HSR Health Services Research

DOI: 10.1111/1475-6773.12741
RESEARCH ARTICLE

Cross-Cultural Validation of the Patient Perception of Integrated Care Survey

Maike V. Tietschert, Federica Angeli, Arno J. A. van Raak, Dirk Ruwaard, and Sara J. Singer

Objective. To test the cross-cultural validity of the U.S. Patient Perception of Integrated Care (PPIC) Survey in a Dutch sample using a standardized procedure.
Data Sources. Primary data collected from patients of five primary care centers in the south of the Netherlands, through survey research from 2014 to 2015.
Study Design. Cross-sectional data collected from patients who saw multiple health care providers during 6 months preceding data collection.
Data collection. The PPIC survey includes 59 questions that measure patient perceived care integration across providers, settings, and time. Data analysis followed a standardized procedure guiding data preparation, psychometric analysis, and included invariance testing with the U.S. dataset.
Principal Findings. Latent scale structures of the Dutch and U.S. survey were highly comparable. Factor "Integration with specialist" had lower reliability scores and noninvariance. For the remaining factors, internal consistency and invariance estimates were strong.
Conclusions. The standardized cross-cultural validation procedure produced strong support for comparable psychometric characteristics of the Dutch and U.S. surveys. Future research should examine the usability of the proposed procedure for contexts with greater cultural differences.
Key Words. Cross-cultural validation, standardization, health system outcome measures, Patient Perception of Integrated Care Survey

Global interest in improving health care systems and the growing number of multicultural research projects call for methods that validly and reliably assess health system outcomes across cultural contexts (Berzon, Hays, and Shumaker 1993; Guillemin, Bombardier, and Beaton 1993; Bullinger et al. 1998; Arah et al. 2006; Sousa and Rojjanasrirat 2011). This need is reinforced by diversifying cultures within national boundaries (Nijkamp and Poot 2015), such as in the United States (Acquadro et al. 2008). In 2000, 18 percent of the

U.S. population on average, and up to 93 percent in some areas, spoke a language other than English at home (Shin and Bruno 2003; Hurtado et al. 2005; Pan and Puente 2005). Because cultural groups may vary in their perception and expression of health, as well as their use and experience of health care, valid and reliable measures should be sensitive to cultural differences (Guillemin, Bombardier, and Beaton 1993). When concepts are studied across cultures, measures should be context sensitive and cross-culturally comparable. This study focuses on the development of measures suitable for cross-cultural comparison.

We adopt a broad definition of culture "as system of symbolic meanings that shapes both social reality and personal experience, mediates between the external and internal parameters of medical systems, and thereby is a major determinant of their content, effects and the changes they undergo" (Kleinman 1978, p. 86). According to this definition, health and health care systems are embedded in cultural systems and hence cannot be examined in isolation (Kleinman 1978). Thus, in performing cross-cultural validation, we aim to account for differences in cultures and systems.

A measure is cross-culturally valid if it "functions as intended and has the same properties as the original" (Epstein, Santo, and Guillemin 2015, p. 436). It requires equivalence on five dimensions (Epstein, Santo, and Guillemin 2015). *Conceptual equivalence* requires domains to be equally relevant and meaningful to the explored concept. *Item equivalence* exists if items are as relevant and acceptable. *Semantic equivalence* requires items to have equal meaning. A measure is *operationally equivalent* if it can be used in the same way by its target population. *Measurement equivalence* requires equal psychometric properties, that is, construct validity, reliability, and responsiveness of both measures.

Conceptual, item, semantic, and operational equivalence are addressed prior to and during the translation and adaptation process, in which the measure is transferred to the new context (Epstein, Santo, and Guillemin 2015). Methods for doing so have been addressed through multiple guidelines such as the WHO guidelines (World Health Organization n.d.) and the IQOLA approach (Bullinger et al. 1998). In comparison, standardization of methods

---

Address correspondence to Maike V. Tietschert, M.Sc., Department of Health Services Research, Maastricht University, Room 0.043, Duboisdomein 30, 6229 GT Maastricht, The Netherlands; e-mail: m.tietschert@maastrichtuniversity.nl. Federica Angeli, Ph.D., Arno J. A. van Raak, Ph.D., and Dirk Ruwaard, Ph.D., are with Department of Health Services Research, Maastricht University, Maastricht, The Netherlands. Sara J. Singer, Ph.D., is with the Department of Health Policy and Management, Harvard T. H. Chan School of Public Health, Boston, MA.

for testing measurement equivalence has received limited attention (Epstein, Santo, and Guillemin 2015). As a result, cross-cultural validation procedures vary greatly in their degree of completeness (Guillemin, Bombardier, and Beaton 1993). Most often psychometric analyses are restricted to reliability and validity of the translated measure but do not assess cross-cultural comparability (measurement equivalence) (Bautista et al. 2016). When psychometric properties from the translated measure are compared with the original, results typically provide limited support for comparability (e.g., Pfeiffer and Manser 2010; Perneger, Staines, and Kundig 2014; Gehring et al. 2015). These inconsistencies hamper investigators' ability to understand how cultural characteristics impact construct formations and perceptions and to draw conclusions about the comparability of study findings across cultures. Hence, in addition to validation processes for the translation and adaptation, standardized, evidence-based procedures for testing measurement equivalence are necessary to increase potential sample populations beyond national boundaries (Epstein, Santo, and Guillemin 2015).

With this study, we aim to promote method standardization for establishing measurement equivalence to complement guidelines for cross-cultural validations. To do so, we present lessons learned from the psychometric analysis of the Patient Perception of Integrated Care (PPIC) survey (Singer et al. 2012) for cross-cultural application. The PPIC survey is a self-administered tool that was successfully transferred from the United States to the Netherlands. It measures integrated care from the patient's perspective, independently of the organizational arrangement of the patient's providers, where integrated care is defined as "patient care that is coordinated across professionals, facilities, and support systems; continuous over time and between visits; tailored to the patients and family members' needs and preferences; and based on shared responsibility between patient and caregivers for optimizing health" (Singer et al. 2011).

The survey was previously translated and adapted for use in the Dutch context and conceptual, item, semantic, and operational equivalence were established (Tietschert et al. 2016). In this study, we assess the transferability of health-related, patient-reported outcome measures for application in cross-cultural comparative research and offer a guide for psychometric testing. We do so by comparing the measurement properties of the Dutch PPIC survey to the previously validated US PPIC survey using a standardized methodology. We also show that the PPIC survey can reliably and validly measure integrated care across U.S. and Dutch contexts.

