB, van Riel PL, van de Laar MA. Calibration of the PROMIS physical function item bank in Dutch patients with rheumatoid arthritis. *PloS one.* 2014;9:e92367.
45. Porter ME. What is value in health care? *N Engl J Med.* 2010;363:2477-2481.
46. Porter ME, Larsson S, Lee TH. Standardizing Patient Outcomes Measurement. *N Engl J Med.* 2016;374:504-506.
47. Porter ME, Lee TH. From Volume to Value in Health Care: The Work Begins. *JAMA.* 2016;316:1047-1048.
48. PROMIS. PROMIS Instrument Development and Validation Scientific Standards (revised May 2013). 2013. Available at: http://www.healthmeasures.net/images/PROMIS/PROMISStandards_Vers2.0_Final.pdf. Accessed May 30, 2018.
49. PROMIS. Minimum requirements for the release of PROMIS instruments after translation and recommandations for futher psychometric evaluation. 2014. Available at: http://www.healthmeasures.net/images/PROMIS/Standards_for_release_of_PROMIS_instruments_after_translation_v8.pdf. Accessed May 30, 2018.
50. Reeve BB, Hays RD, Bjorner JB, Cook KF, Crane PK, Teresi JA, Thissen D, Revicki DA, Weiss DJ, Hambleton RK, Liu H, Gershon R, Reise SP, Lai JS, Cella D, Group PC. Psychometric evaluation and calibration of health-related quality of life item banks: plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical care.* 2007;45:S22-31.

51. Reeve BB, Wyrwich KW, Wu AW, Velikova G, Terwee CB, Snyder CF, Schwartz C, Revicki DA, Moinpour CM, McLeod LD, Lyons JC, Lenderking WR, Hinds PS, Hays RD, Greenhalgh J, Gershon R, Feeny D, Fayers PM, Cella D, Brundage M, Ahmed S, Aaronson NK, Butt Z. ISOQOL recommends minimum standards for patient-reported outcome measures used in patient-centered outcomes and comparative effectiveness research. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2013;22:1889-1905.
52. Riley WT, Rothrock N, Bruce B, Christodolou C, Cook K, Hahn EA, Cella D. Patient-reported outcomes measurement information system (PROMIS) domain names and definitions revisions: further evaluation of content validity in IRT-derived item banks. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2010;19:1311-1321.
53. Rolstad S, Adler J, Ryden A. Response burden and questionnaire length: is shorter better? A review and meta-analysis. *Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research.* 2011;14:1101-1108.
54. Rose M, Bjorner JB, Becker J, Fries JF, Ware JE. Evaluation of a preliminary physical function item bank supported the expected advantages of the Patient-Reported Outcomes Measurement Information System (PROMIS). *Journal of clinical epidemiology.* 2008;61:17-33.
55. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE, Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. *Journal of clinical epidemiology.* 2014;67:516-526.
56. Roy JS, MacDermid JC, Woodhouse LJ. Measuring shoulder function: a systematic review of four questionnaires. *Arthritis Rheum.* 2009;61:623-632.
57. Schoneveld K, Wittink H, Takken T. Clinimetric evaluation of measurement tools used in hand therapy to assess activity and participation. *Journal of hand therapy : official journal of the American Society of Hand Therapists.* 2009;22:221-235; quiz 236.
58. Sendlbeck M, Araujo EG, Schett G, Englbrecht M. Psychometric properties of three single-item pain scales in patients with rheumatoid arthritis seen during routine clinical care: a comparative perspective on construct validity, reproducibility and internal responsiveness. *RMD Open.* 2015;1:e000140.
59. Shauver MJ, Chung KC. The minimal clinically important difference of the Michigan hand outcomes questionnaire. *The Journal of hand surgery.* 2009;34:509-514.
60. Smith E, Hoy DG, Cross M, Vos T, Naghavi M, Buchbinder R, Woolf AD, March L. The global burden of other musculoskeletal disorders: estimates from the Global Burden of Disease 2010 study. *Annals of the rheumatic diseases.* 2014;73:1462-1469.
61. Terwee CB, Roorda LD, de Vet HC, Dekker J, Westhovens R, van Leeuwen J, Cella D, Correia H, Arnold B, Perez B, Boers M. Dutch-Flemish translation of 17 item banks from the patient-reported outcomes measurement information system (PROMIS). *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2014;23:1733-1741.
62. Thoomes-de Graaf M, Scholten-Peeters GG, Schellingerhout JM, Bourne AM, Buchbinder R, Koehorst M, Terwee CB, Verhagen AP. Evaluation of measurement properties of self-administered PROMs aimed at patients with non-specific shoulder pain and "activity limitations": a systematic review. *Quality of life research : an international journal of quality of life aspects of treatment, care and rehabilitation.* 2016;25:2141-2160.
63. van de Ven-Stevens LA, Munneke M, Terwee CB, Spauwen PH, van der Linde H. Clinimetric properties of instruments to assess activities in patients with hand injury: a systematic review of the literature. *Archives of physical medicine and rehabilitation.* 2009;90:151-169.
64. van der Giesen FJ, Nelissen RG, Arendzen JH, de Jong Z, Wolterbeek R, Vliet Vlieland TP. Responsiveness of the Michigan Hand Outcomes Questionnaire--Dutch language version in

