European Journal of General Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/igen20

Evaluation of the psychometric performance of the Spanish and Catalan versions of the patient reported experiences and Outcomes of Safety in Primary Care (PREOS-PC)-Compact questionnaire

Maria A. Fiol-deRoque, José M. Valderas, Jorge Arias de la Torre, Maria J. Serrano-Ripoll, Montserrat Gens-Barberà, Encarna Sánchez-Freire, Francisco M. Martín-Luján, Antonio Olry de Labry & Ignacio Ricci-Cabello

To cite this article: Maria A. Fiol-deRoque, José M. Valderas, Jorge Arias de la Torre, Maria J. Serrano-Ripoll, Montserrat Gens-Barberà, Encarna Sánchez-Freire, Francisco M. Martín-Luján, Antonio Olry de Labry & Ignacio Ricci-Cabello (2024) Evaluation of the psychometric performance of the Spanish and Catalan versions of the patient reported experiences and Outcomes of Safety in Primary Care (PREOS-PC)-Compact questionnaire, European Journal of General Practice, 30:1, 2296573, DOI: 10.1080/13814788.2023.2296573

To link to this article: https://doi.org/10.1080/13814788.2023.2296573

EUROPEAN JOURNAL OF GENERAL PRACTICE
2024, VOL. 30, NO. 1, 2296573
https://doi.org/10.1080/13814788.2023.2296573

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Evaluation of the psychometric performance of the Spanish and Catalan versions of the patient reported experiences and Outcomes of Safety in Primary Care (PREOS-PC)-Compact questionnaire

Maria A. Fiol-deRoque[a,b,c], José M. Valderas[d,e], Jorge Arias de la Torre[f,g,h], Maria J. Serrano-Ripoll[a,b,c], Montserrat Gens-Barberà[i,j], Encarna Sánchez-Freire[k], Francisco M. Martín-Luján[l,m], Antonio Olry de Labry[f,n] and Ignacio Ricci-Cabello[a,b,f]

[a]Research Group in Primary Care and Promotion – Balearic Islands Community (GRAPP-caIB), Health Research Institute of the Balearic Islands (IdISBa), Palma, Spain; [b]Primary Care Research Unit of Mallorca, Balearic Islands Health Services, Palma, Spain; [c]Prevention and Health Promotion Research Network (redIAPP)/Network for Research on Chronicity, Primary Care, and Health Promotion (RICAPPS), Barcelona, Spain; [d]Centre for Research in Health Systems Performance, Yong Loo Lin School of Medicine, National University of Singapore, Singapore, Singapore; [e]Department of Family Medicine, National University Health System, Singapore, Singapore; [f]Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [g]CIBER Biomedical Research Center in Epidemiology and Public Health (CIBERESP), Health Institute Carlos III (ISCIII), Madrid, Spain; [h]Institute of Biomedicine, University of Leon, Leon, Spain; [i]Quality and Patient Safety Central Functional Unit, Gerència d'Atenció Primària Camp de Tarragona, Catalan Institute of Health (ICS), Tarragona, Spain; [j]Research Group in Quality and Patient Safety, Institut Universitari d'Investigació en l'Atenció Primària-IDIAP Jordi Gol, Catalan Institute of Health (ICS), Tarragona, Spain; [k]Quality and Patient Safety Unit, Gerència d'Atenció Primària Catalunya Central, Catalan Institute of Health (ICS), Barcelona, Spain; [l]Primary Healthcare Research Support Unit-Camp de Tarragona, Institut Universitari d'Investigació en l'Atenció Primària-IDIAP Jordi Gol, Catalan Institute of Health (ICS), Tarragona, Spain; [m]Department of Medicine, Faculty of Medicine and Health Sciences, Universitat Rovira i Virgili (URV), Reus, Spain; [n]Research Group in Health and Gender, Andalusian School of Public Health, Granada, Spain

ABSTRACT

Background: Patients provide a unique, irreplaceable, and essential perspective in evaluating patient safety. The suite of Patient Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC) tools are a notable exception to the scarcity of patient-reported patient safety measures. Full evaluation of their performance has only been attempted for the English version, thereby limiting its international applicability.

Objectives: To assess the psychometric performance of the Spanish and Catalan versions of the PREOS-PC-Compact.

Methods: Cross-sectional validation study. We used Classical Test Theory methods to examine scale score distribution, internal consistency, and construct validity; and Item Response Theory (IRT) methods to further explore construct validity.

Results: 3287 patients completed the Spanish version, and 1007 the Catalan version. Similar results were obtained for both versions. Confirmatory Factor Analysis supported a single construct for each scale. The correlations between PREOS-PC-Compact scales and known group analysis suggested adequate construct validity (inconclusive for known groups at the provider level). All four multi-item scales demonstrated adequate internal consistency reliability ($\alpha > 0.7$), which was only confirmed for test-retest reliability for 'Practice activation.' A sample between 60–90 patients per practice was estimated sufficient to produce scores with reliability > 0.7 for all scales except for harm scales. IRT models showed disordered thresholds for 'Practice activation' and 'Harm burden' but showed excellent fit after reducing the response categories.

Conclusion: The Spanish and Catalan versions of the PREOS-PC-Compact are broadly valid and reliable tools to measure patient safety in Spanish primary care centres; confirmation of lower-than-expected test-rest reliability merits further examination .

ARTICLE HISTORY

Received 7 March 2023
Revised 23 November 2023
Accepted 4 December 2023

KEYWORDS

Health care evaluation mechanisms; observational study; patient safety; patient reported outcomes; primary health care

CONTACT Maria A. Fiol-deRoque mariaantonia.fiol@ssib.es Primary Care Research Unit of Mallorca, Balearic Islands Health Services, Carrer de l'Escola Graduada, 3, Palma 07002, Spain; Health Research Institute of the Balearic Islands (IdISBa), Palma 07010, Spain

Supplemental data for this article can be accessed online at https://doi.org/10.1080/13814788.2023.2296573.

Introduction

The Organisation for Economic Co-operation and Development (OECD) reports that approximately 20%–25% of individuals receiving primary and ambulatory care suffer harm, resulting in a direct cost of about 2.5% of total health expenditure due to additional tests, treatments, and healthcare [1]. In Spain, around 3 million adverse events occur in the Primary Care Centres (PCC) [2], resulting in costs of up to 1000 million Euros per year [3]. Consequently, patient safety (defined as 'the prevention of errors and adverse effects to patients associated with healthcare') has become a significant focus for primary care systems worldwide [4].