Establishing the cross-cultural validity of this survey provides opportunity to compare patient-perceived care integration in the United States and

the Netherlands. Both countries invest in various reforms to improve care integration. Examples are accountable care organizations in the United States (Shortell et al. 2015) and bundled payments for chronic care in the Netherlands (Struijs and Baan 2011). Despite these initiatives, it is still poorly understood which approaches improve different aspects of care integration (Tietschert et al. 2016). The PPIC survey can serve as research tool to compare interventions intended to improve care integration and to guide refinement of delivery system innovation. Using the survey in this way would generate knowledge about which interventions can improve integration. Plans for leveraging the cross-culturally validated PPIC survey include comparing the U.S. and Dutch data to identify strengths and weaknesses in patient-perceived care integration of each delivery system in order to foster cross-cultural learning.

METHODS

In the following section, we detail the stepwise procedure used to assess measurement equivalence of the PPIC survey. Because the focus is on the Dutch survey, we provide information about U.S. processes only as needed for comparison. We begin by describing the U.S. survey and proceed with explaining how the different domains of cross-cultural validity were explored.

U.S. Survey—Development and Survey Characteristics

The PPIC survey (version 2.1) was finalized in 2014 to measure patients' perceptions of care integration for application in the United States. The survey was theoretically derived and developed through pilot testing, cognitive testing, and input from an advisory panel of survey measurement and care integration experts, patient representatives, and patients in multiple rounds (Singer et al. 2012; Fryer et al. 2016). Reliability and validity were established (Kerrissey et al. 2017). The survey is designed particularly for administration to patients with complex needs and includes 59 questions about patients' experiences of care across settings, including their primary provider's office, specialists, hospitals, at home, and over time.

The survey measures patient-perceived integration on six dimensions (Kerrissey et al. 2017): (1) *Provider Knowledge of the Patient* assesses the extent to which providers are well informed, up to date, and familiar with the patient's needs and values; (2) *Staff Knowledge About the Patient's Medical History*

measures how familiar staff within in the providers office is with the patient's current and prior medical information; (3) *Specialist Knowledge About the Patient's Medical History* addresses familiarity of the specialist outside the provider's office with the patient's current and prior medical information; (4) *Support for Self-Directed Care* asks to which extent providers encourage, enable, and support the patient to perform self-care; (5) *Support for Medication and Home Health Management* covers how well providers orient patients to their medications and provide support between visits; and (6) *Test Result Communication* addresses the efficacy and timing of efforts made to share test result information with the patient. Integrated care dimensions and corresponding items are available on request. For respondents who had been hospitalized, survey responses also produced an index of integration following hospitalization. Additional items assess demographic characteristic, personal information, general health ratings, and the CAHPS communication construct (Hargraves, Hays, and Cleary 2003).

Testing Conceptual, Item, Semantic, and Operational Equivalence

Dutch Survey—Cultural Translation and Adaptation Process. From November 2013 to August 2014, the U.S. survey was translated and adapted for use in the Netherlands according to the guidelines for translating and culturally adapting survey instruments offered by the World Health Organization (n.d.). The process included a forward–backward translation of the survey from English to Dutch and back to English, aimed at testing semantic equivalence (Chidlow, Plakoyiannaki, and Welch 2014). Thereafter, semantic, operational, and item equivalence were tested in cognitive interviews with respondents who received care from multiple providers (Chidlow, Plakoyiannaki, and Welch 2014). The final Dutch version was produced by adapting survey items based on these tests (Tietschert et al. 2016).

Testing Measurement Equivalence

Dutch Survey Administration: Setting, Data Collection, and Study Population. Differences in U.S. and Dutch health care systems required different approaches to survey administration (Table S1, Appendix SA2). Setting, study population, data collection, and content of the distributed survey packages were similar. Sampling method, survey distributor, and data collection period differed. To test the cross-cultural validity of the Dutch survey, data were collected in five primary care centers in the South of the Netherlands (January

2014–February 2015). The centers varied in size, organizational maturity, and provider mix, but all included multiple providers and a heterogeneous patient mix.

Respondents were selected from the patient population of each primary care center. Several survey dimensions assess interprovider collaboration. Hence, we selected all adult patients (18+) that had been seen by more than one health care provider. The Dutch health care system leverages a gatekeeper concept, in which the patient commonly enters the care process via the general practitioner (GP), who then determines whether referral to a different provider is needed. Because electronic patient records did not allow tracking of referred patients, health care providers other than GPs selected all adult patients that they had seen during the 6 months preceding data collection. These providers—for example, dieticians, physiotherapists, and social workers—worked at the primary care center. We assumed a high likelihood that these patients had also been seen by the GP. The lists were cleaned for duplicates. Patients who were deceased were excluded. Each respondent was anonymized with an individual ID number. The primary care centers distributed the written survey by mail with an accompanying letter explaining the study purpose, that participation was voluntary and how to contact the leading researcher when additional information was desired. When no response was provided, a reminder was sent 3 weeks after the survey was posted. Respondents were thus contacted two times at maximum.

Standardizing the Cross-Cultural Validation of Measurement Equivalence. We hypothesized that assessing comparability of a translated survey with the original requires substantive equivalence of the survey itself and equivalence of the cross-cultural validation procedure. Only then differences or similarities in psychometric results between surveys are unambiguously interpretable. To standardize the cross-cultural validation, the first author composed a manual for analysis of the PPIC survey in partnership with the U.S.-based PPIC development team. In this manual, the validation of the PPIC survey in the United States was reproduced step by step. We reviewed all information relevant to the validation process applied to the U.S. dataset and compiled the information in a single document. To check for accuracy, each section was presented to the person responsible for the respective analysis in the United States. The final manual provided detailed information on the data preparation and psychometric testing. For details on the data preparation, we refer to the technical appendix 1 "Data Preparation." Psychometric testing included (1) analyzing sample

and survey properties, (2) testing the survey latent scale structure, and (3) testing invariance. The manual is available from authors on request.

Psychometric Testing

Sample and Survey Properties. To study comparability between the U.S. and Dutch survey data, we used the same items that were included in the analysis of the U.S. data. To test sample properties, respondents' demographic characteristics were examined. We studied survey properties to detect potential problems related to survey comprehension (Ware and Gandek 1998). We studied item response rate and ran generalized linear regression models to test for influences of respondent characteristics on missing values. We computed frequency distributions of survey items to determine whether all response choices were used. If response choices are underrepresented, this may indicate inadequate translation or insufficient adaptation for the new cultural context being studied (Ware and Gandek 1998).

