Does the Consumer Assessment of Healthcare Providers and Systems Cultural Competence Survey provide equivalent measurement across English and Spanish versions?

Adam C. Carle, MA PhD and
Assistant Professor of Pediatrics, University of Cincinnati School of Medicine, Cincinnati Children's Hospital Medical Center, Assistant Professor of Psychology, University of Cincinnati College of Arts and Sciences, 3333 Burnet Avenue, MLC 7014, Cincinnati, OH 45229, Office Phone: (513) 803-1650, Fax: (513) 636-0171

Robert Weech-Maldonado, Ph.D. [Professor & L.R. Jordan Endowed Chair]
Department of Health Services Administration, University of Alabama at Birmingham, 1675 University Boulevard, 520 Webb, Birmingham, AL 35294, Phone: (205) 996-5838, Fax: (205) 975-6608, rweech@uab.edu

Abstract

BACKGROUND—The English and Spanish versions of the Consumer Assessments of Healthcare Providers and Systems (CAHPS®) Cultural Competence Survey (CAHP-CC) assess patients' experiences with culturally competent care. The possibility exists that even when Spanish and an English speakers experience the same levels of culturally competent care, responses describing their care may differ. This is called measurement bias. To deliver reliable and valid information across language, responses must provide equivalent measurement across versions. In this study, we examined whether measurement bias on the CAHPS-CC impedes valid measurement across the English and Spanish versions.

METHODS—We used multiple group (MG) confirmatory factor analyses (CFA) to examine measurement bias across English (n = 851) and Spanish (n = 113) speakers. Participants came from a 2008 sample of two Medicaid managed care plans, in New York and California.

RESULTS—MG-CFA provided general support for the equivalence of the CAHPS-CC in measuring Doctor Communication-Positive Behaviors, Doctor Communication-Negative Behaviors, Doctor Communication-Preventative Care, Equitable Treatment, and Trust. We did observe statistically significant differences in the thresholds associated with the item asking whether a doctor gave easier to understand instructions. However, analyses indicated that bias did not meaningfully influence conclusions about average experiences using the English and Spanish versions of the CAHPS-CC.

CONCLUSIONS—Our results support the use of the English and Spanish versions of the CAHPS-CC. Though we found some bias, analyses demonstrated that it did not substantively impact conclusions for the studied domains. Health providers can place confidence in the two different CAHPS-CC translations.

Author contact information: Adam C. Carle, PhD, Department of Pediatrics, University of Cincinnati School of Medicine, Department of Psychology, University of Cincinnati College of Arts and Sciences, James M. Anderson Center for Health Systems Excellence, Cincinnati Children's Hospital and Medical Center, 3333 Burnett Ave., Cincinnati, OH 45226, Phone: 513-803-1650, Fax: 513-636-0171, adam.carle.cchmc@gmail.com.

Keywords

Cultural competence; CAHPS®; Spanish; translation; ethnicity; measurement equivalence

Culturally competent care refers to the capacity of healthcare providers at various levels to engage with patients in a safe, patient and family centered, evidence-based, and equitable manner.[1] Given the increasing size of the Spanish speaking Hispanic population in the US,[2] the importance of delivering culturally competent care to this population,[3-8] and the importance of the patient's perspective,[9] it seems self-evident that stakeholders need reliable and valid measures of patients' experiences with culturally competent care. Yet, until recently, few tools have existed to do this, especially for Spanish speaking patients.

In response, a team of investigators developed a new measure of patient's experiences with culturally competent care.[10] The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Cultural Competence Survey (CAHPS-CC) assesses 8 domains of culturally competent care: Doctor Communication-Positive Behaviors; Doctor Communication-Negative Behaviors; Doctor Communication-Health Promotion; Doctor Communication-Alternative Medicine; Shared Decision; Equitable Treatment; Trust; and Access to Interpreter Services. Another paper provides support for the reliability and validity of this survey among patients generally. [10] However, research has not yet examined whether responses to the CAHPS-CC item set provide equivalently reliable and valid measurement across patients responding in English and Spanish.