The lack of appropriate measurement tools impedes efforts to enhance patient safety [5]. Available tools rely on healthcare providers' information, such as safety culture questionnaires or voluntary reporting of safety events by health professionals, while patient-reported tools are scarce. However, over the last decade, patient participation in incident prevention and harm reduction has gained momentum [6]. Patients, as care recipients, often spot different safety incidents and outcomes than staff, being motivated to report errors [7]. This engagement yields positive results in preventing adverse events and increasing awareness of safety risks [6,8–10]. Indeed, patient-reported data drives successful innovative interventions for safer health systems [11–13].

The Patient-Reported Experiences and Outcomes of Safety in Primary Care (PREOS-PC) questionnaire is an innovative patient-centred tool used to measure patient safety in primary care settings [14]. This tool was initially developed in England using a multistage process that involved an expert panel and was informed by two systematic reviews [15,16], four focus groups, and 18 cognitive interviews [14]. A subsequent validation study involving 45 PCCs provided evidence of the questionnaire's validity and reliability [14]. The original version of the questionnaire contained 61 items and was suitable for research purposes, successfully measuring and evaluating patient safety in PCCs in the UK [17]. Two shorter questionnaire versions were subsequently developed: PREOS-PC-Compact (29 items) and PREOS-PC screen (6 items). PREOS-PC-Compact balances high psychometric standards (high reliability and validity) and reduces administrative and patient burden. This could simplify its adoption and utilisation for standard patient safety monitoring in primary care [18,19].

PREOS-PC-Compact was cross-culturally adapted for its use in Spain as part of SinergiAPS ('Sinergias entre profesionales y pacientes para una Atención Primaria Segura'/Synergizing providers and patients for Safer Primary Care) [13], an intervention aimed at improving patient safety in PCCs through the provision of patient feedback. PREOS-PC-Compact (described in detail in Methods and Table 1) is available in Spanish and Catalan. Data from a pilot study involving 493 patients in 10 PCCs in Spain offered incipient evidence supporting the validity and reliability of the Spanish PREOS-PC-Compact [20]. A separate study provided evidence of sensitivity to change in PREOS-PC-Compact scale scores (substantially decreased during the COVID-19 pandemic compared to pre-pandemic scores [21]). However, the available evidence regarding the psychometric properties of the Spanish PREOS-PC-Compact is still limited, as it mainly stems from data from a pilot study and many properties have not yet been analysed. Additional evidence (based on larger samples) is needed to confirm the observed findings further and to examine additional psychometric

Table 1. Structure of the Spanish and Catalan versions of the patient reported experiences and outcomes of safety in primary care PREOS-PC-compact[a].

	PREOS-PC-compact		
Domain	Construct	Rating items	Open-ended items
Practice activation (what does the PCC do to create a safe environment and to ensure safety)	Practice activation	4	1
Patient activation (how pro-active are patients in ensuring safe healthcare delivery)	Patient activation	2	1
Patient experiences of safety problems (safety errors)	Types of safety problems experienced[b]	12	1
Outcomes of patient safety (harm)	Harm severity (health domain-specific harm)	3	1
	Harm burden (health and personal care, and financial needs)	3	
Overall perceptions of patient safety (how safe do patients rate their PCC)	Overall rating of patient safety[c]	1	0
Overall	6 constructs	25	4

[a]Items are based on Likert scale, unless otherwise stated.
[b]Items conceived to be used at the item level (for description, rather than evaluation purposes). Not recommended to combine its items into a summary
[c]Visual Analogue Scale.
PCC: Primary Care centre; PREOS-PC: Patient Reported Experiences and Outcomes of Safety in Primary Care.

properties, such as construct validity. Up to date, the psychometric properties of the Catalan version remain unexplored.

This study aims to analyse the psychometric properties of the Spanish and Catalan versions of the PREOS-PC-Compact.

Methods

Study design

Cross-sectional validation study using data collected as part of SinergiAPS, a randomised clinical trial evaluating a patient feedback intervention to improve patient safety. The SinergiAPS intervention has been described in detail elsewhere [13]. In brief, centres allocated to the intervention arm were fed back information regarding patients' experiences of safety (collected through the PREOS-PC-Compact questionnaire) and were instructed to plan safety improvement actions based on it.

Data collection

Baseline data collection took place between November 2019 and January 2020. The PREOS-PC-Compact questionnaire was administered to 4,555 patients from 59 PCCs (at least 75 per centre) in Mallorca, Barcelona, and Tarragona (Spain). After excluding questionnaires completed in English ($n=27$), those indicating no visits to the PCCs during the last year ($n=89$), and with missing sociodemographic characteristics ($n=145$), we analysed 4294 questionnaires. The characteristics of the participating PCC are detailed in Supplementary Material 1. In each PCC, we sequentially invited all patients in the waiting room to complete the self-administrated PREOS-PC-Compact. Patients were given the option to choose between the Spanish or Catalan version. The questionnaire aimed to gather their perceptions, experiences, and outcomes regarding healthcare safety received from their PCCs over the past 12 months. Additionally, patients were requested to provide sociodemographic and clinical information. For patients under 18, we invited them if they were accompanied by an adult caregiver, who would complete the questionnaire on their behalf. In addition, we provided patients with the opportunity to administer the questionnaire personally. The total number of questionnaires included in this study in each PCC in each language is available in Supplementary Material 2.

Baseline characteristics of the participating PCCs were extracted from electronic health records by information technology specialists from the Balearic Islands' Health Service and the Catalan Institute of Health (see acknowledgements). The information extracted included: the number of registered patients, proportion of female patients, proportion of patients aged > 65, number of Primary Care providers, patients' clinical complexity (GMA [22]), PCC rurality index [23], and rate of avoidable hospital admissions in the previous twelve months [24]. In addition, all Primary Care professionals participating in the SinergiAPS trial were requested to complete the Spanish or Catalan versions of the Medical Office Survey on Patient Safety Culture (MOSPSC) [25,26], a measure of healthcare professionals' perceptions of the PCC's safety climate.

PREOS-PC-Compact

PREOS-PC-Compact contains 29 items organised into five domains (Table 1): centre activation (how the PCC works towards establishing a safe environment and ensuring the safety of its patients – 4 items-scale and an open question); patient activation (how actively do patients ensure safe healthcare? – 2 items scale and an open question); patient experiences of safety problems (safety errors – 12 items not conceived to be combined in a scale, and an open question); outcomes of patient safety (harm – two scales of 3 items each, evaluating severity and burden respectively, and an open question); and overall perception of safety (1 item).

To determine the scale scores, we utilised all available responses without any imputation and expressed them as a percentage of the maximum score across all available items in the scale, with scores ranging from 0 to 100, indicating the higher the score, the greater the safety. In cases where 50% or more of the items in a scale had missing or non-applicable responses, we considered the scale score as missing. The Spanish and Catalan questionnaires and the details about computing the scale scores are available in Supplementary Material 3.

