Check for updates

Received: 25 July 2021 | Revised: 5 January 2022 | Accepted: 10 January 2022

DOI: 10.1111/jep.13662

ORIGINAL PAPER

Journal of Evaluation in Clinical Practice
International Journal of Public Health Policy and Health Services Research
WILEY

Psychometric evaluation of the Patient Assessment of Chronic Illness Care instrument using Item analysis according to the Rasch model

Maren Boecker[1] | Katia Iglesias[2] | Isabelle Peytremann-Bridevaux[3]

[1]Institute of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen University, Aachen, Germany

[2]School of Health Sciences (HEdS-FR), HES-SO University of Applied Sciences and Arts of Western Switzerland, Fribourg, Switzerland

[3]Center for Primary Care and Public Health (Unisanté), University of Lausanne, Lausanne, Switzerland

Correspondence
Maren Boecker, Institute of Medical Psychology and Medical Sociology, University Hospital of RWTH Aachen University, Pauwelsstraße 19, Aachen 52074, Germany.
Email: mboecker@ukaachen.de

Funding information
Department of Public Health of the canton of Vaud

Abstract

Rationale, aims and objectives: The Patient Assessment of Chronic Illness Care (PACIC) instrument assesses patient care in alignment with the chronic care model. The aim of the study was to comprehensively validate the PACIC using the Rasch model. A special focus was placed on the investigation of local dependence (LD), differential item functioning (DIF) and targeting.

Method: This secondary analysis utilized data of 760 patients with a diagnosis of diabetes who had participated in the Swiss CoDiab-VD cohort study. The psychometric properties of the French PACIC-version were evaluated using the Rasch model. DIF was investigated in relation to age, gender, education, year of recruitment into the CoDiab-VD cohort study, type of diabetes and whether patients got an injectable antidiabetic drug or not.

Results: The initial analysis of the PACIC revealed poor fit to the Rasch model (χ^2-$p < 0.001$) with response dependency being the most prominent problem. After combining the items into two testlets (testlet 1: Items 1–11; testlet 2: Items 12–20), good overall model fit was found (χ^2-$p = 0.77$) as well as good reliability (Person Separation Index = 0.85) and targeting. DIF with regard to whether patients got an injectable antidiabetic drug or not was found for testlet 2. However, the size of this DIF was regarded as not being substantial.

Conclusion: The PACIC is a well-targeted, reliable unidimensional instrument to assess patient care in alignment with the chronic care model in patients with diabetes. It is free of substantial DIF. The PACIC-20 sum score can hence be used in clinical practice for individual diagnostic. For evaluation purposes like assessment of change or group evaluations, the usage of the interval-scale level person parameters is recommended as it permits using parametric statistical analyses and provides a more accurate picture about the actual amount of change.

KEYWORDS
differential item functioning, local dependence, patient care, Rasch model, targeting

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.
© 2022 The Authors. *Journal of Evaluation in Clinical Practice* published by John Wiley & Sons Ltd.

1 | INTRODUCTION

Worldwide, chronic diseases constitute a major burden for communities in terms of morbidity, disability, mortality and costs.[1–4] In response to this growing burden, healthcare systems need to change, train healthcare professionals and find innovative models of care and financing schemes, among others. Improving care for chronic patients also constitutes a challenge that healthcare systems must address. Often using the Chronic Care Model (CCM)[5,6] as a framework, integrated care has been implemented across European countries and North America since more than two decades. The Patient Assessment of Chronic Illness Care (PACIC) instrument was developed by Glasgow et al. in 2005.[7] It assesses, from the patient's perspective, in how far chronic care is accomplished in accordance with the CCM.[5,6] Considered in 2009 in a systematic review as one of the 'most appropriate instruments' to assess integrated care,[8] a more recent review showed that more than 200 instruments measuring dimensions of integrated care existed in 2016, one of which being the PACIC.[9]

Several studies performed validation analyses of the PACIC, showing wide variations of the psychometric properties across languages and populations.[10] Whereas different authors agreed on the one-dimensional structure of the instrument, they did not all agree on the number of items to include—all 20 items, such as shown to be the most often used version,[11] or less such as proposed in shorter versions of the PACIC (11[10,12,13] or 13 items[14]). The converging evidence suggesting the unidimensional structure of the PACIC instrument is coherent with the original validation of Glasgow, who despite describing a five-dimensional structure of the PACIC, recommended the use of a unique PACIC global score because 'the intercorrelations among the PACIC scales and the high internal consistency of the total score' may make it difficult for respondents to recognize 'differences among the subscale constructs'.[7]

Up to now, most PACIC validation studies were based on factor analyses and focused on the PACIC's dimensional structure. With the present study, we want to go a step further and investigate additional psychometric properties using Item Response Theory (IRT). In fact, the Rasch model[15] was used which is at the same time a simple, but elegant parametric IRT measurement model. Its application is well suited whenever new patient-reported outcome measures (PROMs) shall be developed, for example,[16,17] or whenever existing PROMs shall be reviewed and where necessary be revised.[18–22] It possesses desirable mathematical properties and allows for a unified approach to simultaneously investigate several measurement issues[23,24] as unidimensionality, local independence, differential item functioning (DIF) and targeting.