Measurement bias refer to the possibility that two people who have had equivalent experiences with culturally competent care nevertheless answer questions about their experiences differently based on whether or not they respond to questions in English or Spanish.[11] Without establishing equivalent measurement, the field cannot discern whether differences in reports of care between English and Spanish speakers result from different care experiences or differences in the way the groups respond. Multiple group confirmatory factor analysis (MG-CFA) provides a potent method for evaluating bias.[12-14] Thus, we used MG-CFA to examine potential measurement bias on the CAHPS-CC across English and Spanish survey versions.

Methods

Participants

Using administrative data provided by the state's plans, we used a stratified random sampling design (based on race/ethnicity and language), to select 3,200 (New York) and 2,800 (California) adults (18-65 year old). The initial sampling frame consisted of: 1,200 White English speakers, 1,200 Black English speakers, 900 Hispanic English speakers, 900 Hispanic Spanish speakers, 900 Asian English speakers, and 900 bilingual Asian speakers (all communication with this group occurred in English. Cost restrictions and the number of Asian languages prevented us from developing numerous separate Asian language surveys).

Data collection occurred in two waves: mailing and follow-up telephone interviews of non-respondents. The mailing included an English survey and a cover letter in English and Spanish. The letter directed Spanish speakers to call an 800 number to request a copy of the Spanish survey materials. Four weeks after the initial mailing, non-respondents received a second mailed survey packet. Telephone follow-ups in English and Spanish started 2 weeks after the second mailing. Remaining non-respondents after the second call attempt received a monetary incentive of $10 to complete the survey. In all, 1,380 individuals completed the survey for an overall response rate of 26%.

We used administrative data to compare responders and non-responders on gender, age, race/ethnicity, primary language, and health plan affiliation. Respondents were more likely to be White (24% versus 20%) and older (39 versus 36 years, and less likely to be Black (18% versus 22%). We observed no other significant differences. Note that using administrative data to compare respondents and non-respondents may have influenced our conclusions regarding non-response bias.

After excluding individuals that did not have a personal doctor or a doctor visit during the last 12 months, the final analytic sample constituted 964 respondents. 851 completed the survey in English and 113 completed the survey in Spanish. See Weech-Maldonado et al.[10] for further methodological details.

Measures

Cultural Competency—The CAHPS Cultural Comparability team developed the CAHPS-CC by: 1) evaluating existing CAHPS surveys to identify existing items that addressed the domains; 2) conducting a literature review in order to identify existing items and instruments; 3) placing a Federal Register notice with a call for measures; 4) reviewing and adapting existing public domain measures; and 5) writing new survey items for each of the domains not addressed in 1 through 4. This resulted in a 49 item draft set. Two independent American Translators Association (ATA) certified translators then conducted two forward translations of the items into Spanish. Subsequently, a committee formed by the two translators and bilingual members of the Comparability team reviewed the translations and reconciled differences. Following translation, cognitive interviews[15] were conducted. Lastly, psychometric analyses evaluated the CAHPS-CC in the sample overall.[10,16,17]

At development's end, the CAHPS-CC consisted of 27 items addressing the extent to which an experience had occurred (rather than evaluating the experience). The items measured 8 constructs: Doctor communication-Positive Behaviors, Doctor Communication- Negative Behaviors, Doctor Communication- Health Promotion, Doctor Communication- Alternative Medicine, Shared Decision Making, Equitable Treatment, Trust, and Access to Interpreter Services. One can view the entire item set at https://www.cahps.ahrq.gov/Surveys-Guidance/Item-Sets/Cultural-Competence.aspx.

Our analyses addressed the Doctor communication-Positive Behaviors, Doctor Communication- Negative Behaviors, Doctor Communication- Health Promotion, Trust, and Equitable Treatment domains only. This occurred because too few individuals used interpreters to create a large enough sample to evaluate the Access to Interpreter Services domain. Additionally, the presence of some bivariate frequencies equal to 0 limited our ability to estimate the polychoric correlation matrix when including all of the remaining items. These "empty" cells occurred as a result of sparse responses in some item's categories.[18] Consistent with the literature, we collapsed categories for polytomous items with this problem [19] and dropped dichotomous items that had this problem. This resolved all estimation problems, but limited our analyses to the five factors listed earlier (and their 19 total items).

