



The Patient Reported Inventory of Self-Management of Chronic Conditions (PRISM-CC): testing for bias across patient characteristics and languages

Ingrid Olsson¹ · George Kephart^{1,2,3} · Tanya Packer^{1,3} · Sabine Björk^{1,4} · Ulf Isaksson^{1,5} · Anna Nordström^{6,7} · Åsa Audulv¹

Received: 3 June 2025 / Accepted: 23 October 2025
© The Author(s) 2026

Abstract

Purpose Developed simultaneously in English and Swedish, the Patient Reported Inventory of Self-Management of Chronic Conditions (PRISM-CC) assesses patients' perceived difficulty managing life with long-term health conditions. This study assessed the comparability of the PRISM-CC across sociodemographic groups, number of health conditions and language (English and Swedish).

Methods Differential item functioning (DIF) and differential test functioning (DTF) were analysed by age, gender, education level, and number of conditions using independent English and Swedish datasets. Language-based DIF and DTF were examined using pooled data. An iterative hybrid ordinal logistic regression approach was applied to identify potential DIF across the PRISM-CC's seven domains. The impact of flagged items on total scores (DTF) was evaluated by comparing test characteristic curves.

Results Few items were flagged for potential DIF in the English, Swedish or pooled data, and only at low cutoff values. The impact of items with potential DIF on DTF was negligible.

Conclusion The absence of meaningful DIF and DTF in either the English or Swedish version of the PRISM-CC or between English and Swedish versions provides further support for the PRISM-CC as a valuable tool for assessing self-management ease and difficulty. These results also demonstrate the value of simultaneous development of instruments in two languages. Further evaluation of DIF is necessary in populations with greater self-management challenges, such as among people with severe disease burden.

Plain language summary

Health providers and researchers often use questionnaires and surveys to help make decisions, such as who needs treatment or what kind of treatment they need. Misleading or wrong decisions can be made if the tools are not well designed. This study was done to test whether a new tool called the Patient Reported Inventory of Self-Management of Chronic Conditions works equally well for men and women, older and younger people, and those with different levels of education and different numbers of conditions. It also tested whether the tool worked well in English and Swedish. The tool has 36 questions answered by patients, with results showing seven areas that people might find hard to manage. Over 1500 people from around the world provided data for this study. The results show that this tool yields comparable results across different types of patients in either English or Swedish.

Keywords Chronic disease · Differential item functioning · Differential test functioning · Measurement invariance · Patient reported outcome measure · Self-management

Extended author information available on the last page of the article

Abbreviations

DIF	Differential item functioning
DTF	Differential test functioning
HAI	Healthy aging initiative
IRT	Item response theory
NA	Not applicable
PRISM-CC	the Patient Reported Inventory of Self-Management of Chronic Conditions
TEDSS	Taxonomy of Everyday Self-management Strategies Framework
TCC	Test characteristic curves

Background

Health systems worldwide struggle to manage the increasing demands of long-term health conditions associated with population ageing and changing disease patterns [1–3]. Prevention, improved efficiency, and effectiveness of interventions are essential to meet this challenge. Improving individuals' ability to better self-manage their long-term health conditions can make a substantial contribution [4], as self-management has been shown to improve quality of life and health outcomes [5–9] while simultaneously decreasing healthcare costs [10]. Supporting patient self-management has, therefore, become internationally recognized as an essential function of healthcare systems [11, 12].

Given the complexity and changing nature of long-term health conditions, and therefore self-management, patients require different support at different times and from different healthcare professionals [7, 13, 14]. Consequently, providing tailored self-management support requires assessment of individual self-management challenges to establish if and what type of self-management support is needed [5, 15, 16]. In the research context, comprehensive self-management assessment measures are needed to describe, facilitate, and evaluate self-management interventions [17–20]. Despite this growing emphasis, self-management practice and science have been hampered by lack of rigorously designed

measurement instruments. Comprehensive and valid measures are critical for both research and clinical practice.

Unfortunately, there are widespread deficiencies in existing self-management measures [17–20]. They vary considerably in their aim, scope, theoretical foundation, and development [17, 19, 20], most are disease-specific, making them less suitable for individuals with multimorbidity [17, 20], and few acknowledge or measure the multi-dimensionality of patient self-management. Lack of suitable measures has necessitated use of related, but non-specific measures of quality of life [6, 21], activity level, or self-efficacy [17], complicating both interpretation and comparison of results [21].

To address these limitations, the Patient Reported Inventory of Self-Management of Chronic Conditions (PRISM-CC) was developed as a generic, multi-dimensional instrument that assesses perceived self-management ease or difficulty across seven self-management domains important for individuals living with one or more long-term health conditions [22, 23]. The PRISM-CC's seven domains are based on a validated theoretical framework: the Taxonomy of Everyday Self-management Strategies (TEDSS). TEDSS measures five goal-oriented domains (Internal, Social Interaction, Activities, Healthy Behaviours, and Disease Controlling domains), and two support-oriented domains (Process and Resource domains) [24] (see Table 1). TEDSS is increasingly being used to describe self-management and self-management support [25–27]. Based on TEDSS, developed iteratively in both English and Swedish and adhering to current guidelines for developing patient-reported outcome measures [28–30], the PRISM-CC is scaled using a multi-dimensional item response theory (IRT) graded response model. Robust construct validity, internal consistency, and test-retest reliability have been established [23, 31, 32].

A crucial next step in the PRISM-CC's validation is to examine whether it performs equivalently across demographic groups and languages. This involves testing for differential item functioning (DIF) and differential test functioning (DTF) to detect potential measurement bias [28]. Such bias could compromise clinical decision-making and

Table 1 The PRISM-CC domains and number of items

Resource (4 items)	Self-perceived ease or difficulty in seeking, pursuing and/or managing needed formal or informal supports and resources.
Process (5 items)	Self-perceived ease or difficulty in seeking information, being aware of choices and making good decisions.
Internal (8 items)	Self-perceived ease or difficulty in creating inner calm by preventing and managing stress, negative emotions, and internal distress.
Activity (5 items)	Self-perceived ease or difficulty in participating in everyday activities (leisure activities, work activities, household chores).
Social Interaction (5 items)	Self-perceived ease or difficulty in disclosing health issues, managing social interactions and relationships.
Healthy Behaviours (5 items)	Self-perceived ease or difficulty maintaining a healthy lifestyle in order to enhance health and limit the risk of lifestyle-related illness
Disease Controlling (4 items)	Self-perceived ease or difficulty in managing health conditions including managing medications and treatments, monitoring symptoms, and limiting complications.

research comparability across studies and groups. Because factors such as age [33, 34], disease progression [35], low levels of education [36, 37], and gender roles [35] are associated with self-management, it is essential that the PRISM-CC remains invariant across these characteristics. A valid measure should perform similarly regardless of these differences. Stated methodologically, DIF occurs when individuals with the same level of the trait being measured have systematically different response probabilities to an item depending on patient characteristics [38]. If one or more items in a measure display DIF, it may lead to DTF, resulting in systematic bias in test scores or variation in measurement error [39, 40].

This study assessed DIF and DTF within and between the English and Swedish versions of the PRISM-CC by age, gender, level of education, number of health conditions and language. Based on the rigorous development, refinement, selection, and translation process used in PRISM-CC item selection, we hypothesized that there would be no meaningful DIF or DTF within or between the English and Swedish PRISM-CC.

Methods

PRISM-CC Development

The PRISM-CC was developed by a multilingual research team. With slight variation in sequence, development occurred simultaneously in English and Swedish, to promote cross-linguistic conceptual equivalence (see Fig. 1).

Item development and selection were also guided by multiple strategies designed to minimize differential item and test functioning (DIF/DTF) [23, 32]. Central to this process was the TEDSS framework, which provided a robust conceptual foundation for operationalizing self-management as perceived ease or difficulty within each domain [23] and provided conceptual guidance for item development and selection. In Phase 1, a preliminary item bank (231 items) was constructed using existing instruments and qualitative data derived from both English and Swedish studies. At each step, items were subject to rigorous face and content validity assessment. In Phase 2, the preliminary item bank was subjected to initial testing and translation to