Survey's Latent Scale Structure. Our main interest was to determine comparability of the U.S. and the Dutch survey. However, imposing the U.S. structure from the outset would not have detected potential alternative and maybe better fitting structures. Based on procedures suggested by Pett, Lackey, and Sullivan (2006), we started with an exploratory factor analysis (EFA) to determine the Dutch survey's latent scale structure. We first computed and visually examined a pairwise correlation matrix to determine whether items had adequate correlations to warrant grouping them. If items are too highly correlated, it may be impossible to determine their unique contribution to a factor, causing problems for factor analysis. Therefore, several sources advise eliminating variables with strong correlations as there is not enough unique contribution by each variable (e.g., Pett, Lackey, and Sullivan 2006). If items had correlations higher than 0.80, we excluded one item based on theoretical considerations and after confirming that the choice did not result in ceiling effects or significant changes to the factor structure. We assessed sampling adequacy by comparing the magnitude of the calculated correlation coefficients to the magnitude of the partial correlation coefficients with the Kaiser–Meyer–Olkin (KMO) Test (Kaiser 1970; Kaiser and Rice 1974). Values larger than 0.60 indicate adequate sample size. We applied Bartlett's test of sphericity (Bartlett 1937) to determine whether items in the correlation matrix are sufficiently related. Under the null hypothesis, the correlation matrix is an identity matrix

with uncorrelated items and factor analysis would be inadvisable. A significant chi-square ($p < .05$) indicates that the correlation matrix is not an identity matrix. We proceeded with an EFA, in which we applied Kaiser's eigenvalue-greater-than-one decision rule (Kaiser 1960). We used Promax Oblique Rotation to consider potential correlations between integrated care dimensions. Items that loaded on multiple factors were assigned to the factor with the highest loading (Arah et al. 2006). Factors' internal consistency was assessed with Cronbach's alpha (Cronbach 1951), which is a reliability estimate of shared variance among items (Ware and Gandek 1998). To examine discriminant validity, Cronbach's alphas were compared to the correlations with all other remaining scales. Cronbach's alphas higher than the remaining correlations indicate unique variance and thus suggest that scales measure distinct factors (Ware and Gandek 1998). To test goodness of fit, we used confirmatory factor analysis (CFA), applying criteria suggested by Brown (2006). Analyses were run on the full sample. To check robustness, we repeated analyses with the split-half procedure in which EFA and CFA were rerun on separate subsamples. The procedure did not yield appreciably different results.

Invariance Testing. To assess comparability between the U.S. and Dutch survey, we tested invariance (2d) of the U.S. and Dutch datasets. The Dutch latent scale (factor) structure was visually compared with the U.S. structure, and an invariance test was performed according to the procedure suggested by Dimitrov (2010). To evaluate model fit, we applied criteria provided by Cheung and Rensvold (2002).

Results of the Cross-Cultural Validation

Sample Properties

A total of 5,991 surveys were distributed, of which 62 were returned undeliverable because respondents had changed address or were deceased. We received back a total of 3,725 surveys (response rate of 62.33 percent). Table 1 details respondent demographics. Eighty-five percent of the respondents were 55 years or older, 52.4 percent were female, and 34 percent patients had attained at most general secondary education or primary vocational education. Respondents were predominantly Dutch (93.6 percent). Almost 82 percent of the respondents had moderate-to-good health, and 81.7 percent completed the survey without help.

Table 1: Respondent Demographics

	Respondents	
	N	*Percentage*
Total	3,725	100
Age (years)		
18–24	21	0.6
25–34	33	0.9
35–44	78	2.1
45–54	373	10.0
55–64	948	25.4
65–74	1,251	33.6
≥75	978	26.3
Gender		
Male	1,714	46.0
Female	1,951	52.4
Level of education		
Low level of education	782	21.0
Middle 1 (general secondary education, primary vocational education)	1,271	34.1
Middle 2 (general secondary education, preuniversity education, secondary vocational education)	689	18.5
High educational level (higher degree of education and university)	513	13.8
Other	364	9.8
Country of origin		
Dutch	3,485	93.6
German	72	1.9
Turkish	10	0.3
Other	107	2.8
Self-reported health		
Excellent	73	2.0
Very good	346	9.3
Good	1,903	51.1
Fair	1,144	30.7
Poor	154	4.1
Had help completing the survey		
Yes	634	17.0
No	3,045	81.7

Survey Properties

We assessed item total and relevant nonresponse. Nonresponse is the absence of a response to an item that the respondent should have answered, that is, not skipped appropriately according to the survey skip pattern. The total nonresponse rate compares the total number of blank responses (not skipped

properly) and the total number of respondents. Of the items included in our analysis, total nonresponse rate per item averaged 5.4 percent. The relative nonresponse rate for a given item compares the number of people who should have answered a question based on the response to the gate question to the number of people who answered. The relative nonresponse averaged 3.3 percent. Older respondents (>75), women and patients who filled in the survey without help had significantly more missing values.

Percentage top-box of items with 4-point scales ranged from 7 percent (Q42) to 93 percent (Q 6). Item means ranged from 1.44 (Q42) to 3.92 (Q6) with standard deviations of at least 0.28 (Table 2).

Survey's Latent Scale Structure

To study appropriateness of exploratory factor analysis, we examined the correlation matrix. Correlations with other items in the proposed scale below 0.30 were revealed for combination question 4/5, question 6, question 9, question 10, and question 40. Combination question 34/35 correlated above the 0.80 threshold with combination question 34/36. As a result, SPSS produced a warning that the matrix was not positive definite, indicating a collinearity problem. The first combination question consisted of the following[1]: *Question 34. In the last 6 months, did this provider or someone in his or her office give you instructions about how to take care of your health? Question 35. In the last 6 months, were the instructions you received easy to follow?* The second combination question included the following[1]: *Question 34. In the last 6 months, did this provider or someone in his or her office give you instructions about how to take care of your health? Question 36. In the last 6 months, how often did the instructions you received help you take care of your health?* Because the first of each pair of items is identical, the combination questions correlated strongly with each other (0.95). As both items had similar means, we excluded combination question 34/35 based on theoretical considerations and in accordance with the U.S. factor analysis. We maintained that if respondents felt instructions were helpful, they would be able to follow these instructions. We tested both items in separate factor analyses, which confirmed that our decision did not affect our factor structure. Excluding items that did not meet the required thresholds from the factor analysis solved associated problems. Item 8 was removed because Bartlett's test of sphericity suggested insufficient sample size.

Twenty-one items were included in the EFA. Kaiser's eigenvalues suggested retention of six factors, where each item loaded on at least one factor with loadings above 0.40. All but five items had cross-loadings higher than

Table 2: Items Response, Mean, and Standard Deviation

Item Number/Item Text	*N*	*Response Scale*	*Percentage Top Box**	*Mean*	*SD*
Combination question 4 and 5. Some offices remind patients about appointments. Before your most recent visit with this provider, did you get a reminder from this provider's office about the appointment? Before your most recent visit with this provider, did you get instructions telling you what to expect or how to prepare for the visit?	2,819	Combination 4-point[†]	9	1.50	0.94
Question 6. In the last 6 months, how often did this provider cancel or change the date of an appointment?	2,878	A/U/S/N[‡]	93	3.92	0.28
Question 9. In the last 6 months, how often did you have to repeat information that you had already provided during the same visit?	2,846	A/U/S/N	79	3.75	0.53
Question 10. In the last 6 months, how often did this provider seem to know the important information about your medical history?	2,827	A/U/S/N	35	2.78	1.10
Question 17. How would you rate this provider's knowledge of your values and beliefs that are important to your health care?	2,878	E/G/F/P[§]	36	3.25	0.65
Combination question 18 and 19. In the last 6 months, did this provider talk with you about setting goals for your health? In the last 6 months, did the care you received from this provider help you meet your goals?	2,753	Combination 4-point[¶]	30	2.48	1.28
Question 21. In the last 6 months, how often did these other staff seem up-to-date about the care you were receiving from this provider?	2,190	A/U/S/N	57	3.42	0.76
Question 22. In the last 6 months, how often did these other staff talk to you about the care you were receiving from this provider?	2,176	A/U/S/N	27	2.58	1.11
Question 23. In the last 6 months, how often did these other staff seem to know the important information about your medical history?	2,153	A/U/S/N	41	3.02	1.00
Question 25. In the last 6 months, when this provider or someone in his or her office ordered a blood test, x-ray, or other test for you, how often did this provider or someone from his or her office follow up to give you those results?	2,636	A/U/S/N	41	2.65	1.30