patients with rheumatoid arthritis. *Archives of physical medicine and rehabilitation.* 2008;89:1121-1126.
65. Veehof MM, Sleegers EJ, van Veldhoven NH, Schuurman AH, van Meeteren NL. Psychometric qualities of the Dutch language version of the Disabilities of the Arm, Shoulder, and Hand questionnaire (DASH-DLV). *Journal of hand therapy : official journal of the American Society of Hand Therapists.* 2002;15:347-354.
66. Wittoek R, Cruyssen BV, Maheu E, Verbruggen G. Cross-cultural adaptation of the Dutch version of the Functional Index for Hand Osteoarthritis (FIHOA) and a study on its construct validity. *Osteoarthritis Cartilage.* 2009;17:607-612.

Appendix A - Item characteristic curves

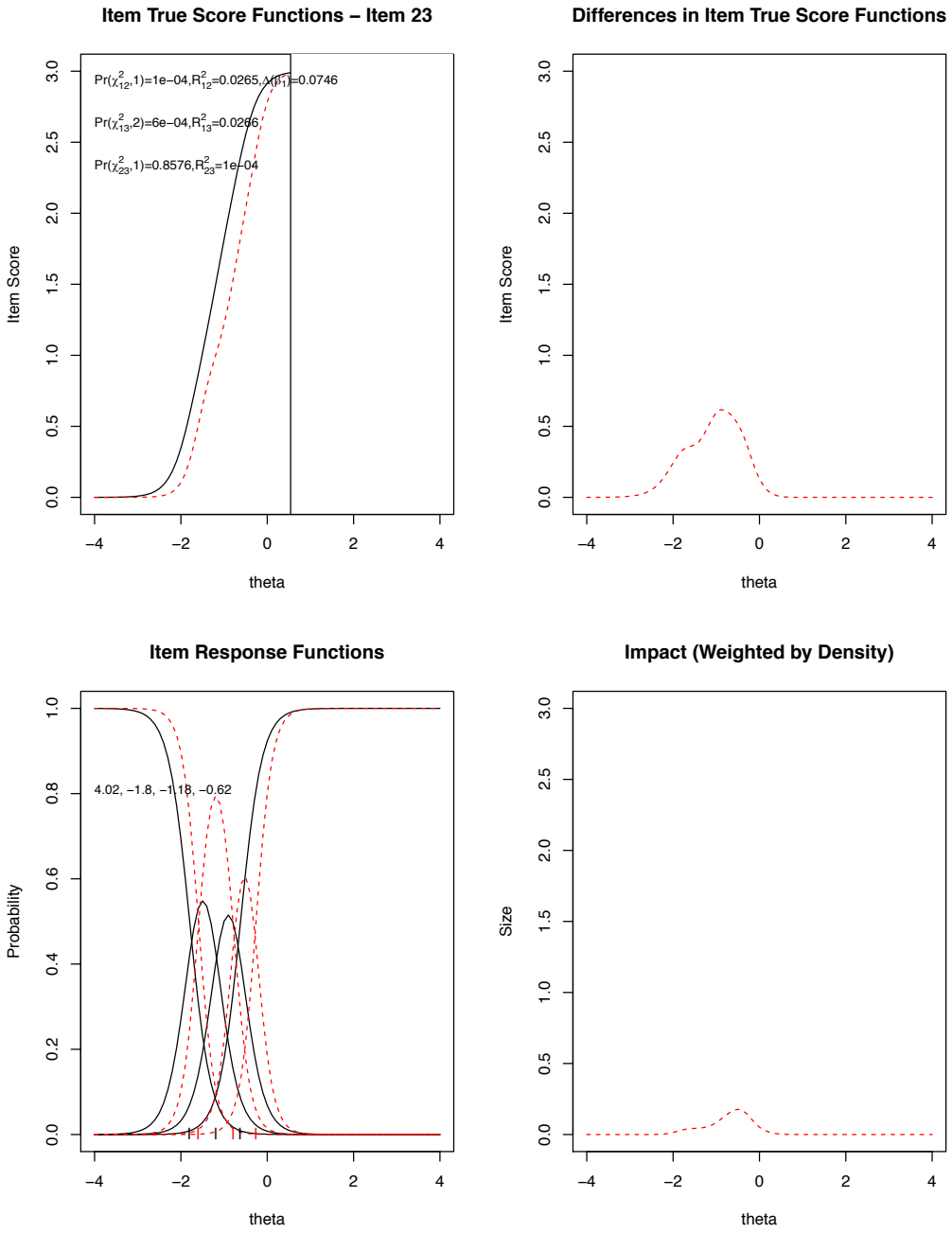

Figure 1: Graphical display of the item PFB19r1 ("Are you able to squeeze a new tube of toothpaste?"), which shows uniform DIF for language. The upper-left graph shows the item characteristic curves (ICC) for the item for Dutch (solid curve) vs. US (dashed curve) scores. The upper-right graph shows the absolute difference between the ICCs, indicating that the difference is mainly at low levels of upper extremity function (theta). The lower left graph shows the item response functions for the two groups. The lower right graph shows the absolute difference between the ICCs, indicating minimal impact.

