

UCSF

UC San Francisco Previously Published Works

Title

Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups

Permalink

<https://escholarship.org/uc/item/2f9176xp>

Journal

Health Services Research, 42(3p1)

ISSN

0017-9124

Authors

Stewart, Anita L
Nápoles-Springer, Anna M
Gregorich, Steven E
et al.

Publication Date

2007-06-01

DOI

10.1111/j.1475-6773.2006.00637.x

Peer reviewed

Interpersonal Processes of Care Survey: Patient-Reported Measures for Diverse Groups

Anita L. Stewart, Anna M. Nápoles-Springer,
Steven E. Gregorich, and Jasmine Santoyo-Olsson

Objective. To create a patient-reported, multidimensional physician/patient interpersonal processes of care (IPC) instrument appropriate for patients from diverse racial/ethnic groups that allows reliable, valid, and unbiased comparisons across these groups.

Data Source/Data Collection. Data were collected by telephone interview. The survey was administered in English and Spanish to adult general medicine patients, stratified by race/ethnicity and language (African Americans, English-speaking Latinos, Spanish-speaking Latinos, non-Latino whites) ($N = 1,664$).

Study Design/Methods. In this cross-sectional study, items were designed to be appropriate for diverse ethnic groups based on focus groups, our prior framework, literature, and cognitive interviews. Multitrait scaling and confirmatory factor analysis were used to examine measurement invariance; we identified scales that allowed meaningful quantitative comparisons across four race/ethnic/language groups.

Principal Findings. The final instrument assesses several subdomains of communication, patient-centered decision making, and interpersonal style. It includes 29 items representing 12 first-order and seven second-order factors with equivalent meaning (metric invariance) across groups; 18 items (seven factors) allowed unbiased mean comparison across groups (scalar invariance). Final scales exhibited moderate to high reliability.

Conclusions. The IPC survey can be used to describe disparities in interpersonal care, predict patient outcomes, and examine outcomes of quality improvement efforts to reduce health care disparities.

Key Words. Quality of care, race, ethnicity, measurement, measurement invariance, factorial invariance, physician-patient communication, physician-patient interaction

Evidence of racial/ethnic disparities in quality of care is accumulating rapidly. Widespread disparities have been observed in many components of quality such as effectiveness of care for chronic and acute health conditions, patient safety, timeliness, communication, respectfulness, and discrimination

(Cooper-Patrick et al. 1999; Collins et al. 2002; Agency for Healthcare Research and Quality 2003, 2004; Smedley, Stith, and Nelson 2003; Weech-Maldonado et al. 2003; Johnson et al. 2004). Because many of these quality of care indicators have been linked to poorer health, they may partially account for race/ethnic disparities in health status (Smedley, Stith, and Nelson 2003). Within Donabedian's structure-process-outcomes quality of care paradigm (Donabedian 1968), most quality indicators in these studies pertain to technical processes of care; only a small proportion address interpersonal aspects of care. Interpersonal care may be a critical pathway to optimal health outcomes for minority, lower socioeconomic status (SES), or limited English proficiency (LEP) patient subgroups considered as priority populations (Foundation for Accountability: FACCT 1997; Clancy and Chesley 2003). Interpersonal quality of care may be as important as technical quality in determining health outcomes (Fung et al. 2005). Thus, research on the role of interpersonal aspects of care is needed, especially among diverse ethnic groups.

Such research has been hampered in several ways by a lack of adequate measures that allow for valid comparisons across groups. First, research on disparities in interpersonal aspects of care has been heavily based on coding audio or videotapes of encounters (Cooper-Patrick et al. 1999; Gallagher, Hartung, and Gregory 2001; Eide et al. 2004), precluding large-scale investigations. Second, although interpersonal processes are multidimensional (Stewart et al. 1999), most patient report measures assess one or two domains such as decision making (Kaplan et al. 1995). In two widely used multidimensional quality-of-care instruments—the Consumer Assessment of Health Plans[®] (CAHPS) and the Primary Care Assessment Survey (PCAS) (Safran et al. 1998)—interpersonal care comprises only a small portion. The CAHPS 2.0 assesses only provider communication and staff helpfulness (Hargraves, Hays, and Cleary 2003) and the PCAS assesses communication and interpersonal treatment (Safran et al. 1998). Third, most patient-reported research on disparities in interpersonal processes is based on single items (Collins et al. 2002; Agency for Healthcare Research and Quality 2003, 2004); although practical,

Address correspondence to Anita L. Stewart, Ph.D., Center for Aging in Diverse Communities, Medical Effectiveness Research Center for Diverse Populations, Institute for Health and Aging, University of California San Francisco, 3333 California Street, Suite 340, San Francisco, CA 94118-1944. Anna M. Nápoles-Springer, Ph.D., and Steven E. Gregorich, Ph.D., are with the Center for Aging in Diverse Communities, Medical Effectiveness Research Center for Diverse Populations, Division of General Internal Medicine, Department of Medicine, University of California San Francisco, CA. Jasmine Santoyo-Olsson, M.S. is with the Institute for Health & Aging, University of California San Francisco, San Francisco, CA.

single items are limited in scope and reliability, have questionable validity, and may result in biased estimates of group differences. Fourth, concepts and measures of interpersonal processes must reflect adequately the concerns of minority, lower SES, or LEP subgroups. Most existing measures were not designed with these groups in mind, and may miss relevant dimensions. Last, measures should have equivalent psychometric properties—factorial invariance—across groups, which allow for meaningful quantitative group comparisons (Meredith 1993; Gregorich 2006).

This article presents a multidimensional, patient-reported interpersonal processes of care (IPC) instrument, designed to be appropriate for four diverse groups (African Americans, English- and Spanish-speaking Latinos, and non-Latino whites).

METHODS

Item Development

Items were developed based on: (1) 19 focus groups stratified by race/ethnicity (African American, Latino, non-Latino white) and language (Spanish, English) (Nápoles-Springer et al. 2005); (2) our original conceptual framework and items (Stewart et al. 1999), (3) literature on quality of care and physician–patient communication, and (4) cognitive interviews with adults representing the same four ethnic/language groups (Nápoles-Springer, Santoyo, O’Brien, and Stewart 2006). Items were developed simultaneously in Spanish and English, aiming for semantic equivalence (Marín and Marín 1991). The item development process yielded 85 items. The measurement model included three broad domains (communication, decision making, and interpersonal style); each had several subdomains (Table 1). Respondents reported on the care they had received from their doctors over the past 12 months. For each item, they were asked how often that type of care had been provided using a five-point scale (1, never; 2, rarely; 3, sometimes; 4, usually; 5, always).