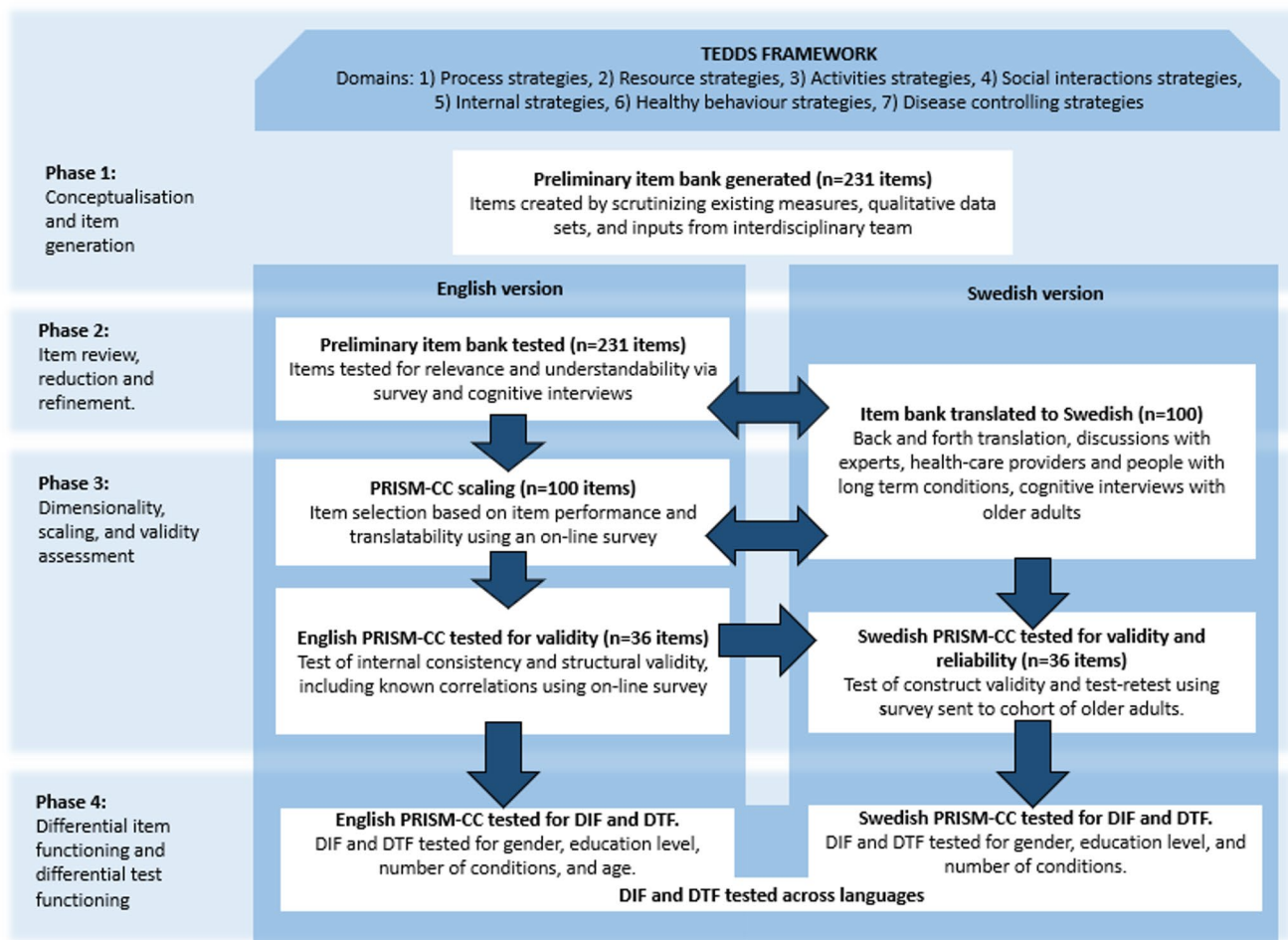


Fig. 1 Overview of the PRISM-CC development process

Swedish. Translation adhered to the Patient-Reported Outcome (PRO) Consortium translations process [41], including forward translation by a professional native speaking Swedish translator fluent in English and back translation by a professional native speaking English translator fluent in Swedish, as well as discussions with an expert group. Some small edits of the translation was made to capture the essence of the items, while making them understandable in Swedish. Cognitive interviews were performed in both languages [22, 31, 32], reducing the item bank to 100 items. These bilingual cognitive interviews provided critical insight into cross-language interpretation and potential sources of DIF, informing item refinement. In Phase 3, additional data collection, followed by psychometric testing, was used to select the final set of items. Then, the English version was scaled and assessed for structural and construct validity [23]. Swedish survey data were then used to assess the performance of the 36 items in Swedish and, in the case of one item showing suboptimal properties, select a replacement item. Structural validity and test-retest reliability were also tested for the Swedish version [31].

This paper reports the results of Phase 4: evaluating DIF and DTF of the PRISM-CC, thus were the English and Swedish datasets originally used for item selection, scaling, and construct validation reanalysed [23, 31].

Study Data

The PRISM-CC was designed as a generic instrument for individuals with long-term health conditions; thus, both datasets employed broad inclusion criteria. However, the recruitment strategies and eligibility criteria differed between the English and Swedish datasets based on the research question and access to participants. Therefore, the English and Swedish samples differ not only by language but also in demographic, clinical characteristics and number of participants.

The English dataset was collected via an online survey promoted through patient organizations, healthcare settings, social media, and recruitment platforms [23]. Eligible participants were adults (≥ 18 years) with one or more self-reported long-term health conditions. In total, 1213 people completed the survey, 158 were excluded due to $> 50\%$ missing item responses, leaving 1055 participants for inclusion in this study. Data collection adhered to the Canadian Tri-Council Policy Statement on Ethical Conduct for Research Involving Humans (including informed consent); ethics approval was given by the Nova Scotia Health Research Ethics Board (Romeo File #: 1025263).

The Swedish dataset was collected in collaboration with the Healthy Aging Initiative (HAI). This population-based study invited all residents in the municipality of Umeå to

participate the year they turned 70 years of age [42]. For this dataset, a follow-up survey was distributed to 1117 participants who had participated in the HAI study in 2018 and 2019 [31]. Since recent information about the individual's health status was unavailable, the survey was sent to all enrolled participants, informing them that only those with one or more long-term health conditions were eligible and should complete the survey. Of the 1117 individuals invited, 542 respondents met the inclusion criteria of having one or more long-term health conditions. After removing individuals with $> 50\%$ missing item responses, 516 participants were included in this study. The HAI study received ethical approval from the Swedish Ethical Review Authority and the Regional Ethics Review Board in Umeå in 2007 (Dnr 2012-85-32 M and Dnr 07-031 M), and complementary ethical applications to add the PRISM-CC data collection and to pool the English and Swedish data were approved in 2020 (Dnr 2020-02387), and 2022 (Dnr 2022-05547-02) respectively. All participants gave informed consent.

The PRISM-CC is comprised of 36 items (4–8 per domain: see Table 1 for domain definitions): items have been described elsewhere [23, 31, 32]. All items are formulated as statements and have a six-option response scale indicating respondents perceived difficulty (ranging from 'cannot do' to 'very easily') or agreement with each item. Low scores indicate more self-management difficulty. A 'not applicable' category was also available. In addition to the PRISM-CC, sociodemographic characteristics (gender, marital status, highest education level and financial status), and the number and type of self-reported long-term health conditions were collected. Age was collected in the English survey, whereas the age of Swedish participants was limited to 72–73 years.

Statistical analysis

Stata version 18 [43] was used to analyse sample demographics, clinical characteristics, and item responses. Differential item (DIF) and test functioning (DTF) were assessed according to patient characteristics: age (only English data set), gender, education, number of reported long-term health conditions, and language. As the analysis required dichotomized variables, age was grouped in two ways to reflect comparisons between age groups with different life experiences and chronic disease patterns: 18–30 versus 31 and over, and 18–60 versus 61 and over. In both datasets, gender was dichotomized as female and male; education as lower education (incomplete elementary school, elementary school, high school) and higher education (bachelor's degree, graduate degree); and number of conditions as 1–2 condition(s) or ≥ 3 conditions. Individuals with missing values for patient characteristic variables were excluded from

associated analyses. As described in prior publications [23, 31], missing values due to item non-response and non-applicable responses in both the English and Swedish data were imputed using chained multiple imputation.

Differential Item and Test Functioning

As the sociodemographic and long-term health condition attributes of the English and Swedish data were substantially different, we first assessed DIF/DTF separately in the English and Swedish data sets regarding age (English only), gender, education and number of conditions. We then used pooled data to assess DIF/DTF by language. To further assess whether DIF/DTF results by language reflect differences in demographics and chronic disease distributions, a sensitivity analysis was conducted with the English data restricted to >40 yrs of age ($n=444$) pooled with the Swedish data. This cutoff was chosen to focus on chronic conditions common in middle and older ages, while also ensuring adequate sample size.

DIF and DTF were assessed for each of the seven PRISM-CC domains by each patient characteristic (gender, level of education, two age groupings (English data), number of conditions, and language) using iterative hybrid ordinal logistic regression [44], implemented with R version 4.3.2 [45] and the “Lordif” R package version 0.3-3 [44]. This method assesses DIF by comparing three nested ordinal logistic regression models to determine whether a patient characteristic is associated with item response probabilities after adjustment for the latent variable (“theta”), represented by the estimated scores from the graded item response theory (IRT) model. The logic of this approach is that among subjects with the same value of theta, association of a patient characteristic with the probability of item response is evidence of DIF. The three ordinal logistic regression models are:

$$\text{Model (1)} \quad P(u_i \geq k) = a_k + \beta_1 \theta$$

$$\text{Model (2)} \quad P(u_i \geq k) = a_k + \beta_1 * \theta + \beta_2 * \text{Group}$$

$$\text{Model (3)} \quad P(u_i \geq k) = a_k + \beta_1 * \theta + \beta_2 * \text{Group} + \beta_3 * \theta * \text{Group}$$

$P(u_i \geq k)$ is the cumulative probability that an item response (u_i) falls in ordinal category k or higher, the a_k are intercepts (i.e. thresholds) between the k categories and β_1 is the slope parameter for estimated scores for the latent variable (θ). In model (2), β_2 measures the degree of “uniform” DIF by a patient characteristic, where the DIF is systematic across level of theta, while in model (3),

β_3 measures the degree of non-uniform DIF by a patient characteristic, where the DIF varies across levels of theta. Assessing improvements in fit between models assesses overall DIF (3 versus 1), uniform DIF (2 versus 1), and non-uniform DIF (3 versus 2). A problem with model estimation is that values of theta represented by IRT scores incorporate any DIF present in the items, thus biasing the estimation of DIF. To address this, the procedure uses an iterative “purification” process where estimates of theta are adjusted to correct for initial estimates of DIF, and then the analyses are repeated. This iterative process continues until the same set of items with evidence of DIF are identified [44, 46].