Unidimensionality is an important prerequisite whenever the responses to the items of a PROM shall be summed up to a valid (ordinal) sum score. Local independence means that items should only be correlated through the latent trait that the PROM is assessing and that therefore there should not be any substantial residual correlations between items.[25] Violations of local independence, also called local dependence (LD), have to be revealed and accounted for as they can lead to the estimation of biased parameters, problems with construct validity and inflate reliability.[25] LD may have very different reasons. It might either indicate multidimensionality or response dependency. The latter may occur when items are linked in some way, for instance when items share some common features like, for example, item content.[25] The investigation of DIF deals with the subject of test fairness and addresses the question of whether external variables influence the way people respond to the items of the PROM. A PROM is deemed to be unfair, to show DIF, if one item or some of the items are not equally difficult to endorse across people of different subgroups despite equal levels of the underlying characteristic. In clinical settings, DIF might be investigated for external variables like, for example, sex, age or diagnostic group. Finally, targeting is related to the question of whether the items of a PROM provide enough measurement information to ensure sufficient precision of the patients' assessment. The investigation of targeting takes advantage of the fact that IRT models map item and person parameters on a common metric. Therefore, it can be examined whether the PROM assesses the part of the dimension with enough measurement precision where the study population is actually located. A special feature of IRT models directly related to the targeting is that for each person, an individual standard error of measurement is estimated which might vary substantially. It depends on the person's location on the assessed continuum of the underlying concept measured by the PROM (i.e., the latent trait) and on where the items are located (i.e., targeting). This is an important difference as compared to classical test theory (CCT) assuming that measurement precision is constant across the scale. Another advantage of the Rasch model is that in case the PROM data comply with the Rasch model's requirements a transformation of the ordinal sum scores into linear, interval-level person parameters can be provided permitting the use of parametric statistical analyses, for example, in the context of measurement of change.

These measurement issues have to be taken into consideration to assess the validity of PROMs and therefore besides the investigation of unidimensionality and reliability, the present study will extend previous findings related to the PACIC by placing a special focus upon the investigation of LD, DIF and targeting. To our knowledge, only one study has so far investigated the PACIC instrument using IRT analysis[14]; while the latter used nonparametric IRT focusing on the number of items and unidimensionality, authors did not investigate targeting, LD or DIF, key elements to consider when validating instruments using IRT. In that context, the aim of the present study is to comprehensively validate the PACIC instrument using the Rasch model.[15]

2 | METHODS

2.1 | Study design

We performed secondary analysis of data from a cross-sectional survey conducted in fall 2017 in the frame of the CoDiab-VD cohort investigating the quality of care of patients with diabetes.[26,27]

2.2 | Sample

Individuals with diabetes were recruited into the CoDiab-VD cohort in 2011–2012 and in 2017 by the means of a paper survey proposed through community-based pharmacies of the canton of Vaud, a large French-speaking canton of Switzerland. Noninstitutionalized adults (≥18 years) with a diagnosis of diabetes for at least 1 year, visiting a participating pharmacy with a diabetes-related prescription and residing in the canton of Vaud were eligible. People with cognitive disorders or with insufficient French skills to understand and fill in a questionnaire, as well as women with gestational diabetes, were excluded. Included individuals from 2011 to 2012 cohort had been followed-up yearly since 2013 by filling in a paper questionnaire sent to their home. In 2017, 790 individuals completed the survey questionnaire; among them were 276 individuals who had been first recruited in 2011–2012 and who were now participating in their follow-up as well as 514 new recruits. Of these 790 individuals, 760 responded to at least five of the 20 PACIC items and were included in the analyses. The protocol of the CoDiab-VD cohort study was approved by the Cantonal Ethics Committee of Research on Human Beings of the Canton of Vaud (CER-VD, protocol numbers 151/11 and PB_2017_00232). This study is registered with ClinicalTrials. gov, identifier NCT01902043. Written informed consent was obtained from participants, and data were kept confidential.

2.3 | Participants' characteristics

The sociodemographic characteristics of the participants considered in this study were the following: year of recruitment into the CoDiab-VD cohort (2011–2012, 2017), age, sex (female, male), civil status (single, married/partnership, divorced/separated/widowed), education (primary, secondary, tertiary), employment status (full-time, part-time, retired, unemployed/benefiting from invalidity insurance/student, stay-at-home), membership in the local diabetes association. Health and health behaviors variables were self-rated health measured with the first question of the SF-12 questionnaire,[28] smoking status (current smoking: yes, no), alcohol consumption measured with the AUDIT-C questionnaire,[29] physical activity measured with the question of the Swiss Health Survey,[30] body mass index (BMI = weight/height^2 = kg/m^2), screening for depression measured with the PRIME-MD.[31,32] Diabetes-related measures considered were type of diabetes (Type 2, other) and antidiabetic treatment (inclusion of injectable drug yes, no).

2.4 | Outcome variable

The main outcome considered was the 20-item French version of the PACIC questionnaire which was translated from the English version[7] and culturally adapted to the local context (see Table S2).[10,27] Designed to assess patient care in alignment with the CCM, the PACIC is composed of the five subdimensions 'Patient Activation' (Items 1–3), 'Delivery System Design/Decision Support' (Items 4–6), 'Goal Setting/Tailoring' (Items 7–11), 'Problem-solving/Contextual' (Items 12–15) and 'Follow-up/Coordination' (Items 16–20).[7] However, as described in the introduction, the use of a unique PACIC global score is recommended.[7] The five response options of the PACIC range from 1 (*never*) to 5 (*always*) with higher scores reflecting higher care congruence with the CCM.