Nature and Sources of Telephone Survey Respondents

Adult patients with at least one visit in the prior 12 months were sampled from a patient database of adult general medicine practices at an academic health center. We recruited approximately 400 patients within each of four groups: African Americans, English-speaking Latinos, Spanish-speaking Latinos, and non-Latino whites. Recruitment and sampling procedures are described elsewhere (Nápoles-Springer, Santoyo, and Stewart 2005). Telephone interviews

Table 1: Hypothesized Measurement Model of Interpersonal Processes of Care (85 Items)

COMMUNICATION
General clarity (five items)
Elicitation of and responsiveness to patient problems, concerns, and expectations (11 items)
Explanations of condition and progress (six items)
Explanations of processes of care (seven items)
Explanations of self-care (six items)
DECISION MAKING
Responsiveness to patient preferences (four items)
Consideration of patients' ability to comply with treatment (four items)
INTERPERSONAL STYLE
Friendliness, courteousness (four items)
Respectfulness (four items)
Discrimination (eight items)
Cultural sensitivity (six items)
Emotional support, reassurance (eight items)
Empowerment (six items)
Friendliness, courteousness of office staff (two items)
Respectfulness of office staff (four items)

(conducted October 2001 through January 2002) lasted about 30 minutes. All procedures were approved by the academic health center's Institutional Review Board.

Methods of Analysis

Items within subdomains were hypothesized to be unidimensional. Associations between items and subdomains represented the first-order structure of the measurement model. Subdomains, in turn, were hypothesized to be unidimensional indicators of the associated three domains, representing the second-order structure. In common factor analysis parlance, subdomains and domains represented first- and second-order common factors; common factors are latent constructs hypothesized to be indirectly observed via responses to items.

The measurement model was assessed in four stages. Initially, within each group, a confirmatory factor analysis (CFA) tested the hypothesized first-order common factor model, with 85 observed items, and 15 common factors (subdomains) in Table 1. Common factors were identified by their respective items (no cross-loadings), common factor variances and covariances were freely estimated, and all item residuals were constrained to be uncorrelated. To identify the model, the factor loading of a single item for each common

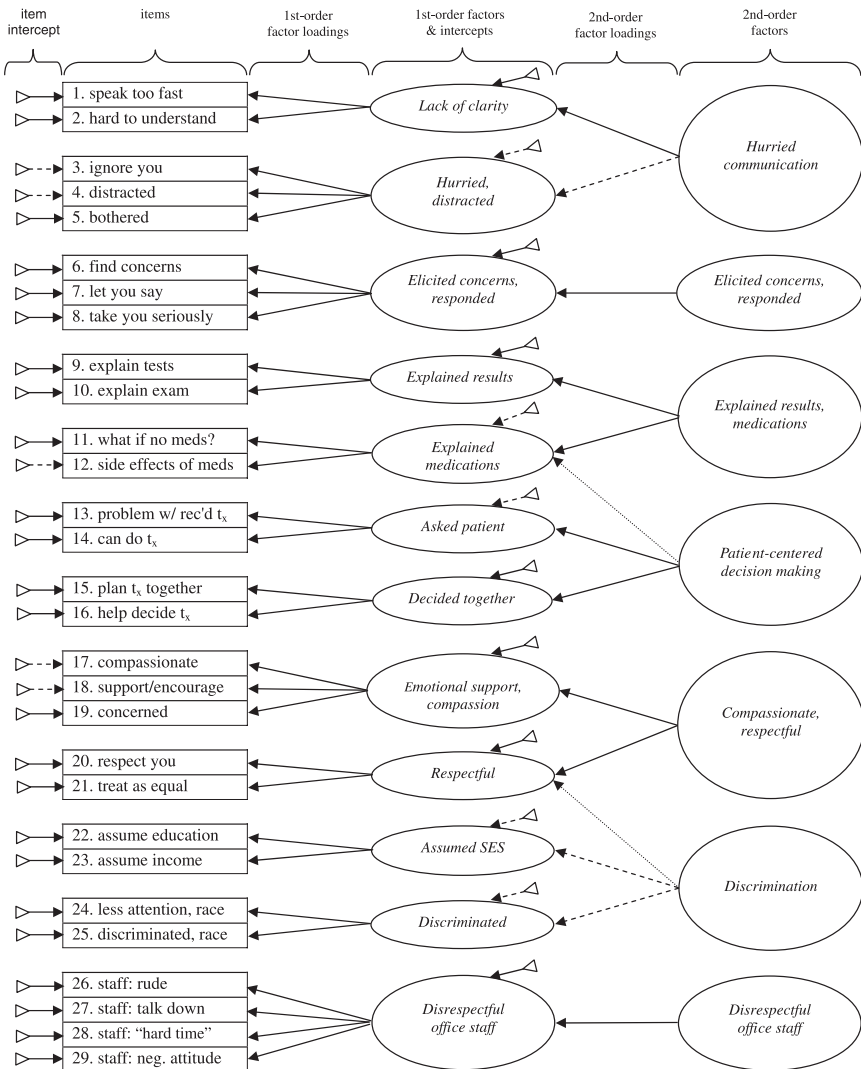
factor was fixed to unity. These models fit poorly within each group, and estimated interfactor correlations were very high, thus, we rejected the hypothesized measurement model and searched for a more appropriate empirical model.

In the second stage, we used multitrait scaling analysis to assess whether each item in a subdomain was linearly related ($r \geq 0.30$, corrected for overlap) to the total score for that subdomain (item convergence), and whether each item correlated at least two standard errors higher with its hypothesized subdomain than with other subdomains (item discrimination) (Hays and Hayashi 1990; Stewart, Hays, and Ware 1992). We analyzed the hypothesized scales within domains—communication, decision making, and interpersonal style—separately for each group. We first eliminated items not meeting the item convergence criterion in all four race/ethnic groups, followed by elimination of nonconvergent items in at least three groups. We then eliminated items not meeting the item discrimination criterion using the same approach. This left 56 items.

In the third stage, data from these 56 items were modeled with multiple-group CFA. These analyses assessed the unidimensionality of items within each revised subdomain across the four groups. Frequently, item sets hypothesized to represent a single subdomain were found to represent multiple highly correlated subdomains, suggesting a higher-order factor structure. In addition, items were dropped via a process of backward elimination, either because they did not have salient loadings (< 0.40) on the hypothesized factor, or loaded on more than one factor. The modified measurement models from each domain were then combined to form a single, empirical measurement model: a second-order factor model with 29 items, 12 first-order common factors, and seven second-order common factors (Figure 1). Detailed results from analysis stages 2 and 3 are not reported.