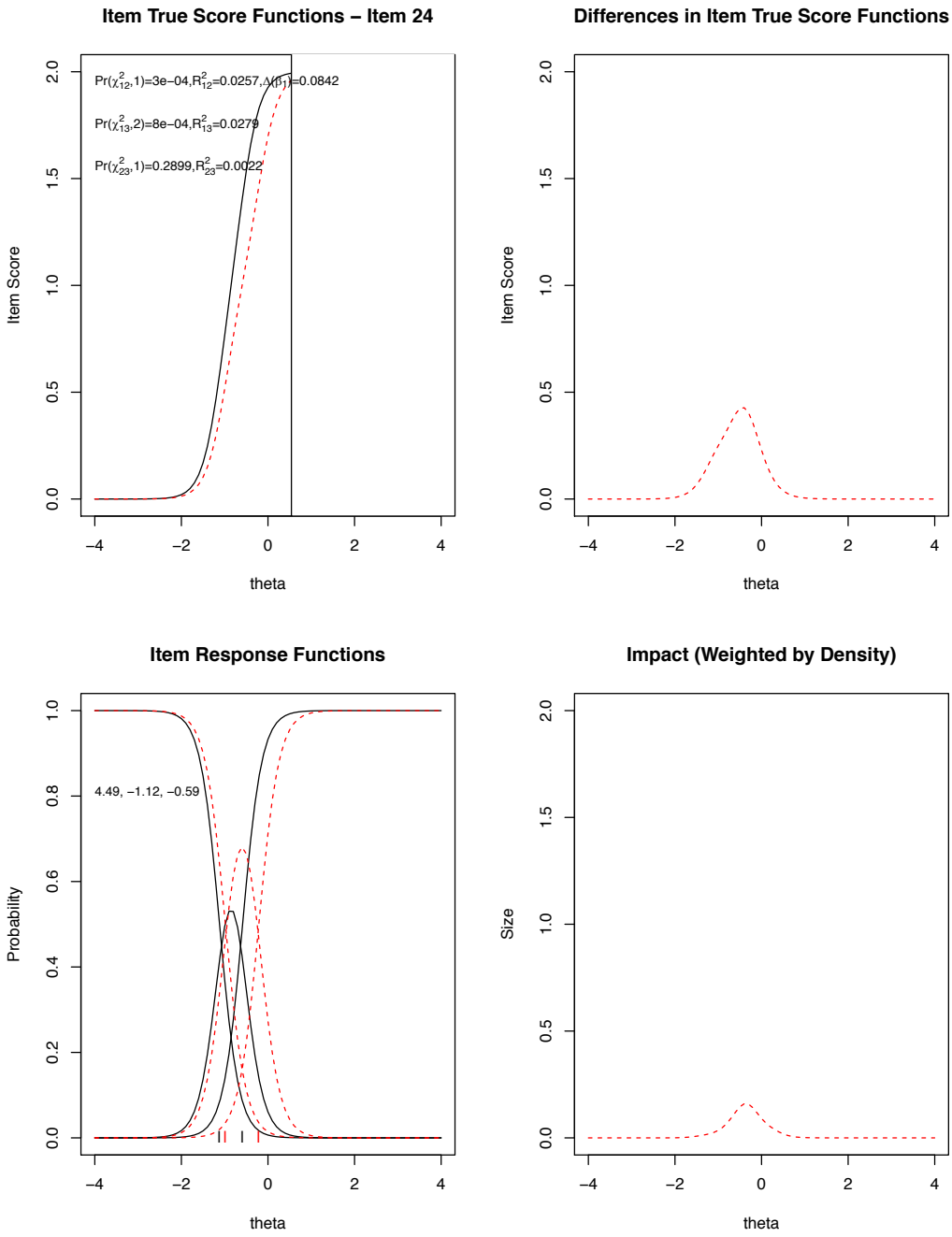

Figure 2: Graphical display of the item PFB20r1 (“Are you able to cut a piece of paper with scissors?”), which shows uniform DIF for language.