2.4.1 | Descriptive statistical analyses

First, we conducted descriptive analyses to characterize the participants (percentage for ordinal or categorical data; means and standard deviation [SD] or median and interquartile interval for continuous data depending on the distribution of data) and check of data quality of each of the 20 PACIC items (distribution by category (floor and ceiling effects), percentage of missing values). The latter were done using Stata 16.1.[33]

2.4.2 | Item analyses according to the Rasch model

The rationale behind the Rasch model is that the items of a PROM vary in terms of item difficulties and the assessed persons vary in terms of how much of the assessed latent trait they 'possess'. In case of the PACIC, an 'easy' item would be an item where many patients fully agree that the care is being carried out in accordance with the CCM, and a difficult item would be an item where only a few patients agree to this. Based on the responses to the PACIC items a person parameter is estimated for each patient. During the process of item analysis according to the Rasch model, it is formally tested whether the persons responding to a set of items, in this case to the PACIC items, respond expectedly to these items. A patient with a high person parameter who reports a high quality of care should, for example, have a high probability to score high on an 'easy' item, and a patient with a low person parameter, should have a low probability to score high on a 'difficult' item.

Generally speaking, the Rasch model is an IRT model with especially desirable measurement properties.[34,35] If a set of item responses fits the Rasch model, the resulting scale has the following properties: (a) unidimensionality; (b) local independence of items (no LD); (c) measurement invariance across groups (no DIF); (d) monotonicity in the relationship between expected item responses and the latent score and (e) homogeneity (the rank order of the probabilities to endorse items is the same across all persons regardless of their level on the latent trait).

For mental health-related scales, it is common to find that some items function differently across subgroups of persons or that not all item responses are locally independent. While the former results in test scores not being fair across the subgroups for which DIF was found (e.g., sex or age group), the latter can lead to inflated reliability values and compromise person parameter estimation.[25] Therefore, if DIF and/or LD is found, it has to be accounted for.

The initial analysis was run including all 20 PACIC items as previous studies had suggested the appropriateness of a unique summated PACIC-score instead of five separate subscores for the subdimensions. Due to the five response options of the PACIC items the polytomous Partial Credit model was used.[36] Each analysis cycle included the investigation of unidimensionality, LD, DIF, fit (item fit and overall fit), ordering of thresholds, reliability and targeting. Unidimensionality of the PACIC items was tested using Smith's test of unidimensionality.[37] Following the recommendations of Marais[38] and Christensen et al.[25] a residual correlation of >0.2 above the average residual correlation was used as critical level for the detection of LD between pairs of items. If too high inter-item correlations beyond what can be attributed to the latent trait (the so-called residual correlations) were found and rather indicated response dependency than multidimensionality, items were merged into testlets by adding them together to absorb the dependency. Using the testlet-strategy results in a bi-factor equivalent solution with a first common (Rasch) factor upon which all items load and secondary factors whose items each share unique variance.[20,39] The explained common variance (EVC) of the first common factor should be >0.9 to consider the scale as unidimensional. Likewise, the latent correlation of the testlets should be high.

DIF was evaluated using analyses of variance (ANOVA), with the significance level set at 5% and applying a Bonferroni correction. Remember that the presence of DIF for a given PROM implies that people respond differently to the item due to their membership to a group (e.g., males vs. females) despite equal levels of the underlying latent trait. In doing so, two types of DIF can be differentiated[23]: (i) uniform DIF, that is, a consistent systematic difference between the investigated DIF groups regarding their responses to a specific item across the whole range of the measured construct; (ii) nonuniform DIF, that is, the response difference between groups varies across the levels of the measured construct with a specific item, for example, being easier for one group at some levels of the construct, but more difficult at other levels. If DIF was found, the impact of DIF was evaluated by computing equated scores.[22] In case the DIF was relevant on sum score level, it was accounted for by splitting the respective item.[40] DIF was evaluated in relation to gender, age group (≤59, 59.1–67.9, 68–74, ≥74.1), year of first recruitment (2011–2012 vs. 2017), education, type of diabetes (Type 2 vs. other) and whether patients received an injectable antidiabetic drug or not. The year of first recruitment was included as a DIF-variable as this data collection was a follow-up survey for those participants who had been recruited in 2011–2012 and the first assessment for those recruited in 2017.

Individual item misfit was determined by item-fit residual values, which were expected to be within the range of −2.5 to +2.5. An overall fit statistic, the item-trait interaction, was investigated using a χ^2 test. Overall model fit is indicated by a nonsignificant χ^2 value ($p > 0.05$; Bonferroni adjusted). The functioning of the response categories of the items was assessed by examining the threshold order. Disordered thresholds might indicate that the patients cannot consistently discriminate between the available response options or they might be caused by LD in the data set. The person separation index (PSI) was calculated, which is a reliability index reflecting the ability of the scale to discriminate reliably between respondents. A PSI ≥ 0.85 indicates a good person separation for the evaluation of individuals and a PSI ≥ 0.70 is regarded as sufficient for group evaluation.[23] Finally, the targeting of the PACIC was investigated. The targeting was assessed graphically based on the person-item threshold distribution map. The targeting is good if the item thresholds cover a broad range of difficulties across the assessed dimension and if the assessed patients are located just within the same range. Additionally, to achieve a well-targeted scale the average of the persons' ability should be close to zero as the average item difficulty is calibrated at zero. Item analysis according to the Rasch model was performed using the software RUMM2030.[41]

3 | RESULTS

3.1 | Sample

Whereas two-thirds of the participants had been recruited in 2017, the others came from the initial 2011–2012 recruitment period. Mean age of participants was 65.8 years, 59.7% were men, 75% reported Type 2 diabetes and 57.7% used antidiabetic treatment that included an injectable drug. Details of the characteristics of the participants are summarized in Table 1.