In the fourth analysis stage, a series of nested multiple-group factor models, based on the empirical model derived from the third stage, were fit to test the invariance of corresponding model parameters across the four groups (Meredith 1993). First- and second-order factors were identified by their respective items and first-order factors. For each common factor, the loading of a single item or first-order factor, as appropriate, was fixed at unity and the corresponding intercept was fixed at 0 to identify the model. All residual variances were constrained to be uncorrelated. Four basic hypotheses were tested via multi-group models: (1) invariance of the item/factor configuration, (2) invariance of first- and second-order factor loading estimates, commonly known as factor pattern or metric invariance (evidence of equivalent factor meaning across groups), (3) invariance of estimated item and first-order factor

Figure 1: Second-Order Factor Structure with 29 Items.



Note. arrow styles represent the following: \longrightarrow parameter is invariant across all 4 groups; \dashrightarrow parameter is not invariant across all 4 groups; $\cdots\rightarrow$ secondary 2nd-order factor loading, or cross-loading.

intercepts, known as strong factorial or scalar invariance (evidence that comparisons of observed means across groups are unbiased), and (4) invariance of estimated item and first-order factor residual variances, known as

strict factorial invariance. These nested models were tested sequentially, first for the first-order and then for the second-order factor structure.

In analysis stages 1, 3, and 4, models were fit to the data with *LISREL* 8.54 using maximum likelihood estimation (Jöreskog and Sörbom 1998). Goodness of fit was assessed by examining model χ^2 and degrees of freedom, the root mean square error of approximation (RMSEA, Steiger 1990), and the comparative fit index (CFI, Bentler 1990). Comparisons between models were aided by the expected cross-validation index (ECVI, Browne and Cudeck 1993). Generally, significant χ^2 tests indicate lack of "exact fit." RMSEA values below 0.05 or 0.06 and CFI values above 0.95 suggest approximate model fit (Browne and Cudeck 1993; Hu and Bentler 1999). In a series of nested models, ECVI reaches a relative minimum value for models with higher expectation of cross validation in independent samples of the same size. Point estimates of RMSEA and ECVI were augmented with 90 percent confidence intervals (Browne and Cudeck 1993). In stage 4, empirical model modifications were guided by *LISREL's* modification indices. Cross-group equality constraints on parameter estimates that contributed the most to lack of fit were subsequently freed and the model reestimated. In deciding when to stop empirical model modifications, we paid particular attention to the RMSEA and ECVI, which adjust for the number of estimated parameters.

Data Preparation and Distributional Assumptions

Because of nonnormal distributions, data were pooled across groups and transformed to normal scores before analysis (Blom 1958). As analyzed in phase 4, the transformed data had median values of absolute skewness and kurtosis equal to 1.2 and 1.3, respectively. Even with ordinal variables, the observed nonnormality will not affect parameter estimates, but can result in overestimation of χ^2 test statistics and underestimation of parameter standard errors (Browne 1984; Muthén and Kaplan 1992). No correction was made to the χ^2 tests because they were expected to be conservatively biased. However, for each model, standard errors were estimated from 200 bootstrap samples. To deal with missing data (<5 percent of all data points), CFA models were fit to covariance matrices and mean vectors estimated by the expectation-maximization (EM) algorithm (Little and Rubin 2002). The EM covariance matrices partialled the effects of respondent age, gender, and education.

Scale Scoring, Reliability, Interscale Correlations, Group Differences, and Readability

Once final scales were selected, we calculated scale scores by averaging non-missing items; scores ranged from 1 to 5 and a higher score indicated higher

frequency of the construct (e.g., higher discrimination scores indicate more discrimination and higher compassion scores indicate more compassion). We calculated internal-consistency reliabilities, the pooled within-groups inter-scale correlation matrix, and mean scale score differences across the four groups. Readability of item stems and instructions was summarized using the Flesch–Kincaid formula.

RESULTS

Of those contacted and eligible, 70 percent responded ($N = 1,664$). This represented 42 percent of the sampling frame (Nápoles-Springer, Santoyo, and Stewart 2005). Men, those aged 18–39 and 75 years and older, and non-Latino whites were slightly underrepresented. A broad age range was obtained and the majority was female (Table 2). Spanish-speaking Latinos were the oldest (mean = 62), had the lowest SES, and were most likely to report fair or poor health (56 percent). English-speaking Latinos were the youngest (mean = 43). The mean number of visits during the prior year was 7 ($SD = 8.5$). Most respondents reported receiving a prescribed medication (84 percent) and a medical test/procedure (89 percent) in the prior year.

Tests of Measurement Models

Stage 1. The hypothesized first-order factor model did not fit well in any group: $\chi^2(3,380, n = 421) = 6,949.77$; $\chi^2(3,380, n = 421) = 6,881.23$; $\chi^2(3,380, n = 364) = 7,298.25$; and $\chi^2(3,380, n = 415) = 7,251.31$, with all p -values $< .001$, for the African American, Latino/English, Latino/Spanish, and white groups. The corresponding RMSEA values ranged from 0.053 to 0.059 suggesting approximate fit, whereas CFI values ranged from 0.780 to 0.839, suggesting poor fit. The hypothesized model was abandoned in favor of the empirical measurement model developed in stages 2 and 3.

Stage 4. Testing factorial invariance of the empirical measurement model across the four groups. Multi-sample CFA models tested the invariance of the 29-item, second-order empirical factor model across the four groups (Figure 1). Although the χ^2 tests of “exact fit” were significant for each of the models, the other indices suggested approximate fit: all RMSEA values were below 0.04 and CFI values were greater than 0.96.

Table 2: Sample Characteristics: Total Sample and by Ethnic and Language Group

Characteristic	Total Sample n = 1,664	African American n = 435	Latino English n = 428	Latino Spanish n = 383	Non-Latino White n = 418	p-Value
Age in years, mean (SD)	51 (18)	50 (16)	43 (16)	62 (17)	49 (17)	<.0001
Sex (% women)	71	76	71	74	63	<.001
Education (%)						<.0001
High school or less	35	38	23	67	14	
Some college	28	36	37	19	17	
College or more	37	25	39	14	68	
Household income (%)						<.0001
<\$25,000	32	40	21	52	21	
\$25,001–\$40,000	19	22	19	24	14	
\$40,001–\$75,000	26	24	35	17	27	
\$75,000 or more	23	14	26	7	39	
Born in the United States (%)	62	96	56	2	88	<.0001
If not: years living in United States, mean (SD)	29 (14)	20 (13)	29 (13)	30 (13)	29 (19)	ns
Limited English proficiency,* (%)	22	4	8	79	3	<.0001
Health insurance (%)						<.0001
Private	62	52	76	44	74	
Self-pay/none, MediCal, MediCal+ Medicare	27	33	18	43	16	
Medicare, Medicare+private	11	15	6	13	10	
Health condition needing ongoing care (%)	63	71	52	63	68	<.0001
Self-rated health fair or poor (%)	38	43	28	56	26	<.0001

*Self-reported speaks English not at all, poorly, or fairly well.