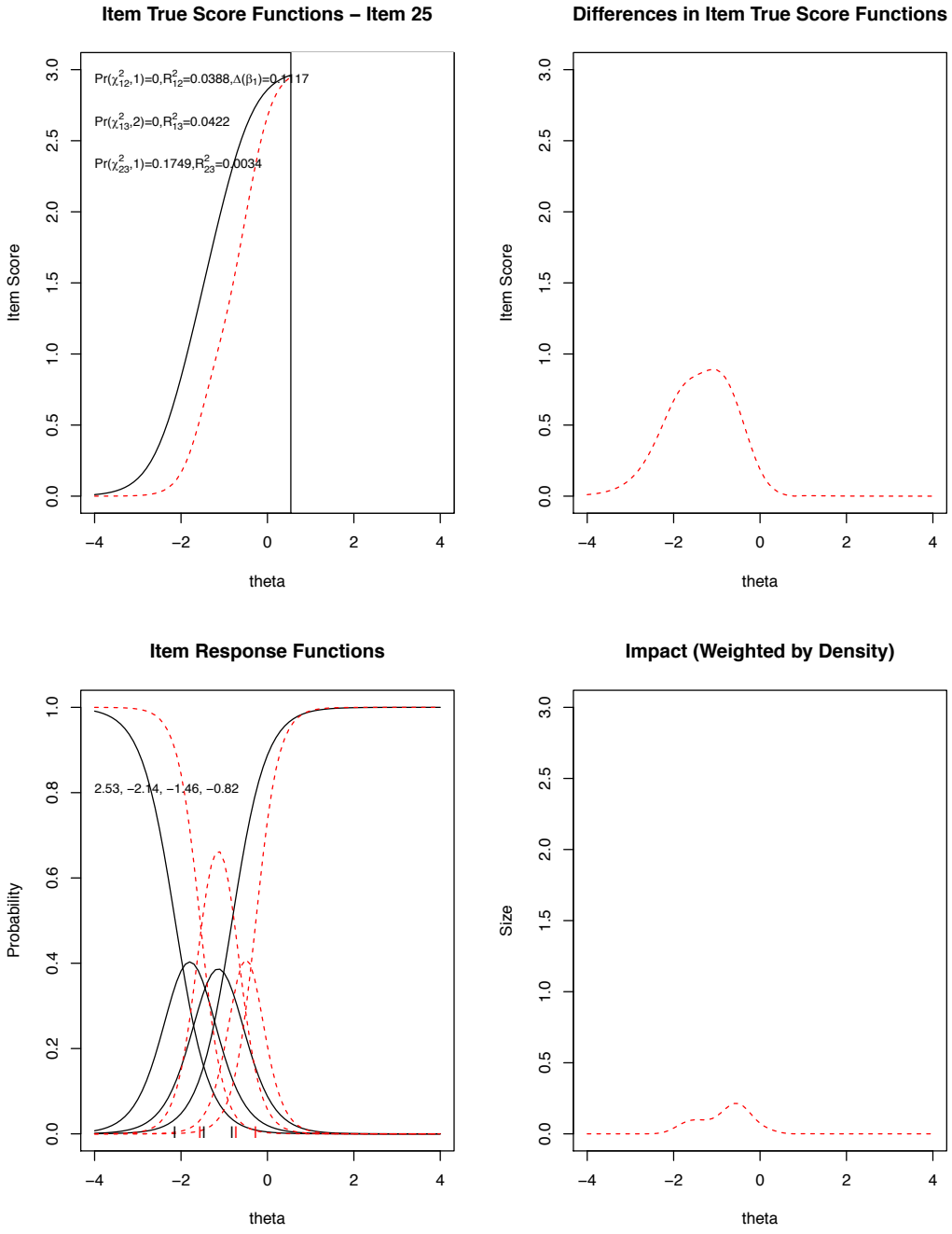

Figure 3: Graphical display of the item PFB21r1 (“Are you able to pick up coins from a table top?”), which shows uniform DIF for language.