3.1.1 | Descriptive statistical analyses of items

Descriptive results of all 20 PACIC items are provided in Table 2. The latter shows that 86.2% of the participants responded to all 20 items and that the number of missing values per item varied between 0.9% and 3.2%. Across all 20 PACIC items, 3.68% of the participants had extreme values with 0.66% of them reporting the highest possible level of CCM care congruency (overall ceiling effect).

3.1.2 | Item analysis according to the Rasch model

The results of the item analyses according to the Rasch model are presented in Table 3. The initial analysis of the PACIC-20 revealed poor fit to the Rasch model (χ^2 -$p < 0.001$). Apart from Item 15 ('Asked how my chronic condition affects my life') all items showed reversed thresholds and seven items displayed misfit (Items 6, 7, 8, 10, 16, 18, 20). DIF was found for four items: Item 6 with regard to whether patients received an injectable antidiabetic drug or not, Item 14 with regard to the type of diabetes, Item 16 with regard to age group and Item 20 with regard to year of recruitment. The test of unidimensionality revealed that the percentage of significant *t* tests was slightly above 5%, in fact, 6.8% with the lower bound of the confidence interval being 5.2%. Additionally, there was strong evidence of LD (see Table S2). The highest residual correlation (0.61) was found for the pair of Items 10 ('Encouraged to go to a specific group or class to help me cope with my diabetes') and Item 17

TABLE 1 Characteristics of participants (n = 760 individuals with diabetes)

Year of recruitment into the CoDiab-VD cohort	
2011–2012	35.3%
2017	64.7%
Age (n = 760)	Mean (SD): 65.8 (12.5)
Age Group 1: ≤59	26.1%
Age Group 2: 59.1–67.9	23.4%
Age Group 3: 68–74	25.8%
Age Group 4: ≥74.1	24.7%
Sex (n = 760)	
Women	40.3%
Men	59.7%
Civil status (n = 750)	
Single	10.5%
Married/partnership	59.5%
Divorced/separated/widowed	30.0%
Education (n = 720)	
Primary	15.6%
Secondary	53.1%
Tertiary	31.4%
Employment status (n = 744)	
Full-time	18.4%
Part-time	7.3%
Retired	62.5%
Unemployment/benefiting from invalidity insurance/student	9.0%
Stay-at-home	2.8%
Membership in the local diabetes association (n = 750)	14.3%
Self-reported health[a] (n = 754)	
Excellent/very good	14.5%
Good	62.9%
Medium/poor	22.7%
Current smoking (n = 740)	18.8%
Risky or excessive alcohol consumption[b] (n = 731)	41.7%
Physically inactive[c] (n = 743)	29.3%
BMI (n = 731)	
Overweight or obese	80.0%
Screened positive for depression[d] (n = 751)	32.8%
Type of diabetes (n = 760)	
Type 2	72.8%
Other	27.2%
Antidiabetic treatment including an injectable drug (n = 756)	57.7%

[a]Measured with the first question of the Short Form Health Survey-12 (SF-12).

[b]Measured with the Alcohol Use Disorders Identification Test-Consumption (AUDIT-C). A score ≥4 for men and ≥3 for women is considered as a risky or excessive alcohol consumption.

[c]Measured with questions from the Swiss Health Survey: active: ≥150 min of moderate physical activity or ≥ two intense activities per week; partly active: 30 to 149 min of moderate physical activity or one intense activity per week; inactive: <30 min of moderate physical activity and <one intense activity per week

[d]Measured with the PRIME-MD.

('Encouraged to attend programs in the community that could help me'), which are indeed very similar regarding item content. However, most of the found LD was across items within the respective five subdomains, for example, Item pair 7 and 8 (residual correlation = 0.32; 'Asked to talk about my goals in caring for my condition' and 'Helped to set specific goals to improve my eating or exercise'). The overall pattern of results rather suggested response dependency being the cause of LD instead of multidimensionality.

Different strategies were used to account for the found LD, like combining successively items with the highest residual correlation. However, the most promising approach was the creation of five testlets, which corresponded to the five subdomains of the PACIC as theoretically defined by Glasgow. This procedure improved significantly the overall fit to the Rasch model (χ^2-p = 0.11). Besides the good overall fit, unidimensionality, as well as the absence of further LD, could be demonstrated. Unidimensionality was also supported by the found ECV (0.96) indicating a strong first common factor upon which all items load. However, some deviations from the Rasch model could still be observed. The testlets 'Goal Setting' (fit residual = −3.01) and 'Follow-up/Coordination' (fit residual = 2.96) displayed item misfit. The latter testlet additionally showed DIF with regard to whether patients got injections or not—like the testlet 'Delivery System Design/Decision Support', which also displayed DIF with regard to year of recruitment. Whereas three of the testlets had ordered thresholds, the thresholds of the two testlets 'Patient Activation' and 'Delivery System Design/Decision Support' were still disordered.

