Validation of Instruments to Evaluate Primary Healthcare from the Patient Perspective: Overview of the Method

Validation des instruments d'évaluation des soins primaires du point de vue des patients : aperçu de la méthode

JEANNIE L. HAGGERTY, PHD
Department of Family Medicine, McGill University
Montreal, QC

FREDERICK BURGE, MD, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

MARIE-DOMINIQUE BEAULIEU, MD, MSC
Chaire Dr Sadok Besrour en médecine familiale
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

RAYNALD PINEAULT, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

CHRISTINE BEAULIEU, MSC
St. Mary's Research Centre, St. Mary's Hospital Center
Montreal, QC

JEAN-FRÉDÉRIC LÉVESQUE, MD, PHD
Centre de recherche du Centre hospitalier de l'Université de Montréal
Montréal, QC

DARCY A. SANTOR, PHD
School of Psychology, University of Ottawa
Ottawa, ON

Jeannie L. Haggerty et al.

DAVID GASS, MD
Department of Family Medicine, Dalhousie University
Halifax, NS

BEVERLEY LAWSON, MSC
Department of Family Medicine, Dalhousie University
Halifax, NS

Abstract

Patient evaluations are an important part of monitoring primary healthcare reforms, but there is little comparative information available to guide evaluators in the choice of instruments or to determine their relevance for Canada.

Objective: To compare values and the psychometric performances of validated instruments thought to be most pertinent to the Canadian context for evaluating core attributes of primary healthcare.

Method: Among validated instruments in the public domain, we selected six: the Primary Care Assessment Survey (PCAS); the Primary Care Assessment Tool – Short Form (PCAT-S); the Components of Primary Care Index (CPCI); the first version of the EUROPEP (EUROPEP-I); the Interpersonal Processes of Care Survey, version II (IPC-II); and part of the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS). We mapped subscales to operational definitions of attributes. All were administered to a sample of adult service users balanced by English/French language (in Nova Scotia and Quebec, respectively), urban/rural residency, high/low education and overall care experience. The sample was recruited from previous survey respondents, newspaper advertisements and community posters. We used common factor analysis to compare our factor resolution for each instrument to that of the developers.

Results: Our sample of 645 respondents was approximately balanced by design variables, but considerable effort was required to recruit low-education and poor-experience respondents. Subscale scores are statistically different by excellent, average and poor overall experience, but interpersonal communication and respectfulness scores were the most discriminating of overall experience. We found fewer factors than did the developers, but when constrained to the number of expected factors, our item loadings were largely similar to those found by developers. Subscale reliability was equivalent to or higher than that reported by developers.

Conclusion: These instruments perform similarly in the Canadian context to their original development context, and can be used with confidence. Interpersonal and respectfulness scores are most discriminating of excellent, average or poor overall experience and are crucial dimensions of patient evaluations.

Résumé

L'évaluation des soins de santé primaires par les patients constituent un aspect important du suivi des réformes des soins de santé primaires, mais il existe peu d'information comparative pour orienter les évaluateurs dans le choix d'instruments ou pour déterminer leur pertinence au Canada.

Objectif : Comparer la valeur et la performance psychométrique des instruments valides et considérés comme les plus pertinents au contexte canadien pour l'évaluation des caractéristiques centrales des soins de santé primaires.

Méthode : Nous avons choisi six instruments validés du domaine public, soit : Primary Care Assessment Survey (PCAS); Primary Care Assessment Tool – version courte (PCAT-S); Components of Primary Care Index (CPCI); la première version de l'EUROPEP (EUROPEP-I); Interpersonal Processes of Care Survey, version II (IPC-II); et une partie du Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS). Les sous-échelles ont été mises en relation avec les définitions opérationnelles des caractéristiques. Tous ces instruments ont été appliqués a un échantillon d'adultes utilisateurs de services qui était équilibré en fonction de la langue (anglais : Nouvelle-Écosse ou français : Québec), du lieu de résidence urbain ou rural, du niveau de scolarité élevé ou faible et de l'expérience générale des soins. L'échantillon provenait de répondants recrutés grâce à un sondage antérieur, à des annonces dans les journaux et à des affiches dans des lieux communautaires. Nous avons procédé a une analyse factorielle normale pour comparer nos structures factorielles à celles des concepteurs, et ce, pour chacun des instruments.

Résultats : L'échantillon de 645 répondants était globalement équilibré en fonction des paramètres du devis, mais des efforts considérables ont été nécessaires pour trouver des répondants avec un faible niveau de scolarité et une faible expérience des soins. Les scores des sous-échelles sont statistiquement différents en fonction d'une expérience générale excellente, moyenne ou faible, mais les scores présentant la discrimination la plus forte pour l'expérience générale étaient ceux de la communication interpersonnelleé. Nous avons observé moins de facteurs que ne l'avaient fait les concepteurs, mais en forçant le nombre de facteurs au nombre attendu, la correspondance de nos items était très similaire à celle observée par les concepteurs. La fiabilité des sous-échelles était équivalente ou plus élevée que celle indiquée par les concepteurs.