Analytical Approach

Measurement bias—We probed for measurement bias following the method described by Millsap and Yun-Tien[20] and Carle.[12] To evaluate overall fit, we used fit index levels identified by the literature.[21,22] Fit evaluation focused on the index set. We used the chi-square difference test ($\Delta\chi^2$) to test for bias. After identifying bias using this omnibus test, we used item level comparisons to identify bias' source.[12] All analyses used Mplus (6.1),[18] its theta parameterization and robust weighted least squares estimator and missing data

estimation capability. Given the number of models tests and consistent with the literature, we used a more conservative alpha of 0.01 for all significance tests.[12] We evaluated bias' substantive impact on substantive conclusions by comparing the pattern and size of mean differences from a model ignoring measurement bias to a model incorporating measurement bias, as described by Carle.[12]

Results

Demographics

Table 1 presents the analytic sample's descriptive statistics.

Evaluating Measurement Bias

We initially tested a 5 factor model's fit (Model 1) across the English and Spanish groups. This model fit well (RMSEA = 0.05, CFI = 0.98, TLI = 0.98). We then tested Model 2, which constrained the loadings to equality across groups. This model also fit well (RMSEA = 0.04, CFI = 0.99, TLI = 0.99) and the constraints did not result in statistically significant misfit ($\Delta\chi^2 = 12.7$, 13, $n = 633$, p = 0.23). This indicated no statistically significant bias in the loadings. We next examined bias in the thresholds. Thresholds give the level of the latent variable present before a respondent is more likely than not to respond in a given category. Model 3 constrained the thresholds to equality across the groups. The threshold's equivalence led to statistically significant misfit ($\Delta\chi^2 = 138.6$, 34, $n = 964$, p < 0.01), revealing bias in at least one threshold. Follow-up analyses indicated bias only in the thresholds of the "easy to understand instructions" items. The final partially invariant model relaxed the ill-fitting constraints. Summarily, we found no differences in the loadings and differences in only one item's thresholds. Table 2 presents the final partially invariant measurement model.

Evaluating the Influence of Measurement Bias

Statistically significant measurement bias may not substantively influence scores.[23,24] To evaluate bias' influence, we compared model-based estimates from the final partially invariant measurement model incorporating measurement differences to estimates from a model ignoring bias. Any differences in the pattern of mean differences would indicate influence. For example, White's had a mean of 0 on each factor (for statistical identification). Thus, we could first evaluate whether the means for each factor and group differed from Whites by examining whether their means differed significantly from 0. If we observed differences, we could then examine changes (if any) in these differences across the models. None of the means for Spanish respondents (Doctor Communication-Positive = -0.062, z = -0.565; Doctor Communication-Negative = -0.052, z = -0.278; Health Promotion = -0.092, z -0.609; Trust = 0.234, z = 2.107; and Equitable Treatment = -0.137, z = -0.393) showed statistically significant differences relative to English respondents. Under the model adjusting for bias, we also observed no statistically significant mean differences, providing support for the hypothesis that bias does not substantially influence mean-based conclusions for these factors.

Discussion

In this study, we investigated whether the CAHPS-CC provides sufficiently equivalent measurement across individuals responding in English and Spanish. Despite best efforts at survey translation, the possibility exists that two people with equivalent cultural competence experiences who answered the CAHPS-CC in different languages may have responded to questions about their experiences differently. Our results indicate that the CAHPS-CC has

equivalent measurement properties across individuals responding in English and Spanish for the domains included in our analyses.

We used MG-CFA and probed for bias across language (Spanish and English) in a sample of Medicaid patients in New York and California. Though we found some statistically significant measurement bias, further analyses demonstrated that the observed bias did not influence mean-based comparative conclusions across language when using the CAHPS-CC. These findings highlight the importance of evaluating whether measurement bias exists and whether any observed, statistically significant bias substantively influences decisions.