We used changes in McFadden pseudo- R^2 between the three ordinal logistic regression models as the criteria to detect items with DIF [47]. This is preferable to the use of likelihood ratio χ^2 tests, in which sensitivity depends on sample size and the precision of scores from the graded response model [48], and the Mantel-Haenszel (M-H) technique, which cannot be used to assess non-uniform DIF [49]. There is debate and disparate guidance in the literature as to cutoffs for changes in R^2 indicative of significant DIF [47, 49–52]. Accordingly, we used progressively more stringent cutoffs. Initially, we used a cutoff of > 0.035 which has been used to differentiate negligible from moderate DIF and is more conservative than a criteria of > 0.13, which has also been used [53]. However, this cutoff did not flag any items as having DIF for any patient characteristic in either the English or Swedish data. We thus used even more conservative proposed criteria (> 0.01 and > 0.006) to flag items with potential DIF [49]. A concern with low cutoffs is the risk of Type I error (false-positive identification of DIF) [50]. However, this is not a concern since the goal was to flag items for potential DIF, rather than to select items.

The most important analysis focused on the assessment of DTF, since the magnitude of bias in final scores resulting from DIF matters when using a measure [39, 54]. The effect of item DIF on scores can be washed out, counteracted or compounded by the influence of other items. For all domains where items were estimated to have DIF, DTF was assessed using two types of plots [44]. First, to assess the magnitude of bias by patient characteristics, we plotted and compared test characteristic curves showing differences in the expected total scores by level of theta, for each patient characteristic due to DIF. In the absence of DIF, the curves would be identical, and differences between the curves show the magnitude of DTF due to DIF. Second, to assess the distribution of bias for individual scores resulting from DIF, we plotted the bias for each subject (measured as the difference between “naïve” scores that ignored DIF and “purified scores” that accounted for DIF) by “naïve” (unpurified) levels of theta. In the absence of clinical criteria for a meaningful level of bias, we classified meaningful bias due

to DIF based on the 25th percentiles of the standard error of estimated theta values across the seven domains, which ranged from 0.24 to 0.33 in the English data and from 0.24 to 0.30 in the Swedish data. Accordingly, we conservatively defined meaningful bias as greater than 0.24.

Results

As noted, because of differences in recruitment, the sociodemographic characteristics of the English and Swedish participants differed (Table 2). Compared to the Swedish subjects, English subjects were younger, more likely to be female, highly educated, and to report greater perceived impact of their conditions on daily life.

Few items were flagged for potential DIF in either the English or Swedish data, and only at low cutoff values for changes in R^2 (Table 3). In the English data, our analysis tested for uniform and non-uniform DIF in 36 items by four different patient characteristics. Despite the large number of tests, potential DIF was flagged for only three items across three domains (Social Interaction, Healthy Behaviours, Internal): two by gender and only one of the two tested age grouping (i.e. age dichotomized as 18–60 vs. 61 and over). However, the effect sizes were small, with two items flagged for uniform DIF based on changes in R^2 slightly greater than 0.01 and an additional item flagged for uniform DIF with a change in R^2 of 0.008. Somewhat more items were flagged for potential DIF in the Swedish data. With tests for 36 items

across three different subject characteristics, nine items were flagged for potential DIF by gender, education level or number of health conditions. Again, all had small effect sizes, except for one item in the disease-controlling domain with a change in R^2 of 0.02 for uniform DIF (Table 3).

In both the Swedish and the English data, the estimated bias in measurement from items flagged for DIF (see Table 3), as assessed through DTF analysis, was very small. The items flagged for DIF contributed almost no difference in expected test scores across the full range of theta (see Fig. 2 English, and Fig. 3 Swedish). The box-whisker plots in the left panel of the figures show that: (1) in the English data (for both domains where items were flagged for potential DTF), the 10th and 90th percentiles of bias in subject scores were about ± 0.05 ; and (2) in the Swedish data, the 10th and 90th percentiles of bias were approximately ± 0.06 for the Disease Controlling domain and ± 0.04 for the Healthy Behaviours domain. These values, and the minimum and maximum biases in subject scores shown in the plots, are considerably below our cutoff for significant bias (0.24).

Having established negligible DIF and DTF in either the English or Swedish data, we proceeded to pool the two datasets and assess DIF and DTF by language. Only four items (across three domains) were flagged for potential DIF by language, two of which had changes in R^2 indicating uniform DIF greater than 0.02. DTF analysis indicated negligible bias in domain scores resulting from DIF in these two items (Fig. 4). Bias in individual scores was greater than

Table 2 Sociodemographic characteristics of the English and Swedish data

	Swedish total sample (n=516)	Swedish female (n=256)	Swedish male (n=249)	English total sample (n=1055)	English female (n=719)	English male (n=300)
Characteristic	n (%)	n (%)	n (%)	n (%)	n (%)	n (%)
<i>Age</i>						
18–30	0	0	0	354 (33.6)	222 (30.9)	112 (37.3)
31–60	0	0	0	328 (31.1)	238 (33.1)	78 (26.0)
>60	516 (100)	256 (100)	249 (100)	268 (25.4)	184 (25.6)	83 (27.7)
Missing	0	0	0	105 (10.0)	75 (10.4)	27 (9.0)
<i>Number of diseases*</i>						
1–2 disease(s)	260 (50.4)	116 (45.3)	140 (56.2)	740 (70.1)	491 (68.3)	235 (78.3)
≥3 diseases	256 (49.6)	140 (54.7)	109 (43.8)	315 (29.9)	228 (31.7)	65 (21.7)
<i>Perceived impact of disease(s) on life</i>						
Extremely	24 (4.7)	13 (5.1)	11 (4.4)	260 (24.8)	181 (25.2)	69 (23.0)
Quite a bit	92 (17.8)	52 (20.3)	38 (15.3)	379 (36.1)	272 (37.8)	87 (29.0)
Moderately	176 (34.1)	90 (35.2)	81 (32.5)	290 (27.6)	193 (26.8)	91 (30.3)
A little bit	139 (26.9)	69 (27.0)	69 (27.7)	107 (10.2)	61 (8.5)	46 (15.3)
Not at all	73 (14.2)	26 (10.2)	44 (17.7)	11 (1.1)	6 (0.8)	5 (1.7)
Missing	12 (2.3)	6 (2.3)	6 (2.4)	8 (0.4)	6 (0.9)	2 (0.7)
<i>Level of education</i>						
Lower education	263 (51.0)	116 (45.3)	146 (58.6)	231 (21.9)	157 (21.8)	60 (20.0)
Higher education	235 (45.5)	136 (53.1)	99 (39.8)	815 (77.3)	557 (77.5)	236 (78.7)
Missing	18 (3.5)	4 (1.6)	4 (1.6)	9 (0.9)	5 (0.7)	4 (1.3)

Table 3 Items flagged for potential Uniform, Non-Uniform and overall differential item functioning by variable and data source (English, Swedish and Pooled)

Domain	Item	Variable	Change in McFadden R^2		
			1 vs. 2 ¹	2 vs. 3 ²	1 vs. 3 ³
<i>English data</i>					
Social Interaction	Soc4: I devote time and attention to those who are dear to me	Gender	0.012	<0.001	0.012
Healthy Behaviours	Hea3: I find ways to train my brain to keep mentally fit.	Age ⁴	0.011	<0.001	0.012
Internal	Int4: I have and use ways to recover after a bad day	Gender	0.008	<0.001	0.008
<i>Swedish data</i>					
Disease Controlling	Dis1: When problems with my health arise, I understand what to do to manage my condition(s).	Gender	0.020	<0.001	0.020
Healthy Behaviours	Hea5S: I maintain healthy behaviours even when I have a lot to do.	Gender	0.010	0.007	0.016
Healthy Behaviours	Hea3: I find ways to train my brain to keep mentally fit.	Gender	0.015	<0.001	0.015
Process	Pro4: I try different things to find out what works best for me.	# Conditions	0.002	0.013	0.015
Healthy Behaviours	Hea3: I find ways to train my brain to keep mentally fit.	Education	0.011	<0.001	0.011
Social Interaction	Soc5: When problems with my health arise, I stay in touch with people who are important to me.	Education	0.010	<0.001	0.010
<i>Pooled data</i>					
Resource	Res4: When I need to, I find people to help me understand information I receive about my condition(s)	Language	0.027	<0.001	0.028
Process	Pro4: I try different things to find out what works best for me.	Language	0.021	0.007	0.028
Process	Pro2: I make informed decisions	Language	<0.001	0.009	0.009
Internal	Int7: I focus on the positives	Language	0.007	<0.001	0.007

¹Uniform DIF, ²Non-uniform DIF, ³ Overall DIF, ⁴ Age grouped 18–60 versus 61 and over. Reported change in R^2 between models is after “purification” of latent variable for DIF. Within each data set, items are sorted by largest change in McFadden R^2 .

observed in the separate English and Swedish analyses (10th and 90th percentiles: ± 0.15 for the Resource domain and -0.15 to 0.19 for the Process domain), but were still below the 0.24 cutoff we established for meaningful bias. The sensitivity analysis, limiting the English sample to ages >40 , produced highly similar results to analyses using the full sample. The same items were flagged for potential DIF, and they contributed negligible bias to domain scores.