Conclusion : Le rendement de ces instruments dans le contexte canadien est similaire à celui obtenu de leur contexte original, et ils peuvent être utilisés en toute confiance. Les scores de la communication interpersonnelle et du respect sont les plus discriminants d'une expérience générale excellente, moyenne ou faible, et ils constituent des aspects importants de l'évaluation par les patients.

———

INSTRUMENTS TO EVALUATE CARE FROM THE PATIENT'S PERSPECTIVE HAVE BEEN developed and validated elsewhere, but no comparative information is available on their performance in the Canadian context to guide researchers and policy makers in selecting one instrument over another. Our objective was to compare validated instruments thought to be most pertinent to the Canadian context. Specifically, we aimed to compare subscales from different instruments for the same attribute of care and to ensure that the instruments' reported psychometric properties were similar in the Canadian context. Program evaluators

could then be confident of the tools' applicability to that context, and if different instruments were used, either at different times or in different jurisdictions, our results would provide a common benchmark for comparing relative scores. In this paper, we report in detail on how we selected the instruments, identified and recruited the study sample and administered the instruments. We provide general descriptive results and compare psychometric properties with those reported by the instrument developers.

Method

Ethical approval for this study was obtained from the Research Centre of the Université de Montréal Hospital (Quebec) and the Capital Health Research Ethics Board (Nova Scotia).

Identification and selection of instruments

We conducted an electronic search of the Medline and CINAHL databases in spring 2004 using as keywords "primary healthcare," "outcome and process assessment," "questionnaires" and "psychometrics." From identified instruments, we eliminated those used to screen for illnesses, functional health status or perceived outcomes of care for specific conditions (migraines, mental healthcare). We identified additional instruments by consulting with colleagues

TABLE 1. Subscales selected from six instruments retained for the study and their correspondence to attributes of PHC, in the order used in the study questionnaire, showing subscale as named by the instrument developer (number of items shown in parentheses). The last row names the scales excluded from this study.

Attribute of Care to which Subscale was Mapped	Primary Care Assessment Survey **PCAS**	Primary Care Assessment Tool **PCAT**	**EUROPEP**	Components of Primary Care Index **CPCI**	Interpersonal Processes of Care **IPC**	Veterans Affairs National Outpatient Community Services Survey **VANOCSS**
Accessibility	Organizational access (6)	First-contact access (4); First-contact utilization (3)	Organization of care (7)			
Relational Continuity	Contextual knowledge of patient (5); Visit-based continuity (2)	Ongoing care (4)		Accumulated knowledge (8); Preference for regular physician (5)		
Interpersonal Communication	Communication (6); Trust (8)		Clinical behaviour (16)	Interpersonal communication (6)	Elicitation, responsiveness, explanations (6); Patient-centred decision-making (4)	
Respectfulness	Interpersonal treatment (5)				Emotional support (4); Non-hurried, attentive (6); Perceived discrimination (4); Respectfulness (4); Respectfulness of office staff (4)	
Comprehensiveness of Services		Comprehensiveness (services available) (4)		Comprehensive care (6)		

TABLE 1. Continued

Attribute of Care to which Subscale was Mapped	Primary Care Assessment Survey PCAS	Primary Care Assessment Tool PCAT	EUROPEP	Components of Primary Care Index CPCI	Interpersonal Processes of Care IPC	Veterans Affairs National Outpatient Community Services Survey VANOCSS
Whole-Person Care (Community-Oriented Care)		Community orientation (3)		Community context (2)		
Management Continuity (Coordination)	Integration (6)	Coordination (4)		Coordination of care (8)		Overall coordination of care (6); Specialty provider access (4)
Subscales excluded from the Study	**Financial access (2); Physical examination (1); Preventive counselling (7)**	**Culturally competent (3); Coordination (information systems) (3); Family-centredness (3)**		**Advocacy (9); Family context (3)**	**Cultural sensitivity (2); Doctor's sensitivity to language (3); Office staff's sensitivity to language (2); Empowerment (3); Explain medications (2); Self-care (2)**	**Visit-based scales: Access/timeliness (7); Coordination of care at visit (5); Courtesy (2); Emotional support (4); Patient education information (7); Preferences (5)**

and scanning reference lists in published papers. When several instruments were derived from or inspired by a common instrument, for example, the General Practice Assessment Questionnaire derived from the Primary Care Assessment Survey, we retained only the parent instrument. We identified 13 unique validated instruments, on which we then obtained psychometric information from available publications or from the instrument developers.