Based on the LD pattern of the initial analysis and on the results of the principal component analysis of the residuals of the five-testlet analysis, we combined the first three testlets (Items 1–11) into one testlet and the last two testlets (Items 12–20) into a second testlet. This combination resulted in good overall model fit (χ^2-p = 0.77), unidimensionality, item fit and no more evidence of LD. The latent correlation between the two testlets was 0.89 and 94% of the variance was common, also strongly supporting unidimensionality of the PACIC. As the items were combined into two testlets in this analysis, it was possible to additionally calculate a conditional χ^2 test of fit to

TABLE 2 Distribution of the 20 PACIC items

	Percentages					
	Response categories					
Item	Never[a] 1	Generally not 2	Sometimes 3	Most of the time 4	Always[b] 5	Missing values
1	23.6%	10.5%	16.4%	20.4%	29.1%	1.3%
2	34.5%	14.1%	16.9%	16.4%	18.2%	2.1%
3	20.1%	11.0%	16.1%	21.9%	30.8%	2.1%
4	47.6%	14.9%	17.1%	10.7%	9.7%	1.3%
5	7.5%	3.9%	11.1%	37.2%	40.2%	3.2%
6	13.0%	9.8%	21.7%	24.9%	30.8%	1.6%
7	26.4%	15.6%	22.6%	20.5%	14.9%	2.0%
8	26.1%	16.6%	25.7%	19.5%	12.1%	1.7%
9	48.2%	13.7%	14.1%	12.4%	11.7%	2.0%
10	60.8%	12.6%	13.6%	7.0%	6.1%	2.1%
11	26.8%	16.6%	23.6%	18.9%	14.1%	1.2%
12	12.3%	6.8%	13.5%	30.7%	36.8%	2.6%
13	36.7%	10.8%	15.6%	17.9%	19.0%	1.5%
14	36.3%	12.1%	17.8%	19.3%	14.6%	1.7%
15	27.4%	18.0%	22.6%	16.5%	15.6%	2.0%
16	49.7%	16.4%	17.1%	9.7%	7.0%	2.4%
17	64.7%	12.5%	13.9%	5.4%	3.5%	2.8%
18	49.6%	13.8%	18.5%	7.9%	10.2%	1.8%
19	23.1%	9.3%	18.2%	22.2%	27.2%	0.9%
20	36.0%	11.6%	17.1%	16.4%	18.8%	1.5%

Abbreviation: PACIC, Patient Assessment of Chronic Illness Care.

[a]Floor effect.

[b]Ceiling effect.

investigate whether the two testlets assess the same underlying construct.[42] This test was not significant ($p = 0.63$) which is another indicator for the unidimensionality. Moreover, the thresholds of the two new testlets were now ordered. The PSI, a reliability index, was 0.85 and consequently high enough for individual use. Figure 1 shows that the PACIC-20 was well-targeted for our sample with a mean patients' location close to zero and most patients located in the same range as the item threshold locations. Only at the higher and lower end of the dimension patients cannot be assessed with enough measurement precision. However, DIF with regard to whether patients received an injectable antidiabetic drug or not was found for testlet2 (Items 12–20). Patients receiving an injectable antidiabetic drug because of their diabetes tended to report higher levels of CCM care congruency for this testlet than patients not receiving an injectable antidiabetic drug given the same person parameter. To investigate the impact of this found DIF, equated PACIC-scores were compared for patients receiving an injectable antidiabetic drug versus not receiving one. A slight difference between the equated PACIC-scores was found across the entire dimension with the biggest difference being 2.5 score points (for further details see Figure S1). We decided not to adjust for the DIF (see discussion).

A score-to-measure transformation table is provided transforming the nonlinear PACIC-20 raw score into interval-level person estimates (see Table S3).

4 | DISCUSSION

The present study is the first study validating the PACIC-20 instrument using the Rasch model with a special focus on the examination of unidimensionality, possible response dependencies between items (LD), the impact of external variables influencing the way people respond to items (DIF) and whether the items provide enough measurement information to ensure sufficient precision of the patients' assessment (targeting). Based on a sample of 760 patients with diabetes, the PACIC-20 revealed poor fit to the Rasch model with LD

TABLE 3 Overall model fit of the PACIC-20

	Overall item-trait interaction			Unidimensionality *t* test		Testlet analysis			Relia-bility	Targeting				Item misfit	Differential item functioning (DIF)
										Item location		Person location			
Analysis	χ^2	df[a]	*p* value	test (%)	LB95CI[b] (%)	Latent r[c]	ECV[d]	χ^2-*p* value[e]	PSI[f]	Mean	SD[g]	Mean	SD[g]	Item number	Item number (source & kind of DIF)
Original PACIC-20	481.5	180	0.000**	6.81%	5.2%				0.92	0.00	0.55	−0.30	1.08	6, 7, 8, 10, 16, 18, 20	6 (receiving injectable antidiabetic drug; nu), 14 (type of diabetes; u), 16 (age; u), 20 (year of recruitment; u)
PACIC-20 (5 testlets)	57.06	45	0.11	2.95%	1.3%	0.84	0.96		0.87	0.00	0.29	−0.15	0.80	Testlet 'Goal setting', testlet 'Follow-up/ Coordination'	Testlet 'Follow-up/ Coordination' (receiving injectable antidiabetic drug; u), testlet 'Delivery Sytem Design/Decision Support' (receiving injectable antidiabetic drug; u & nu & year of recruitment; nu)
PACIC-20 (2 testlets)	13.41	18	0.77	2.6%	0.9%	0.89	0.94	0.63	0.85	0.00	0.12	−0.19	0.65		Testlet 2 (receiving injectable antidiabetic drug; u)[h]

Abbreviations: DIF, differential item functioning; nu, noununiform DIF; PACIC, Patient Assessment of Chronic Illness Care; u, uniform DIF

**Bonferroni-adjusted $p = 0.0025$.

[a]Degrees of freedom.

[b]Lower bound of 95% confidence interval.