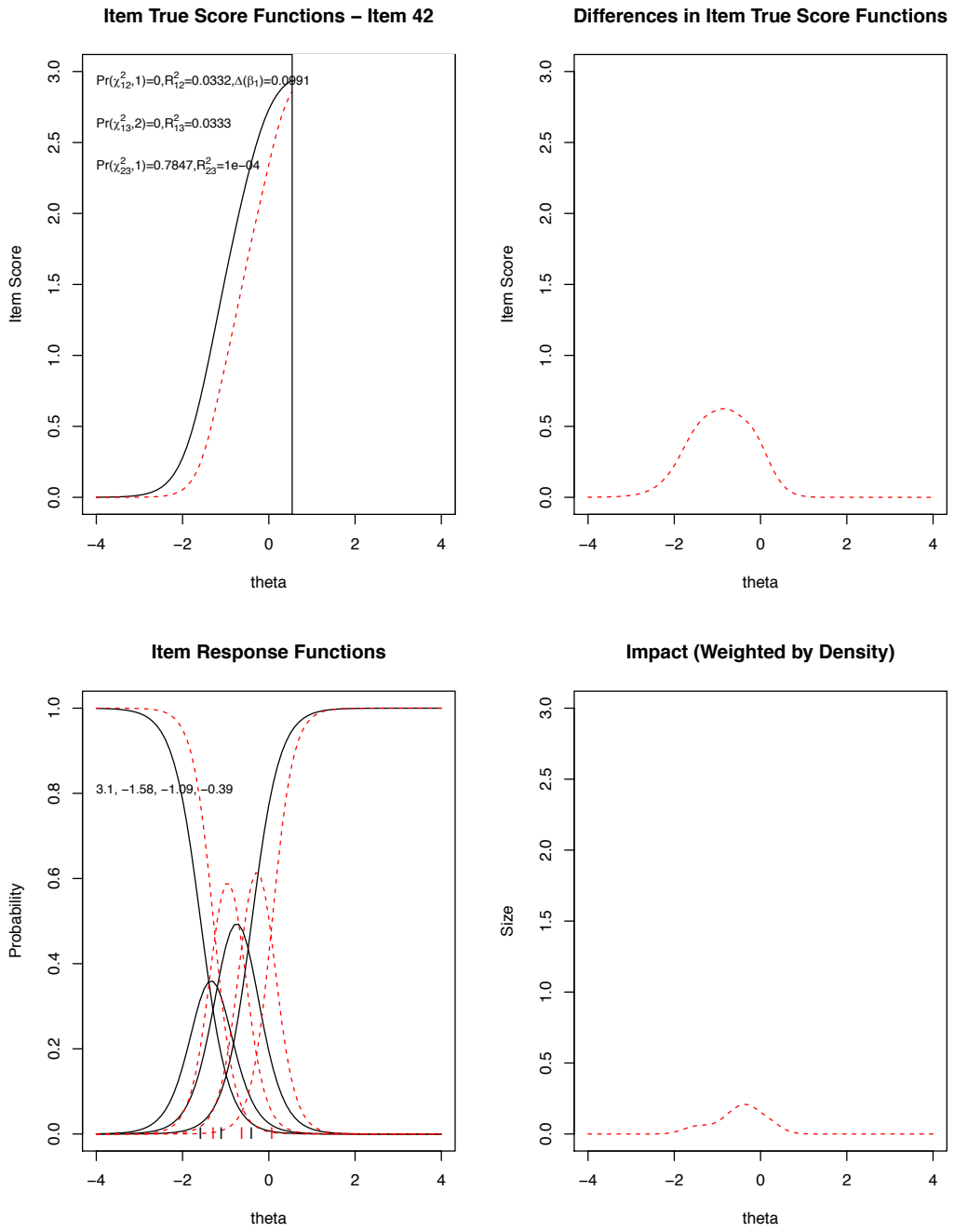

Figure 4: Graphical display of the item PFC43 (“Are you able to use your hands, such as for turning faucets, using kitchen gadgets, or sewing?”), which shows uniform DIF for language.