Discussion

No evidence of meaningful DIF or DTF in the PRISM-CC by gender, level of education, number of conditions, age (18–60 vs. 61 and over) or language was found. The low level of DIF and DTF builds confidence in using PRISM-CC with diverse patient populations and in cross-language studies. Failure to appropriately assess patient-reported outcome measures for DIF and/or DTF before making group comparisons may lead to significant interpretational challenges, as differences can result from how the two groups interpret the latent construct and not a real group difference [54]. Evaluating a measure’s equivalence is crucial to ensure fair patient assessment [55]. Clinicians can be confident that use of the PRISM-CC is a consistent measure of self-management ease or difficulty, regardless of the age, gender, education or multimorbidity of patients. It can, therefore, be

used to identify areas of concern and tailor care. Researchers can also be confident in comparisons between studies by variables assessed in our study and expect that comparisons by other variables are also unlikely to be biased.

As noted, a finding of DIF and DTF within a measurement tool identifies potential biases. In contrast, a finding of no or limited DIF or DTF strengthens the confidence in the tool’s results. The PRISM-CC stands out as one of only a few self-management measures assessed for DIF and/or DTF [56–58]. Where DIF has been assessed for self-management measures [59–66], and self-reported health measures more generally [57, 67, 68], DIF by gender, education and across disease groups have been found. The Patient Activation Measure [69, 70], one of the most widely used measures to assess self-management, has shown DIF across gender and education [59–63], and across different disease groups [64, 65]. DTF was also found for the Health Education Impact Questionnaire across disease groups in two of eight subscales [66]. Self-management measures not tested for DIF and/or DTF should be used cautiously when comparing patient groups. As noted, confirmation of a lack of bias has important implications for both clinicians and researchers.

The findings of this study suggest that a thorough development process can limit the risk of inherent biases in a measurement tool. We attribute the remarkably low level of DIF and DTF detected in PRISM-CC to rigorous application

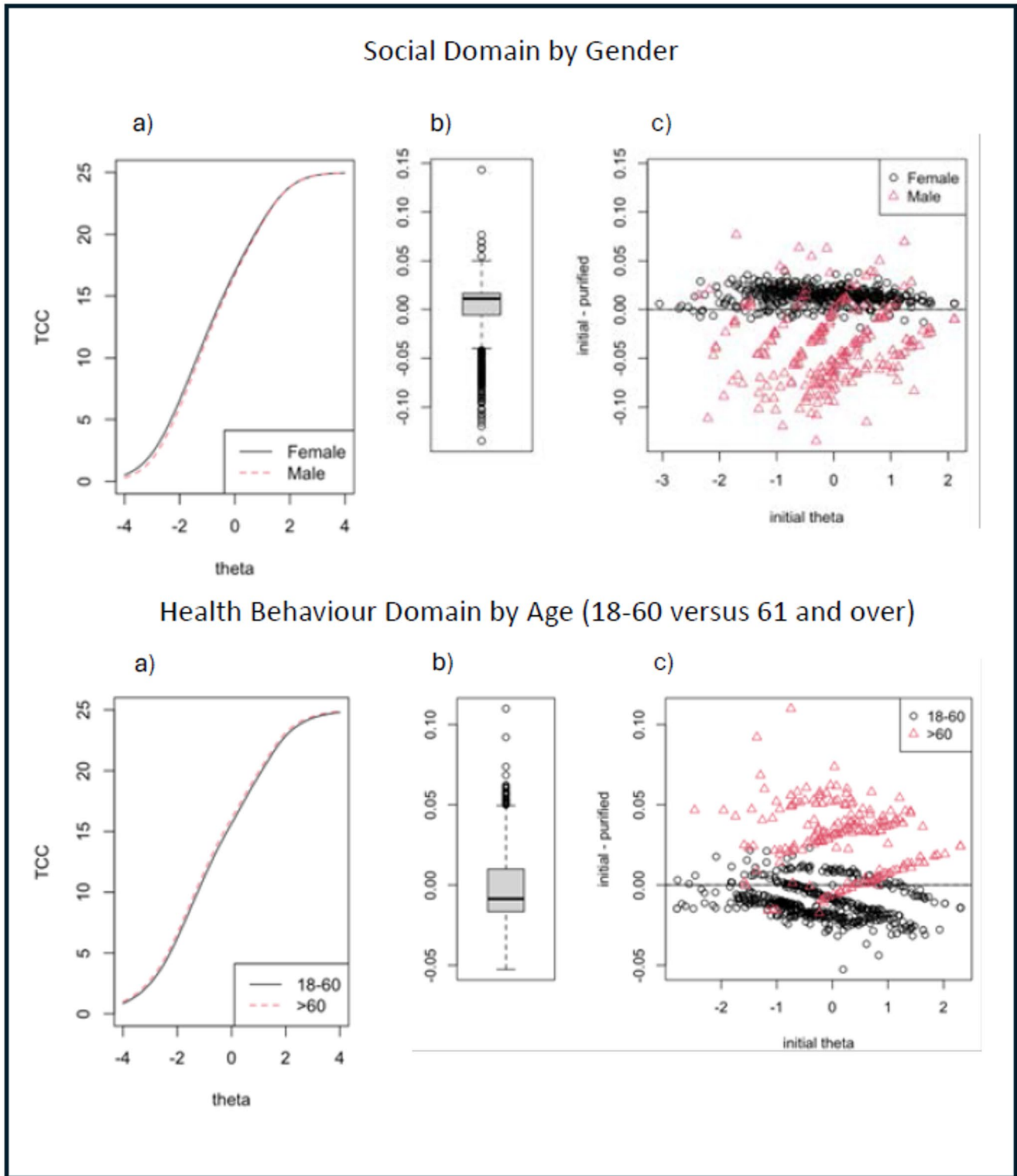


Fig. 2 DTF analysis of bias for the two domains with the largest DIF in the English dataset. The left-hand panels (labelled “a”) show the difference in test characteristic curves (i.e. expected summative test scores

by level of theta) due to DIF. The two right-hand panels (labelled “b” and “c”) show the estimated distribution of bias in individual scores due to DIF