Three instruments were visit-based, and the other 10 were retrospective, addressing usual care. We excluded the visit-based instruments and one that focused exclusively on satisfaction with all healthcare received – the Patient Satisfaction Questionnaire (PSQ-18) (Marshall and Hays 1994). Each researcher independently ranked the remaining nine instruments according to their relevance in the Canadian context, and we retained the six highest-ranked instruments: the Primary Care Assessment Survey (PCAS) (Safran et al. 1998); the Primary Care Assessment Tool – Short Form (PCAT-S) (Shi et al. 2001); the Components of Primary Care Index (CPCI) (Flocke 1997); the first version of the European general practice evaluation instrument EUROPEP-I (Grol et al. 2000); the Interpersonal Processes of Care Survey – 18-item version (IPC-II) (Stewart et al. 2007); and the Veterans Affairs National Outpatient Customer Satisfaction Survey (VANOCSS) (Borowsky et al. 2002). Permission to use the instruments was obtained from all instrument developers.

Because our objective was to compare measures by attribute of care, we further retained only the subscales of attributes measured in more than one instrument. For example, we dropped the Advocacy subscale from the CPCI because it is measured only in this instrument. The six instruments and the subscales retained for this study are listed in Table 1.

Study population

Our target population was English- and French-speaking adult primary healthcare (PHC) users in Canada, undifferentiated by age, health condition, geographic location or level of functional literacy. Eligible subjects were adults (≥18 years) with a regular source of PHC that they had consulted in the previous 12 months. We maximized the statistical efficiency for conducting subgroup comparisons by balancing our study sample by English/French language, urban/rural location and educational level. We also stratified by excellent, average and poor primary care experience based on a single screening question: "Overall, has your experience of care from your regular family doctor or medical clinic been excellent, poor or average?" The sample size was designed to provide statistical power for factor analysis of up to 150 items with 25 subjects in each sampling cell.

We used educational achievement as a proxy for literacy above and below high-school reading level. Because the association between literacy and education varies as a function of age, we used an age-sensitive cut-off for high school: completed high school, if under age 45; completed 10 years, for ages 45 to 55; and less than eight years, if over age 55 (Smith and Haggerty 2003). Urban location was defined as residing in a census metropolitan area; rural, in areas more than one hour's travel from a metropolitan area; and remote (Quebec only), in areas more than four hours' travel from the nearest metropolitan area. Questionnaires were administered exclusively in English in Nova Scotia and in French in Quebec.

Subjects were recruited by various means. Our goal was to achieve representativeness of the sampling strata, not of the population as a whole. We initially used a sampling frame of persons from previous PHC surveys who had agreed to future contact: 647 from a 2002 clinic-based survey in Quebec (Haggerty et al. 2007) and 1,247 from a 2005 telephone survey in Nova Scotia. Eligibility for different strata was determined from screening questions administered by telephone or e-mail.

Owing to difficulties in recruiting low-literacy participants and those with poor experience of care, we obtained ethical approval in Quebec only to expand recruitment strategies to newspaper advertisements, then community posters in laundromats, grocery stores, recreation centres and health centres and, finally, word of mouth. All participants were offered compensation for completing the questionnaire.

Data collection

The study questionnaire consisted of the retained subscales from the six selected instruments (153 items, 28 specific to care from multiple providers), as well as socio-demographic and utilization information (total, 198 questions). Instruments were formatted in their original form, and subscales were grouped by instrument family in the sequence shown in Table 1. The VANOCSS was placed last because it was specific to those who had seen multiple providers.

Participants were offered either paper-based or online response modalities. To maximize response, we used a protocol of two reminder postcards or e-mails at two-week intervals, followed by a second posting of the questionnaire, then phone calls. A subset of participants

completed the questionnaire in a group setting where they could be observed directly and then participated in a 30- to 45-minute discussion; these results are reported elsewhere (Haggerty, Beaulieu et al. 2011). Data were collected between February and July 2005.

Analysis

We analyzed our recruitment descriptively by substrata in terms of the success of different recruitment strategies and differential response rates.

The details of individual subscales are presented in the attribute-specific papers elsewhere in this special issue of the journal. We expressed the value of each subscale as the mean of the values of the items. Thus, the mean of several items with a 1-to-5 Likert response scale varied between 1 and 5 regardless of the number of items in the subscale. We calculated the internal reliability of each subscale using Cronbach's alpha.

We used one-way ANOVA tests to determine whether subscale means differed significantly among respondents with poor, average and excellent overall experience of care; we used discriminant analysis to examine the magnitude of the Fisher linear discriminant (F-test) as an indicator of which subscale score best differentiates between the groups. We also conducted exploratory factor analysis using common factor analysis for each instrument, to determine how many factors emerged with an eigenvalue >1. We repeated the analysis forcing the number of factors found by instrument developers, then examined whether item loading accorded with that identified by the developer. Factor analysis used only observations with no missing values on any item (listwise missing), but we repeated the analyses, imputing for missing values by using either maximum likelihood within the subscale (Jöreskog and Sörbom 1996) or the developer's suggested imputation algorithm.