The cross-cultural adaptation process of PREOS-PC-Compact into Spanish involved setting up a panel of 6 national experts in patient safety from different backgrounds, including PHC doctors, academics, and managers, to determine its content validity. This was followed by translation and back-translation of the questionnaire, cognitive testing with eight patients, and a pilot study with 10 PCCs and 500 patients [20]. The Catalan version was adapted similarly, including questionnaire forward and back translation from English and cognitive debriefing with patients. More details are available in the Supplementary Material 4.

Psychometric properties

Acceptability of items. We evaluated the number and patterns of missing data as a proxy for the acceptability of items [27].

Floor and ceiling effect. We analysed the proportion of responses that fell in the minimum and maximum categories for each item in each scale to assess floor and ceiling effects. We established a threshold of 80% or more of the responses falling into the maximum or minimum category, respectively [28].

Reliability. We evaluated internal consistency by examining inter-item correlations (with coefficients ≥0.3 indicating good consistency) [29] and Cronbach's α (α ≥ 0.7 [30] indicating good reliability). We assessed test-retest reliability using one-way random effects intra-class correlations (ICC) (with ICC of at least 0.7 [31] indicating good test-retest reliability), utilising data from a subset of patients who completed the PREOS-PC-Compact twice in the same language version, approximately two weeks apart.

Structural and construct validity. To assess structural validity, we independently conducted a Confirmatory Factor Analysis (CFA) on each multi-item scale with more than 2 items, that is, 'Practice Activation,' 'Harm Severity' and 'Harm Burden' with each item loading only on the corresponding scale without restrictions to correlate with the rest of the items on the scale and excluding the observations with missing values. We examined goodness-of-fit statistics, including the Satorra-Bentler Chi-squared statistic, comparative fit index (CFI), and standardised root-mean residual (SRMR). For model evaluation, we followed Hu and Bentler's recommendation, using a combinational rule of CFI > 0.95 and SRMR < 0.09 [32]. We also evaluated construct validity (known-group [33]) by (1) testing pre-specified patient safety differences (at the patient level – according to patients' gender, age, number of visits, and clinical complexity) by bivariate linear regression analysis; (2) examining the correlation between practice level patient safety scores and PCCs characteristics (mean number of patients per doctor, clinical complexity (GMA), *the* proportion of patients aged 65 or over, safety climate score (MOSPSC), and avoidable hospital admissions), and (3) examining the correlations amongst the PREOS-PC-Compact scales. All the a priori hypothesised relationships are available at Supplementary Material 5.

Practice-level precision and discrimination between PCCs. To evaluate how effectively each scale measures safety at the PCC level, we computed the standard error of a PCC mean score to measure the precision of measurement. A standard error of 5 points on the 0–100 scale indicates strong precision [28]. Additionally, we utilised the between-practice ICC coefficient to compute the reliability coefficient, which measures the ability to discriminate between PCCs [28].

$$Reliability\ of\ PCC\ mean = \frac{n_{per\ prectice} \times SD^2_{of\ PCC\ effect}}{n_{per\ prectice} \times SD^2_{o\ PCC\ effect} + SD^2_{within\ PCC}}$$

We estimated the sample size required to achieve reliable discrimination (at the 0.7 level) between PCC scores based on the between-practice ICC coefficient [28].

$$Responses\ needed\ for\ 0.7\ reliability = \frac{0.7 \times SD^2_{within\ PCC}}{0.3 \times SD^2_{of\ PCC\ effect}}$$

One parameter item response Theory (IRT) models. We used IRT models [34] to analyse data from participants who responded to all items in each scale. We evaluated the unidimensionality assumption by conducting a post-hoc principal components analysis (PCA) on the item residuals and the local independence by examining correlations between item residuals. We considered assumptions satisfied if the first eigenvalue of the PCA was ≤2, and residual correlations were ≤0.40 [35]. We reported the fit of each item to the scale according to the IRT model using the infit mean square statistic, where values between 0.50 and 1.50 were considered productive for measurement [36]. Additionally, we used person-item maps [37] to inspect item difficulties and the spread of *the* response category threshold for each model. Items with the same difficulty parameters are indicative of redundant information. Thresholds are the specific positions that separate response categories, ensuring equal chances of selection for each response (determined at the median probability point).

We conducted analyses using Stata\SE version 15.1, except the IRT models, which we constructed using the eRm package for R version 4.1.0.

Ethical approval was granted by the Balearic Islands Ethics Committee (CEI IB 3686/18) and by IDIAP-Jordi Gol Research Ethics Committee (CEIC-IJG 19/116-P, 24/07/2019).

Results

Characteristics of the participating patients

Of the 4294 respondents, 3287 completed the Spanish PREOS-PC-Compact version and 1007 the Catalan PREOS-PC-Compact version (Table 2). Among those completing the Spanish version, 2127 (64.71%) were female, with a mean age of 51.83 years (SD = 18.53).

Most respondents were from Spain (2813; 85.58%) and without a university degree (1949; 59.29%). Around half (1778, 54.09%) reported multiple long-term conditions and 60.69% (1995) reported a 'very good'/'good' health status. Patients completing the Catalan version were more likely to be aged 65, retired, from Spain and with university studies.

Psychometric properties

The psychometric characteristics of the Spanish and Catalan versions are presented in Tables 3 and 4.

Acceptability of items

The median item 'Non-Applicable-Don't know' (N/A) response rate for the Spanish version was 2.62% (interquartile range 1.82%–3.36%), indicating a high acceptability. The Catalan version presented similar acceptability, showing a median item N/A response rate of 1.99% (interquartile range 1.69%–2.73%). In both versions the items showing less acceptability pertained to the 'Patient activation' construct, with an N/A response rate higher than 8% for both items.

Table 2. Sociodemographic and clinical characteristics of the respondents.