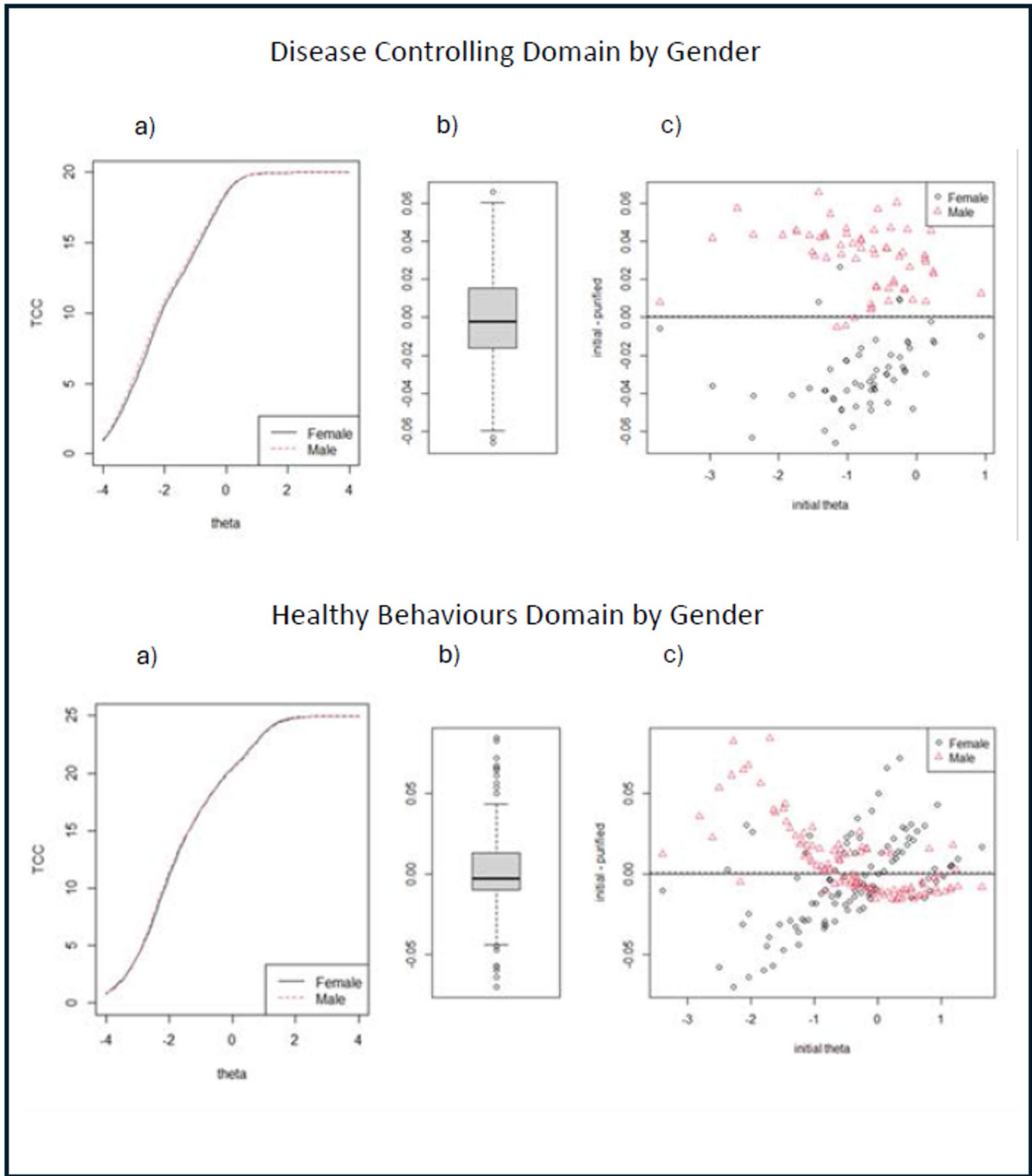


Fig. 3 DTF analysis of bias for the two domains with the largest DIF in the Swedish dataset. The left-hand panels (labelled “a”) show the difference in test characteristic curves (i.e. expected summative

test scores by level of theta) due to DIF. The two right-hand panels (labelled “b” and “c”) show the estimated distribution of bias in individual scores due to DIF

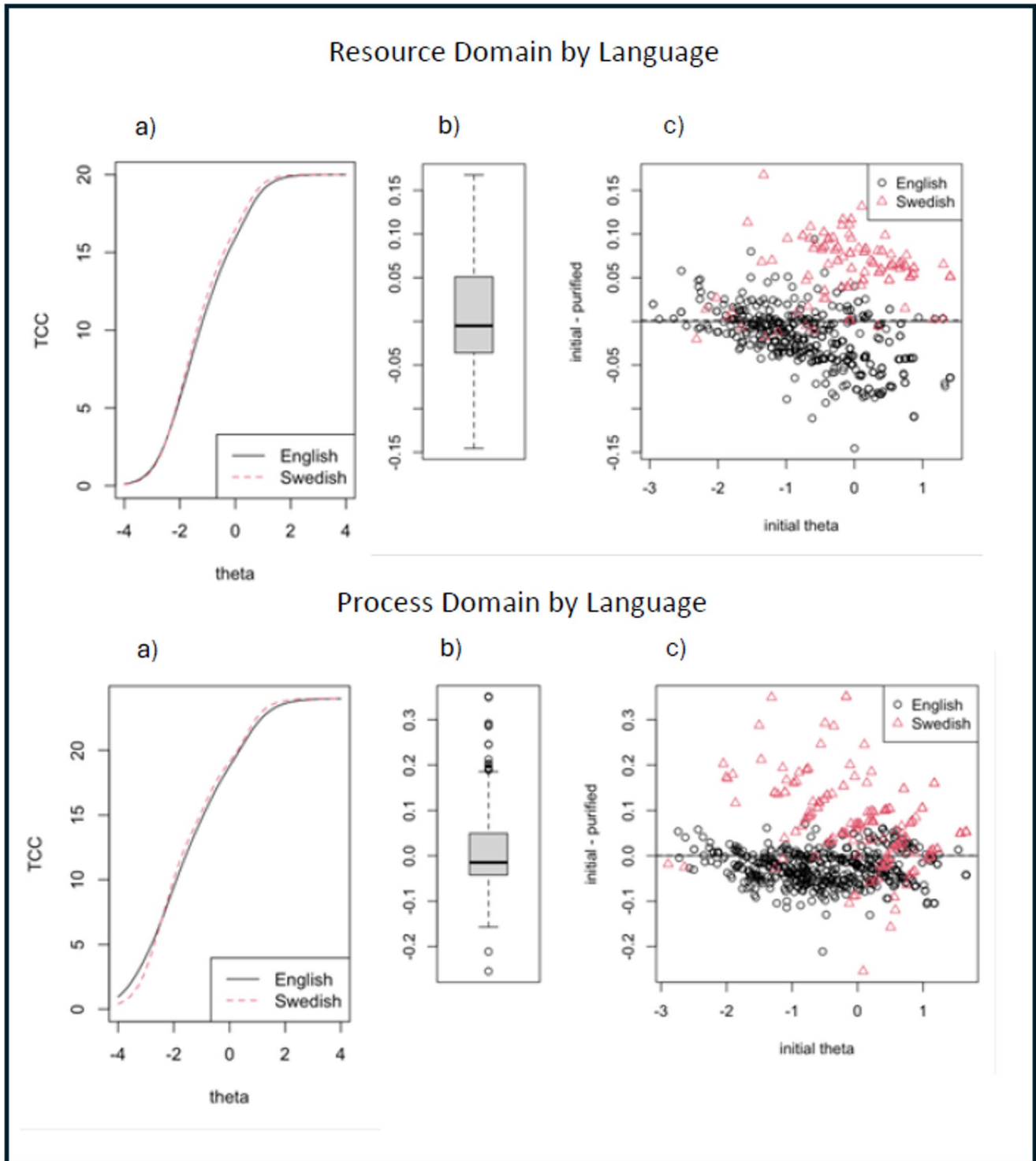


Fig. 4 DTF analysis of bias for the two domains with the largest DIF across languages. The left-hand panels (labelled “a”) show the difference in test characteristic curves (i.e. expected summative test scores

by level of theta) due to DIF. The two right-hand panels (labelled “b” and “c”) show the estimated distribution of bias in individual scores due to DIF

of a number of practices, many of which are recommended in guidelines for developing patient-reported measures [28, 29]. We propose that avoiding DIF and DTF in measures is best addressed at a measure's conceptual and development phase. We further suggest that collaborations across countries and disciplines can lead to the development of stronger instruments. The PRISM-CC is unusual in that it was simultaneously developed in two languages by a multi-disciplinary (nursing, occupational therapy, physiotherapy, and health services researchers) and multilingual (English, Swedish, Portuguese, and Mandarin first language speakers) team [23, 31]. Translatability was extensively used to inform item wording and selection. Further, patients were engaged in the item development and selection process. Cognitive interviewing was used to identify differences in patient interpretation, which contributed to item revision and selection.

In this study, DIF/DTF was tested for four demographic variables (age, education, gender and number of conditions) and two languages. However, other variables have also been found to associate with self-management behaviours. For example, self-efficacy underpins many self-management interventions and has been shown to be strongly associated with self-management [34, 37]. Likewise, impaired functional ability such as decreased strength or sensory impairments [71], disease progression, years living with a condition [33], cognitive impairment [35], family and social support [33, 72, 73], health literacy [34, 74, 75], or mental health [76–78] have all been related to self-management. Future research to better understand if these variables are associated with bias in PRISM-CC scores would further strengthen confidence in its use.

Limitations in the available data may have impacted this study's ability to detect meaningful DIF and DTF. For example, the assessment of DIF and/or DTF by disease duration and type of condition was not possible. This, along with different categorizations of education and number of conditions should be a priority for future research since it may be important for clinical practice and research. As previously noted, the Swedish sample did not allow assessment of DIF and DTF by age. Using a close-to representative sample of older adults age 72–23, meant the number of participants with severe self-management challenges may have been underpowered to detect DIF in this subgroup—a limitation that future clinical studies should also address. Despite this analytical limitation, the findings are generalizable to a similar cohort of older adults in Sweden. Findings from the large English dataset, a convenience sample that specifically recruited participants living with a chronic condition, may be more similar and generalizable to clinical populations.

The Swedish and English samples differed considerably in their demographic and chronic disease attributes, which is both a limitation and a strength. The key limitation is that isolation of DIF/DTF by language is challenging. We addressed this by first ensuring there was no meaningful DIF within either sample by age, education and number of conditions before pooling to test for language differences. An alternative approach would be to select matched samples from the two data sources before testing for DIF/DTF, but there was not sufficient overlap in subject characteristics to permit this. As a crude alternative, a sensitivity analyses where only participants > 40 years of age in the English data were pooled with the Swedish data yielded nearly identical results. The assessment of DIF and DTF between the English and Swedish samples is a key strength, however, if framed as global assessment covering differences by language, culture, sociodemographic and health attributes. The fact that no meaningful DIF or DTF was found in two considerably different samples adds strength to the findings that there is little DIF or DTF in the PRISM-CC items or domain scores.

Conclusion

The absence of meaningful DIF and DTF in the PRISM-CC underscores its significance as a valuable tool for assessing self-management difficulty. It demonstrates its utility for application across diverse demographic groups and languages. Combined with previous evidence of its strong psychometric properties in Swedish and English, the PRISM-CC is a promising instrument for research and clinical practice.