[c]Correlation.

[d]Explained common variance.

[e]Conditional test of fit χ2-p value.

[f]Person separation index.

[g]Standard deviation.

[h]It was not adjusted for the found DIF (see text for more details).

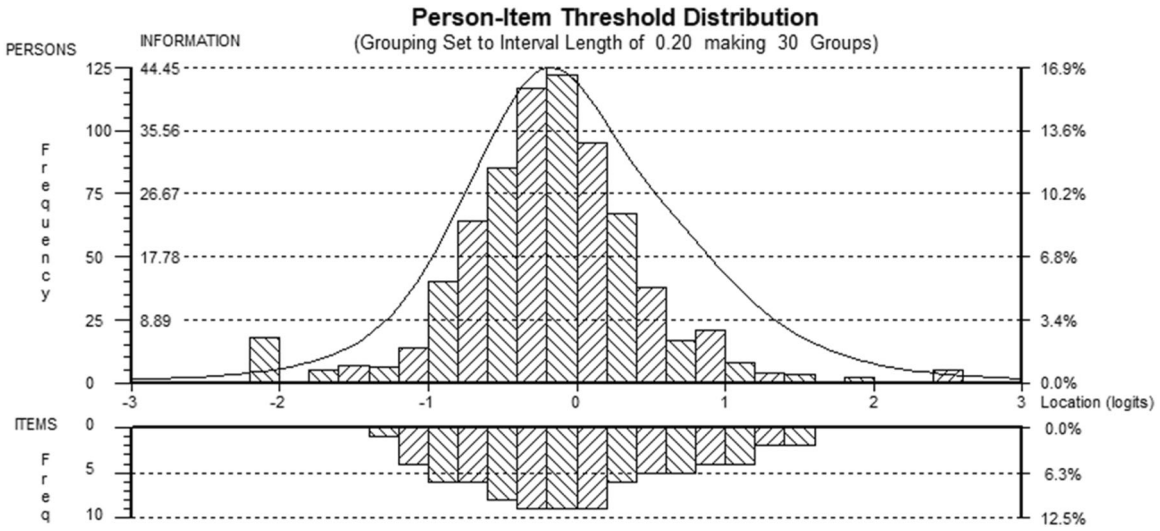

FIGURE 1 Person-item threshold map. The person-item threshold map pictures the distributions of person parameter locations (upper part of the figure) and item thresholds (lower part of the figure). It illustrates how person parameters and item threshold locations are distributed along the patient-reported care congruency with the Chronic Care Model dimension, with higher values indicating higher care congruency (persons) and higher difficulty to score in the direction of higher care congruency (items) respectively. The black line above the person parameters represents the test information curve. Where test information is high the patients can be assessed with high measurement precision. At both ends of the dimension, items are actually missing. Therefore, the measurement precision here is very low

being the most prominent problem. Forming two testlets (testlet1: three first subdimensions [Items 1–11]; and testlet 2: the last two subdimensions [Items 12–20]) resulted in good overall model fit, unidimensionality, item fit and no more evidence of LD. The initial problem of disordered thresholds within the PACIC-20 was solved with the two testlets indicating that the disordering was due to the massive dependence between item responses (LD) and not to difficulties of the patients to consistently discriminate between the available response options. The PSI of 0.85 was high enough for individual use. Overall, the targeting of the PACIC-20 was very good for this sample of patients with diabetes. Only at the higher and lower end of the dimension patients cannot be assessed with sufficient measurement precision as item thresholds are missing in these areas.

In the initial analysis, DIF had been found for four single items. However, this DIF was not pronounced enough to persist once testlets were formed. Consequently, the PACIC-20 has no DIF related to gender, age group, year of recruitment, education and type of diabetes (Type 2 vs. other). Particularly having no DIF related to the year of first recruitment might indicate good measurement stability of the construct. However, for testlet 2 DIF with regard to whether patients received an injectable antidiabetic drug or not was found with patients receiving an injectable antidiabetic drug reporting higher level of care congruency than patients not receiving one given the same person level of perceived congruency with the CCM. The biggest difference between persons with or without injectable antidiabetic drug treatment being 2.5 score points, the splitting for DIF is not adviced for the following reasons. First, the difference in equated scores between persons with or without injectable antidiabetic drug treatment cannot be considered as clinically important considering the possible scale range of the PACIC-score. Second, the differences found might reflect real differences in their experience of care for testlet2 in the sense that patients receiving an injectable antidiabetic drug might get a more profound medical supervision because of the greater severity of the disease. This is supported when looking more closely to the items showing the DIF with regard to receiving an injectable antidiabetic drug. In fact, the DIF is related to the items from the "Follow-up/Coordination"-subdimension (Items 16–20) which reflect how far the attending doctor coordinates and supervises the treatments of other health services. This could indicate a slight difference in treatment of patients receiving an injectable antidiabetic drug—with the difference being especially pronounced with regard to the 'Follow-up/Coordination'-aspect.