	Spanish version	Catalan version
Total respondents	3287	1007
Sex		
Male	1160 (35%)	376 (37%)
Female	2127 (65%)	631(63%)
Age*		
Mean (SD); [range]	51.83 (18.53); [1–97]	53.43 (19.29); [1–95]
< 18	74 (2%)	26 (3%)
18–29	360 (11%)	103 (10%)
30–44	739 (22%)	205 (20%)
45–64	1192 (36%)	326 (32%)
≥ 65	922 (28%)	347 (34%)
Country of origin*		
Spain	2813 (85%)	971 (96%)
EU countries	123 (4%)	12 (1%)
Non-EU countries	351 (11%)	24 (2%)
Educational level*		
No qualifications	777 (24%)	232 (23%)
Non-university studies	1949 (59%)	568 (56%)
Degree, degree equivalent and above	561 (17%)	207 (21%)
Visits to their primary care centre during the last 12 months		
1–5	1844 (56%)	569 (56%)
6–10	786 (24%)	258 (26%)
>10	657 (20%)	180 (18%)
Employment situation*		
Working	1647 (50%)	468 (46%)
Retired	1043 (32%)	362 (36%)
Other (unemployed, student)	597 (18%)	177 (18%)
Health status		
Very good/ Good	1995 (61%)	643 (64%)
Fair /Bad /Very bad	1292 (39%)	364 (36%)
Number of long-term conditions		
Mean (SD); [range]	2.19 (2.11); [0–16]	2.09 (2.03); [0–11]
0	862 (26%)	280 (28%)
1	647 (20%)	202 (20%)
2–3	1002 (30%)	293 (29%)
>3	776 (24%)	232 (23%)

Differences between the Spanish and Catalan sample were evaluated by *t*-test or Chi square.
*Differences were found for age, country of origin, educational level and employment situation ($p < .05$).

Floor and ceiling effect

In both versions, we observed a ceiling effect for the item 'GP took patient's concerns seriously?' (80% of patients reported 'Always'); and for all the items in the scales 'harm severity' and 'harm burden' (more than 90% of patients reported 'not at all' for all the items). Accordingly, we observed a ceiling effect for both harm scales, with 89.81% and 88.51% of respondents with the highest possible scores in the Spanish version and 97.26% and 97.39% of respondents achieving the highest possible scores in the Catalan version (Table 3).

Reliability

The four multi-item scales, the Spanish version demonstrated adequate internal consistency (Cronbach's α, 0.71–0.85) and adequate homogeneity (inter-item correlations, 0.37–0.74) (Table 4). The test-retest was performed on a sub-sample of 55 patients of the 386 who initially consented (retest response rate 14.2%). The test-retest ICC was above the 0.70 standard only for the 'Practice activation' scale.

The Catalan version showed lower reliability features, with three scales presenting suboptimal internal consistency ('Practice Activation' (α = 0.67), 'Harm severity' (α = 0.69) and 'Harm burden' (α = 0.62)). We could not explore test-retest ICC for the Catalan version due to insufficient sample size.

Structural and construct validity

The CFA on 'Practice Activation,' 'Harm Severity' and 'Harm Burden' scales provided evidence for high structural validity in both versions (Table 3). The models met Hu and Bentler's criteria, suggesting adequate goodness-of-fit, although they failed to meet the Chis-squared Satorra–Bentler scaled statistic (data available in Supplementary Material 6). The Spanish version presented moderately high internal consistency coefficients while the Catalan version presented suboptimal consistency coefficients for some scales. In both cases, the results of the CFA

Table 3. Psychometric characteristics of the Spanish ($n = 3287$) and Catalan ($n = 1007$) versions of PREOS-PC-Compact scales and items.

		N/A – "Don't know" responses %		Min category response %		Max category response %		Alpha if item deleted		Factor loading from the CFA (95% CI)[a]	
Items	Categories	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat
Practice Activation (unidimensional)											
GP available when needed? (b1_p1)	5 (Never-Always)	0.33	0.30	0.33	0.50	67.20	68.62	0.675	0.606	0.501 (0.458; 0.544)	0.578 (0.492; 0.664)
GP took patient's concerns seriously? (b1_p2)	5 (Never-Always)	0.18	0.20	0.18	0.20	80.96	80.54	0.635	0.584	0.571 (0.534; 0.608)	0.615 (0.537; 0.693)
GP told the patient about side effects? (b1_p3)	5 (Never-Always)	2.86	2.48	7.91	9.83	56.34	54.12	0.644	0.628	0.658 (0.620; 0.695)	0.523 (0.441; 0.605
GP encouraged the patient to talk about healthcare concerns? (b1_p4)	5 (Never-Always)	4.50	3.48	16.88	13.41	43.47	46.57	0.611	0.583	0.721 (0.687; 0.755)	0.603 (0.536; 0.670)
Patient activation											
Raise a concern? (b2_p1)	5 (Never-Always)	11.90	13.01	32.34	38.53	21.42	20.56	–	–	–	–
Make a suggestion? (b2_p2)	5 (Never-Always)	10.37	8.24	40.80	48.86	13.87	13.11	–	–	–	–
Harm severity (unidimensional)											
Physical health (b5_p2)	4 (Yes, extreme-Not at all)	2.98	1.99	0.82	0.10	90.36	94.46	0.743	0.608	0.755 (0.679; 0.832)	0.667 (0.488; 0.845)
Mental health (b5_p3)	4 (Yes, extreme-Not at all)	2.62	1.99	0.52	0.10	93.58	95.03	0.754	0.626	0.735 (0.650; 0.819)	0.632 (0.430; 0.834)
Limitations doing usual activities (b5_p4)	4 (Yes, extreme-Not at all)	1.83	1.79	0.55	0.00	93.25	94.64	0.715	0.548	0.790 (0.714; 0.866)	0.675 (0.503; 0.846)
Harm burden (unidimensional)											
Healthcare needs (b5_p5)	4 (Yes, extreme-Not at all)	1.98	1.89	0.76	0.50	90.36	92.45	0.723	0.432	0.766 (0.707; 0.826)	0.716 (0.551; 0.881)
Personal care needs (b5_p6)	4 (Yes, extreme-Not at all)	1.79	1.69	0.49	0.00	95.59	97.22	0.722	0.622	0.762 (0.676; 0.846)	0.450 (0.255; 0.645)
Financial needs (b5_p7)	4 (Yes, extreme-Not at all)	1.86	1.69	0.85	0.40	91.33	92.75	0.738	0.487	0.737 (0.664; 0.810)	0.630 (0.460; 0.802)
General perception											
Overall rating of patient safety (visual analogue scale with scores from 0-10)	N/A	N/A	N/A	0.15	0.20	34.16	35.65	N/A	N/A	N/A	N/A

Cat: Catalan; GP: general practitioner; N/A: not applied; Spa: Spanish.
Unidimensionality was assessed with a factor analysis. Factor loadings >0.4 are considered moderate, and loadings > 0.6 are considered good.
[a]Standarised coefficients with 95% confidence intervals.

indicated that each scale measures a singular construct and that the items can be aggregated to generate summary scores.

Whereas the results from the patient-level known group analyses generally supported the construct validity of the scales from the Spanish version, the results from the practice-level analyses were largely inconclusive. In the Catalan version most of the hypothesised differences between groups of patients or PCC characteristics did not reach statistical significance (Table 5).