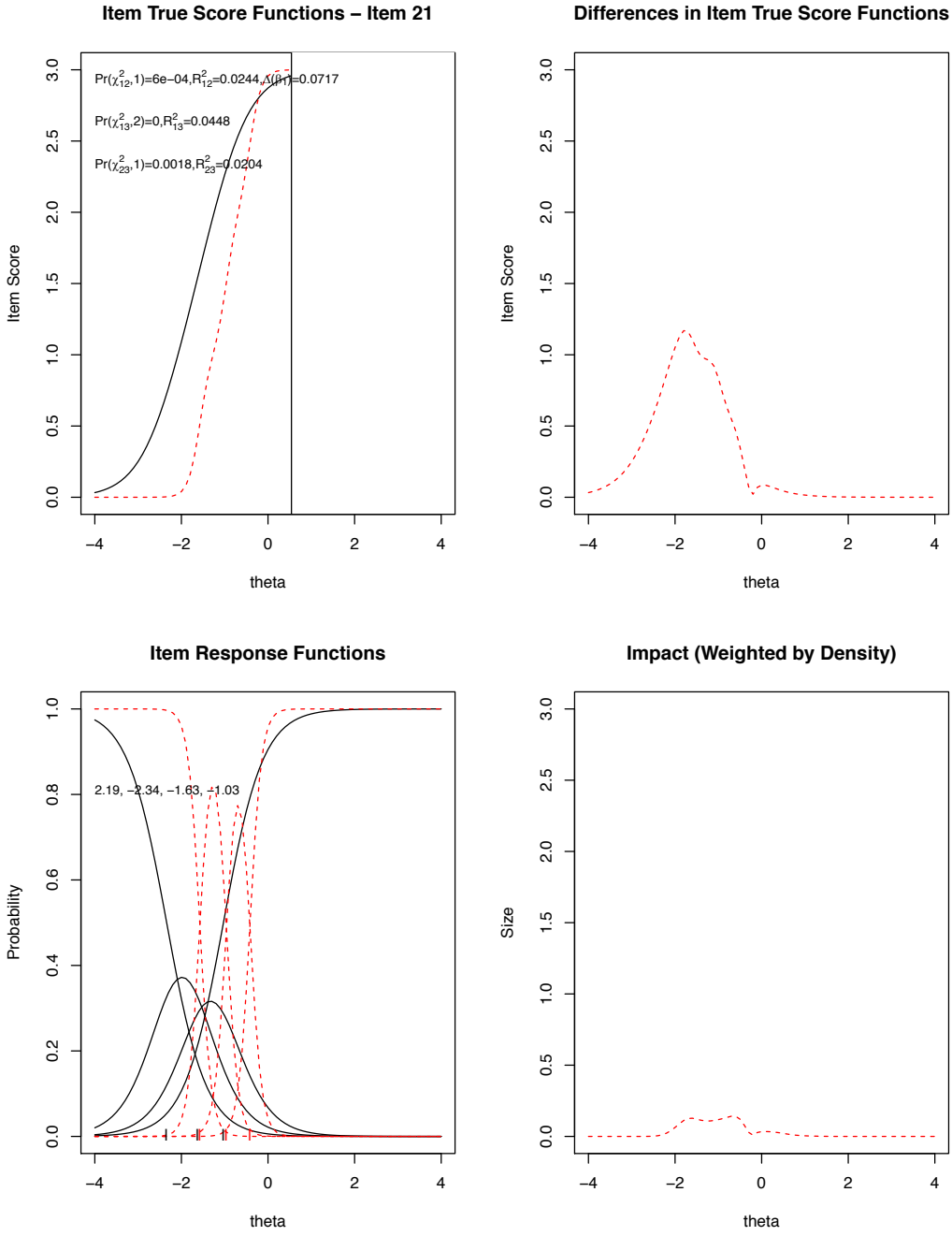

Figure 5: Graphical display of the item PFB16r1 (“Are you able to press with your index finger (for example ringing a doorbell)?”), which shows non-uniform DIF for language.

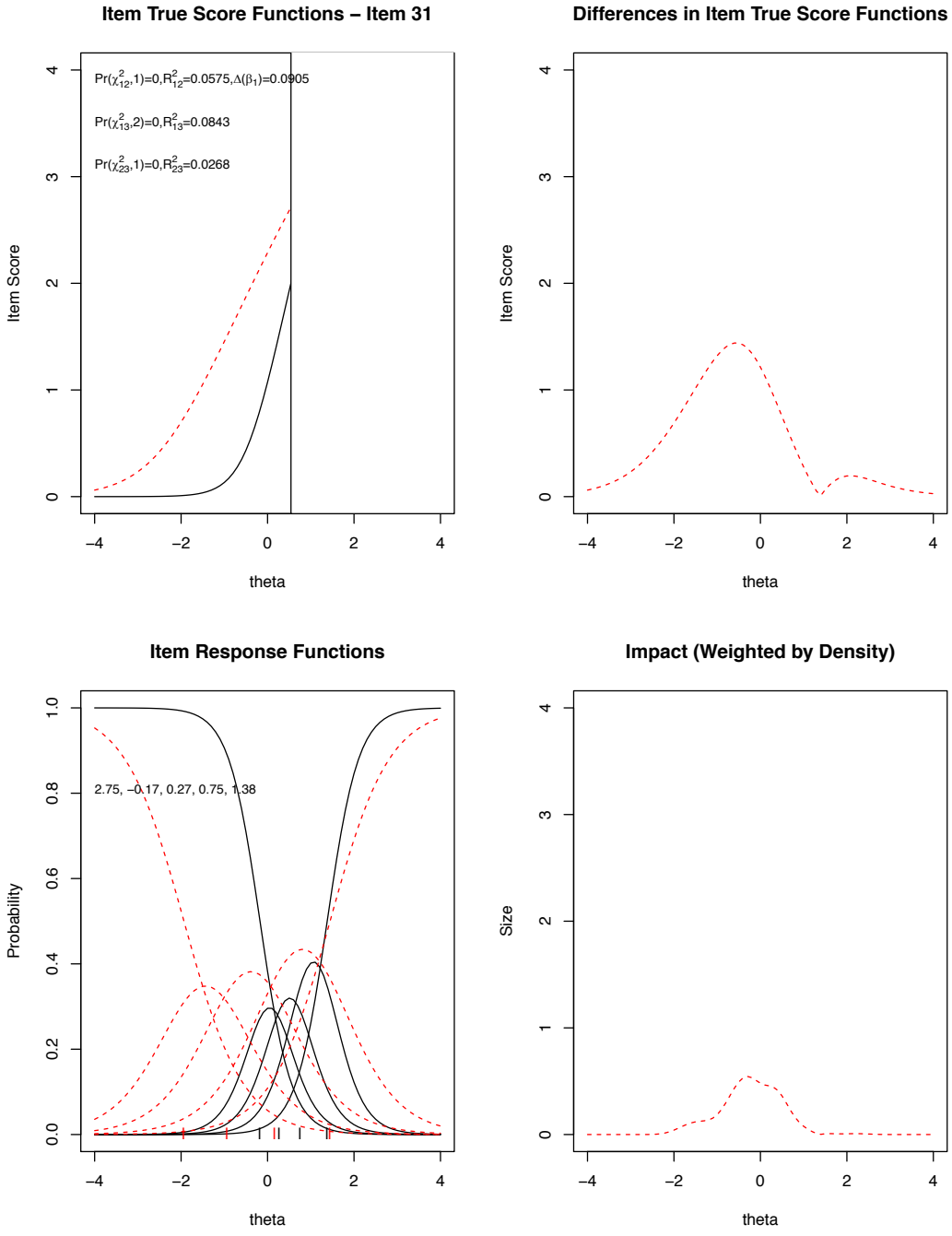

Figure 6: Graphical display of the item PFB28r1 (“Are you able to lift 10 pounds (5 kg) above your shoulder?”), which shows non-uniform DIF for language.