Regarding the dimensionality of the PACIC-20, we found that the PACIC-20 can indeed be regarded as a unidimensional scale including all 20 items if the huge amount of response dependency across items is accounted for. The found high ECV of the common factor and the high latent correlation between the testlets also support the assumption of unidimensionalty. The found unidimensionality is in line with previous studies, which however disagreed regarding the number of items to include. The only other study using an IRT model, in this case Mokken analysis—a nonparametric IRT model, also found that the PACIC-20 was unidimensional.[13] However, Gibbons et al. had to exclude seven items as these items violated the invariant item ordering. Because of software restrictions, they could not investigate LD and did not investigate DIF. Future studies using IRT should further investigate unidimensionality and LD in diverse patient populations to reappraise whether unidimensionality and the same testlet structure as found in the present study can be replicated across different patient populations. Methodological studies

comparing different Rasch approaches like the separate evaluation of the five subdimensions versus the testlet approach as in the present study versus using a multidimensional Rasch model approach might also help to better understand the structure of the PACIC. Additionally, future studies should clarify whether the item difficulties of the PACIC-20 are invariant across different diseases (no DIF).

5 | CONCLUSION

Once accounting for massive response dependency, the PACIC-20 appears to be a valid unidimensional instrument to assess patient care in alignment with the chronic care model in patients with diabetes. Clinicians can hence continue to use the ordinal PACIC-20 sum score in clinical practice for individual diagnostic. However, whenever patient care in alignment with the chronic care model should be followed up in the course of treatment (assessment of change) or should be investigated in quality assurance or other studies the usage of the interval-scale level person parameters is strongly recommended. This permits using parametric statistics and provides a more accurate picture about the actual amount of change.

ACKNOWLEDGMENTS

We would like to thank all patients with diabetes who participated in the study. We would also like to thank Emilie Antille Zuercher for data acquisition and help with the database. The cohort CoDiab-VD is supported by the Department of Public Health of the canton of Vaud ('Programme cantonal Diabète'). Open access funding enabled and organized by Projekt DEAL.

CONFLICT OF INTERESTS

The authors declare that there are no conflict of interests.

DATA AVAILABILITY STATEMENT

The metadata from the CoDiab-VD datasets are available in a public repository (CoDiab-VD: Cohort of Patients with Diabetes in the Canton of Vaud (Switzerland)), doi:10.16909/dataset/18. Data request can be made via the data repository.

ORCID

Maren Boecker http://orcid.org/0000-0003-2356-7803
Katia Iglesias http://orcid.org/0000-0003-1308-1631
Isabelle Peytremann-Bridevaux http://orcid.org/0000-0002-6514-8781

REFERENCES

1. World Health Organization. Global Health Estimates 2015: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2015; 2016. http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html
2. GBD. Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1545-1602.
3. GBD. Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388(10053):1459-1544.
4. Muka T, Imo D, Jaspers L, et al. The global impact of non-communicable diseases on healthcare spending and national income: a systematic review. *Eur J Epidemiol*. 2015;30(4):251-277.
5. Wagner E, Austin B, Davis C, Hindmarsh M, Schaefer J, Bonomi A. Improving chronic illness care: translating evidence into action. *Health Aff (Millwood)*. 2001;20:64-78.
6. Epping-Jordan J, Pruitt S, Bengoa R, Wagner E. Improving the quality of health care for chronic conditions. *Quality and Safety in Health Care*. 2004;13:299-305.
7. Glasgow R, Wagner E, Schaefer J, Mahoney L, Reid R, Green S. Development and validation of the Patient Assessment of Chronic Illness Care. *Med Care*. 2005;43:436-444.
8. Vrijhoef HJM, Berbee R, Wagner EH, Steuten LMG. Quality of integrated chronic care measured by patient survey: identification, selection and application of most appropriate instruments. *Health Expect*. 2009;12(4):417-429. doi:10.1111/j.1369-7625.2009.00557.x
9. Bautista MAC, Nurjono M, Lim YW, Dessers E, Vrijhoef HJ. Instruments measuring integrated care: a systematic review of measurement properties: instruments measuring integrated care. *Milbank Q*. 2016;94(4):862-917. doi:10.1111/1468-0009.12233
10. Iglesias K, Burnand B, Peytremann-Bridevaux I. PACIC Instrument: disentangling dimensions using published validation models. *Int J Qual Health Care*. 2014;26(3):250-260.
11. Arditi C, Iglesias K, Peytremann-Bridevaux I. The use of the Patient Assessment of Chronic Illness Care (PACIC) instrument in diabetes care: a systematic review and meta-analysis. *Int J Qual Health Care*. 2018;30(10):743-750. doi:10.1093/intqhc/mzy091
12. Cramm JM, Strating MM, Tsiachristas A, Nieboer AP. Development and validation of a short version of the Assessment of Chronic Illness Care (ACIC) in Dutch Disease Management Programs. *Health Qual Life Outcomes*. 2011;9(1):49. doi:10.1186/1477-7525-9-49
13. Gugiu PC, Coryn C, Clark R, Kuehn A. Development and evaluation of the short version of the Patient Assessment of Chronic Illness Care instrument. *Chronic Illn*. 2009;5(4):268-276. doi:10.1177/1742395309348072
14. Gibbons CJ, Small N, Rick J, Burt J, Hann M, Bower P. The Patient Assessment of Chronic Illness Care produces measurements along a single dimension: results from a Mokken analysis. *Health Qual Life Outcomes*. 2017;15(1):61. doi:10.1186/s12955-017-0638-4
15. Rasch G. *Probabilistic Models for Some Intelligence and Attainment Tests*. Danmarks Paedagogiske Institut; 1960:1.
16. Kallinger S, Scharm H, Boecker M, Forkmann T, Baumeister H. Calibration of an item bank in 474 orthopedic patients using Rasch analysis for computer-adaptive assessment of anxiety. *Clin Rehabil*. 2019;33(9):1468-1478. doi:10.1177/0269215519846225
17. Gecht J, Mainz V, Boecker M, et al. Development of a short scale for assessing economic environmental aspects in patients with spinal diseases using Rasch analysis. *Health Qual Life Outcomes*. 2017;15(1):196. doi:10.1186/s12955-017-0767-9
18. Upegui-Arango LD, Forkmann T, Nielsen T, et al. Psychometric evaluation of the Interpersonal Needs Questionnaire (INQ) using item analysis according to the Rasch model. *PLOS ONE*. 2020;15(8):e0232030. doi:10.1371/journal.pone.0232030
19. Vaganian L, Bussmann S, Boecker M, et al. An item analysis according to the Rasch model of the German 12-item WHO Disability Assessment Schedule (WHODAS 2.0). *Qual Life Res*. 2021;30(10):2929-2938. doi:10.1007/s11136-021-02872-8

