Evaluating Measurement Equivalence across Race and Ethnicity on the CAHPS® Cultural Competence Survey

Adam C. Carle, MA 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

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

Quyen Ngo-Metzger, MD, MPH, and
Associate Clinical Professor, Department of Medicine, University of California, Irvine School of Medicine, 100 Theory Drive, Suite 110, Irvine, CA 92697-5800, Qngo-metzger@hrsa.gov

Ron D. Hays, Ph.D.
Department of Medicine, University of California, Los Angeles, 911 Broxton Avenue, Room 110, Los Angeles, CA 90024, drhays@ucla.edu

Abstract

BACKGROUND—The Consumer Assessments of Healthcare Providers and Systems (CAHPS®) Cultural Competence Survey assesses patients' experiences with culturally competent care. This study evaluates the equivalence of responses to this survey across different racial and ethnic subgroups. In this study, we examined whether measurement bias on the CAHPS Cultural Competence Survey impedes valid measurement across White, Black, and Hispanic patients.

METHODS—We used multiple group (MG) confirmatory factor analyses (CFA) to examine possible measurement bias across non-Hispanic White (n = 146), non-Hispanic Black (n = 148), and Hispanic (n = 339) adults. Participants came from two Medicaid managed care plans, one in New York and the other in California in 2008.

RESULTS—MG-CFA provided general support for the equivalence of the CAHPS Cultural Competence Survey in measuring doctor communication, health promotion and perceived trust across groups. However, we observed statistically significant differences in the thresholds associated with the Doctor Communication-Positive Behaviors. Nevertheless, sensitivity analyses indicated that measurement bias did not meaningfully influence conclusions about average experiences with culturally competent care across non-Hispanic White, non-Hispanic Black, and Hispanic patients in our sample.

CONCLUSIONS—Our results support the use of the CAHPS Cultural Competence Survey across non-Hispanic White, non-Hispanic Black, and Hispanic patients. Though we found some statistically significant measurement bias, sensitivity analyses demonstrated that measurement bias does not substantively influence conclusions based on patients' responses. Health providers at various levels can place confidence in the CAHPS Cultural Competence Survey and use it in diverse populations to evaluate patients' experiences with culturally competent care.

Keywords

Cultural competence; CAHPS®; race; ethnicity; measurement equivalence

Culturally competent medical care has the potential to reduce disparities in racial and ethnic differences in patients' experiences with their medical care.[1] Though multiple definitions exist,[2] 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.[3] Yet, until recently, few tools have existed to measure cultural competency.

The Consumer Assessment of Healthcare Providers and Systems (CAHPS®) Cultural Competence Survey (CC) assesses 8 aspects 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.[4] However, research has not yet examined whether the CAHPS-CC item set provides equivalently reliable and valid measurement across patients with different racial and ethnic backgrounds.

Measurement bias refer to the possibility that two people who have had equivalent experiences with culturally competent care will nevertheless answer questions about their experiences differently based on some characteristic such as their race or ethnicity.[5] They should respond similarly, but they do not. Without establishing equivalent measurement, the field cannot discern whether differences in reports and ratings of care between subgroups result from different care experiences or differences in the way the groups interpret or respond to the survey.[6,7] In this study, we used MG multiple group confirmatory factor analysis (MG-CFA)[6,8–10] to examine measurement bias on the CAHPS-CC.

Methods

Participants

Participants came from a field test of the CAHPS-CC conducted in 2008 among a stratified random sample (based on race/ethnicity and language) of 6,000 adult (aged 18-64) Medicaid (a US health program for individuals with low incomes and resources) managed care enrollees in two health plans: New York (3,200) and California (2,800). 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 Asian non-English speakers.

Data collection consisted of a 2-wave mailing with follow-up telephone interview of non-respondents. The first 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 the Spanish survey materials (13% mail response rate; n = 722). Four weeks after the initial mailing, non-respondents received a second mailed survey packet. Telephone follow-ups (English and Spanish) started 2 weeks after the second mailing. We offered a $10 monetary incentive

to non-respondents remaining after the second call (14% phone response rate; n = 489). These steps resulted in a 26% response rate overall (n=1,380).

Using administrative data, we compared responders and non-responders on gender, age, race/ethnicity, primary language, and health plan. Respondents were more likely White (24% versus 20%) and older (39 vs. 36 on average), and less likely Black (18% vs. 22%). We observed no other significant differences. After excluding individuals without a personal doctor or a doctor visit during the last 12 months, the final analytic sample constituted 991 respondents: 146 non-Hispanic White (hereafter White), 148 non-Hispanic Black (hereafter Black), 339 Hispanic, 173 Asian, 182 Other Race/Ethnicity, and 3 Missing Race/Ethnicity.