Results

Recruitment of study population

Of the 647 Quebec residents in the initial sampling frame, the first 208 who met the eligibility criteria for specific strata were selected for telephone contact; 168 had still-active telephone numbers, and 38% (62/168) agreed to participate. Of these, 85% (53/62) returned the questionnaire. Of the 1,247 persons in Nova Scotia, 290 had provided e-mail addresses and were invited by e-mail without being pre-screened, of which 112 (38.5%) responded to the questionnaire. The final overall response rates were similar. While the telephone strategy (Quebec) was more resource-intensive, the resulting sample corresponded more closely to the desired design; the e-mail strategy (Nova Scotia) oversampled high-education respondents (91% vs. the 50% desired).

We had difficulty recruiting eligible subjects with low education and/or poor experience of care from their regular provider. Advertising in local newspapers (Quebec only) was most cost-effective in urban areas. Posters in laundromats, grocery stores, community recreation centres and credit unions were effective for reaching low-education participants in urban areas, but not rural areas. This method required few resources but provided a steady trickle

of responses. In both provinces, peer recruitment by word of mouth (snowball sampling) was the most effective strategy for targeted recruitment in rural areas and among people with low educational attainment.

Table 2 presents the final sample size and distribution by sampling design variables. The sample distribution was more balanced in the design variables in Quebec (French) than in

TABLE 2. Final recruitment of study subjects by design variables; original aim was for 25 subjects per cell

Prior Experience with Primary Care	French (n=302, 46%)				English (n=343, 53%)				Total
	Urban (n=148, 49%)		Rural (n=154, 51%)		Urban (n=203, 59%)		Rural (n=140, 41%)		
	Low Education	High Education	Low Education	High Education	Low Education	High Education	Low Education	High Education	
Excellent	31	31	28	32	24	66	11	41	264 (41%)
Average	22	31	28	31	14	57	11	39	233 (36%)
Poor	9	24	17	18	10	32	16	22	148 (23%)
Total	62 (21%)	86 (28%)	73 (24%)	81 (27%)	48 (14%)	155 (45%)	38 (11%)	102 (30%)	

Nova Scotia (English). Additionally, the Nova Scotia sample was in better health than the Quebec sample and more likely to be affiliated to a family doctor and for a longer time, to concentrate care among fewer unique family physicians and to have shorter waits for care (details presented elsewhere in this special issue of the journal, Haggerty, Bouharaoui et al. 2011).

Of the 645 respondents, 130 (20.2%) responded to the online version of the questionnaire: 25% in urban areas and 14% in rural areas ($\chi^2=11.6$, $p=.0007$). Of the high-education participants, 26.9% responded online, compared to 7.2% of low-education participants ($\chi^2=34.9$, $p<.0001$). There was no difference in subscale scores by response modality after controlling for language, geographic location and educational status.

Table 3 presents the sample characteristics and compares them with respect to their reported overall experience of care. Compared to those participants with just average or poor experience, those with excellent experience are more likely to be in better health, to be affiliated with a physician rather than a clinic (and with longer affiliations), to have seen fewer unique physicians in the year and to report shorter waits for appointments.

TABLE 3. Characteristics of the study sample and comparison of subjects by overall experience of care

Characteristic	Total (n=645)	Overall Experience of Care			Test for Difference
		Excellent (n=264)	Average (n=232)	Poor (n=149)	
Personal Characteristics					
Average age	48.0 (14.9)	48.4 (14.9)	47.6 (14.3)	47.8 (15.8)	NS
Per cent female	64.6 (414)	63.7 (167)	65.8 (152)	64.6 (95)	NS
Per cent indicating health status as good or excellent	37.8 (241)	43.0 (113)	37.3 (85)	29.9 (43)	$x^2=6.9$; *df* 2 $p=.03$

TABLE 3. Continued

| | Overall Experience of Care | | | | |
Characteristic	Total (n=645)	Excellent (n=264)	Average (n=232)	Poor (n=149)	Test for Difference
Per cent with disability	31.6 (200)	29.5 (77)	32.4 (73)	33.8 (50)	NS
Per cent with chronic health problem*	61.6 (392)	61.1 (160)	60.5 (138)	64.4 (94)	NS
Healthcare Use					
Regular provider: Physician	94.1 (607)	97.4 (257)	92.7 (215)	90.6 (135)	x^2=9.2; df 2
Clinic only	5.9 (38)	2.7 (7)	7.3 (17)	9.4 (14)	p=.01
Mean number of years of affiliation	11.2 (9.0)	11.9 (10)	11.3 (8.5)	9.7 (7.8)	NS
Mean number of primary care visits in last 12 months	6.3 (7.0)	7.1 (8.3)	4.9 (4.6)	7.1 (7.3)	F=6.9; df 2 p=.001
Mean number of unique general or family physicians seen	2.0 (1.3)	1.8 (1.1)	2.0 (1.5)	2.3 (1.5)	F=8.3; df 2 p=.003
Usual wait time for appointment Less than 2 days 2 to 7 days 7 days to 2 weeks 2 weeks to 4 weeks More than 4 weeks	35.2 (220) 32.6 (204) 11.8 (74) 9.3 (58) 11.0 (69)	47.3 (123) 28.5 (74) 9.2 (24) 5.8 (15) 9.2 (24)	30.6 (68) 37.4 (83) 9.0 (20) 11.7 (26) 11.3 (25)	20.3 (29) 32.9 (47) 21.0 (30) 11.9 (17) 14.0 (20)	x^2=45; df 8 p<.0001
Usual wait time in waiting room before clinical visit Less than 15 minutes 15 to 29 minutes 30 to 59 minutes More than an hour	34.7 (218) 38.8 (244) 19.9 (125) 6.7 (42)	37.6 (99) 39.2 (103) 19.0 (50) 4.2 (11)	38.7 (87) 35.6 (80) 18.7 (42) 7.1 (16)	22.7 (32) 43.3 (61) 23.4 (33) 10.6 (15)	x^2=15.8; df 6 p=.02