Continued

Table 2 *Continued*

Item Number/Item Text	*N*	*Response Scale*	*Percentage Top Box**	*Mean*	*SD*
Question 26. In the last 6 months, how often did you have to request your test results before you got them?	2,645	A/U/S/N	75	2.59	1.22
Question 27. In the last 6 months, how often were your test results presented in a way that was easy to understand?	2,664	A/U/S/N	69	3.57	0.73
Combination question 29 and 30. In the last 6 months, did this provider or someone in his or her office ask you about these things that make it hard for you to take care of your health? In the last 6 months, did you and this provider or someone in his or her office come up with a plan to help you deal with the things that make it hard for you to take care of your health?	736	Combination 4-point**	27	2.28	1.28
Question 31. In the last 6 months, how often did this provider or someone in his or her office help you identify the most important things for you to do for your health?	3,599	A/U/S/N	25	2.36	1.19
Question 33. In the last 6 months, how often did this provider or someone in his or her office help you get these services at home to take care of your health?	391	A/U/S/N	23	2.04	1.26
Combination question 34 and 35. In the last 6 months, did this provider or someone in his or her office give you instructions about how to take care of your health? In the last 6 months, how often were you able to follow these instructions about taking care of your health?	3,456	Combination 4-point††	15	1.88	1.18
Combination question 34 and 36. In the last 6 months, did this provider or someone in his or her office give you instructions about how to take care of your health? In the last 6 months, how often did the instructions you received help you take care of your health?	3,454	Combination 4-point‡‡	11	1.79	1.08
Question 37. In the last 6 months, if you had any trouble taking care of your health at home, would you know who to ask for help?	3,588	D/S/N§§	50	2.20	0.87
Question 39. In the last 6 months, how often did this provider or someone in his or her office talk with you about how you were supposed to take your medicine?	3,300	A/U/S/N	83	2.87	1.27

Continued

Table 2 *Continued*

Item Number/Item Text	*N*	*Response Scale*	*Percentage Top Box**	*Mean*	*SD*
Question 40. There are many reasons why people may not always be able to take their medicines as prescribed. In the last 6 months, how often were you able to take your medicine as prescribed?	3,351	A/U/S/N	83	3.79	0.53
Question 41. In the last 6 months, how often did this provider or someone in his or her office talk with you about what to do if you have a bad reaction to your medicine?	3,301	A/U/S/N	29	2.29	1.29
Question 42. In the last 6 months, how often did this provider or someone in his or her office contact you between visits to see how you were doing?	3,586	A/U/S/N	7	1.44	0.90
Question 46. In general, how often does the provider named in Question 1 seem informed and up-to-date about the care you get from specialists?	1,822	A/U/S/N	60	3.40	0.85
Question 47. In general, how often do you have to remind the provider named in Question 1 about care you receive from specialists?	1,816	A/U/S/N	66	3.50	0.81
Question 49. In general, how often does the provider named in Question 1 talk with you about the medicines prescribed by these specialists?	1,013	A/U/S/N	17	2.06	1.15
Question 50. These questions ask about care you received from the specialist you saw most often in the last 6 months outside the office of the provider named in question 1. When you see this specialist, does he or she seem to know enough information about your medical history?	1,804	D/S/N	65	2.55	0.65
Question 52. When you see this specialist, how often does this specialist seem to know your important test results from other providers?	1,722	A/U/S/N	26	2.46	1.16

*Percentage top box is the percent of responses in the most favorable category.
†YesReminder&YesInstruction/YesReminder&NoInstructions/NoReminder&YesInstruction/NoReminder&NoInstruction.
‡Always/Usually/Sometimes/Never.
§Poor/Fair/Good/Excellent.
¶NoTalk/Talk&NoHelp/Talk&SomeHelp/Talk&DefinitelyHelp.
**NoAsk/Ask&NoPlan/Ask&SomePlan/Ask&DefinitelyPlan.
††NoInstruc/Instruc&NeverSomeFollow/Instruc&UsuallyFollow/Instruc&AlwaysFollow.
‡‡NoInstruc/Instruc&NeverSomeHelp/Instruc&UsuallyHelp/Instruc&AlwaysHelp.
§§Yes, definitely/Yes, somewhat/No.

0.30 but primary loadings were substantially higher for most of the items (Table 3).

We calculated Cronbach's alpha to assess factors' internal consistency. Factor 1 (mostly consistent with *Support for Self-Directed Care*), factor 3 (*Staff Knowledge of Patient's Medical History*), and factor 5 (mostly consistent with *Test Result Communication*) had acceptable to strong consistency (>0.70). Factor 2 (*Support for Medication and Home Health Management*) and factor 4 (mostly consistent with *Provider Knowledge of Patient*) had moderate consistencies (0.66 & 0.63). Cronbach's alpha for factor 6 (Specialist Knowledge of Patient's Medical History) was lower at 0.44. We explored the effect on reliability of reallocating items. Specifically, we reallocated items with high loadings on more than one factor to the factor for which the item exhibited the second highest loadings. This reallocation of items did not result in improvements, and removing items from the factor to which they were originally assigned reduced Cronbach's alphas. We therefore adopted the structure that was proposed based on the factor analysis. Comparing Cronbach's alphas with the correlation of that scale with all other remaining scales supported the discriminant validity of factors (Table S3, Appendix SA2).

We tested model fit of the identified scales using confirmatory factor analysis and calculated goodness-of-fit indices (Brown 2006) (Table S4, Appendix SA2). The Root Mean Squared Error of Approximation (RMSEA) met the acceptable threshold of 0.06 or lower (0.04). The Standardized Root Mean Square Residual (SRMR) met the acceptable threshold of 0.08 or lower (0.05). The Comparative fit index (CFI) and Tucker Lewis index (TLI) met the conventional threshold of 0.90 but were somewhat lower than the conservative threshold of 0.95 (CFI 0.93 and TLI 0.91). Because the missing at random assumption is the default setting for these tests but respondents older than 75 and female respondents had significantly more missing values, we repeated analyses for subsamples of male respondents younger than 75. Scores remained consistent for all but one sample, and the threshold was met for the overall subsample (CFI 0.93 and TLI 0.91).