Acknowledgements The authors wish to thank all participants. We want to thank Jonathan Bergman from the Healthy Ageing Initiative (HAI) and Julia Andersson for supporting the Swedish data collection and recruitment. For funding and support, we want to acknowledge the Medical Faculty at Umeå University and the Swedish National Graduate School for Competitive Science on Ageing and Health (SWEAH), funded by the Swedish Research Council. The English data collection was funded by the Canadian Institutes for Health Research (CIHR Award 152932) and the Nova Scotia Health Research Foundation (Award 222). Special thanks to our patient advisors, all research staff and the many administrators and practitioners at Nova Scotia health who supported the development of the PRISM-CC. The authors wish to thank all participants. This study was accomplished within the Swedish National Graduate School for Competitive Science on Ageing and Health (SWEAH), funded by the Swedish Research Council. Special appreciation to Jonathan Bergman from the Healthy Ageing Initiative (HAI) for supporting participant recruitment. Special thanks to Julia Andersson for her help during data collection.

Author contributions All authors contributed to the study's conception and design. IO, ÅA and SB performed material preparation and data collection. IO and GK performed analysis. IO, ÅA and GK wrote the first draft of the manuscript, and all authors commented on previ-

ous versions. All authors read and approved the final manuscript.

Funding Open access funding provided by Umeå University. This research was funded by a strategic research resource from the Faculty of Medicine, Umeå University and by grants from Erik and Anne Marie Detlof's Foundation, Umeå University. The funding bodies have not participated in the design, data collection, analysis, or interpretation of data or in writing the manuscript.

Data availability The dataset generated and analysed during the current study is not publicly available because participant consent included restrictions on the use of the data due to patients' privacy concerns. Limited availability is possible. Researchers wishing for information may contact the corresponding author Åsa Audulv.

Declarations

Conflict of interest The authors have no relevant financial interests to disclose. However, the PRISM-CC has been developed and is owned by TP, GK and ÅA.

Ethical approval The Healthy Ageing Initiative (HAI) study has received ethical approval from the Swedish Ethical Review Authority and the Regional Ethics Review Board in Umeå in 2007 (Dnr 2012-85-32 M- and dnr 07-031 M), and a complementary ethical application for this project was approved by the Regional Ethics Review Board in Umeå in 2020 (Dnr 2020-02387). This study was performed in agreement with the principles of the Declaration of Helsinki.

Consent to participate Written informed consent was obtained from all participants.

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Vollset, S. E., et al. (2024). Burden of disease scenarios for 204 countries and territories, 2022–2050: A forecasting analysis for the global burden of disease study 2021. *Lancet*, *403*(10440), 2204–2256. [https://doi.org/10.1016/S0140-6736\(24\)00685-8](https://doi.org/10.1016/S0140-6736(24)00685-8)
- Anderson, M., Revie, C. W., Stryhn, H., Neudorf, C., Rosehart, Y., Li, W., Osman, M., Buckridge, D. L., Rosella, L. C., & Wodchis, W. P. (2019). Defining 'actionable' high- costhealth care use: Results using the Canadian Institute for health information population grouping methodology. *International Journal for Equity in Health*, *18*(1), 171. <https://doi.org/10.1186/s12939-019-1074-3>
- Palladino, R., Lee, T., Ashworth, J., Triassi, M., M., & Millett, C. (2016). Associations between multimorbidity, healthcare utilisation and health status: Evidence from 16 European countries. *Age and Ageing*, *45*(3), 431–435. <https://doi.org/10.1093/ageing/afw044>
- Institute of Medicine (US) Committee on the Crossing the Quality Chasm: Next Steps Toward a New Health Care System (2004). In: Adams K, Greiner AC, Corrigan JM (Ed.). *The 1st Annual Crossing the Quality Chasm Summit: A Focus on Communities*. Washington (DC), National Academies Press (US). Retrieved February 3, 2025, from <https://www.ncbi.nlm.nih.gov/books/NBK215518/>
- Camargo-Plazas, P., Robertson, M., Alvarado, B., Paré, G. C., Costa, I. G., & Duhn, L. (2023). Diabetes self-management education (DSME) for older persons in Western countries: A scoping review. *PLoS One*, *18*(8), 0288797. <https://doi.org/10.1371/journal.pone.0288797>
- Timmermans, L., Golder, E., Decat, P., Foulon, V., Van Hecke, A., & Schoenmakers, B. (2023). Characteristics of self-management support (SMS) interventions and their impact on quality of life (QoL) in adults with chronic diseases: An umbrella review of systematic reviews. *Health Policy*, *135*, 104880. <https://doi.org/10.1016/j.healthpol.2023.104880>
- Lancey, A., & Slater, C. E. (2023). Heart failure self-management: A scoping review of interventions implemented by allied health professionals. *Disability and Rehabilitation*, *46*(21), 4848–4859. <https://doi.org/10.1080/09638288.2023.2283105>
- Zhao, Q., Chen, C., Zhang, J., Ye, Y., & Fan, X. (2021). Effects of self-management interventions on heart failure: Systematic review and meta-analysis of randomized controlled trials - Reprint. *International Journal of Nursing Studies*, *116*, 103909. <https://doi.org/10.1016/j.ijnurstu.2021.103909>
- Allegrante, J. P., Wells, M. T., & Peterson, J. C. (2019). Interventions to support behavioral Self-Management of chronic diseases. *Annual Review of Public Health*, *40*, 127–146. <https://doi.org/10.1146/annurev-publhealth-040218-044008>
- Talboom-Kamp, E., Ketelaar, P., & Versluis, A. (2021). A National program to support self-management for patients with a chronic condition in primary care: A social return on investment analysis. *Clinical eHealth*, *4*, 45–49. <https://doi.org/10.1016/j.ceh.2021.02.001>
- Zhou, Y., Dai, X., Ni, Y., Zeng, Q., Cheng, Y., Carrillo-Larco, R. M., Yan, L. L., & Xu, X. (2023). Interventions and management on multimorbidity: An overview of systematic reviews. *Ageing Research Reviews*, *87*, 101901. <https://doi.org/10.1016/j.arr.2023.101901>
- World Health Organization (2015). WHO global strategy on people-centred and integrated health services: interim report. Geneva: World Health Organization. Retrieved February 3, 2025, from https://iris.who.int/bitstream/handle/10665/155002/WHO_HIS_SDS_2015.6_eng.pdf?sequence=1&isAllowed=y:%20World%20Health%20organization;%202015.%20%20Contract%20No.:%20WHO/HIS/SDS/2015.6
- Ghasemiardekani, M., Willetts, G., Hood, K., & Cross, W. (2024). The effectiveness of chronic disease management planning on self-management among patients with diabetes at general practice settings in australia: A scoping review. *BMC Primary Care*, *25*(1), 75. <https://doi.org/10.1186/s12875-024-02309-4>
- Keddy, A. C., Packer, T. L., Audulv, Å., Sutherland, L., Sampalli, T., Edwards, L., & Kephart, G. (2021). The team assessment of Self-Management support (TASMS): A new approach to Uncovering how teams support people with chronic conditions. *Healthcare Management Forum*, *34*(1), 43–48. <https://doi.org/10.1177/0840470420942262>
- Byrne, G., Keogh, B., & Daly, L. (2022). Self-management support for older adults with chronic illness: Implications for nursing practice. *British Journal of Nursing (Mark Allen Publishing)*, *31*(2), 86–94. <https://doi.org/10.12968/bjon.2022.31.2.86>