* Percentage indicating they had been told by a doctor that they had any of the following: high blood pressure, diabetes, cancer, depression, arthritis, respiratory disease, heart disease.

Comparison of instrument scores

The subscale scores grouped by PHC attribute are presented in Table 4. Several points are noteworthy. First, with few exceptions, the score distributions are skewed; the median is higher than the mean, indicating that the mass of the distribution is in the positive zone of assessment. Second, the back-to-back placement of scores demonstrates the challenge of comparing scores even within the same group of respondents, let alone between groups or jurisdictions. Third, the subscale means differ significantly by levels of overall experience, as shown in the last two columns of Table 4. All subscales scores, except the VANOCSS Specialty Provider Access, distinguish between poor and excellent care; the vast majority, between poor and average and between average and excellent care. The magnitude of the Fisher test shows that subscales for interpersonal communication and respectfulness provide the greatest discrimination among poor, average and excellent overall experience of care. Average Fisher test values are 66.5 and 55.8, respectively, compared to average values in the 20s and 30s for other attribute families.

Psychometric properties

In Table 5, the subscales are grouped within their instrument families. Note that the Cronbach's alphas reported by the developers are similar to those observed in our sample. For factor analysis, with the exception of the EUROPEP, the number of factors observed

by common factor analysis was approximately half that expected (item loading available on request). When we constrained the factor resolution to the number of factors found by the instrument developer, the item loading corresponded generally to that identified by the developer. The observed factor solutions deviated most from the expected for the CPCI and PCAT-S instruments. The deviation for the CPCI may be explained by halo effects related to the instrument's format and response scale, and for the PCAT-S, by problems related to missing values – a case that merits additional explanation.

TABLE 4. Subscale values, grouped by attribute of care, showing comparison of statistically significant differences in mean values by overall experience of care

| Instrument | Developer's Subscale Name | # Items | Likert Response Range | Raw Values | | | Heathcare Experience* | *F*-Test of Discrimination |
				Mean	Median	SD	Poor Average Excellent	
Accessibility								
PCAS	Organizational Access	6	1 to 6	3.97	4.00	0.92	4,85 5,73 6,74	62.6
PCAT	First-Contact Accessibility	4	1 to 4	2.68	2.75	0.78	4,51 5,32 6,45	11.8
PCAT	First-Contact Utilization	3	1 to 4	3.73	4.00	0.48	8,71 8,94 9,44	13.8
EUROPEP	Organization of Care	7	1 to 5	3.61	3.71	0.90	5,01 6,27 7,59	29.9
Comprehensiveness								
PCAT	Comprehensiveness (services available)	4	1 to 4	3.32	3.50	0.74	7,35 7,52 8,08	4.9
CPCI	Comprehensive Care	6	1 to 6	4.86	5.00	1.10	6,73 7,54 8,43	35.5
Interpersonal Communication								
PCAS	Communication	6	1 to 6	4.66	4.83	1.05	5,85 6,98 8,45	101.4
PCAS	Trust	8	1 to 5	4.01	4.13	0.71	6,28 7,28 8,46	96.9
CPCI	Interpersonal Communication	6	1 to 6	4.59	4.83	1.16	5,87 6,90 8,16	57.2
EUROPEP	Clinical Behaviour	16	1 to 5	4.14	4.33	0.83	6,48 7,61 8,86	79.9