First-order measurement model. A series of first-order CFA factor models were fit (Models 1–3f, Table 3). Model 1 allowed all parameter estimates to be freely estimated across groups to test first-order configural invariance. The contribution of each group to the overall χ^2 was nearly equal and the RMSEA and CFI values provided evidence of the same item clusterings in each group. Model 2 constrained corresponding first-order factor loadings to be equal across groups, testing metric invariance. Based upon approximate fit indices, this model fit as well or better than the configural invariance model and suggested that the 12 first-order factors had the same meanings across all four groups. Model 3 further constrained corresponding item intercepts to be equal across groups, testing scalar invariance. This model resulted in a worsening of fit, suggesting that some intercept values were not equivalent across groups. Modification indices suggested that model 2 imposed six equality constraints on item intercepts that should be freed: three each for the Latino/Spanish (items q18, q17, and q4) and white groups (items q3, q12, and q17) (Table 4). Models 3a–3f represented partial scalar invariance models: each freely estimated one additional item intercept parameter (Byrne, Shavelson, and Muthén 1989; Steenkamp and Baumgartner 1998).

For model 3f, the values of RMSEA (0.034) and CFI (0.970) suggested good approximate fit. Further, the point estimate of ECVI approached that for model 2, which was arguably the best-fitting model considered. Model 3f suggested that unbiased group comparisons of means were possible if analyses were restricted to those items demonstrating invariant factor loadings and item intercepts (Steenkamp and Baumgartner 1998).

On the left-hand side of Figure 1, items with invariant intercept estimates have arrows with solid lines. A further model (not shown) tested the invariance of item residual variances across groups—a test of partial strict factorial invariance. Although indices suggested approximate model fit, their point values were well outside of the 90 percent confidence intervals of the previous models (RMSEA = 0.048; ECVI = 2.23), thus the model was rejected. Further empirical modifications were considered to identify a well-fitting partial strict first-order factorial invariance model, but after 10 such modifications (freeing 10 cross-group equality constraints on residual variance estimates) the resulting RMSEA and ECVI values still suggested worse fit than model 3f (i.e., RMSEA = 0.038; ECVI = 1.94). Therefore, attempts to specify a model with invariant item residual variances were abandoned.

Table 3: Invariance of 18-Item Short Form Factor Structure across Four Groups

<i>Model</i>	χ^2	<i>df</i>	<i>RMSEA (90% CI)</i>	<i>ECVI (90% CI)</i>	<i>CFI</i>
<i>First-order factor structure</i>					
1. Configural invariance model	1,758.46	1,244	0.032 (0.028–0.035)	1.84 (1.78–1.92)	0.975
2. Metric invariance model	1,842.63	1,295	0.032 (0.029–0.036)	1.83 (1.77–1.91)	0.974
3. Scalar factorial invariance model	2,128.73	1,346	0.038 (0.035–0.041)	1.95 (1.87–2.03)	0.964
(a) Free intercept: item #18, LS	2,053.29	1,345	0.036 (0.033–0.039)	1.90 (1.83–1.98)	0.966
(b) Free intercept: item #3, W	2,030.33	1,344	0.036 (0.032–0.039)	1.89 (1.82–1.97)	0.967
(c) Free intercept: item #12, W	2,010.10	1,343	0.035 (0.032–0.038)	1.88 (1.81–1.96)	0.968
(d) Free intercept: item #17, W	1,993.72	1,342	0.035 (0.031–0.038)	1.87 (1.80–1.95)	0.969
(e) Free intercept: item #17, LS	1,975.36	1,341	0.034 (0.031–0.037)	1.86 (1.79–1.94)	0.969
(f) Free intercept: item #4, LS	1,958.37	1,340	0.034 (0.031–0.037)	1.85 (1.78–1.93)	0.970
<i>Second-order factor structure</i>					
4. Configural invariance model	2,151.63	1,472	0.034 (0.031–0.037)	1.81 (1.73–1.89)	0.968
5. Metric invariance model	2,360.24	1,493	0.038 (0.035–0.041)	1.91 (1.83–1.99)	0.959
(a) Free loading: <i>discriminated</i> , W	2,253.84	1,492	0.036 (0.033–0.038)	1.84 (1.77–1.93)	0.964
(b) Free loading: <i>assumed SES</i> , LE	2,217.24	1,491	0.035 (0.032–0.038)	1.82 (1.75–1.90)	0.965
(c) Free loading: <i>hurried, distracted</i> , LS	2,202.97	1,490	0.034 (0.031–0.037)	1.82 (1.74–1.90)	0.966
(d) Free loading: <i>discriminated</i> , LS	2,194.25	1,489	0.034 (0.031–0.037)	1.81 (1.74–1.89)	0.966
6. Partial scalar factorial invariance model	2,335.09	1,500	0.037 (0.034–0.040)	1.88 (1.81–1.97)	0.961
(a) Free intercept: <i>explained medications</i> , W	2,268.61	1,499	0.036 (0.033–0.039)	1.85 (1.77–1.93)	0.963
(b) Free intercept: <i>hurried, distracted</i> , W	2,243.17	1,498	0.035 (0.032–0.038)	1.83 (1.75–1.91)	0.965
(c) Free intercept: <i>asked patient</i> , W	2,220.02	1,497	0.035 (0.032–0.038)	1.82 (1.74–1.90)	0.965

Note. group identifiers: LE, English-speaking Latinos; LS, Spanish-speaking Latinos; W, whites; RMSEA, root mean square error of approximation; ECVI, expected cross-validation index; CFI, comparative fit index.