Invariance Test: Comparability of the Dutch and U.S. Surveys

To assess the comparability of the Dutch and the U.S. PPIC survey, we examined factorial invariance. Configural invariance requires model configurations, that is, the patterns of free and fixed model parameters, to be equal across compared groups (Dimitrov 2010). Factor structures of both datasets

Table 3: Patient-Level Exploratory Factor Analysis

Scales Names		*Staff Knowledge of Patient's Medical History*	*Support for Self-Directed Care*	*Test Result Communication*	*Provider Knowledge of the Patient*	*Support for Medication and Home Health Management*	*Specialist Knowledge of Patient's Medical History*
	Factor numbers U.S.\ Dutch	Factor 3	Factor 1	Factor 5	Factor 4	Factor 2	Factor 6
Question 21. In the last 6 months, how often did these other staff seem up-to-date about the care you were receiving from this provider?	***1***	**0.79**					
Question 22. In the last 6 months, how often did these other staff talk to you about the care you were receiving from this provider?	***1***	**0.73**				0.47	
Question 23. In the last 6 months, how often did these other staff seem to know the important information about your medical history?	***1***	**0.83**					
Combination question 18 and 19. In the last 6 months, did this provider talk with you about setting goals for your health? In the last 6 months, did the care you received from this provider help you meet your goals?	***2***		**0.53**			0.5	

Continued

Table 3 *Continued*

Scales Names		*Staff Knowledge of Patient's Medical History*	*Support for Self-Directed Care*	*Test Result Communication*	*Provider Knowledge of the Patient*	*Support for Medication and Home Health Management*	*Specialist Knowledge of Patient's Medical History*
Combination question 29 and 30. In the last 6 months, did this provider or someone in his or her office ask you about these things that make it hard for you to take care of your health? In the last 6 months, did you and this provider or someone in his or her office come up with a plan to help you deal with the things that make it hard for you to take care of your health?	***2***		**0.83**			0.44	
Question 31. In the last 6 months, how often did this provider or someone in his or her office help you identify the most important things for you to do for your health?	***2***	0.46	**0.8**			0.49	
Question 33. In the last 6 months, how often did this provider or someone in his or her office help you get these services at home to take care of your health?	***2***		**0.71**	0.48		0.4	

Continued

Table 3 *Continued*

Scales Names		*Staff Knowledge of Patient's Medical History*	*Support for Self-Directed Care*	*Test Result Communication*	*Provider Knowledge of the Patient*	*Support for Medication and Home Health Management*	*Specialist Knowledge of Patient's Medical History*
Combination question 34 and 36. In the last 6 months, did this provider or someone in his or her office give you instructions about how to take care of your health? In the last 6 months, how often did the instructions you received help you take care of your health?	***2***		**0.8**			0.33	
Question 26. In the last 6 months, how often did you have to request your test results before you got them?	***3***			**0.9**			
Question 25. In the last 6 months, when this provider or someone in his or her office ordered a blood test, x-ray, or other test for you, how often did this provider or someone from his or her office follow up to give you those results?	***3***			**0.9**			
Question 27. In the last 6 months, how often were your test results presented in a way that was easy to understand?	***3***	0.45			**0.51**		

Continued

Table 3 *Continued*

Scales Names		*Staff Knowledge of Patient's Medical History*	*Support for Self-Directed Care*	*Test Result Communication*	*Provider Knowledge of the Patient*	*Support for Medication and Home Health Management*	*Specialist Knowledge of Patient's Medical History*
Question 17. How would you rate this provider's knowledge of your values and beliefs that are important to your health care?	***4***				**0.65**		
Question 46. In general, how often does the provider named in Question 1 seem informed and up-to-date about the care you get from specialists?	***4***				**0.75**		
Question 47. In general, how often do you have to remind the provider named in Question 1 about care you receive from specialists?	***4***				**0.74**		
Question 39. In the last 6 months, how often did this provider or someone in his or her office talk with you about how you were supposed to take your medicine?	***5***					**0.75**	

Continued

Table 3 *Continued*

Scales Names		*Staff Knowledge of Patient's Medical History*	*Support for Self-Directed Care*	*Test Result Communication*	*Provider Knowledge of the Patient*	*Support for Medication and Home Health Management*	*Specialist Knowledge of Patient's Medical History*
Question 41. In the last 6 months, how often did this provider or someone in his or her office talk with you about what to do if you have a bad reaction to your medicine?	***5***					**0.77**	
Question 42. In the last 6 months, how often did this provider or someone in his or her office contact you between visits to see how you were doing?	***5***			0.46		**0.51**	
Question 49. In general, how often does the provider named in Question 1 talk with you about the medicines prescribed by these specialists?	***5***					**0.68**	
Question 52. When you see this specialist, how often does this specialist seem to know your important test results from other providers?	***6***						**0.81**

Continued

Table 3 *Continued*

Scales Names		*Staff Knowledge of Patient's Medical History*	*Support for Self-Directed Care*	*Test Result Communication*	*Provider Knowledge of the Patient*	*Support for Medication and Home Health Management*	*Specialist Knowledge of Patient's Medical History*
Question 50. These questions ask about care you received from the specialist you saw most often in the last 6 months outside the office of the provider named in question 1. When you see this specialist, does he or she seem to know enough information about your medical history?	***6***						**0.8**
†Question 37. In the last 6 months, if you had any trouble taking care of your health at home, would you know who to ask for help?			**0.58**				

Notes: Factor names are borrowed from U.S. factor model. Not all factors are identical but highly consistent with U.S. factor model (see Table 5).
Factor loadings less than 0.40 are suppressed.
Items are listed by hypothesized U.S. latent scale structure (bold italic numbers).
Bold numbers present the primary loading of each survey item.
†Item did not load in the U.S. factor analysis.

showed a high degree of overlap (Table 4). Of 21 items that were included in the analysis, 19 items loaded on the same factors in both datasets.

Next, we ran separate confirmatory factor analyses with the Dutch and the U.S. datasets, including the 19 overlapping items to test model fit for each dataset. The results are displayed in Tables S5 and S6 in the Appendix SA2. Factor loadings and CFI, TLI, and RMSEA met the conventional thresholds. Factor loading for item 26 in the U.S. dataset was somewhat lower than the threshold of 0.40. Nevertheless, configural invariance was confirmed for all factors but factor 6, for which factor loadings differed by 0.45 between datasets (items Q50).