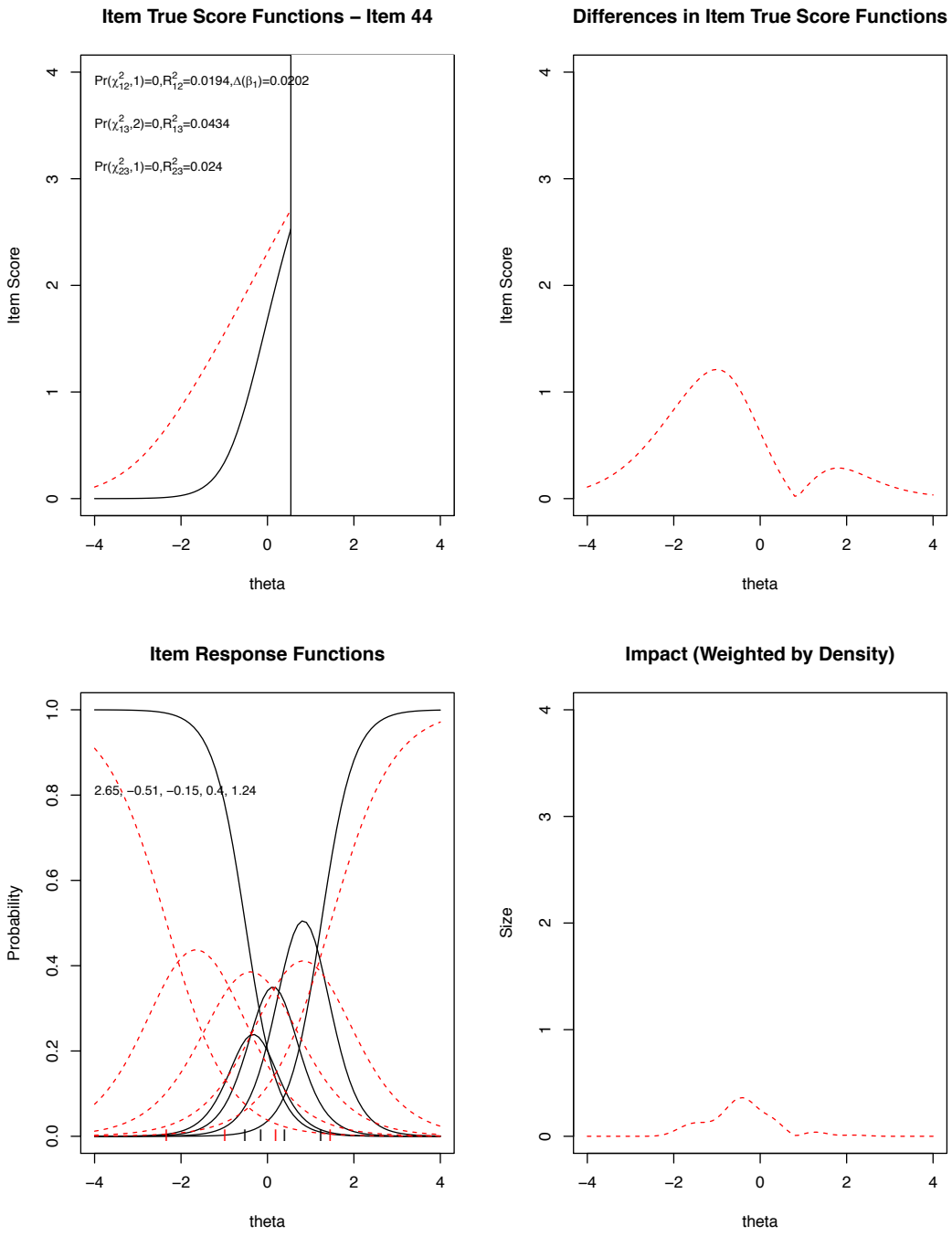

Figure 7: Graphical display of the item PFM2 (“Are you able to lift a heavy painting or picture to hang on your wall above eye-level?”), which shows non-uniform DIF for language.