Second-order measurement model. Next, a series of second-order CFA models were fit. With model 3f defining the first-order structure of the measurement model, configural, metric, and scalar invariance of the second-order factor structure were investigated. Model 4 assessed second-order configural invariance. Associated approximate fit indices suggested that each group had the same clusterings of first- and second-order factors. Model 5 tested the invariance of corresponding second-order factor loadings across groups (i.e., second-order metric invariance). Relative to model 4, the fit of this model was poor and model modifications were considered. Based upon modification indices, a set of models relaxed some cross-group equality constraints to test partial second-order metric invariance. The last of these models, 5d, freed four second-order factor loading estimates: the loadings for *discriminated due to race/ethnicity* in the white and Latino/Spanish groups (models 5a and 5d); *assumed SES* in the Latino/English group (5b); and *hurried/distracted* in the Latino/Spanish group (5c; Table 3). The approximate fit of model 5d was similar to the well fitting, but less parsimonious model 4. Model 6 constrained all corresponding first-order factor intercepts to be equal across groups, except for those associated with the four second-order factor loadings freely estimated in model 5d. This model assessed partial second-order scalar invariance. The associated RMSEA, ECVI, and CFI values suggested a worsening of fit relative to model 5d. Again, modification indices guided model modifications. The final model, 6c, freely estimated three additional first-order factor intercepts, for *explained medications*, *hurried/distracted*, and *asked patient* for whites. The fit of this most parsimonious model suggested a reasonable approximation and compared well with other models.

Common metric standardized factor loadings and intercept estimates from model 6c are summarized in Table 4. Cross-loadings of *explained medications* on *patient-centered decision making*, and *respectful* on *discriminated* equaled 0.38 and -0.47 , respectively (not tabled).

Eighteen items representing seven constructs met the scalar invariance criterion, either at the first- or second-order level (underlined items in Table 4). The fit of model 6c suggested that group mean comparisons based upon these 18 items would be approximately unbiased. Final items for the full 29-item survey and the 18-item "short form" are presented in Table 5.

Reliability, Reading Level, and Group Differences

For six of the seven short-form scales, internal-consistency reliability was above 0.70 in the total sample (range 0.65–0.90); the reliability for *lack of clarity*

Table 4: Partial Strong Factorial Invariance of the 18-Item Short Form

DOMAIN		
Second-Order Factor		
First-Order Factor		
#. Item*	Factor Loading	Intercept
COMMUNICATION		
Hurried communication		
<i>Lack of clarity</i>	0.63	0
1. . . . <u>did doctors speak too fast?</u>	0.83	0
2. . . . <u>did doctors use words that were hard to understand?</u>	0.59	0
<i>Hurried, distracted</i>	1.10, 0.71 _s	0.03, - 0.25 _s , 0.40 _w
3. . . . did doctors ignore what you told them	0.70	0, - 0.19 _w
4. . . . did doctors appear to be distracted when they were with you?	0.68	- 0.01, - 0.18 _s
5. . . . did doctors seem bothered if you asked several questions?	0.70	- 0.05
Elicited concerns, responded		
<i>Elicited concerns, responded</i>	1.0	0
6. . . . <u>did doctors really find out what your concerns were?</u>	0.72	0
7. . . . <u>did doctors let you say what you thought was important?</u>	0.75	0.02
8. . . . <u>did doctors take your health concerns very seriously?</u>	0.80	0.01
Explained results, medications		
<i>Explained results</i>	0.86	0
9. . . . <u>did doctors explain your test results such as blood tests, X-rays, or cancer screening tests?</u>	0.79	0
10. . . . <u>did doctors clearly explain the results of your physical exam?</u>	0.89	- 0.01
<i>Explained medications</i>	0.46	0.08, - 0.33 _w
11. . . . did doctors tell you what could happen if you didn't take a medicine that they prescribed for you?	0.75	0
12. . . . did doctors tell you about side effects you might get from a medicine?	0.80	- 0.06, 0.18 _w
DECISION MAKING		
Patient-centered decision making		
<i>Asked patient</i>	0.92	0, - 0.20 _w
13. . . . did doctors ask if you would have any problems following what they recommended?	0.78	0

Table 4. *Continued*

DOMAIN		
Second-Order Factor		
<i>First-Order Factor</i>		
#. Item*	Factor Loading	Intercept
14. . . did doctors ask if you felt you could do the recommended treatment?	0.75	0
<i>Decided together</i>	0.84	- 0.05
15. . . <u>did you and your doctors work out a treatment plan together?</u>	0.76	0
16. . . <u>did doctors ask if you would like to help decide your treatment?</u>	0.78	0.01
INTERPERSONAL STYLE		
Compassionate, respectful		
<i>Emotional support, compassion</i>	0.99	0
17. . . were doctors compassionate?	0.75	0, 0.22 _s , 0.18 _w
18. . . did doctors give you support and encouragement?	0.85	0.01, 0.47 _s
19. . . <u>were doctors concerned about your feelings?</u>	0.76	0.10
<i>Respectfulness</i>	0.57	0.04
20. . . <u>did doctors really respect you as a person?</u>	0.79	0
21. . . <u>did doctors treat you as an equal?</u>	0.73	0.02
Discrimination		
<i>Assumed socioeconomic status</i>	0.54, 0.99 _e	0, 0.07 _e
22. . . did doctors make assumptions about your level of education?	0.88	0
23. . . did doctors make assumptions about your income?	0.79	0.02
<i>Discriminated due to race/ethnicity</i>	1.02, 1.60 _s , 0.25 _w	0.09, 0.58 _s , - 0.20 _w
24. . . <u>did doctors pay less attention to you because of your race or ethnicity?</u>	0.82	0
25. . . <u>did you feel discriminated against because of your race or ethnicity?</u>	0.86	0.01
Disrespectful office staff		
<i>Disrespectful office staff</i>	1.0	0
26. . . <u>was office staff rude to you?</u>	0.81	0
27. . . <u>did office staff talk down to you?</u>	0.80	0.01
28. . . <u>did office staff give you a hard time?</u>	0.84	0
29. . . <u>did office staff have a negative attitude toward you?</u>	0.87	0.01

*Item stems were Over the past 12 months, how often. . .

Note: Underlined items indicate the 18-item short form (items meeting the scalar invariance criterion across all four groups). Tabled parameter estimates without a subscript were equivalent (invariant) across two or more groups. Other parameter estimates are subscripted to indicate the group to which they correspond: e, English-speaking Latino; s, Spanish-speaking Latino; w, White.

Table 5: Interpersonal Processes of Care Survey

Communication: The following questions are about your experiences talking with your doctors at (name of site) over the past 12 months.

Hurried communication

1. How often did doctors speak too fast? (SF)*
2. How often did doctors use words that were hard to understand? (SF)
3. How often did doctors ignore what you told them?
4. How often did doctors appear to be distracted when they were with you?
5. How often did doctors seem bothered if you asked several questions?