We then examined metric invariance, which exists if equal factor loadings across groups indicate equivalent relationships between a latent factor and its indicators derived through CFA (Dimitrov 2010). To test for metric invariance, we combined the U.S. and Dutch databases. To run a difference test, we computed two models with increasing restrictions. The 0-model had no parameters constrained to be equal. Model 1 had factor loadings constrained to be equal (Dimitrov 2010). Because of chi-square's sensitivity to large sample size, we followed the criteria suggested by Cheung and Rensvold (2002) and used CFI and RMSEA difference tests (Chen 2007). When all six factors were included in the models, no convergence could be achieved. We then continued to test for partial invariance by excluding Factor 6 (*Specialist knowledge of Patient's Medical History*) from the model due to large difference in factor loadings in the Dutch and U.S. datasets. In the new models, acceptable thresholds of a negative ΔCFI value higher than −0.01 (−0.005) were met. These results were supported by the ΔRMSEA and ΔSRMR (Table 5).

DISCUSSION

To our knowledge, this is the first study in which psychometric characteristics of a culturally transferred patient survey were assessed following a standard procedure. Through a standardized validation of measurement equivalence, we demonstrated the validity and reliability of the PPIC survey for measuring integrated patient care across U.S. and Dutch contexts and cultures. Item nonresponse for the Dutch survey indicated that items were generally comprehensible and consistent with the U.S. survey (5.4 percent vs. 4.8 percent). Responses to the Dutch survey were also sufficiently varied, with percentage top-box scores that were about the same as in the U.S. data (none over 79 percent) for all but three items (83 percent and 93 percent). Psychometric

Table 4: Comparison of Latent Scale Structure, U.S. and Dutch Datasets

Item Number/Item Text	*Factor Names*	*Factors U.S.*	*Factors Dutch*
Question 21. In the last 6 months, how often did these other staff seem up-to-date about the care you were receiving from this provider?	Staff Knowledge of Patient's Medical History	1	3
Question 22. In the last 6 months, how often did these other staff talk to you about the care you received from this provider?	Staff Knowledge of Patient's Medical History	1	3
Question 23. In the last 6 months, how often did these other staff seem to know the important information about your medical history?	Staff Knowledge of Patient's Medical History	1	3
Combination question 18 and 19. In the last 6 months, did this provider talk with you about setting goals for your health? In the last 6 months, did the care you received from this provider help you meet your goals?	Support for Self-Directed Care	2	1
Combination question 29 and 30. In the last 6 months, did this provider or someone in his or her office ask you about these things that make it hard for you to take care of your health? In the last 6 months, did you and this provider or someone in his or her office come up with a plan to help you deal with the things that make it hard for you to take care of your health?	Support for Self-Directed Care	2	1

Continued

Table 4 *Continued*

Item Number/Item Text	*Factor Names*	*Factors U.S.*	*Factors Dutch*
Question 31. In the last 6 months, how often did this provider or someone in his or her office help you identify the most important things for you to do for your health?	Support for Self-Directed Care	2	1
Question 33. In the last 6 months, how often did this provider or someone in his or her office help you get these services at home to take care of your health?	Support for Self-Directed Care	2	1
Combination question 34 and 36. In the last 6 months, did this provider or someone in his or her office give you instructions about how to take care of your health? In the last 6 months, how often did the instructions you received help you take care of your health?	Support for Self-Directed Care	2	1
Question 26. In the last 6 months, how often did you have to request your test results before you got them?	Test Result Communication	3	5
Question 25. In the last 6 months, when this provider or someone in his or her office ordered a blood test, x-ray, or other test for you, how often did this provider or someone from his or her office follow up to give you those results?	Test Result Communication	3	5
Question 27. In the last 6 months, how often were your test results presented in a way that was easy to understand?	Test Result Communication	3	4

Continued

Table 4 *Continued*

Item Number/Item Text	*Factor Names*	*Factors U.S.*	*Factors Dutch*
Question 17. How would you rate this provider's knowledge of your values and beliefs that are important to your health care?	Provider Knowledge of Patient	4	4
Question 46. In general, how often does the provider named in Question 1 seem informed and up-to-date about the care you get from specialists?	Provider Knowledge of Patient	4	4
Question 47. In general, how often do you have to remind the provider named in Question 1 about care you receive from specialists?	Provider Knowledge of Patient	4	4
Question 39. In the last 6 months, how often did this provider or someone in his or her office talk with you about how you were supposed to take your medicine?	Support for Medication and Home Health Management	5	2
Question 41. In the last 6 months, how often did this provider or someone in his or her office talk with you about what to do if you have a bad reaction to your medicine?	Support for Medication and Home Health Management	5	2
Question 42. In the last 6 months, how often did this provider or someone in his or her office contact you between visits to see how you were doing?	Support for Medication and Home Health Management	5	2
Question 49. In general, how often does the provider named in Question 1 talk with you about the medicines prescribed by these specialists?	Support for Medication and Home Health Management	5	2

Continued

Table 4 *Continued*

Item Number/Item Text	*Factor Names*	*Factors U.S.*	*Factors Dutch*
Question 50. When you see this specialist, does he or she seem to know enough information about your medical history?	Specialist Knowledge of Patient's Medical History	6	6
Question 52. When you see this specialist, how often does this specialist seem to know all your test results from other providers?	Specialist Knowledge of Patient's Medical History	6	6
Question 37. In the last 6 months, if you had any trouble taking care of your health at home, would you know who to ask for help?		0*	1

*Orphan item; that is, item that did not load on a factor in the U.S. factor analysis.

Table 5: Difference Test for Partial Invariance Excluding Factor 6 from the Model

							90% CI for RMSEA	
Model	χ^2	*df*	*CFI*	*TLI*	*SRMR*	*RMSEA*	*LL*	*UL*
MODEL 0	1043.935	218	0.939	0.924	0.058	0.04	0.037	0.042
MODEL 1	1106.076	230	0.934	0.922	0.061	0.04	0.038	0.042
Δ	62.141	12	−0.005	−0.002	0.003	0		

assessment suggested a six-dimensional measurement framework that largely overlapped the six dimensions that were produced from the U.S. dataset. The empirical support for this model is strong for five of the six factors. Item factor loadings and goodness of fit of the six-factor model met standard statistical criteria. Internal consistency was moderate to good for five factors (0.6–0.84), with the lowest factor being somewhat lower than in the U.S. data (0.68–0.84). As for the U.S. survey, discriminant validity was supported by Cronbach's alphas greater than correlations with all other remaining scales, indicating that dimensions are conceptually distinct. Invariance testing, which examines the "extent to which score properties and interpretations generalize to and across population groups, settings and tasks" (Dimitrov 2010 p. 123; Messick 1995) confirmed strong partial invariance for five of the six PPIC scales and factor loadings between the Dutch and U.S. dataset. The extremely consistent results suggest that the constructs of the PPIC survey transcend cultural differences between Dutch and U.S. contexts and are promising for the transferability of such measures to assess integration outcomes across different populations and health systems.