The correlations between PREOS-PC-Compact scales in the Spanish version (Table 6) suggested an adequate construct validity of the scales, being in line with our pre-specified hypothesis (e.g. strong positive correlation between 'Harm severity' and 'Harm burden' ($r = 0.54$); but not a significant correlation between 'Patient activation' and 'Overall rating').

In the Catalan version, the correlations between PREOS-PC scales also indicated adequate construct validity (e.g. strong positive correlation between 'Harm severity' and 'Harm burden' ($r = 0.48$) as well as between 'Practice activation' and 'Overall rating' ($r = 0.46$) (Table 6).

Practice-level precision and discrimination between PCCs

In the Spanish version, the observed mean sample size per PCC ranged between 50 and 55 answers by scale. PCC mean scores demonstrated high precision for all

Table 4. Distribution of scores and reliability of the Spanish and Catalan patient reported experiences and outcomes of safety in primary care (PREOS-PC) scales.

Scales	N		Mean (SD)		Respondents with the lowest possible score (%)		Respondents with the highest possible score (%)		Internal consistency				Test-retest reliability[c]		PCC mean scores					
									Cronbach's α[a], mean		Inter-item correlation[b], standardised mean		ICC[d] (95%CI)		Precision (standard error)		Between-practices reliability		*N* of responses needed for 0.7 reliability	
	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat	Spa	Cat
Practice activation	3284	1005	81.20 (19.89)	81.94 (18.79)	0.03	0.50	33.80	32.04	0.705	0.667	0.374	0.334	0.832 (0.703; 0.904)[e] $p < .001$		2.603	3.749	0.731	0.644	48	30
Patient Activation	3028	936	38.91 (36.88)	34.21 (37.38)	34.91	43.48	13.74	13.03	0.854	0.865	0.745	0.762	0.402 (−0.0450; 0.662)[e] $p = .036$		5.071	7.503	0.613	0.760	76	16
Harm severity	3218	992	96.53 (12.13)	97.26 (9.88)	0.31	0.81	89.81	90.93	0.808	0.688	0.584	0.423	0.069 (−0.319; 0.382) $p = .354$		1.632	2.053	0.425	0.114	172[f]	419[f]
Harm burden	3237	991	96.13 (12.48)	97.39 (9.32)	0.34	0.10	88.51	90.92	0.800	0.617	0.572	0.350	0.055 (−0.329; 0.369) $p = .383$		1.677	1.941	0.332	0.040	258[f]	1275[f]
Overall rating of patient safety	3287	1007	84.65 (16.30)	85.70 (15.27)	0.15	0.20	34.16	64.35	N/A	N/A	N/A	N/A	0.411 (0.027; 0.648)[e] $p = .015$		2.156	3.082	0.595	0.543	88	46

Cat: Catalan; CI: confidence interval; ICC: intercluster correlation coeficient; N/A: non-applicable (single item scales); PCC: Primary Care Centres; SD: standard deviation; Spa: Spanish.
[a]An α > 0.7 represents adequate internal consistency, and an α > 0.8 is considered good.
[b]Inter-item correlations above 0.3 are considered adequate.
[c]Based on data from 55 patients who completed the questionnaire again after two weeks.
[d]Two-way random-effects intra-class correlation coefficients and 95% confidence intervals.
[e]$p < .05$.
[f]Responses needed to have sufficient cases reporting harm.

the scales, with standard errors ranging from 1.63 ('Harm Burden') to 5.07 ('Patient activation'), being just above the predefined limit for the scale of 'Patient activation.'

The between-practices reliability coefficients were < 0.70 except for 'Practice activation' (0.73). This indicates that a sample size of 60–90 patients per PCC would be adequate to generate well-discriminating scores (reliability ≥ 0.7) for all scales except the harm scales (severity and burden), for which larger samples (172 and 258, respectively) would be necessary.

For the Catalan version the between-practices reliability for 'Practice activation' was of 0.76. However, the rest of the coefficient were lower than 0.70 (especially for both harm scales), indicating that, with the available sample (about 23 questionnaires per PCC), the PCC mean score does not discriminate well between PCCs regarding patient perceptions of safety. A sample size of 16–46 patients per PCC would be adequate to generate well-discriminating scores for most scales while for the harm scales (severity and burden) larger samples would be necessary (419 and 1275, respectively).

One parameter IRT models

The unidimensionality and local independence assumptions were met in each model (goodness of fit indices are available at Supplementary Material 6). Results from the IRT models showed that all items had infit mean square statistics considered productive for measurement (Supplementary Material 7A,B). The item difficulty estimates ranged from −0.75 (b1_p2) to 1.12 (b1_p4) in the Spanish version, and from −0.79 (b1_p2) to 0.88 (b5_p6) for the Catalan version, without items with similar difficulty parameters. Regarding response category thresholds, in both versions all items from 'Patient activation' and 'Harm severity' had properly ordered response categories. However, in 'Practice activation' three items had disordered category thresholds (b1_p2, b1_p3 and b1_p4) and in 'Harm burden,' one item had disordered category thresholds (b5_p7). Disordered thresholds could indicate response options are not well-calibrated. Item Characteristic Curves and person-item maps are available in Supplementary Material 7A,B).

Reduction of scales

Based on IRT results, we rescored all items from the 'Practice activation' and 'Harm burden' scales (those with disordered thresholds) in both language versions. We combined the response options, leaving both in

Table 5. Known group Analysis based on characteristics of patients and Primary Care Centres.