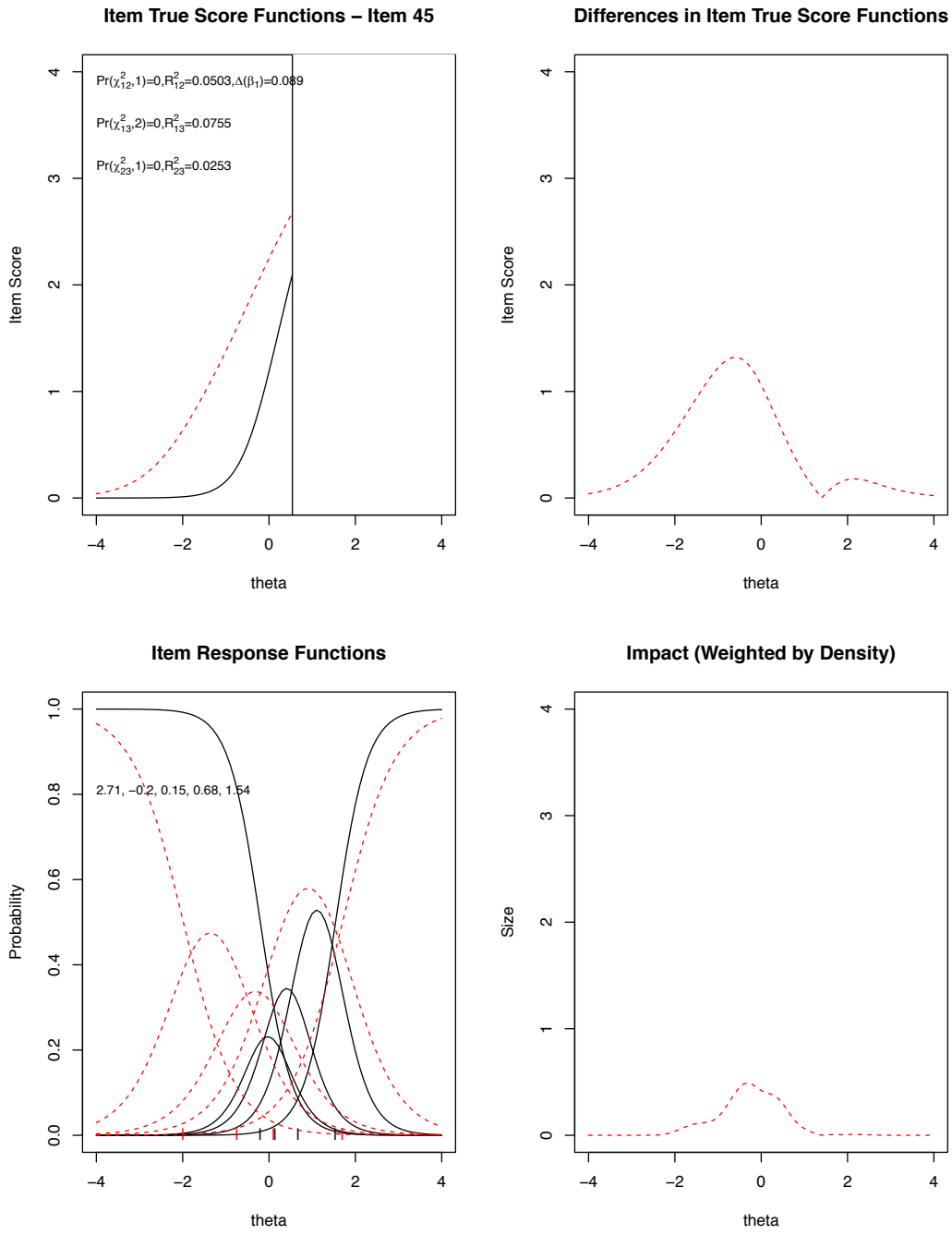

Figure 8: Graphical display of the item PFM16 (“Are you able to pass a 20-pound (10 kg) turkey or ham to other people at the table?”), which shows non-uniform DIF for language.

SAMENVATTING

Doelstelling De Patient Peported Outcomes Measurement Information System (PROMIS) item bank bovenste extremiteit v2.0 is ontwikkeld in de USA, maar is nog niet beschikbaar voor toepassing in Nederland of Vlaanderen. Het doel van deze studie was het bepalen van de cross-culturele validiteit en construct-validiteit van de Nederlands-Vlaamse PROMIS item bank bovenste extremiteit (v2.0), bij Nederlandse patiënten met een musculoskeletale aandoening van de bovenste extremiteit.

Methode Na vertaling werden de Nederlands-Vlaamse PROMIS item bank en verschillende "legacy" vragenlijsten afgenomen bij volwassen patiënten die een polikliniek voor orthopedie bezochten. Cross-culturele validiteit werd onderzocht door middel van Differential Item Functioning (DIF) analyses voor taal (Nederlands vs Engels). Construct validiteit werd onderzocht door het testen van hypotheses over correlaties van de item bank met de "legacy" vragenlijsten.

Resultaten De scores van 205 Nederlandse patiënten werden vergeleken met die van 246 Amerikaanse participanten. Acht items vertoonden minimale DIF voor taal, wat resulteerde in een voldoende cross-culturele validiteit. De Nederlands-Vlaamse item bank bovenste extremiteit had een matige correlatie met het Nederlands-Vlaamse PROMIS pijn intensiteit item ($r = -0.43$) en sterke correlaties (allen ≥ 0.7) met de andere "legacy" vragenlijsten.

Conclusie De Nederlands-Vlaamse PROMIS item bank bovenste extremiteit (v2.0) heeft een voldoende cross-culturele validiteit en een voldoende construct validiteit. De volgende stap is verdere validatie door middel van volledige item bank calibratie.

Klinische relevantie Na volledige item bank calibratie, kan de Nederlands-Vlaamse PROMIS item bank bovenste extremiteit (v2.0) dienen als basis voor "short forms" en computer adaptieve tests, wat resulteert in een hoge meetprecisie bij een lage belasting voor de patiënt. Dit zal de kwaliteit en efficiency van uitkomstmetingen in de gezondheidszorg verbeteren, bij Nederlandse en Vlaamse patiënten met klachten aan de bovenste extremiteit.