Among the Asian subgroup, too little variation in item responses occurred, resulting in a large amount of bivariate frequencies of zero. This in turn led to an inestimable model for this group. Thus, we excluded Asians from the analysis. We excluded Other Race/Ethnicity individuals from our analyses given the heterogeneity of racial groups this category captured. Relatedly, due to small samples sizes within each group constituting the "Other" group, we could not include each of these groups separately. Thus we examined measurement bias across White, Black, and Hispanic individuals only.

Measures

Cultural Competency

The CAHPS Cultural Comparability team developed the CAHPS-CC in several steps: 1) evaluating existing CAHPS surveys to identify existing items addressing the domains of interest; 2) conducting a literature review in order to identify relevant existing instruments or item sets; 3) placing a Federal Register notice with a call for measures; 4) reviewing and adapting publically available measures; and 5) writing new items for each of domain not addressed in 1–4. This resulted in a 49 item draft set.

Subsequently, two independent American Translators Association (ATA) certified translators conducted two forward translations of the survey into Spanish. A committee formed by the two translators and bilingual members of the comparability team reviewed the translations and reconciled any differences. Following translation, conducted cognitive interviews occurred.[11] Lastly, the team conducted psychometric analyses to evaluate the CAHPS-CC in the sample overall.[4]

At item development's end, the CAHPS-CC included 27 items. These 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. Too few individuals used interpreters to create a large enough sample to evaluate the Access to Interpreter Services domain in this analysis. Consequently, our analyses included 23 items.

Race and ethnicity

Respondents self reported their race and ethnicity.

Analytical Approach

Measurement bias

We examined measurement invariance following the method described by Millsap and Yun-Tien.[12] This method uses a series of nested models with increasing equivalence constraints on the measurement parameters across groups to evaluate measurement bias. We used fit

index levels (RMSEA, CFI, & TLI) identified by the literature.[13,14] Fit evaluation focused on the index set. After identifying bias using omnibus fit criteria, we used item level comparisons to identify bias' source and modify the model accordingly.[6] Constraints that led to significantly decreased fit identified measurement bias. We subsequently freed these constraints to develop a partial invariance model that directly modeled measurement bias.

All analyses used Mplus (6.1),[15] its theta parameterization and robust weighted least squares estimator and missing data estimation capability. Consistent with the literature, we used a more conservative alpha of 0.01 for all significance tests, given the number of tested models.[6] We evaluated bias's influence on substantive conclusions by comparing a model ignoring bias to a model incorporating measurement bias, as described by Carle.[6]

Results

Demographics

Table 1 presents the descriptive statistics for the analytic sample. A visual comparison of our sample's demographics with the general Medicaid population evidenced generally similar distributions, excepting for the variables for which we oversampled (e.g., race).

Evaluating Measurement Bias

Given previous research, we initially tested a 7 factor model's fit (Model 1)[4] across Whites, Blacks, and Hispanics. Though we achieved good fit when estimating the model in the sample ignoring group status (RMSEA = 0.04; TLI = 0.99; CFI = 0.91), we encountered problems when attempting to fit the model using MG-CFA. This occurred for several reasons. First, upon splitting the sample into groups, we observed several bivariate frequencies equal to 0, limiting our ability to estimate the polychoric correlation matrix.[15] These 0's occurred primarily as a result of sparse responses in some categories and items, thus we collapsed categories for those items. [16] This resolved the problem for all but one item "did this doctor use a condescending…tone"). Thus, we dropped it from our model. Second, we experienced difficulty fitting the baseline model due to the fact that three of the factors (Shared Decision Making, Equitable Treatment and Alternative Medicine) each had only two indicators per factor, resulting in an unstable model. Thus, we had to drop these factors from our model, resulting in a 4 factor model (Doctor Communication-Positive Behaviors, Doctor Communication-Negative Behaviors, Doctor Communication-Health Promotion, and Trust). The modified baseline model (Model 1b) fit well (RMSEA = 0.056, CFI = 0.99, TLI = 0.99). Given good fit, we tested Model 2, which constrained the loadings to equality across groups. These constraints did not result in statistically significant measurement bias ($\Delta_\chi^2 = 28.73$, 24, $n = 633$, $p = 0.23$).

Model 3 constrained the thresholds to equality across the groups. Thresholds indicate the level of the latent trait present before (on average) respondents are more likely than not to endorse a given category. Model 3 revealed statistically significant measurement bias in at least one threshold ($\Delta_\chi^2 = 141.72$, 24, $n = 633$, $p < 0.01$). Univariate indicated bias four items' thresholds: "listens carefully," "spend enough time," "show respect," and "easy to understand instructions." The pattern of bias was sometimes similar and sometimes different across Hispanics and Blacks relative to Whites (see Table 2). The final partially invariant model (see Table 2 for values) relaxed the equality constraints for these four items' thresholds.