16. NHS England. Supported self-management. Retrieved February 17 (2025). from <https://www.england.nhs.uk/personalisedcare/supported-self-management/>
17. Lawless, M. T., Tieu, M., Chan, R. J., Hendriks, J. M., & Kitson, A. (2023). Instruments measuring Self-Care and Self-Management of chronic conditions by Community-Dwelling older adults: A scoping review. *Journal of Applied Gerontology: the Official Journal of the Southern Gerontological Society*, 42(7), 1687–1709. <https://doi.org/10.1177/07334648231161929>
18. Wientzek, R., Brückner, R. M., Schönenberg, A., & Prell, T. (2023). Instruments for measuring self-management and self-care in geriatric patients - a scoping review. *Frontiers in Public Health*, 11, 1284350. <https://doi.org/10.3389/fpubh.2023.1284350>
19. Hudon, É., Hudon, C., Lambert, M., Bisson, M., & Chouinard, M. C. (2021). Generic Self-Reported questionnaires measuring Self-Management: A scoping review. *Clinical Nursing Research*, 30(6), 855–865. <https://doi.org/10.1177/1054773820974149>
20. Packer, T. L., Fracini, A., Auduly, Å., Alizadeh, N., van Gaal, B. G. I., Warner, G., & Kephart, G. (2018). What we know about the purpose, theoretical foundation, scope and dimensionality of existing self-management measurement tools: A scoping review. *Patient Education and Counseling*, 101(4), 579–595. <https://doi.org/10.1016/j.pec.2017.10.014>
21. Hansen, C. W., Esbensen, B. A., de Thurah, A., Christensen, R., de Wit, M., & Cromhout, P. F. (2022). Outcome measures in rheumatology applied in self-management interventions targeting people with inflammatory arthritis A systematic review of outcome domains and measurement instruments. *Seminars in Arthritis and Rheumatism*, 54, 151995. <https://doi.org/10.1016/j.semarthrit.2022.151995>
22. Auduly, Å., Hutchinson, S., Warner, G., Kephart, G., Versnel, J., & Packer, T. L. (2021). Managing everyday life: Self-management strategies people use to live well with neurological conditions. *Patient Education and Counseling*, 104(2), 413–421. <https://doi.org/10.1016/j.pec.2020.07.025>
23. Kephart, G., Packer, T., Auduly, Å., Chen, Y. T., Robinson, A., Olsson, I., & Warner, G. (2022). Item selection, scaling and construct validation of the Patient-Reported inventory of Self-Management of chronic conditions (PRISM-CC) measurement tool in adults. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment Care and Rehabilitation*, 31(9), 2867–2880. <https://doi.org/10.1007/s11136-022-03165-4>
24. Auduly, Å., Ghahari, S., Kephart, G., Warner, G., & Packer, T. L. (2019). The taxonomy of everyday Self-management strategies (TEDSS): A framework derived from the literature and refined using empirical data. *Patient Education and Counseling*, 102(2), 367–375. <https://doi.org/10.1016/j.pec.2018.08.034>
25. Wright, H., Turner, A., Ennis, S., Percy, C., Loftus, G., Clyne, W., Matouskova, G., & Martin, F. (2022). Digital Peer-Supported Self-Management intervention codesigned by people with long COVID: Mixed methods Proof-of-Concept study. *JMIR Formative Research*, 6(10), e41410. <https://doi.org/10.2196/41410>
26. Hutchinson, S., Lauckner, H., Stilwell, C., & Meisner, B. A. (2022). Leisure and leisure education as resources for rehabilitation supports for chronic condition Self-Management in rural and remote communities. *Frontiers in Rehabilitation Sciences*, 3, 889209. <https://doi.org/10.3389/fresc.2022.889209>
27. Cadel, L., Overall, A. C., Packer, T. L., Hitzig, S. L., Patel, T., Lofters, A. K., & Guilcher, S. J. T. (2020). Exploring the perspectives on medication self-management among persons with spinal cord injury/dysfunction and providers. *Research in Social & Administrative Pharmacy: RSAP*, 16(12), 1775–1784. <https://doi.org/10.1016/j.sapharm.2020.01.014>
28. PROMIS® Instrument Development and Validation Scientific Standards Version 2.0 (revised (May 2013). Retrieved March 24, 2025, from https://www.healthmeasures.net/images/PROMIS/PR OMISStandards_Vers2.0_Final.pdf
29. Gagnier, J. J., Lai, J., Mokkink, L. B., & Terwee, C. B. (2021). COSMIN reporting guideline for studies on measurement properties of patient-reported outcome measures. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment Care and Rehabilitation*, 30(8), 2197–2218. <https://doi.org/10.1007/s11136-021-02822-4>
30. Mokkink, L. B., de Vet, H. C. W., Prinsen, C. A. C., Patrick, D. L., Alonso, J., Bouter, L. M., & Terwee, C. B. (2018). 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*, 27(5), 1171–1179. <https://doi.org/10.1007/s11136-017-1765-4>
31. Olsson, I., Kephart, G., Packer, T., Björk, S., Isaksson, U., Chen, Y-T., Nordström, A., & Auduly, Å. (2025). Structural validity and test-retest reliability of the patient reported inventory of Self-Management of chronic conditions (PRISM-CC) in a Swedish population of seventy-year-olds with long-term health conditions. *Journal of Patient-Reported Outcomes*, 9(1), 59. <https://doi.org/10.1186/s41687-025-00892-3>
32. Packer, T., Kephart, G., Auduly, Å., Keddy, A., Warner, G., Peacock, K., & Sampalli, T. (2020). Protocol for development, calibration and validation of the Patient-Reported inventory of Self-Management of chronic conditions (PRISM-CC). *BMJ Open*, 10(9), e036776. <https://doi.org/10.1136/bmjopen-2020-036776>
33. Cong, Z., Huo, M., Jiang, X., & Yu, H. (2024). Factors associated with the level of self-management in elderly patients with chronic diseases: A pathway analysis. *BMC Geriatrics*, 24(1), 377. <https://doi.org/10.1186/s12877-024-04956-9>
34. Dinh, T. T. H., & Bonner, A. (2023). Exploring the relationships between health literacy, social support, self-efficacy and self-management in adults with multiple chronic diseases. *BMC Health Services Research*, 23(1), 923. <https://doi.org/10.1186/s12913-023-09907-5>
35. Nguyen, T. N. M., Whitehead, L., Saunders, R., & Dermody, G. (2022). Systematic review of perception of barriers and facilitators to chronic disease self-management among older adults: Implications for evidence-based practice. *Worldviews on Evidence-based Nursing*, 19(3), 191–200. <https://doi.org/10.1111/wvn.12563>
36. Adjei Boakye, E., Varble, A., Rojek, R., Peavler, O., Trainer, A. K., Osazuwa-Peters, N., & Hinyard, L. (2018). Sociodemographic factors associated with engagement in diabetes Self-management education among people with diabetes in the united States. *Public Health Reports (Washington D C : 1974)*, 133(6), 685–691. <https://doi.org/10.1177/0033354918794935>
37. Ni, Z., Wang, Y., Jiang, N., & Jiang, X. (2022). Predictors of health self-management behaviour in Kazakh patients with metabolic syndrome: A cross-sectional study in China. *PLoS One*, 17(12), e0278190. <https://doi.org/10.1371/journal.pone.0278190>
38. Osterlind, S., & Everson, H. (2009). *Differential Item Functioning*. Thousand Oaks, California.
39. Chalmers, R. P., Counsell, A., & Flora, D. B. (2016). It might not make a big DIF: Improved differential test functioning statistics that account for sampling variability. *Educational and Psychological Measurement*, 76(1), 114–140. <https://doi.org/10.1177/0013164415584576>
40. Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is Now, and where it is going. *Language Assessment Quarterly*, 4(2), 223–233. <https://doi.org/10.1080/15434300701375832>
41. Eremenco, S., Pease, S., Mann, S., Berry, P., & PRO Consortium's Process Subcommittee. (2017). Patient-Reported outcome