20. Pomeroy IM, Tennant A, Mills RJ, TONiC Study Group, Young CA. The WHOQOL-BREF: a modern psychometric evaluation of its internal construct validity in people with multiple sclerosis. *Qual Life Res*. 2020;29(7):1961-1972. doi:10.1007/s11136-020-02463-z
21. Prodinger B, Hammond A, Tennant A, Prior Y, Tyson S. Revisiting the disabilities of the arm, shoulder and hand (DASH) and QuickDASH in rheumatoid arthritis. *BMC Musculoskelet Disord*. 2019;20(1):41.
22. Christensen KB, Thorborg K, Hölmich P, Clausen MB. Rasch validation of the Danish version of the shoulder pain and disability index (SPADI) in patients with rotator cuff-related disorders. *Qual Life Res*. 2019;28(3):795-800.
23. Tennant A, Conaghan PG. The Rasch measurement model in rheumatology: what is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism*. 2007;57(8):1358-1362.
24. Christensen KB, Kreiner S, Mesbah M, eds. *Rasch Models in Health*. John Wiley & Sons, Inc; 2013.
25. Christensen KB, Makransky G, Horton M. Critical Values for Yen's Q_3: identification of local dependence in the Rasch model using residual correlations. *Applied Psychological Measurement*. 2017;41(3):178-194.
26. Antille-Zuercher E, Peytremann-Bridevaux I, Carron TCohorte CoDiab-VD: Caractéristiques des patient·e·s diabétiques vaudois·e·s et évaluation de leur prise en charge: Etat des lieux en 2017 et comparaison avec 2011-2012: Institut universitaire de médecine sociale et préventive. 2019;164. doi:10.16908/ISSN.1660-7104/293
27. Zuercher E, Bordet J, Burnand B, Peytremann-Bridevaux I. CoDiab-VD: protocol of a prospective population-based cohort study on diabetes care in Switzerland. *BMC Health Serv Res*. 2015;15(1):329. doi:10.1186/s12913-015-0991-0
28. Ware J, Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-233.
29. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. *Arch Intern Med*. 1998;158(16):1789-1795. doi:10.1001/archinte.158.16.1789
30. Swiss Health Survey. https://www.bfs.admin.ch/bfs/fr/home/statistiques/sante/enquetes/sgb.html#1002476496
31. Spitzer RL, Williams JB, Kroenke K, et al. Utility of a new procedure for diagnosing mental disorders in primary care: the PRIME-MD 1000 study. *JAMA*. 1994;272(22):1749-1756.
32. Whooley MA, Avins AL, Miranda J, Browner WS. Case-finding instruments for depression: two questions are as good as many. *J Gen Intern Med*. 1997;12(7):439-445.
33. StataCorp. Stata Statistical Software: Release 16. StataCorp LLC; 2019.
34. Mesbah M, Kreiner S. Rasch Models for Ordered Polytomous Items. In: Christensen KB, Kreiner S, Mesbah M, eds. *Rasch Models in Health*. John Wiley & Sons, Inc; 2013:27-42.
35. Kreiner S. The Rasch Model for Dichotomous Items. In: Christensen KB, Kreiner S, Mesbah M, eds. *Rasch Models in Health*. John Wiley & Sons, Inc; 2013:5-26.
36. Masters GN. A Rasch model for partial credit scoring. *Psychometrika*. 1982;47(2):149-174.
37. Smith EV. Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *J Appl Meas*. 2002;3(2):205-231.
38. Marais I. Local dependence. In: Christensen KB, Kreiner S, Mesbah M, eds. *Rasch Models in Health*. John Wiley & Sons, Inc; 2013:111-130.
39. Andrich D. Components of variance of scales with a bifactor subscale structure from two calculations of α. *Educ Meas Issue Prac*. 2016;35(4):25-30. doi:10.1111/emip.12132
40. Hagquist C, Bruce M, Gustavsson JP. Using the Rasch model in nursing research: an introduction and illustrative example. *Int J Nurs Stud*. 2009;46(3):380-393.
41. Andrich D, Sheridan B, Luo G. Rumm 2030: Rasch Unidimensional Measurement Models (software). Perth, Western Australia: RUMM Laboratory; 2015.
42. Andrich D. The polytomous Rasch model and the equating of two instruments. In: Christensen KB, Kreiner S, Mesbah M, eds. *Rasch Models in Health*. John Wiley & Sons, Inc.; 2013:163-196. doi:10.1002/9781118574454.ch10

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Boecker M, Iglesias K, Peytremann-Bridevaux I. Psychometric evaluation of the Patient Assessment of Chronic Illness Care instrument using item analysis according to the Rasch model. *J Eval Clin Pract*. 2022;28:670-679. doi:10.1111/jep.13662