Evaluating the Influence of Measurement Bias

Statistically significant bias does not necessarily indicate that bias would substantively influence conclusions.[17] To evaluate bias' influence, we compared model-based estimates

that resulted from the final partially invariant measurement model incorporating measurement differences to estimates that resulted from a model ignoring bias. Any differences in the pattern of mean differences would indicate bias' 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. Ignoring bias, none of the means across Blacks (Doctor Communication-Positive $M_{Black} = .42$, $z = 1.37$; Doctor Communication-Negative $M_{Black} = -0.73$, $z = -2.37$; Health Promotion $M_{Black} = -0.3$, $z = -1.643$; Trust $M_{Black} = -0.15$, $z = -0.76$) or Hispanics (Doctor Communication-Positive $M_{Hispanic} = 0.136$, $z = 0.517$; Doctor Communication-Negative $M_{Hispanic} = -0.24$, $z = -1.23$; Health Promotion $M_{Hispanic} = -0.14$, $z = -0.81$; Trust $M_{Hispanic} = .12$, $z = 0.73$) differed from Whites. Under the model adjusting for bias, Blacks' and Hispanics' means still did not differ significantly from the means for Whites, supporting the hypothesis that bias did not substantively influence mean-based conclusions.

Discussion

In this study, we evaluated whether the CAHPS Cultural Competence Survey provide sufficiently equivalent measurement across people of different racial and ethnic backgrounds.? In answer, yes. We used MG-CFA and probed for bias across Whites, Blacks, and Hispanics in a sample of Medicaid patients in New York and California. Though we found some statistically significant measurement bias, sensitivity analyses indicated that the observed measurement bias did not influence conclusions. These findings highlight the importance of both evaluating whether measurement bias exists and whether any observed, statistically significant measurement bias has the potential to substantively influences decisions based the measure's scores.

These findings provide preliminary support for the use of the CAHPS-CC to measure experiences culturally competent care across White, Black, and Hispanic patients. Scores on the measure correspond to the underlying constructs similarly across groups. Patients' reports should also have similar reliability. And, while some differences appear to exist in the levels of Doctor Communication-Positive present before Black and Hispanics will likely endorse some of the categories measuring the Doctor Communication-Positive construct, these differences do not appear to substantively influence mean-based conclusions.

Before closing, we note some limitations. First, due to sparse categories, we had to collapse some item categories and drop some subscales. Therefore we could not fully examine bias. Second, our data came from a sample of two state's Medicaid enrollees. Our findings may not generalize to the Medicaid or other populations. Third, the fit indices we used may not have been robust enough to identify misfit. Fourth, limited response rates may affect our findings' validity. Finally, sample sizes precluded us from including Asians or separating Hispanics or the other groups into finer grained groups (e.g., by acculturation, education, or other culturally relevant variables) to address these potential confounds with race and ethnicity. Future research in a larger, more diverse sample can and should address these issues before reaching firm conclusions about measurement bias on the CAHPS-CC.

Summarily, we used MG-CFA to examine whether measurement bias influences conclusions regarding 4 of 8 CAHPS-CC subscales across Whites, Blacks and Hispanics. Though we found some statistically significant bias, analyses demonstrated that bias does not substantively influence conclusions based on patients' responses for these subscales, indicating preliminary support that stakeholders can place confidence in the CAHPS-CC when used among White, Black and Hispanic groups.

References

1. 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]

2. Betancourt JR, Green AR, Carrillo JE, Ananeh-Firempong O. Defining cultural competence: a practical framework for addressing racial/ethnic disparities in health and health care. Public Health Reports. 2003; 118(4):293. [PubMed: 12815076]

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

4. 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.

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

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

7. Weech-Maldonado, R.; Weidmer, BO.; Morales, LS.; Hays, RD. Cross-Cultural Adaptation of Survey Instruments: The CAHPS Experience. In: Cynamon, M.; Kulka, R., editors. Seventh Conference on Health Survey Research Methods; Hyattsville, MD: DHHS; 2001. p. 75-82.

8. 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]

9. Carle AC. Tolerating Inadequate Alcohol Dependence Measurement: Cross-cultural Invalidity of Alcohol Dependence across Hispanics and Caucasians in 2001 and 2002. Addictive Behaviors. 2008 Online First(Journal Article).

10. Carle AC. Cross-cultural validity of alcohol dependence across Hispanics and non-Hispanic Caucasians. Hispanic Journal of Behavioral Sciences. 2008; 30(1):106–120.

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

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

13. 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.

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

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

16. 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 Suppl 3(11):S115–S123.