The high degree of comparability that was achieved also supports the usefulness of a standardized procedure for measurement equivalence and shows the potential for extending guidelines for cross-cultural validations. In doing so, the methodology applied in this study can function as a template for future studies that aim to validate measurement equivalence in cross-cultural work. Adding standardized validation procedures, such as applied in this study, to best practices of cross-cultural adaptations can advance research by facilitating a better understanding of how culture influences constructs and perceptions of health system outcomes. This is a prerequisite for studying bigger samples and comparing health system changes and their effects across cultures and populations in the future research. It will enable policy makers and health care organizations to share evidence about how to achieve healthcare

system improvements and promote change. It also provides the opportunity to take into account the cultural diversity within countries when evaluating health system outcomes and to include population groups particularly vulnerable for inequalities within healthcare systems who would otherwise be left out (Arah et al. 2006; Hsia and Shen 2015). Replication of our methodology is required to confirm the usability of this standardized validation procedure. Particularly, the impact of standardization on cross-cultural evaluations in countries with greater culture differences should receive attention, because the generalizability of our findings may be limited by the relative similarity between the U.S. and Dutch cultures (Hofstede 1985). Future studies should explore survey properties at organizational level, for which our sample size did not allow. The results of our analysis confirm that the PPIC survey dimensions are relevant conceptualizations of integrated care for patients in the Netherlands. Whether these scales are exhaustive cannot be concluded from this analysis. We did, however, mitigate this uncertainty by asking respondents during the translation and adaptation process whether survey items missed any aspects. Respondents mentioned minor aspects relating to items but did not identify dimensions that our items did not cover. Due to system differences, implementing the U.S. and Dutch survey took different approaches. Nevertheless, psychometric results support comparability of the survey data across both countries.

Invariance analysis should be expanded to include residual invariance and intercept invariance, which we were unable to evaluate because the U.S. data were corrected for life orientation, which measures respondents' levels of optimism. These items that measured respondents' life orientation were included in the U.S. PPIC survey to correct for heterogeneity in the sample population. In the Netherlands, the survey was administered in a smaller geographical location with a more homogenous culture and context. Therefore, life-orientation items were not included in the Dutch version to reduce respondents' burden. Not including life-orientation questions in the Dutch survey did not seem to impact the comparability across the Dutch and the U.S. data. Most probably this is a result of the more homogenous population provided by the smaller geographical location. We advise future studies to include items that assess respondent's life orientation. However, if shortening of the survey is required and the study population has homogenous characteristics, excluding the life-orientation items could be contemplated. For the Dutch PPIC survey, possibilities for improvement are suggested for the factor related to integration with the specialist, indicated by the small Cronbach's alpha and lack of invariance. Adding items could increase the validity of this factor.

Together with the U.S.-based research team, we discussed the possibility for extending the factor with question 51 "*When you see this specialist, how often do you have to repeat information that you have already given to the provider named in Question 1?"* This item was excluded from the U.S. factor analysis based on insufficient covariance coverage and hence also from the analysis of the Dutch data. Including this item results in a small improvement in Cronbach's alfa from 0.438 to 0.467. This may provide an opportunity to expand the dimension "Integration with specialist."

CONCLUSION

The standardized cross-cultural validation procedure demonstrated that five of the six scales in the PPIC survey are reliable and valid for use within the Dutch context. Strong invariance with the U.S. survey supports the applicability of the survey for cross-cultural assessment of care integration as perceived by the patient. Scale *Specialist Knowledge About the Patient's Medical History* warrants further examination and potentially improvement. Replication studies across Europe are currently underway.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: Multiple entities provided financial support. The Netherlands Organization for Scientific Research (NWO Graduate Program) funded by the primary researcher in the project. The Netherlands Organization for Health Research and Development (ZonMw; project number 204000025) funded the data collection in the Netherlands. The Commonwealth Fund (Award #20130024), Kaiser Permanente Community Benefit Fund, and Leape Foundation for Patient Safety supported development of the U.S. PPIC survey and U.S. data collection. The authors are grateful to the participating primary care centers that provided access to their data. In order to preserve anonymity, we refrain from identifying the centers. Centers have been informed that results of the study will be published in scientific journals. There are no arrangements that require the centers to review and approve the manuscripts. Centers have been informed through presentations and fact sheets about the study results. The authors thank Niels Hameleers for his support

during the data collection and preparation. The authors are highly appreciative for the support of the U.S.-based PPIC team, their cooperativeness in sharing their data, and their valuable insights about the validation of the PPIC survey. In addition to co-author Dr. Sara Singer, we would like to thank Dr. Jonathan Clark, Dr. Mark Friedberg, Dr. Ashley Fryer, Micheala Kerrissey, and Dr. Maria Orlando Edelen for their advice. There is no conflict of interest or other disclosures to be reported.

Disclosure: None.

Disclaimer: None.

NOTE

1. Item wording comes from the original U.S. questionnaire.

REFERENCES

Acquadro, C., K. Conway, A. Hareendran, and N. Aaronson. 2008. "Literature Review of Methods to Translate Health-Related Quality of Life Questionnaires for Use in Multinational Clinical Trials." *Value in Health* 11 (3): 509–21.

Arah, O. A., A. Ten Asbroek, D. M. Delnoij, J. S. De Koning, P. J. Stam, A. H. Poll, B. Vriens, P. F. Schmidt, and N. S. Klazinga. 2006. "Psychometric Properties of the Dutch Version of the Hospital-Level Consumer Assessment of Health Plans Survey® Instrument." *Health Services Research* 41 (1): 284–301.

Bartlett, M. S. 1937. "Properties of Sufficiency and Statistical Tests." *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* 160 (901): 268–82.

Bautista, M. A. C., M. Nurjono, Y. W. Lim, E. Dessers, and H. J. Vrijhoef. 2016. "Instruments Measuring Integrated Care: A Systematic Review of Measurement Properties." *Milbank Quarterly* 94 (4): 862–917.

Berzon, R., R. Hays, and S. Shumaker. 1993. "International Use, Application and Performance of Health-Related Quality of Life Instruments." *Quality of Life Research* 2 (6): 367–8.

Brown, T. 2006. *Confirmatory Factor Analysis for Applied Research*, p. 475. New York: Guilford Press.

Bullinger, M., J. Alonso, G. Apolone, A. Leplège, M. Sullivan, S. Wood-Dauphinee, B. Gandek, A. Wagner, N. Aaronson, and P. Bech. 1998. "Translating Health Status Questionnaires and Evaluating Their Quality: The IQOLA Project Approach." *Journal of Clinical Epidemiology* 51 (11): 913–23.

Chen, F. F. 2007. "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Structural Equation Modeling* 14 (3): 464–504.

Cheung, G. W., and R. B. Rensvold. 2002. "Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance." *Structural Equation Modeling* 9 (2): 233–55.

Chidlow, A., E. Plakoyiannaki, and C. Welch. 2014. "Translation in Cross-Language International Business Research: Beyond Equivalence." *Journal of International Business Studies* 45 (5): 562–582.