- (PRO) consortium translation process: Consensus development of updated best practices. *Journal of Patient-Reported Outcomes*, 2(1), 12. <https://doi.org/10.1186/s41687-018-0037-6>
42. Nordström, A., Bergman, J., Björk, S., Carlberg, B., Johansson, J., Hult, A., & Nordström, P. (2020). A multiple risk factor program is associated with decreased risk of cardiovascular disease in 70-year-olds: A cohort study from Sweden. *PLoS Medicine*, 17(6), e1003135. <https://doi.org/10.1371/journal.pmed.1003135>
 43. StataCorp. (2023). *Stata statistical software: Release 18*. College station. StataCorp LLC.
 44. Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic Regression/Item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39(8), 1–30. <https://doi.org/10.18637/jss.v039.i08>
 45. R Core Team R: A Language and Environment for Statistical Computing. Version 4.3.2. R Foundation for Statistical Computing, Vienna, Austria.
 46. Crane, P. K., Gibbons, L. E., Jolley, L., & van Belle, G. (2006). Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. *Medical Care*, 44(11 Suppl 3), 115–123. <https://doi.org/10.1097/01.mlr.0000245183.28384.ed>
 47. Cuevas, M., & Cervantes, V. H. (2012). Differential item functioning detection with logistic regression. *Mathématiques et sciences humaines. Mathematics and Social Sciences*, 199, 45–59. <https://doi.org/10.4000/msh.12274>
 48. Gómez-Benito, J., Hidalgo, M. D., & Zumbo, B. D. (2013). Effectiveness of combining statistical tests and effect sizes when using logistic discriminant function regression to detect differential item functioning for polytomous items. *Educational and Psychological Measurement*, 73(5), 875–897. <https://doi.org/10.1177/0013164413492419>
 49. Gesicki, A. (2015). Decision rules based on hypothesis tests and effect sizes for logistic regression differential item functioning [dissertation on the Internet]. Vancouver: University of British Columbia. Retrieved April 7, 2025, from <https://central.bac-lac.gc.ca/.item?id=TC-BVAU-54871>
 50. Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, 14(4), 329–349. https://doi.org/10.1207/S15324818AME1404_2
 51. French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement*, 67(3), 373–393. <https://doi.org/10.1177/0013164406294781>
 52. Gómez-Benito, J., Hidalgo, M. D., & Padilla, J. L. (2009). Efficacy of effect size measures in logistic regression: An application for detecting DIF. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, 5(1), 18–25. <https://doi.org/10.1027/1614-2241.5.1.18>
 53. Zumbo, B. D., & Department of National Defense. (1999). *A Handbook on the Theory and Methods of Differential Item Functioning (DIF): Logistic Regression Modeling as a Unitary Framework for Binary and Likert-Type (Ordinal) Item Scores*. Ottawa, ON: Directorate of Human Resources Research and Evaluation. Retrieved March 10, 2025, from <https://faculty.educ.ubc.ca/zumbo/DIF/handbook.pdf>
 54. Teresi, J. A., Ramirez, M., Lai, J. S., & Silver, S. (2008). Occurrences and sources of differential item functioning (DIF) in patient-reported outcome measures: Description of DIF methods, and review of measures of depression, quality of life and general health. *Psychology Science Quarterly*, 50(4), 538.
 55. McHorney, C. A., & Fleishman, J. A. (2006). Assessing and Understanding measurement equivalence in health outcome measures. Issues for further quantitative and qualitative inquiry. *Medical Care*, 44(11 Suppl 3), S205–S210. <https://doi.org/10.1097/01.mlr.0000245451.67862.57>
 56. Chen, Y., Lu, M., & Jia, L. (2023). Psychometric properties of self-reported measures of self-management for chronic heart failure patients: A systematic review. *European Journal of Cardiovascular Nursing*, 22(8), 758–764. <https://doi.org/10.1093/eurjcn/zvad028>
 57. Wee, P. J. L., Kwan, Y. H., Loh, D. H. F., Phang, J. K., Puar, T. H., Østbye, T., Thumboo, J., Yoon, S., & Low, L. L. (2021). Measurement properties of Patient-Reported outcome measures for diabetes: Systematic review. *Journal of Medical Internet Research*, 23(8), e25002. <https://doi.org/10.2196/25002>
 58. Ng, Q. X., Liau, M. Y. Q., Tan, Y. Y., Tang, A. S. P., Ong, C., Thumboo, J., & Lee, C. E. (2024). A systematic review of the reliability and validity of the patient activation measure tool. *Healthcare (Basel Switzerland)*, 12(11), 1079. <https://doi.org/10.3390/h12111079>
 59. Eyles, J. P., Ferreira, M., Mills, K., Lucas, B. R., Robbins, S. R., Williams, M., Lee, H., Appleton, S., & Hunter, D. J. (2020). Is the patient activation measure a valid measure of osteoarthritis self-management attitudes and capabilities? Results of a Rasch analysis. *Health and Quality of Life Outcomes*, 18(1), 121. <https://doi.org/10.1186/s12955-020-01364-6>
 60. Hung, M., Carter, M., Hayden, C., Dzierzon, R., Morales, J., Snow, L., Butler, J., Bateman, K., & Samore, M. (2013). Psychometric assessment of the patient activation measure short form (PAM-13) in rural settings. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment Care and Rehabilitation*, 22(3), 521–529. <https://doi.org/10.1007/s11136-012-0168-9>
 61. Graffigna, G., Barello, S., Bonanomi, A., Lozza, E., & Hibbard, J. (2015). Measuring patient activation in Italy: Translation, adaptation and validation of the Italian version of the patient activation measure 13 (PAM13-I). *BMC Medical Informatics and Decision Making*, 15, 109. <https://doi.org/10.1186/s12911-015-0232-9>
 62. Maindal, H. T., Sokolowski, I., & Vedsted, P. (2009). Translation, adaptation and validation of the American short form patient activation measure (PAM13) in a Danish version. *BMC Public Health*, 9, 209. <https://doi.org/10.1186/1471-2458-9-209>
 63. Ngooi, B. X., Packer, T. L., Kephart, G., Warner, G., Koh, K. W., Wong, R. C., & Lim, S. P. (2017). Validation of the patient activation measure (PAM-13) among adults with cardiac conditions in Singapore. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment Care and Rehabilitation*, 26(4), 1071–1080. <https://doi.org/10.1007/s11136-016-1412-5>
 64. Lightfoot, C. J., Wilkinson, T. J., Memory, K. E., Palmer, J., & Smith, A. C. (2021). Reliability and validity of the patient activation measure in kidney disease: Results of Rasch analysis. *Clinical Journal of the American Society of Nephrology: CJASN*, 16(6), 880–888. <https://doi.org/10.2215/CJN.19611220>
 65. Packer, T. L., Kephart, G., Ghahari, S., Auduly, Å., Versnel, J., & Warner, G. (2015). The patient activation measure: A validation study in a neurological population. *Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment Care and Rehabilitation*, 24(7), 1587–1596. <https://doi.org/10.1007/s11136-014-0908-0>
 66. Schuler, M., Musekamp, G., Bengel, J., Nolte, S., Osborne, R. H., & Faller, H. (2014). Measurement invariance across chronic conditions: a systematic review and an empirical investigation of the Health Education Impact Questionnaire (heiQ™). *Health and quality of life outcomes*, 12, 56. <https://doi.org/10.1186/1477-7525-12-56>
 67. Taple, B. J., Chapman, R., Schalet, B. D., Brower, R., & Griffith, J. W. (2022). The impact of education on depression assessment:

- Differential item functioning analysis. *Assessment*, 29(2), 272–284. <https://doi.org/10.1177/1073191120971357>
68. Hagquist, C., & Andrich, D. (2017). Recent advances in analysis of differential item functioning in health research using the Rasch model. *Health and Quality of Life Outcomes*, 15(1), 181. <https://doi.org/10.1186/s12955-017-0755-0>
69. Hibbard, J. H., Stockard, J., Mahoney, E. R., & Tusler, M. (2004). Development of the patient activation measure (PAM): Conceptualizing and measuring activation in patients and consumers. *Health Services Research*, 39(4 Pt 1), 1005–1026. <https://doi.org/10.1111/j.1475-6773.2004.00269.x>
70. Hibbard, J. H., Mahoney, E. R., Stockard, J., & Tusler, M. (2005). Development and testing of a short form of the patient activation measure. *Health Services Research*, 40(6 Pt 1), 1918–1930. <https://doi.org/10.1111/j.1475-6773.2005.00438.x>
71. Jin, Y., Bratzke, L., & Baumann, L. C. (2021). Helping persons with multiple chronic conditions overcome barriers to self-management. *The Nurse Practitioner*, 46(3), 20–28. <https://doi.org/10.1097/01.NPR.0000733676.28520.db>
72. King, D. K., Glasgow, R. E., Toobert, D. J., Strycker, L. A., Estabrooks, P. A., Osuna, D., & Faber, A. J. (2010). Self-efficacy, problem solving, and social-environmental support are associated with diabetes self-management behaviors. *Diabetes Care*, 33(4), 751–753. <https://doi.org/10.2337/dc09-1746>
73. Qama, E., Rubinelli, S., & Diviani, N. (2022). Factors influencing the integration of self-management in daily life routines in chronic conditions: A scoping review of qualitative evidence. *BMJ Open*, 12(12), e066647. <https://doi.org/10.1136/bmjopen-2022-066647>
74. Ladner, J., Alshurafa, S., Madi, F., Nofal, A., Jayasundera, R., Saba, J., & Audureau, E. (2022). Factors impacting self-management ability in patients with chronic diseases in the united Arab Emirates, 2019. *Journal of Comparative Effectiveness Research*, 11(3), 179–192. <https://doi.org/10.2217/ceer-2021-0177>
75. Dineen-Griffin, S., Garcia-Cardenas, V., Williams, K., & Benrimoj, S. I. (2019). Helping patients help themselves: A systematic review of self-management support strategies in primary health care practice. *PloS One*, 14(8), e0220116. <https://doi.org/10.1371/journal.pone.0220116>
76. Kim, S., Choi, M., Lee, J., Kim, H., Song, K., & Park, H. J. (2022). Type D personality, cognitive illness perception, depression, approach coping, and self-management among older adults in long-term care hospitals: Structural equation modeling. *Geriatric Nursing (New York N Y)*, 48, 150–157. <https://doi.org/10.16/j.gerinurse.2022.09.011>
77. Alexandre, K., Campbell, J., Bugnon, M., Henry, C., Schaub, C., Serex, M., Elmers, J., Desrichard, O., & Peytremann-Bridevaux, I. (2021). Factors influencing diabetes self-management in adults: An umbrella review of systematic reviews. *JBIG Evidence Synthesis*, 19(5), 1003–1118. <https://doi.org/10.11124/JBIES-20-00020>
78. Schulman-Green, D., Jaser, S. S., Park, C., & Whittemore, R. (2016). A metasynthesis of factors affecting self-management of chronic illness. *Journal of Advanced Nursing*, 72(7), 1469–1489. <https://doi.org/10.1111/jan.12902>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Ingrid Olsson¹  · George Kephart^{1,2,3}  · Tanya Packer^{1,3}  · Sabine Björk^{1,4}  · Ulf Isaksson^{1,5}  · Anna Nordström^{6,7}  · Åsa Audulv¹ 

✉ Åsa Audulv
asa.audulv@umu.se

¹ Department of Nursing, Umeå University, Umeå, Sweden

² Department of Community Health and Epidemiology, Dalhousie University, Halifax, Canada

³ School of Health Administration, Dalhousie University, Halifax, Canada

⁴ Department of Epidemiology and Global Health, Umeå University, Umeå, Sweden

⁵ Artic Centre, Umeå University, Umeå, Sweden

⁶ Department of Medical Sciences, Rehabilitation Medicine, Uppsala University & University Hospital, Uppsala, Sweden

⁷ School of Sport Sciences, UiT The Arctic University of Norway, Tromsø, Norway