Elicited concerns, responded

6. How often did doctors really find out what your concerns were? (SF)
7. How often did doctors let you say what you thought was important? (SF)
8. How often did doctors take your health concerns very seriously? (SF)

Explained results, medications

9. How often did doctors explain your test results such as blood tests, X-rays, or cancer screening tests? (SF)
10. How often did doctors clearly explain the results of your physical exam? (SF)
11. How often did doctors tell you what could happen if you didn't take a medicine that they prescribed for you?
12. How often did doctors tell you about side effects you might get from a medicine?

Decision making: These questions are about how you and your medical doctors decide about your health care. Please continue to think about your experiences over the past 12 months.

Patient-centered decision making

13. How often did doctors ask if you would have any problems following what they recommended?
14. How often did doctors ask if you felt you could do the recommended treatment?
15. How often did you and your doctors work out a treatment plan together? (SF)
16. If there were treatment choices, how often did doctors ask if you would like to help decide your treatment? (SF)

Interpersonal style: The following questions are about the personal interactions between you and your doctors over the past 12 months.

Compassionate, respectful

17. How often were doctors compassionate?
18. How often did doctors give you support and encouragement?
19. How often were doctors concerned about your feelings? (SF)
20. How often did doctors really respect you as a person? (SF)
21. How often did doctors treat you as an equal? (SF)

Discrimination

22. How often did doctors make assumptions about your level of education?
23. How often did doctors make assumptions about your income?
24. How often did doctors pay less attention to you because of your race or ethnicity? (SF)
25. How often did you feel discriminated against by doctors because of your race or ethnicity? (SF)

Disrespectful office staff

26. How often was office staff rude to you? (SF)
27. How often did office staff talk down to you? (SF)
28. How often did office staff give you a hard time? (SF)
29. How often did office staff have a negative attitude toward you? (SF)

Note: Response choices: 1, never; 2, rarely; 3, sometimes; 4, usually; 5, always.

*SF indicates that item is part of the 18-item short form.

was 0.65 (Table 6). Within all four groups, reliabilities were also generally high (range 0.61–0.91); of 28 coefficients, 24 were above 0.70. The Flesch–Kincaid grade level of the item stems and instructions for the 29-item survey was 8.6; for the 18-item short form it was 5.8.

We observed significant mean group differences on all seven short-form scales, although no group consistently had the lowest or highest scores. Overall, the worst scores were observed for *decided together* with a total sample mean of 3.13; Spanish-speaking Latinos had the lowest scores (2.84) and whites the highest (3.31; Table 6). Mean scores on *discriminated due to race/ethnicity* were near 1.0 (less discrimination) for all groups; African Americans experienced the most discrimination followed by English-speaking Latinos. Latinos (Spanish- and English-speaking) scored the lowest on all three communication scales.

DISCUSSION

Research on race/ethnic disparities in interpersonal aspects of care has been limited by a lack of measures that reflect the multidimensional nature of these processes and allow valid, unbiased comparisons across diverse groups. This study helps fill this gap by conceptualizing and operationalizing interpersonal processes as multidimensional. We provide a patient-reported survey developed through a sequence of qualitative and quantitative studies, with Spanish and English versions developed in parallel, with evidence of reliability and validity, and demonstrating scalar invariance of a subset of items in each domain. The final empirically based framework shared many features of the hypothesized model, which in turn was based on previous work done in this area. However, it had fewer subdomains, in part because hypothesized constructs were interrelated in more complex ways than originally thought.

The IPC Survey should facilitate research to explore how specific aspects of interpersonal care affect various health outcomes and whether interpersonal care explains disparities in such outcomes. The 29-item survey performed well within each group and can be used for within-group studies. The 18-item short form can be used to make unbiased mean comparisons across the four groups represented in this sample. Analyses of determinants and outcomes of IPC using these measures are forthcoming.

One notable finding was the relatively high scores overall. Relatively good processes could be the true state in these practices located in a major medical teaching university serving a highly diverse population. Also, many

Table 6: Reliability and Means for 18-Item Short-Form Scales: Total Sample and by Ethnic and Language Group

Domain Scale	Total Sample, n = 1,664		African American n = 435		Latino English n = 428		Latino Spanish n = 383		White n = 418	
	α^* (Range)	Percent Missing	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	p-Value
Communication										
<i>Lack of clarity</i> (-)	0.65 (0.61-0.72)	1.6	1.84 (0.94)	1.78 (0.90)	1.94 (0.91)	2.01 (1.10)	1.66 (0.80)	<.0001		
Elicited concerns, responded (+)	0.80 (0.79-0.82)	1.0	4.14 (0.93)	4.29 (0.90)	4.02 (0.93)	4.08 (1.06)	4.18 (0.80)	<.0001		
<i>Explained results</i> (+)	0.81 (0.78-0.85)	1.9	4.11 (1.14)	4.27 (1.08)	3.93 (1.18)	4.05 (1.28)	4.17 (0.99)	<.001		
Decision making										
<i>Decided together</i> (+)	0.75 (0.71-0.77)	4.7	3.13 (1.43)	3.17 (1.40)	3.17 (1.40)	2.84 (1.56)	3.31 (1.32)	<.0001		
Interpersonal style										
Compassionate, respectful (+)	0.71 (0.59-0.82)	1.7	4.08 (0.98)	4.30 (0.93)	4.05 (0.97)	3.96 (1.04)	3.98 (0.95)	<.0001		
<i>Discriminated due to race/ethnicity</i> (-)	0.79 (0.73-0.85)	2.3	1.23 (0.62)	1.35 (0.77)	1.28 (0.62)	1.17 (0.60)	1.11 (0.40)	<.0001		
Disrespectful office staff (-)	0.90 (0.89-0.91)	1.6	1.66 (0.88)	1.56 (0.81)	1.91 (1.00)	1.39 (0.76)	1.73 (0.83)	<.0001		

* α , internal-consistency reliability (range across groups shown in parentheses).

Note: (-) high score = worse IPC; (+) high score = better IPC. Regular typeface scales = second-order factors; italicized scales = first-order factors.

F-tests for Table 6

F(3, 1,634) = 10.01 Lack of clarity.

F(3, 1,648) = 6.97 Elicited concerns, responded.

F(3, 1,628) = 6.77 Explained results.

F(3, 1,581) = 7.19 Decided together.

F(3, 1,632) = 10.40 Compassionate, respectful.

F(3, 1,622) = 12.56 Discriminated.

F(3, 1,633) = 28.06 Disrespectful office staff.

patients had attended these clinics for over a decade and may have found providers with whom they were comfortable. Despite the relatively high scores, there is room for improvement, particularly with respect to patient-centered decision making.