Cronbach, L. J. 1951. "Coefficient Alpha and the Internal Structure of Tests." *psychometrika* 16 (3): 297–334.

Dimitrov, D. M. 2010. "Testing for Factorial Invariance in the Context of Construct Validation." *Measurement and Evaluation in Counseling and Development* 43 (2): 121–49.

Epstein, J., R. M. Santo, and F. Guillemin. 2015. "A Review of Guidelines for Cross-Cultural Adaptation of Questionnaires Could Not Bring out a Consensus." *Journal of Clinical Epidemiology* 68 (4): 435–41.

Fryer, A.-K., M. W. Friedberg, R. W. Thompson, and S. J. Singer. 2016. "Achieving Care Integration from the Patients' Perspective: Results from a Care Management Program." *Healthcare* 4 (1): 36–44.

Gehring, K., A. C. Mascherek, P. Bezzola, and D. L. Schwappach. 2015. "Safety Climate in Swiss Hospital Units: Swiss Version of the Safety Climate Survey." *Journal of Evaluation in Clinical Practice* 21 (2): 332–8.

Guillemin, F., C. Bombardier, and D. Beaton. 1993. "Cross-Cultural Adaptation of Health-Related Quality of Life Measures: Literature Review and Proposed Guidelines." *Journal of Clinical Epidemiology* 46 (12): 1417–32.

Hargraves, J., 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 (6p1): 1509–28.

Hofstede, G. 1985. "The Interaction Between National and Organizational Value Systems [1]." *Journal of Management Studies* 22 (4): 347–57.

Hsia, R. Y., and Y. C. Shen. 2015. "Percutaneous Coronary Intervention in the United States: Risk Factors for Untimely Access." *Health Services Research* 51 (2): 592–609.

Hurtado, M. P., J. Angeles, S. A. Blahut, and R. D. Hays. 2005. "Assessment of the Equivalence of the Spanish and English Versions of the CAHPS Hospital Survey on the Quality of Inpatient Care." *Health Services Research* 40 (6p2): 2140–61.

Kaiser, H. F. 1960. "The Application of Electronic Computers to Factor Analysis." *Educational and Psychological Measurement* 20 (1): 141–51.

———. 1970. "A Second Generation Little Jiffy." *Psychometrika* 35 (4): 401–15.

Kaiser, H. F., and J. Rice. 1974. "Little Jiffy, Mark IV." *Educational and Psychological Measurement* 34: 111–117.

Kerrissey, M., J. Clark, W. Jiang, P. Ramsay, S. Shortell, L. Casalino, M. Friedberg, and S. J. Singer. 2017. "Medical Group Structural Integration May Not Ensure That Care Is Integrated, from The Patient's Perspective." *Health Affairs* 36 (5): 885–892.

Kleinman, A. 1978. ""Concepts and a Model for the Comparison of Medical Systems as Cultural Systems." *Social Science & Medicine. Part B: Medical Anthropology* 12: 85–93.

Messick, S. 1995. "Validity of Psychological Assessment: Validation of Inferences from Persons' Responses and Performances as Scientific Inquiry into Score Meaning." *American Psychologist* 50 (9): 741.

Nijkamp, P., and J. Poot. 2015. "Cultural Diversity: A Matter of Measurement."

Pan, Y., and M. d. l. Puente. 2005. "*Census Bureau Guideline for the Translation of Data Collection Instruments and Supporting Materials: Documentation on How the Guidelin Was Developed.*" Research Report Series (Survey Methodology #2005-06). Washington, DC: U.S. Bureau of the Census, Statistical Research Division.

Perneger, T. V., A. Staines, and F. Kundig. 2014. "Internal Consistency, Factor Structure and Construct Validity of the French Version of the Hospital Survey on Patient Safety Culture." *BMJ Quality & Safety* 23 (5): 389–97.

Pett, M. A., N. R. Lackey, and J. J. Sullivan. 2006. *Making Sense of Factor Analysis: The Use of Factor Analysis for Instrument Development in Health Care Research.* Thousand Oaks, CA: Sage.

Pfeiffer, Y., and T. Manser. 2010. "Development of the German Version of the Hospital Survey on Patient Safety Culture: Dimensionality and Psychometric Properties." *Safety Science* 48 (10): 1452–62.

Shin, H. B., and R. Bruno. 2003. "Language Use and English-Speaking Ability: 2000" [accessed on November 1, 2003]. Available at https://www.census.gov/prod/2003pubs/c2kbr-29.pdf

Shortell, S. M., C. H. Colla, V. A. Lewis, E. Fisher, E. Kessell, and P. Ramsay. 2015. "Accountable Care Organizations: The National Landscape." *Journal of Health Politics, Policy and Law* 40 (4): 647–68.

Singer, S. J., J. Burgers, M. Friedberg, M. B. Rosenthal, L. Leape, and E. Schneider. 2011. "Defining and Measuring Integrated Patient Care: Promoting the Next Frontier in Health Care Delivery." *Medical Care Research and Review* 68 (1): 112–27.

Singer, S. J., M. W. Friedberg, M. V. Kiang, T. Dunn, and D. M. Kuhn. 2012. "Development and Preliminary Validation of the Patient Perceptions of Integrated Care Survey." *Medical Care Research and Review* 70 (2): 143–164.

Sousa, V. D., and W. Rojjanasrirat. 2011. "Translation, Adaptation and Validation of Instruments or Scales for Use in Cross-Cultural Health Care Research: A Clear and User-Friendly Guideline." *Journal of Evaluation in Clinical Practice* 17 (2): 268–74.

Struijs, J. N., and C. A. Baan. 2011. "Integrating Care through Bundled Payments—Lessons from The Netherlands." *New England Journal of Medicine* 364 (11): 990–1.

Tietschert, M., F. Angeli, A. van Raak, S. Singer, and D. Ruwaard. 2016. "Translating the Patient Perception of Integrated Care Survey to Measure Integrated Care in the Netherlands: Combining Equivalence and Contextualization Approaches for Optimal Results." *International Journal of Integrated Care* 16 (3): 1–9.

Ware, J. E., and B. Gandek. 1998. "Methods for Testing Data Quality, Scaling Assumptions, and Reliability: The IQOLA Project Approach." *Journal of Clinical Epidemiology* 51 (11): 945–52.

World Health Organization. n.d. "Process of Translation and Adaptation of Instruments" [accessed on 28 September, n.d.]. Available at http://www.who.int/substance_abuse/research_tools/translation/en/

SUPPORTING INFORMATION

Additional supporting information may be found online in the supporting information tab for this article:

Appendix SA1: Author Matrix.
Appendix SA2: Data Preparation.
Table S1: Survey Implementation in the U.S. and the Netherlands.
Table S2: Combined Items Due to Conceptual Link.
Table S3: Component Correlation Matrix.
Table S4: Goodness of Fit Indices.
Table S5: Standardized Factor Loadings and Standard Errors by Groups.
Table S6: Goodness of Fit Indexes.