TABLE 4. Continued

Instrument	Developer's Subscale Name	# Items	Likert Response Range	Raw Values			Heathcare Experience*	F-Test of Discrimination
				Mean	Median	SD	Poor Average Excellent	
IPC-II	Communication (elicited concerns, responded)	3	1 to 5	4.12	4.33	0.87	6,55 7,54 8,77	61.2
IPC-II	Communication (explained results, medications)	4	1 to 5	3.96	4.25	1.00	6,24 7,03 8,39	42.9
IPC-II	Decision-Making (patient-centred decision-making)	4	1 to 5	3.17	3.25	1.26	4,71 4,89 6,30	16.8
Management Continuity								
PCAS	Integration	6	1 to 6	4.45	4.67	1.00	5,74 6,64 7,80	42.7
PCAT	Coordination	4	1 to 4	3.27	3.50	0.80	6,61 7,38 8,30	18.0
CPCI	Coordination of Care	8	1 to 6	4.30	4.38	1.00	5,80 6,27 7,34	34.4
VANOCSS	Coordination of Care (overall), number of problems	6	0 to 6	2.51	2.00	1.88	5,05 5,66 6,47	4.7
VANOCSS	Specialty Provider Access (number of problems)	4	0 to 4	0.62	0.00	0.90	–	
Relational Continuity								
PCAS	Visit-Based Continuity	2	1 to 6	5.17	5.50	1.05	7,54 8,44 8,72	15.5
PCAS	Contextual Knowledge	5	1 to 6	3.96	4.10	1.14	4,67 5,55 6,93	60.5
PCAT	Ongoing Care	4	1 to 4	3.15	3.25	0.70	5,94 6,89 8,06	46.1
CPCI	Accumulated Knowledge	8	1 to 6	4.50	4.75	1.24	5,84 6,58 7,99	45.1
CPCI	Patient Preference for Regular Physician	5	1 to 6	4.84	5.00	1.00	6,86 7,53 8,27	29.3
Respectfulness								
PCAS	Interpersonal Treatment	5	1 to 6	4.72	4.90	1.08	5,90 7,14 8,56	97.3
IPC-II	Hurried Communication	5	1 to 5	4.20	4.37	0.71	7,01 7,90 8,68	48.1

TABLE 4. Continued

| Instrument | Developer's Subscale Name | # Items | Likert Response Range | Raw Values | | | Heathcare Experience* | F-Test of Discrimination |
				Mean	Median	SD	Poor Average Excellent	
IPC-II	Interpersonal Style (compassionate, respectful)	5	1 to 5	4.21	4.60	0.9	6,57 7,86 9,01	60.4
IPC-II	Interpersonal Style (respectful office staff**)	4	1 to 5	4.51	5.00	0.73	8,05 8,84 9,14	17.4
Whole-Person Care – Community Context								
PCAT	Community Orientation	3	1 to 4	2.47	2.50	0.86	3,75 4,60 5,79	24.8
CPCI	Community Context	2	1 to 6	4.23	4.50	1.56	5,07 6,21 7,46	30.2

* Means by group only presented where difference statistically significant at $p<.01$.
** Subscale reversed as well as normalized; raw value indicates frequency of disrespectful behaviour. Consequently, the normalized score of 10 = never disrespectful, 0 = always disrespectful.

The PCAT-S offers five response options to desirable characteristics in PHC: 1 = definitely not; 2 = probably not; 3 = probably; 4 = definitely, and "don't know/not sure." Processed classically, the "don't know" response counts as a missing value, yielding us only 146/645 valid observations for factor analysis. The developer suggests replacing this latter response with a value of 2 (probably not) for respondents with at least 50% true values within the subscale, based on the logic that when patients are unsure of service attributes at the clinic, this reflects negatively on the provider. Using the developer's replacement algorithm yielded 470 observations, and the factor resolution corresponded more closely to that of the developer, although the grouping of items in factors 3 and 6 (Table 5) persisted, suggesting a construct overlap between first-contact accessibility and community orientation, and between first-contact accessibility and ongoing care (details available on request).

TABLE 5. Reported and observed internal consistency (Cronbach alpha) and factor resolution by instrument, showing observed factors with eigenvalue >1 and factor solution when constrained to expected number

Instrument and Subscale (number items)	Mapped Attributes	Reported Alpha	Observed Alpha	Solution of Expected Number of Factors (eigen) Subscales
Primary Care Assessment Survey (PCAS)				*Expected=6, Observed=4 (n=377)*
Organizational Access (6)	Accessibility	.84	.83	(17.45) Communication + Interpersonal Treatment
Visit-Based Continuity (2)	Relational Continuity	—	.69	(1.98) Contextual Knowledge
Contextual Knowledge (5)	Relational Continuity	.92	.91	(1.48) Integration