These findings support the use of the CAHPS-CC to measure patients' experiences with culturally competent care across Spanish and English speaking patients. Scores on the measure correspond to and estimate the underlying CAHPS-CC constructs similarly whether or not patients answer in Spanish or English. Patients' reports should have similar reliability across responses in either language and mean-based estimates should correspond to similar levels of the domain across English and Spanish respondents.

Before concluding, we note some study limitations. Due to sparse categories and relatively small within group sample sizes, we had to collapse some item categories and drop three domains (Shared Decision Making, Alternative Medicine, and Access to Interpreter Services). Therefore we could not examine bias in the full set of thresholds and for all of the domains. Also, our data came from a sample of Medicaid managed care enrollees in two states. New York and California's Medicaid populations may not generalize to the full Medicaid population. Additionally, we only investigated bias across language using Medicaid patients; our findings may not generalize to other populations. Likewise, we did not have measures of other potentially relevant variables (e.g., income, language ability) that might have influenced our results. Moreover, due to sample size restrictions, we could not further split our groups to examine additional for which we did have measures (e.g., race and ethnicity). Future research in larger, more diverse samples can address all of these issues.

Summarily, we used MG-CFA to examine whether measurement bias influences conclusions based on the patients' responses to the CAHPS-CC depending on whether they answer the survey in Spanish or English. Though we found some statistically significant measurement bias, our analyses demonstrated that this measurement bias does not substantively influence mean-based conclusions based on patients' responses. CAHPS-CC users can place confidence in efforts to compare the cultural competence experiences of English and Spanish speakers using the CAHPS-CC on the studied domains.

Acknowledgments

References

1. National Quality Forum. Endorsing a Framework and Preferred Practices for Measuring and Reporting Culturally Competent Care Quality. Washington DC: 2008.

2. Bureau USC. Annual Estimates of the Population by Sex, Race, and Hispanic Origin for the United States: April 1, 2000 to July 1, 2007 (NC-EST2007-03). 2008. http://www.census.gov/popest/national/asrh/NC-EST2007-srh.html

3. Lambert BL, Street RL, Cegala DJ, Smith DH, Kurtz S, Schofield T. Provider-patient communication, patient-centered care, and the mangle of practice. Health communication. 1997; 9(1):27–43.

4. McWhinney I. The need for a transformed clinical method. Communicating with medical patients. 1989; 9:25–40.

5. Ngo-Metzger, Q.; Telfair, J.; Sorkin, D., et al. Cultural Competency and Quality of Care: Obtaining the Patient's Perspective. New York, NY: Commonwealth Fund; 2006.

6. Weech-Maldonado R, Morales LS, Elliott M, Spritzer K, Marshall G, Hays RD. Race/ethnicity, language, and patients' assessments of care in Medicaid managed care. Health services research. 2003; 38(3):789–808. [PubMed: 12822913]

7. Weech-Maldonado R, Dreachslin J, Dansky K, De Souza G, Gatto M. Racial/ethnic diversity management and cultural competency: the case of Pennsylvania hospitals. Journal of healthcare management/American College of Healthcare Executives. 47(2):111. [PubMed: 11933599]

8. Nápoles-Springer AM, Santoyo J, Houston K, Pérez-Stable EJ, Stewart AL. Patients' perceptions of cultural factors affecting the quality of their medical encounters. Health Expectations. 2005; 8(1):4–17. [PubMed: 15713166]

9. Stewart AL, Nápoles-Springer AM. Advancing health disparities research: can we afford to ignore measurement issues? Medical care. 2003; 41(11):1207–1220. [PubMed: 14583684]

10. Weech-Maldonado, R.; Carle, AC.; Weidmer, B.; Ngo-Metzger, Q.; Hays, RD. Working Paper. Department of Health Services Administration, University of Alabama at Birmingham; 2010. Assessing Cultural Competency from the Patient's Perspective: The CAHPS Cultural Competency (CC) Item Set.