Although there were significant group differences in interpersonal processes, these differences did not consistently favor any group. Our finding that African Americans obtained the best scores on two communication scales are consistent with results on provider communication from a CAHPS Medicaid Managed Care survey (Weech-Maldonado et al. 2003), and our finding that whites had the highest scores on *decided together* is consistent with one other study (Cooper-Patrick et al. 1999).

We envision four broad applications of the IPC Survey. First, the short-form facilitates comparative population- or clinic-based studies of disparities in interpersonal processes (e.g., by race/ethnicity). Another application is to determine if the measures predict technical processes (e.g., procedures or tests) or patient outcomes (e.g., patient adherence or satisfaction) (Stewart 1995; Blanchard and Lurie 2004; Fung et al. 2005).

A third application is to use the IPC Survey as an outcome of quality improvement policies (e.g., provider training). There is evidence that race/ethnic concordance of physicians and patients is associated with better communication (Saha et al. 1999) and more participatory decision making (Cooper-Patrick et al. 1999). If such findings are replicated with these measures, systems of care might be more likely to diversify their health care professional staff and develop targeted interventions to improve specific aspects of interpersonal processes (Cegala, Post, and McClure 2001). Finally, the IPC Survey could be useful to administrators as measures of outcomes of continuous quality improvement (e.g., to monitor disparities or provide feedback to physicians on interpersonal care).

We recommend continued validation research on the IPC Survey across a range of groups and settings. We empirically eliminated several conceptually relevant subdomains such as empowerment and cultural sensitivity that warrant continued measurement efforts. Because power differentials between patients and physicians place vulnerable patients at a unique disadvantage, empowerment may be an outcome of quality care. Cultural sensitivity was found to be multidimensional in our qualitative analyses (Nápoles-Springer et al. 2005), possibly explaining why our efforts to derive a unidimensional scale were unsuccessful. Cultural sensitivity may be difficult to measure because it is manifested through a broad spectrum of behaviors and attitudes (e.g., respect, compassion) toward nonwhite patients (Clancy and Stryer 2001).

Our results should be interpreted in light of several limitations. The study was conducted within a single university-based system of care in a geographic area known for diversity. We did not include Asian Americans or other ethnic subgroups. The measurement models were tested and modified using a single data set, thus results are provisional, conditional on future replication in independent samples. To achieve invariance across four groups, we eliminated items that worked well within some of the groups (e.g., were culture-specific). Future studies might supplement invariant scales with ethnic-specific scales when evidence suggests that a construct may help explain disparities in that group. Because we used telephone administration to accommodate persons with limited literacy or English proficiency, we do not know how well self-administration would work.

Numerous reports and policy statements call for quality measures that are relevant, valid, and unbiased across ethnic and linguistic groups to assess possible quality-of-care disparities (Bethell et al. 2003; Fortier and Bishop 2003; Beach et al. 2004). Although, we demonstrated the methodological complexities associated with doing so, the IPC Survey should help to fill this gap, and may prove useful in assessing quality of care disparities in other settings and ethnic groups.

ACKNOWLEDGMENTS

This study was supported by grant R01 HS10599 from the Agency for Healthcare Research and Quality; it was also supported in part by grant P30-AG15272 from the National Institute on Aging, the National Institute of Nursing Research, and The National Center on Minority Health and Health Disparities, National Institutes of Health. We appreciate the care with which the Public Research Institute designed and conducted the computer-assisted telephone interview survey. We are indebted to Dr. A. Eugene Washington who provided the initial spark and unfailing enthusiasm. We thank Dr. Eliseo Pérez-Stable for his support and guidance along the way.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Agency for Healthcare Research and Quality. 2003. *National Healthcare Disparities Report*. Rockville, MD: [Accessed on April 1, 2005]. Available at http://www.qualitytools.ahrq.gov/disparitiesreport/download_report.aspx

- . 2004. *2004 National Healthcare Disparities Report (No. 05-0014)*. Rockville, MD: U.S. Department of Health and Human Services.
- Beach, M. C., L. A. Cooper, K. A. Robinson, E. G. Price, T. L. Gary, M. W. Jenckes, A. Gozu, C. Smarth, A. Palacio, C. J. Feuerstein, E. B. Bass, and N. R. Powe. 2004. *Strategies for Improving Minority Healthcare Quality: Evidence Report/Technology Assessment No. 90 (No. 04-E008-02)*. Rockville, MD: Agency for Healthcare Research and Quality.
- Bentler, P. M. 1990. "Comparative Fit Indexes in Structural Models." *Psychological Bulletin* 107 (2): 238–46.
- Bethell, C., K. Carter, D. Lansky, B. Latzke, and K. Gowen. 2003. *Measuring and Interpreting Health Care Quality across Culturally-Diverse Populations: A Focus on Consumer-Reported Indicators of Health Care Quality*. New York: Foundation for Accountability. [Accessed on April 30, 2005]. Available at <http://www.markle.org/resources/faact/index.php>
- Blanchard, J., and N. Lurie. 2004. "R-E-S-P-E-C-T: Patient Reports of Disrespect in the Health Care Setting and Its Impact on Care." *Journal of Family Practice* 53 (9): 721–30.
- Blom, G. 1958. *Statistical Estimates and Transformed Beta Variables*. New York: John Wiley & Sons Inc.
- Browne, M. W. 1984. "Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures." *British Journal of Mathematical and Statistical Psychology* 37 (part 1): 62–83.
- Browne, M. W., and R. Cudeck. 1993. "Alternative Ways of Assessing Model Fit." In *Testing Structural Equation Models*, edited by K. A. Bollen and J. S. Long, pp. 136–62. Newbury Park, CA: Sage Publications.
- Byrne, B. M., R. J. Shavelson, and B. Muthén. 1989. "Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance." *Psychological Bulletin* 105 (3): 456–66.
- Cegala, D. J., D. M. Post, and L. McClure. 2001. "The Effects of Patient Communication Skills Training on the Discourse of Older Patients during a Primary Care Interview." *Journal of the American Geriatrics Society* 49 (11): 1505–11.
- Clancy, C. M., and F. D. Chesley. 2003. "Strengthening the Health Services Research to Reduce Racial and Ethnic Disparities in Health Care." *Health Services Research* 38 (5, part 1): xi–xviii.
- Clancy, C. M., and D. B. Stryer. 2001. "Racial and Ethnic Disparities and Primary Care Experience." *Health Services Research* 36 (6): 979–86.
- Collins, K. S., D. L. Hughes, M. M. Doty, B. L. Ives, J. N. Edwards, and K. Tenney. 2002. *Diverse Communities, Common Concerns: Assessing Health Care Quality for Minority Americans. Findings from the Commonwealth Fund 2001 Health Care Quality Survey*. New York: The Commonwealth Fund.
- Cooper-Patrick, L., J. J. Gallo, J. J. Gonzales, H. T. Vu, N. R. Powe, C. Nelson, and D. E. Ford. 1999. "Race, Gender, and Partnership in the Patient–Physician Relationship." *Journal of the American Medical Association* 282 (6): 583–9.
- Donabedian, A. 1968. "Promoting Quality through Evaluating the Process of Patient Care." *Medical Care* 6: 181–202.