	Harm severity (Impact on health)		Harm burden (Impact on personal care, and financial needs)	
Patient characteristics, β (95% CI)[a]	**Spa (*n* = 3287)**	**Cat (*n* = 1007)**	**Spa (*n* = 3287)**	**Cat (*n* = 1007)**
Gender				
Male	1	1	1	1
Female	−0.631 (−1.545; 0.176)	−0.935 (−2.214; 0.216)	−1.029 (−1.870 − 0.143)*	−0.826 (−1.907; 0.214)
Age	0.059 (0.037; 0.081)Ł	0.022 (−0.007; 0.051)	0.055 (0.034; 0.078)Ł	0.026 (−0.005; 0.056)
Visits to the PCC/year				
1–5	1	1	1	1
6–10	−0.715 (−1.658; 0.225)*	−2.403 (−3.984; −0.912)*	−1.335 (−2.417; −0.350)*	−0.864 (−2.161; 0.320)
>10	−2.837 (−4.199; −1.520)*	−3.791 (−6.246; −1.718)*	−2.450 (−3.736; −1.219)*	−3.907 (−5.992; −1.800)*
Discordant multimorbidity				
No	1	1	1	1
Yes	−1.297 (−2.280; −0.447)*	−2.262 (−3.627; −0.907)*	−1.174 (−2.058; −0.276)*	−1.051 (−2.283; 0.121)
PCC characteristics, *r* (95% CI)[b]	**Spa (*n* = 59)**	**Cat (*n* = 43)**	**Spa (*n* = 59)**	**Cat (*n* = 43)**
Quota of assigned patients per doctor	−0.386 (−0.584; −0.1480)*	−0.130 (−0.453; 0.180)	−0.141 (−0.391; 0.153)	−0.107 (−0.440; 0.210)
PCC proportion of patients aged >65	0.292 (0.027; 0.522)Ł	0.166 (−0.147; 0.486)	0.088 (−0.203; 0.354)	0.033 (−0.287; 0.337)
Patient complexity index (GMA)[c]	0.003 (−0.338; 0.331)	0.003 (−0.332; 0.342)	−0.067 (−0.388; 0.250)	−0.067 (−0.394; 0.238)
Patient safety culture (MOSPSC, SIPS)	0.047 (−0.237; 0.297)	0.001 (−0.281; 0.310)	0.049 (−0.215; 0.310)	0.092 (−0.217; 0.395)
Avoidable hospitalisations	0.276 (−0.003; 0.533)Ł	0.048 (−0.276; 0.421)	−0.052 (−0.349; 0.216)	0.028 (v0.310; 0.313)

CI: confidence interval; SIPS: Synthetic index of Patient Safety; MOSPSC: Medical Office Survey on Patient Safety Culture; PCC: Primary Care Centre.
[a]bivariate linear regression, non-parametric bootstrapped percentile confidence interval, based on bootstrap samples of 1000.
[b]spearman correlation coefficient, non-parametric bootstrapped percentile confidence interval, based on bootstrap samples of 1000.
[c]complexity measured using the Adjusted Morbidity Groups Index (higher scores indicating higher complexity levels).
*Statistically significant difference ($p < .05$) favouring (supporting) our pre-specified hypothesis.
Ł Statistically significant difference ($p < .05$) against (i.e. in opposite direction of) our pre-specified hypothesis.

Table 6. Correlations between practice level PREOS-PC scale scores.

Spanish	1	2	3	4	5
1. Practice activation	1				
2. Patient Activation	0.143*	1			
3. Harm severity	0.202*⊣	−0.093*⊣	1		
4. Harm burden	0.206*⊣	−0.082*⊣	0.538*⊣	1	
5. Overall rating of patient safety	0.445*⊣	−0.032⊣	0.232*⊣	0.248*⊣	1
Catalan					
1. Practice activation	1				
2. Patient Activation	0.067&	1			
3. Harm severity	0.198*⊣	−0.117*⊣	1		
4. Harm burden	0.191*⊣	−0.177*⊣	0.477*⊣	1	
5. Overall rating of patient safety	0.463*⊣	−0.095$	0.243*⊣	0.260*⊣	1

Values indicate Spearman pairwise correlations coefficients.
⊣: observed correlations matching pre-specified hypothesis.
*$p \geq .001$; $p \geq .01$; &$p \leq .05$.

three response categories: the current extreme categories and a central one that combines the adjacent response options. The psychometric characteristics of the modified 'Practice activation' and 'Harm burden' scales and their items are presented in Supplementary Material 8. For both versions, the two showed scales were unidimensional, productive for measurements and presented a similar structural validity to the original one. The modified 'Practice activation' scale's internal consistency was just below adequate for both versions (α = 0.677 [Spanish] and α = 0.671 [Catalan]). Results from the IRT models for both modified scales showed that all items had properly ordered response categories. According to these results, reducing the response items of these scales seems appropriate.

Discussion

Main findings

In this study, we conducted a detailed psychometric evaluation of the Spanish and Catalan versions of the PREOS-PC-Compact questionnaire. The results support the acceptability of both versions. They also support their reliability, validity, and internal consistency.

Comparison with existing literature

As far as we know, the PREOS-PC is currently the only validated questionnaire available in Spanish and Catalan to assess patient safety in primary care from patients' perspectives. While other questionnaires exist in different languages, such as the Primary Care Patient Measure of Safety questionnaire or the ASK-ME-questionnaire [38,39], they primarily focus on contributory factors to safety incidents rather than actual experiences of safety events and harm.

Our Classical Test Theory results are consistent with our previous exploratory study involving 493 patients [20], confirming the validity and reliability of the

questionnaire. Moreover, when comparing the Spanish and Catalan versions with the original English PREOS-PC-Compact [19], we observed similar psychometric properties. Some of the CFA coefficients in the Spanish version were marginally higher than the English one (indicating better performance), except for the 'Practice Activation' scale, which was lower in the Spanish version (with values below 0.6 for the first two items).

Implications

The PREOS-PC-Compact questionnaire discriminates among various user groups and types of primary care centres, making it a valuable tool for monitoring patient safety in different contexts. However, it's important to consider the minimum sample size necessary to ensure precision and discrimination, particularly for scales measuring less frequent events like harm.

We did not find relevant differences between the Spanish and Catalan versions (which may be explained by the proximity between the two languages and the shared cultural context in Spain) supports the combined use of both versions in Catalan-speaking regions of Spain.

Strengths and limitations

This study benefits from several methodological strengths, including a large dataset of completed questionnaires in both languages, allowing for thorough analyses. Additionally, combining Classical Test Theory with IRT models represents a state-of-the-art psychometric evaluation. Moreover, we explored various features, including comparing psychometric properties in different languages.

However, certain limitations should be acknowledged. First, we could not obtain precise data regarding the response rate at the questionnaire level (as it was not accurately recorded). Our previous study shows that the response rate to PREOS-PC-Compact administered in PCCs is around 77% [20]. Sometimes, response bias is unlikely to introduce major systematic biases in psychometric studies [40]. However, future research should examine the characteristics of non-respondents. Second, the PREOS-PC aims to capture the experiences of all groups attending the PCC, including children. In our sample, 2.5% of the questionnaires correspond to under 18 patients. Consequently, some of these questionnaires may have been answered on behalf of someone else. Third, concerning the CFA, the three models fail to meet the Chis-squared Satorra–Bentler scaled statistic. Fourth, we observed skewed score distributions for some items and scales. Skew is common in questionnaires evaluating patients' perspectives on medical care and does not necessarily limit the ability to reliably distinguish PCCs and patient subgroups with substantial sample sizes such as ours. Fifth, we addressed missing data by list-wise deletion for consistency with the methodology used in previous studies evaluating the psychometric performance of the PREOS-PC questionnaire [19,20] (thus allowing comparability across studies). Given the low proportion of missing data in this study (7.6% on the scale with the highest proportion), we feel that it is unlikely that having used alternative approaches (such as multiple imputation or full-information maximum likelihood) would have yielded substantially different results. Finally, the small sample of patients who completed the questionnaire after two weeks reduced our ability to analyse the test-retest reliability.