TABLE 5. Continued

Instrument and Subscale (number items)	Mapped Attributes	Reported Alpha	Observed Alpha	Solution of Expected Number of Factors (eigen) Subscales
Communication (6)	Interpersonal Communication	.95	.96	(1.06) Organizational Access
Trust (8)	Interpersonal Communication	.86	.88	(0.90) 4/8 Trust
Interpersonal Treatment (5)	Respectfulness	.95	.96	(0.65) 4/8 Trust
Integration (6)	Management Continuity	.92	.93	(0.51) Visit-based continuity
Primary Care Assessment Tool (PCAT)				*Expected=6, Proposed=3 (n=470)*
First-Contact Utilization (3)	Accessibility / Comprehensiveness	TBD	.68	(5.01) Coordination
First-Contact Access	Accessibility		.72	(1.40) 3/4 Ongoing Care
Comprehensiveness (services available) (4)	Comprehensiveness of Services		.72	(0.86) Comprehensive Services
Ongoing Care (4)	Relational Continuity		.73	(0.63) 2/4 First-Contact Access + 1/4 Ongoing Care (telephone)
Coordination (4)	Management Continuity		.76	(0.51) Community Orientation + 2/4 First-Contact Access
Community Orientation (3)	Whole-Person Care		.65	(0.40) First-Contact Utilization
Components of Primary Care Instrument (CPCI)				*Expected=6, Proposed=3 (N=487)*
Comprehensive Care (6)	Comprehensiveness of Services	.79	.83	(13.75) Community Context + 6/8 Coordination + 1/5 Preference
Accumulated Knowledge (8)	Relational Continuity	.88	.91	(1.29) 7/8 Accumulated Knowledge + 1/6 Communication
Preference for Regular Physician (5)	Relational Continuity	.71	.68	(1.15) 5/6 Comprehensive
Interpersonal Communication (6)	Interpersonal Communication	.75	.83	(0.93) 5/6 Communication
Coordination of Care (8)	Management Continuity	.92	.74	(0.85) 4/5 Preference + 1/8 Coordination
Community Context (2)	Whole-Person Care	—	.82	(0.51) 2/8 Coordination
EUROPEP				*Expected=2, Proposed=2 (n=355)*
Organization of Care (7)	Accessibility	.87	.89	(13.62) Clinical Behaviour
Clinical Behaviour (16)	Interpersonal Communication	.96	.97	(1.56) Organization of Care
Interpersonal Processes of Care (IPC-II)				*Expected=6, Proposed=3 (n=536)*
Elicit Concerns, Respond (3)	Interpersonal Communication	.80	.86	(11.92) Compassionate + (3/5) Non-Hurried, Attentive
Explain Results, Medications (4)	Interpersonal Communication	.81	.88	(2.61) Decision-Making
Decision-Making (4)	Interpersonal Communication	.75	.91	(1.36) Respectful Office Staff
Non-Hurried, Attentive (5)	Respectfulness	.65	.95	(0.79) Explain Results
Compassionate, Respectful (5)	Respectfulness	.71	.95	(0.57) Non-Hurried, Attentive (3/5 load equally with factor 1)
Respectful Office Staff (5)	Respectfulness	.90	.93	(0.39) Elicit Concerns
Veterans Affairs Outpatient Community Services Survey				*NB: Dichotomous scoring of items, factor analysis not applicable*
Management Continuity (6)	Overall Coordination of Care		NA	
Management Continuity (4)	Specialty Provider Access		NA	

Discussion and Conclusion

We found that relevant subscales from generic PHC evaluation instruments demonstrate general psychometric properties in a Canadian sample that are similar to those observed in the United States and Europe, where the instruments were developed. Despite important differences in PHC organization among countries, our results suggest that Canadian program evaluators and researchers can confidently rely on the reported psychometric properties of these instruments for evaluating PHC attributes.

Almost all the subscales demonstrate skewed distribution, regardless of whether the response type is reporting or rating. We would expect the skewing to be even more extreme in a representative sample of the population that was not selected to balance the sample by overall experience of care, as ours was. This skewing has been demonstrated consistently in other studies (Crow et al. 2002) and is a major challenge in program evaluation. Qualitative studies suggest that patients are reluctant to report negative assessments of care even when not entirely satisfied, unless clear responsibility can be attributed to the source of the negative experience (Collins and O'Cathain 2003). This means that positive assessments will reflect a mix of experiences ranging from only adequate to excellent, and therefore have low sensitivity and specificity. Negative assessments, by contrast, tend to be true negatives, indicating good specificity of negative scores. Thus, in reports to decision-makers about PHC performance, it may be more informative and accurate to report the percentage of less-than-positive scores, rather than masking the negative scores within generally positive average scores.

Our recruitment experience illustrates the difficulty of including low-literacy subjects in surveys of healthcare experience. These patients are not reached easily by written material. Yet their participation in evaluations is important, because low literacy is an independent health risk (Smith and Haggerty 2003), and subjects will be more dependent than high-literacy subjects on their doctors' actions and advice (Bostick et al. 1994; Fiscella et al. 2002; Breitkopf et al. 2004; Willems et al. 2005). We found that for the most part, these instruments function equivalently in low-literacy and high-literacy responders (Haggerty, Bouharaoui et al. 2011), further highlighting the importance of reaching these patient groups.

All the instrument subscales distinguish between different levels of overall experience of care, but interpersonal communication and respectfulness are the most discriminating. This finding has important policy implications. The implication for policy makers is that public support for proposed healthcare innovations will suffer if reforms interfere with providers' capacity to attend to interpersonal communication and respectfulness. These attributes were not targeted for accountability within the Health Accords (Health Canada 2003) nor for renewal in the Primary Health Care Transition Fund (Health Canada 2007), but they are of critical importance to patients, and it is crucial to ensure that reforms not be implemented at their expense.