- Eide, H., V. Quera, P. Graugaard, and A. Finset. 2004. "Physician–Patient Dialogue Surrounding Patients' Expressions of Concern: Applying Sequence Analysis to Rias." *Social Science and Medicine* 59 (1): 145–55.
- Fortier, J. P., and D. Bishop, (Eds.). 2003. *Setting the Agenda for Research on Cultural Competence in Health Care: Final Report*. Rockville, MD: U.S. Department of Health and Human Services Office of Minority Health and Agency for Healthcare Research and Quality.
- Foundation for Accountability: FACCT. 1997. "Hedis Outcomes May Not be the Only Ones You Need to Benchmark." *Healthcare Benchmarks* 4 (6): 81–3.
- Fung, C. H., M. N. Elliott, R. D. Hays, K. L. Kahn, D. E. Kanouse, E. A. McGlynn, M. D. Spranca, and P. G. Shekelle. 2005. "Patients' Preferences for Technical Versus Interpersonal Quality When Selecting a Primary Care Physician." *Health Services Research* 40 (4): 957–77.
- Gallagher, T. J., P. J. Hartung, and S. W. Gregory Jr. 2001. "Assessment of a Measure of Relational Communication for Doctor–Patient Interactions." *Patient Education and Counseling* 45 (3): 211–8.
- Gregorich, S. 2006. "Do Self Report Instruments Allow Meaningful Comparisons Across Diverse Population Groups? Testing Measurement Invariance Using the Confirmatory Factor Analysis Framework." *Medical Care*, 44 (11, Suppl. 3): in press.
- Hargraves, J. L., R. D. Hays, and P. D. Cleary. 2003. "Psychometric Properties of the Consumer Assessment of Health Plans Study (CAHPS) 2.0 Adult Core Survey." *Health Services Research* 38 (6, part 1): 1509–27.
- Hays, R. D., and T. Hayashi. 1990. "Beyond Internal Consistency Reliability: Rationale and User's Guide for Multitrait Scaling Analysis Program on the Microcomputer." *Behavior Research Methods, Instruments and Computers* 22: 167–75.
- Hu, L. T., and P. M. Bentler. 1999. "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Structural Equation Modeling* 6: 1–55.
- Johnson, R. L., S. Saha, J. J. Arbelaez, M. C. Beach, and L. A. Cooper. 2004. "Racial and Ethnic Differences in Patient Perceptions of Bias and Cultural Competence in Health Care." *Journal of General Internal Medicine* 19 (2): 101–10.
- Jöreskog, K. G., and D. Sörbom. 1998. *New Statistical Features in Lisrel 8.30*. Chicago: Scientific Software.
- Kaplan, S. H., B. Gandek, S. Greenfield, W. Rogers, and J. E. Ware. 1995. "Patient and Visit Characteristics Related to Physicians' Participatory Decision-Making Style." *Medical Care* 33 (12): 1176–87.
- Little, R. J. A., and D. B. Rubin. 2002. *Statistical Analysis with Missing Data*. 2d edition. New York: Wiley.
- Marín, G., and B. V. Marín. 1991. *Research with Hispanic Populations*. Newbury Park, CA: Sage Publications.
- Meredith, W. 1993. "Measurement Invariance, Factor Analysis and Factorial Invariance." *Psychometrika* 58 (4): 525–43.

- Muthén, B., and D. A. Kaplan. 1992. "A Comparison of Some Methodologies for the Factor Analysis of Non-Normal Likert Variables: A Note on the Size of the Model." *British Journal of Mathematical and Statistical Psychology* 45: 19–30.
- Nápoles-Springer, A. M., J. Santoyo, K. Houston, E. J. Pérez-Stable, and A. L. Stewart. 2005. "Patients' Perceptions of Cultural Factors Affecting the Quality of Their Medical Encounters." *Health Expectations* 8: 4–17.
- Nápoles-Springer, A. M., J. Santoyo, and A. L. Stewart. 2005. "Recruiting Ethnically Diverse General Internal Medicine Patients for a Telephone Survey on Physician–Patient Communication." *Journal of General Internal Medicine* 20 (5): 438–43.
- Nápoles-Springer, A. M., J. Santoyo, H. O'Brien, and A. L. Stewart. 2006. "Using Cognitive Interviews to Develop Surveys in Diverse Populations." *Medical Care* 44 (11, Suppl. 3): in press.
- Safran, D. G., M. 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.
- Saha, S., M. Komaromy, T. D. Koepsell, and A. B. Bindman. 1999. "Patient–Physician Racial Concordance and the Perceived Quality and Use of Health Care." *Archives of Internal Medicine* 159 (9): 997–1004.
- Smedley, B. D., A. Y. Stith, and A. R. Nelson, (Eds.). 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academies Press.
- Steenkamp, J. B. E. M., and H. Baumgartner. 1998. "Assessing Measurement Invariance in Cross-National Consumer Research." *Journal of Consumer Research* 25: 78–90.
- Steiger, J. H. 1990. "Structural Model Evaluation and Modification: An Interval Estimation Approach." *Multivariate Behavioral Research* 25 (2): 173–80.
- Stewart, M. A. 1995. "Effective Physician–Patient Communication and Health Outcomes: A Review." *Canadian Medical Association Journal* 152 (9): 1423–33.
- Stewart, A. L., R. D. Hays, and J. E. Ware. 1992. "Methods of Constructing Health Measures." In *Measuring Functioning and Well-Being: The Medical Outcomes Study Approach*, edited by A. L. Stewart and J. E. Ware, pp. 67–85. Durham, NC: Duke University Press.
- Stewart, A. L., A. Nápoles-Springer, E. J. Pérez-Stable, S. F. Posner, A. B. Bindman, H. L. Pinderhughes, and A. E. Washington. 1999. "Interpersonal Processes of Care in Diverse Populations." *Milbank Quarterly* 77 (3): 305–39.
- Weech-Maldonado, R., L. S. Morales, M. Elliott, K. Spritzer, G. Marshall, and R. D. Hays. 2003. "Race/Ethnicity, Language, and Patients' Assessments of Care in Medicaid Managed Care." *Health Services Research* 38 (3): 789–807.