Conclusion

This study suggests that the Spanish and Catalan versions of the PREOS-PC-Compact are valid and reliable tools to measure the level of patient safety in PCCS accurately from the patients' perspective. Further refinements could help optimise and strengthen its measurement properties.

Acknowledgements

We thank Mr. Jerónimo Fiol Renaudin for his support in data management–related tasks. We thank Dr. Luke Mounce for his advice on IRT analysis. We thank the technology specialists from the Research in Health Information Platform (Health Research Institute of the Balearic Islands (IdISBa)) and the Catalan Institute of Health for their support with the extraction of the data used in this study. Finally, we are deeply thankful to all the patients who completed the PREOS-PC questionnaire.

Authors contribution

IR-C, JMV, JAT and MAF-dR contributed to the concept and design of the study. MJS-R, MAF-dR, MG-B, ES-F were responsible for conducting data collection. Analysis and interpretation of data were performed by MAF-dR, JMV, JAT and IR-C. MAF-dR and IR-C wrote the first draft of the manuscript and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Ethics statement

This study was performed in line with the principles of the Declaration of Helsinki. The study was approved by Balearic

Islands Ethics Committee (CEI IB 3686/18) and by IDIAP-Jordi Gol Research Ethics Committee (CEIC-IJG 19/116-P, 24/07/2019).

Intellectual property

The suite of PREOS-PC questionnaires is licenced by Oxford University Innovation Ltd. Free licences of the PREOS-PC questionnaire can be requested at https://innovation.ox.ac.uk/outcome-measures/patient-reported-experiences-outcomes-safety-primary-care-preos-pc/.

Consent to participate

Informed consent was obtained from all individual participants included in the study.

Consent for publication

Additional informed consent was obtained from all individual participants for whom identifying information is included in this article.

Disclaimer

The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, and approval of the manuscript; and decision to submit the manuscript for publication.

Disclosure statement

IR-C and JMV co-developed the PREOS-PC questionnaire, which is licenced by Oxford Innovation Ltd. The other authors report no conflict of interest.

Funding

This study was supported by a personal award (Miguel Servet Fellowship CP17/00017) to IR-C awarded by the Instituto de Salud Carlos III (Spanish Ministry of Sciences and Innovation) and co-funded by Fondo Europeo de Desarrollo Regional (FEDER). MJS-R and MAF-dR were funded by Instituto de Investigación Sanitaria de las Islas Baleares, grant number FOLIUM17/10 (co-funded by ITS-2017 and PO FSE 2014-2020) and FOLIUM19/05 (founded by ITS-2019-003), respectively. The payment of the open access publication fees has been supported by the 2023 LIBERI programm from IdISBa. Any grant or award did not grant the rest of authors to develop this work.

ORCID

Maria A. Fiol-deRoque http://orcid.org/0000-0001-8566-0929
José M. Valderas http://orcid.org/0000-0002-9299-1555
Jorge Arias de la Torre http://orcid.org/0000-0001-6908-9611
Maria J. Serrano-Ripoll http://orcid.org/0000-0002-1869-1132
Montserrat Gens-Barberà http://orcid.org/0000-0003-2633-9494
Encarna Sánchez-Freire http://orcid.org/0000-0002-1945-5115
Francisco M. Martín-Luján http://orcid.org/0000-0003-0359-3588
Antonio Olry de Labry http://orcid.org/0000-0001-5448-1370
Ignacio Ricci-Cabello http://orcid.org/0000-0002-4725-8274

Data availability statement

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.

References

[1] Slawomirski L, Auraaen A, Klazinga N. 2018. The economics of patient safety in primary and ambulatory care: flying blind. Paris: OECD Publishing. doi: 10.1787/baf425ad-en.
[2] Aranaz-Andrés JM, Aibar C, Limón R, et al. A study of the prevalence of adverse events in primary healthcare in Spain. Eur J Public Health. 2012;22(6):1–12. doi: 10.1093/eurpub/ckr168.
[3] Antoñanzas Villar F. Approach to the costs of non-security in the national health system. [aproximación a los costes de la no seguridad en el sistema nacional de]. Rev Esp Salud Publica. 2013;87(3):283–292. doi: 10.4321/S1135-57272013000300008.
[4] Lorincz CY, Drazen E, Sokol PE, et al. Research in ambulatory patient safety 2000–2010: A 10-year review. Chicago IL. 2011. [cited 2023 Dec 7]. https://psnet.ahrq.gov/issue/research-ambulatory-patient-safety-2000-2010-10-year-review.
[5] Spencer R, Campbell SM. Tools for primary care patient safety: a narrative review. BMC Fam Pract. 2014;15(1):166. doi: 10.1186/1471-2296-15-166.
[6] Park M, Giap TT. Patient and family engagement as a potential approach for improving patient safety: a systematic review. J Adv Nurs. 2020;76(1):62–80. doi: 10.1111/jan.14227.
[7] Vincent C, Davis R. Patients and families as safety experts. CMAJ. 2012;184(1):15–16. doi: 10.1503/cmaj.111311.
[8] Severinsson I, Holm AL. Patient's role in their own safety-a systematic review of patient involvement in safety. OJN. 2015;05(07):642–653. doi: 10.4236/ojn.2015.57068.
[9] Valderas Martinez J, Ricci-Cabello N, Prasopa-Plazier N, et al. Patient engagement: WHO technical series on safer primary care. World Health Organisation. 2016 [cited 2023 Dec 7]. https://www.who.int/publications-detail-redirect/9789241511629.
[10] Giap TT, Park M. Implementing patient and family involvement interventions for promoting patient safety: a systematic review and meta-analysis. J Patient Saf. 2021;17(2):131–140. doi: 10.1097/PTS.0000000000000714.
[11] Sheard L, Marsh C, O'Hara J, et al. Exploring how ward staff engage with the implementation of a patient safety intervention: a UK-based qualitative process evaluation. BMJ Open. 2017;7(7):e014558. doi: 10.1136/bmjopen-2016-014558.
[12] Hernan AL, Giles SJ, Beks H, et al. Patient feedback for safety improvement in primary care: results from a feasibility study. BMJ Open. 2020;10(6):e037887. doi: 10.1136/bmjopen-2020-037887.