17. 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 for the Full Sample

Variable	%	*n*
Race/Ethnicity		
Hispanic	34.2	339
White	14.7	146
Black	14.9	148
Asian	17.5	173
Other	18.4	182
Missing	0.3	3
Self-rated Health		
Excellent	11	109
Very Good	17.9	177
Good	32.5	322
Fair	22.9	227
Poor	7	69
Missing	8.8	86
Age		
18–24	14.9	148
25–34	15.6	155
35–44	21.8	216
45–54	24.2	240
55–64	15.5	154
Missing	7.9	79
Gender		
Female	67.1	665
Male	25.2	250
Missing	7.7	76
Education		
8th grade or less	13.1	130
Some high school	18.3	181
High school graduate or GED	26.9	267
Some college or 2-year degree	24.3	241
4-year college graduate or more	8.4	83
Missing	9	89
Spanish Survey	11.8	117

Table 2

Final Partial Measurement Invariance Model

Doctor Communication-Positive	Loadings
Explain understandably	1.00
Listen carefully	1.52
Spend enough time	1.15
Show respect	1.18
Understandable instructions	0.45
Doctor Communication-Negative	
Interrupt	1.00
Talk too fast	1.56
Health Promotion	
Talk about healthy diet	1.00
Talk about exercise	1.48
Talk about stress	0.73
Asked about depression	−0.77
Trust	
Can tell Dr. anything	1.00
Trust Dr. with medical care	2.55
Feel Dr. tells you the truth	1.17
Feel Dr. cares about your health	1.92
How often felt Dr. cared	−1.70
Explain Understandably	Thresholds
Never-Almost Never	−3.88
Almost Never-Sometimes	−3.32
Sometimes-Usually	−2.06
Usually-Almost Always	−1.31
Always	−0.53
Listen carefully	
Never or Almost Never-Sometimes	−4.78
	−3.34 (Hispanic)
	−3.01 (White)
Sometimes-Usually	−3.05 (Black)
	−2.57 (Hispanic)
	−1.85 (White)
Usually-Almost Always	−1.77 (Black)
	−1.61 (Hispanic)
	−0.62 (White)
Always	−0.70 (Black)
Spend enough time	

Doctor Communication-Positive	Loadings
Never-Almost Never	−4.11 (Hispanic)
	−4.77 (White)
	−4.24 (Black)
Almost Never-Sometimes	−3.34 (Hispanic)
	−3.45 (White)
	−3.15 (Black)
Sometimes-Usually	−2.11 (Hispanic)
	−2.00 (White)
	−1.31 (Black)
Usually-Almost Always	−1.29 (Hispanic)
	−0.85 (White)
	−0.79 (Black)
Always	−0.62 (Hispanic)
	0.34 (White)
	0.31 (Black)
Interrupt	
Never-Almost Never	0.86
Almost Never-Sometimes	1.61
Sometimes-Usually	2.41
Usually-Almost Always	2.64
Always	2.91
Talk too fast	
Never-Almost Never	1.46
Almost Never-Sometimes	2.31
Sometimes-Usually	3.19
Usually-Almost Always	3.59
Always	4.13
Show respect	
Never or Almost Never-Sometimes	−4.1
Sometimes-Usually	−2.77
Usually-Almost Always	−2.31 (Hispanic)
	−1.71 (White)
	−1.24 (Black)
Always	−1.62 (Hispanic)
	−0.49 (White)
	−0.90 (Black)
Understandable instructions	
Did not talk - Never	−1.24 (Hispanic)
	−1.98 (White)

Doctor Communication-Positive	Loadings
	−1.05 (Black)
	−1.09 (Hispanic)
	−1.47 (White)
Never or Almost Never-Sometimes	−1.01 (Black)
	−0.88 (Hispanic)
	−1.00 (White)
Sometimes-Usually	−0.66 (Black)
	−0.62 (Hispanic)
	−0.60 (White)
Usually-Almost Always	−0.38 (Black)
	−0.34 (Hispanic)
	0.06 (White)
Always	−0.08 (Black)
Talk about healthy diet	
Yes, Definitely-Yes Somewhat	−0.04
Yes Somewhat-No	0.88
Talk about exercise	
Yes, Definitely-Yes Somewhat	0.05
Yes Somewhat-No	1.38
Talk about stress	
Yes, Definitely-Yes Somewhat	−0.54
Yes Somewhat-No	0.14
Asked about depression	
Yes-No	0.29
Can tell Dr. anything	
Yes, Definitely-Yes Somewhat	0.03
Yes Somewhat-No	1.28
Trust Dr. with medical care	
Yes, Definitely-Yes Somewhat	1.93
Yes Somewhat-No	4.66
Feel Dr. tells you the truth	
Yes, Definitely-Yes Somewhat	1.46
Yes Somewhat-No	2.54
Feel Dr. cares about your health	
Yes, Definitely-Yes Somewhat	0.84
Yes Somewhat-No	2.95
How often felt Dr. cared	
Never-Almost Never	−3.72
Almost Never-Sometimes	−3.09

Doctor Communication-Positive	Loadings
Sometimes-Usually	−1.85
Usually-Almost Always	−1.05
Always	−0.14