Correspondence may be directed to: Jeannie L. Haggerty, Associate Professor, Department of Family Medicine, McGill University, St. Mary's Research Centre, Hayes Pavilion – Suite 3734, 3830 Lacombe Ave., Montreal QC H3T 1M5; tel.: 514-345-3511 ext. 6332; fax: 514-734-2652; E-mail: jeannie.haggerty@mcgill.ca.

REFERENCES

Borowsky, S.J., D.B. Nelson, J.C. Fortney, A.N. Hedeen, J.L. Bradley and M.K. Chapko. 2002. "VA Community-Based Outpatient Clinics: Performance Measures Based on Patient Perceptions of Care." *Medical Care* 40(7): 578–86.

Bostick, R.M., J.M. Sprafka, B.A. Virnig and B.A. Potter. 1994. "Predictors of Cancer Prevention Attitudes and Participation in Cancer Screening Examinations." *Preventive Medicine* 23: 816–26.

Breitkopf, C.R., J. Catero, J. Jaccard and A.B. Berenson. 2004. "Psychological and Sociocultural Perspectives on Follow-up of Abnormal Papanicolaou Results." *Obstetrics and Gynecology* 104: 1347–54.

Collins, K. and A. O'Cathain. 2003. "The Continuum of Patient Satisfaction – from Satisfied to Very Satisfied." *Social Science and Medicine* 57: 2465–70.

Crow, R., H. Gage, S. Hampson, J. Hart, A. Kimber, L. Storey et al. 2002. "The Measurement of Satisfaction with Healthcare: Implications for Practice from a Systematic Review of the Literature." *Health Technology Assessment* 6: 1–244.

Fiscella, K., M.A. Goodwin and K.C. Stange. 2002. "Does Patient Educational Level Affect Office Visits to Family Physicians?" *Journal of the National Medical Association* 94: 157–65.

Flocke, S. 1997. "Measuring Attributes of Primary Care: Development of a New Instrument." *Journal of Family Practice* 45(1): 64–74.

Grol, R., M. Wensing and Task Force on Patient Evaluations of General Practice. 2000. *Patients Evaluate General/Family Practice: The EUROPEP Instrument.* Nijmegen, Netherlands: Centre for Quality of Care Research, Raboud University.

Haggerty, J.L., R. Pineault, M.-D. Beaulieu, Y. Brunelle, J. Gauthier, F. Goulet et al. 2007. "Room for Improvement: Patient Experience of Primary Care in Quebec Prior to Major Reforms." *Canadian Family Physician* 53: 1056–57.

Haggerty, J.L., C. Beaulieu, B. Lawson, D.A. Santor, M. Fournier and F. Burge. 2011. "What Patients Tell Us about Primary Healthcare Evaluation Instruments: Response Formats, Bad Questions and Missing Pieces." *Healthcare Policy* 7 (Special Issue): 66–78.

Haggerty, J.L., F. Bouharaoui and D.A. Santor. 2011. "Differential Item Functioning in Primary Healthcare Evaluation Instruments by French/English Version, Educational Level and Urban/Rural Location." *Healthcare Policy* 7 (Special Issue): 47–65.

Health Canada. 2003. *First Ministers' Health Accord on Health Care Renewal.* Ottawa: Author.

Health Canada. 2007. *Primary Health Care Transition Fund.* Retrieved May 10, 2011. <http://www.hc-sc.gc.ca/hcs-sss/prim/phctf-fassp/index-eng.php>.

Jöreskog, K.G. and D. Sörbom. 1996. *LISREL 8: User's Reference Guide.* Lincolnwood, IL: Scientific Software International.

Marshall, G.N. and R.D. Hays. 1994. *The Patient Satisfaction Questionnaire – Short Form (PSQ-18).* Report no. P-7865. Santa Monica, CA: Rand.

Safran, D.G., J. Kosinski, A.R. Tarlov, W.H. Rogers, D.A. Taira, N. Lieberman and J.E. Ware. 1998. "The Primary Care Assessment Survey: Tests of Data Quality and Measurement Performance." *Medical Care* 36(5): 728–39.

Shi, L., B. Starfield and J. Xu. 2001. "Validating the Adult Primary Care Assessment Tool." *Journal of Family Practice* 50(2): n161w–n171w.

Smith, J.L. and J. Haggerty. 2003. "Literacy in Primary Care Populations: Is It a Problem?" *Canadian Journal of Public Health* 94: 408–12.

Stewart, A.L., A.M. Nápoles-Springer, S.E. Gregorich and J. Santoyo-Olsson. 2007. "Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups." *Health Services Research* 42(3 Pt. 1): 1235–56.

Willems, S., S. De Maesschaick, M. Deveugele, A. Derese and J. De Maeseneer. 2005. "Socio-Economic Status of the Patient and Doctor–Patient Communication: Does It Make a Difference?" *Patient Education and Counselling* 56: 139–46.