[13] Serrano-Ripoll MJ, Ripoll J, Llobera J, et al. Development and evaluation of an intervention based on the provision of patient feedback to improve patient safety in Spanish primary healthcare centres: study protocol. BMJ Open. 2019;9(12):e031367. doi: 10.1136/bmjopen-2019-031367.
[14] Ricci-Cabello I, Pons-Vigues M, Berenguera A, et al. Patients' perceptions and experiences of patient safety in primary care in England. Fam Pract. 2016;33(5):535–542. doi: 10.1093/fampra/cmw046.
[15] Ricci-Cabello I, Goncalves DC, Campbell S, et al. Patients' experiences of patient safety in primary care in England: a systematic review and meta-synthesis. 42 North American Primary Care Research Group (NAPCRG) annual meeting, New York. 2014.
[16] Ricci-Cabello I, Goncalves DC, Rojas-Garcia A, et al. Measuring experiences and outcomes of patient safety in primary care: a systematic review of available instruments. Fam Pract. 2015;32(1):106–119. doi: 10.1093/fampra/cmu052.
[17] Ricci-Cabello I, Reeves D, Bell BG, et al. Identifying patient and practice characteristics associated with patient-reported experiences of safety problems and harm: a cross-sectional study using a multilevel modelling approach. BMJ Qual Saf. 2017;26(11):899–907. doi: 10.1136/bmjqs-2016-006411.
[18] Litchfield I, Gill P, Avery T, et al. Influences on the adoption of patient safety innovation in primary care: a qualitative exploration of staff perspectives. BMC Fam Pract. 2018;19(1):72. doi: 10.1186/s12875-018-0761-2.
[19] Mounce LTA, Salema NE, Gangannagaripalli J, et al. Development of 2 short patient-report questionnaires of patient. J Patient Saf. 2022;18(3):161–170. doi: 10.1097/PTS.0000000000000880.
[20] Serrano-Ripoll MJ, Llobera J, Valderas JM, et al. Cross-cultural adaptation, validation, and piloting of the patient reported experiences and outcomes of safety in primary care questionnaire for its use in Spain. J Patient Saf. 2022;18(2):102–110. doi: 10.1097/PTS.0000000000000819.
[21] Fiol-deRoque MA, Serrano-Ripol MJ, Gens-Barbera M, et al. Impact of the COVID-19 pandemic on patient-reported patient safety in primary care. Aten Primaria. 2021;53(Suppl 1):102222. doi: 10.1016/j.aprim.2021.102222.
[22] Monterde D, Vela E, Clèries M. Adjusted morbidity groups: a new multiple morbidity measurement of use in primary care. Aten Primaria. 2016;48(10):674–682. doi: 10.1016/j.aprim.2016.06.003.
[23] Prieto-Lara E, Ocaña-Riola R. Updating rurality index for small areas in Spain. Soc Indic Res. 2010;95(2):267–280. doi: 10.1007/s11205-009-9459-0.
[24] Martínez-Lizaga N, Montes Y, Rodrigo I, et al. Metodología del atlas de variaciones en hospitalizaciones potencialmente evitables en el sistema nacional de salud. Atlas Var Pract Med Sist Nac Salud. 2011;4(2):371–378.
[25] Torijano-Casalengua ML, Olivera-Cañadas G, Astier-Peña MP, et al. Validation of a questionnaire to assess patient safety culture in spanish primary health care professionals. Aten Primaria. 2013;45(1):21–37. doi: 10.1016/j.aprim.2012.07.003.
[26] Hernandez Vidal N, Satue Gracia EM, Basora GJ, et al. Translation, adaptation and validation in catalan of a questionnaire about patient safety culture: the MOSPSC (medical office survey on patient safety culture) questionnaire. J Health Qual Res. 2019;34(5):248–257. doi: 10.1016/j.jhqr.2019.05.012.
[27] Sepucha KR, Matlock DD, Wills CE, et al. It's valid and reliable is not enough: critical appraisal of reporting of measures in trials evaluating patient decision aids. Med Decis Making. 2014;34(5):560–566. doi: 10.1177/0272989X14528381.
[28] Ricci-Cabello I, Avery AJ, Reeves D, et al. Measuring patient safety in primary care: the development and validation of the "patient reported experiences and outcomes of safety in primary care" (PREOS-PC). Ann Fam Med. 2016;14(3):253–261. doi: 10.1370/afm.1935.
[29] Streiner DL, Norman GR. Health measurement scales: a practical guide to their development and use. 5th ed. Oxford: Oxford Academic; 2015. [cited 2023 Dec 5]. doi: 10.1093/med/9780199685219.001.0001.
[30] Nunnally JC, Berstein IH. Psychometric theory. (3rd ed.). New York: McGraw-Hill; 1994.
[31] Valderas JM, Ferrer M, Mendívil J, et al. Development of EMPRO: a tool for the standardized assessment of patient-reported outcome measures. Value Health. 2008;11(4):700–708. doi: 10.1111/j.1524-4733.2007.00309.x.
[32] Hu LT, Bentler PM. Fit indices in covariance structure modeling: sensitivity to underparameterized model misspecification. Psychological Methods. 1998;3(4):424–453. doi: 10.1037/1082-989X.3.4.424.
[33] Netemeyer RG, Bearden WO, Sharma S. Scaling procedures: issues and applications. Thousand Oaks (CA): Sage Publications; 2003. doi: 10.4135/9781412985772.
[34] Nguyen TH, Han HR, Kim MT, et al. An introduction to item response theory for patient-reported outcome measurement. Patient. 2014;7(1):23–35. doi: 10.1007/s40271-013-0041-0.
[35] Linacre J. A user's guide to Winsteps. 450–452. 2023 [cited 2023 Nov 13]. https://www.winsteps.com/a/Winsteps-Manual.pdf
[36] Wright B, Linacre J. Reasonable mean-square fit values. RMT. 1994;8(3):370.
[37] Wright B, Stone M. Measurement essentials. Wilmington (DE): Wide Range, Inc; 1999.
[38] Giles SJ, Parveen S, Hernan AL. Validation of the primary care patient measure of safety (PC PMOS) questionnaire. BMJ Qual Saf. 2019;28(5):389–396. doi: 10.1136/bmjqs-2018-007988.
[39] Stahl K, Reisinger A, Groene O. Assessing patient experience with patient safety in primary care: development and validation of the ASK-ME-questionnaire. BMJ Open. 2022;12(4):e049237. doi: 10.1136/bmjopen-2021-049237.
[40] Groves RM. Nonresponse rates and nonresponse bias in household surveys. J Public Opinion Quarterly. 2006;70(5):646–675. doi: 10.1093/poq/nfl033.