11. Mellenbergh GJ. Item bias and item response theory. International Journal of Educational Research. 1989; 13:127–143.

12. Carle A. Mitigating systematic measurement error in comparative effectiveness research in heterogeneous populations. Medical Care. 2010; 48(6):S68. [PubMed: 20473212]

13. Carle AC. Assessing the adequacy of self-reported alcohol abuse measurement across time and ethnicity: cross-cultural equivalence across Hispanics and Caucasians in 1992, non-equivalence in 2001–2002. BMC Public Health. 2009; 9:60. [PubMed: 19228373]

14. Carle AC. Tolerating Inadequate Alcohol Dependence Measurement: Cross-cultural Invalidity of Alcohol Dependence across Hispanics and Caucasians in 2001 and 2002. Addictive Behaviors. 2009; 34:43–50. [PubMed: 18801620]

15. Willis, G. Cognitive interviewing: a tool for improving questionnaire design. Sage Publications, inc; 2005.

16. Weech-Maldonado R, Carle AC, Weidmer B, Hurtado M, Ngo-Metzger Q, Hays RD. The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Cultural Competence (CC) Item Set. Medical Care. In Press.

17. Carle AC, Weech-Maldonado R. Evaluating Measurement Equivalence across Race and Ethnicity on the CAHPS® Cultural Competence Survey. Medical Care. In Press.

18. Muthén, LK.; Muthén, BO. Mplus User's Guide. Los Angeles, CA: Muthén & Muthén; 2009.

19. Crane PK, Gibbons LE, Jolley L, van Belle G. Differential Item Functioning Analysis With Ordinal Logistic Regression Techniques: DIFdetect and difwithpar. Medical Care. Special Issue: Measurement in a multi-ethnic society. 2006; 44(11, Suppl 3):S115–S123.

20. Millsap RE, Yun-Tein J. Assessing factorial invariance in ordered-categorical measures. Journal of Multivariate Behavioral Research. 2004; 39:479–515.

21. Hu L, Bentler P. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal. 1999; 6(1):1–55.

22. Hu L, Bentler PM. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological methods. 1998; 3(4):424–453.

23. Millsap, RE. Statistical Approaches to Measurement Invariance. Routledge; 2011.

24. Millsap RE, Kwok O-M. Evaluating the Impact of Partial Factorial Invariance on Selection in Two Populations. Psychological methods. 2004; 9(1):93–115. [PubMed: 15053721]

Table 1

Descriptive Statistics across Race and Ethnicity and across Survey Language for Analytic Sample

Variable	Hispanic	White	Black	Asian	Other	English	Spanish
Self-rated Health							
Excellent	5.53	1.33	2.65	1.22	1.33	10.51	1.55
Very Good	7.41	3.1	3.54	3.54	1.88	18.69	0.88
Good	11.62	5.2	4.87	8.19	5.64	31.97	3.65
Fair	9.4	4.87	3.76	5.09	1.99	20.69	4.42
Poor	2.88	1.66	1.22	1	0.77	6.75	0.88
Age							
18-24	6.57	1.64	2.96	3.29	1.75	15.33	0.88
25-34	5.81	3.29	3.5	2.85	1.42	15.33	1.64
35-44	8.43	3.4	3.4	5.04	3.29	20.37	3.29
45-54	9.97	5.15	4.6	3.83	2.74	22.89	3.4
55-64	6.13	2.41	1.64	3.94	2.52	14.68	2.19
Gender							
Female	29.07	11.58	11.69	12.24	7.87	63.72	8.96
Male	7.87	4.37	4.48	6.67	3.83	24.92	2.4
Education							
8th grade or less	9.65	0.55	0.22	2.55	1.33	8.54	5.88
Some high school	8.09	3.1	2.99	2.99	2.77	18.18	1.88
High school graduate or GED	9.53	5.54	5.43	5.43	3.55	27.61	2
Some college or 2-year degree	8.09	5.1	6.1	4.66	2.77	25.72	1
4-year college graduate or more	1.99	1.89	1.44	3	0.99	8.65	0.55
Survey Language							
English	24.12	14.73	14.93	17.46	16.65	-	-
Spanish	10.09	0	0	0	0	-	-

Table 2

Final Partially Invariant Measurement Model.

Doctor Communication-Positive	Loadings
Explain understandably	1.79
Listen carefully	2.85
Spend enough time	1.90
Show respect	1.91
Understandable instructions	0.87
Doctor Communication-Negative	
Interrupt	1.20
Talk too fast	1.91
Equitable Treatment	
Treated Unfairly Due to Race/Ethnicity	1.18
Treated Unfairly Due to Insurance Type	2.71
Health Promotion	
Talk about healthy diet	1.92
Talk about exercise	2.17
Talk about stress	0.96
Asked about depression	0.99
Trust	
Can tell Dr. anything	1.01
Trust Dr. with medical care	2.65
Feel Dr. tells you the truth	1.41
Feel Dr. cares about your health	2.05
How often felt Dr. cared	1.95
Explain Understandably	Thresholds
Never-Almost Never	-3.85
Almost Never-Sometimes	-3.36
Sometimes-Usually	-2.09
Usually-Almost Always	-1.24
Always	-0.51
Listen carefully	
Never or Almost Never-Sometimes	-5.12
Sometimes-Usually	-3.44
Usually-Almost Always	-2.20
Always	-1.14
Spend enough time	
Never-Almost Never	-4.07
Almost Never-Sometimes	-3.23
Sometimes-Usually	-1.86

Usually-Almost Always	-0.94
Always	-0.17
Interrupt	
Never-Almost Never	0.69
Almost Never-Sometimes	1.44
Sometimes-Usually	2.24
Usually-Almost Always	2.52
Always	2.76
Talk too fast	
Never-Almost Never	1.22
Almost Never-Sometimes	2.09
Sometimes-Usually	3.04
Usually-Almost Always	3.46
Always	3.89
Show respect	
Never or Almost Never-Sometimes	-3.62
Sometimes-Usually	-2.50
Usually-Almost Always	-1.66
Always	-0.94
Understandable instructions	
Did not talk	-1.52 (White)
	-0.58 (Hispanic)
Never	-1.41 (White)
	-0.47 (Hispanic)
Never or Almost Never-Sometimes	-1.34 (White)
	-0.40 (Hispanic)
Sometimes-Usually	-1.00 (White)
	-0.24 (Hispanic)
Usually-Almost Always	-0.54
Always	-0.15
Talk about healthy diet	
Yes, Definitely-Yes Somewhat	-0.15
Yes Somewhat-No	1.07
Talk about exercise	
Yes, Definitely-Yes Somewhat	-0.20
Yes Somewhat-No	1.22
Talk about stress	
Yes, Definitely-Yes Somewhat	-0.67
Yes Somewhat-No	0.03
Asked about depression	

Yes-No	-0.40
Can tell Dr. anything	
Yes, Definitely-Yes Somewhat	0.02
Yes Somewhat-No	1.11
Trust Dr. with medical care	
Yes, Definitely-Yes Somewhat	1.90
Yes Somewhat-No	4.44
Feel Dr. tells you the truth	
Yes, Definitely-Yes Somewhat	1.53
Yes Somewhat-No	2.77
Feel Dr. cares about your health	
Yes, Definitely-Yes Somewhat	1.00
Yes Somewhat-No	2.90
How often felt Dr. cared	
Never-Almost Never	0.12
Almost Never-Sometimes	1.07
Sometimes-Usually	2.02
Usually-Almost Always	3.22
Always	3.79
Treated Unfairly Due to Race/Ethnicity	
Never-Almost Never	1.68
Almost Never-Sometimes	2.04
Sometimes-Usually or more	2.59
Always	
Treated Unfairly Due to Insurance Type	
Never-Almost Never	2.92
Almost Never-Sometimes	3.57
Sometimes-Usually or more	4.76
	Means
Doctor Communication-Positive	0.00
Doctor Communication-Negative	0.00
Health Promotion	0.00
Equitable Treatment	0.00
Trust	0.00