

# Functional Assessment and Neuropsychiatric Inventory Questionnaires: Measurement Invariance Across Hispanics and Non-Hispanic Whites

Philip Sayegh, MA, \*,<sup>1</sup> and Bob G. Knight, PhD<sup>2</sup>

<sup>1</sup>Departments of Psychology and Preventive Medicine and

<sup>2</sup>Davis School of Gerontology and Department of Psychology, University of Southern California, Los Angeles, California.

\*Address correspondence to Philip Sayegh, MA, Department of Psychology, University of Southern California, 3620 S. McClintock Ave-SGM 501, Los Angeles, CA 90089–1061. E-mail: [psayegh@usc.edu](mailto:psayegh@usc.edu)

Received October 30, 2012; Accepted March 4, 2013  
Decision Editor: Rachel Pruchno, PhD

**Purpose of the Study:** We tested the ethnic-group measurement invariance of 2 commonly used informant-report scales of patients' dementia symptoms: the Functional Assessment Questionnaire (FAQ), a measure of functional abilities, and the Neuropsychiatric Inventory Questionnaire (NPI-Q), a measure of behavioral and psychological symptoms of dementia. **Design and Methods:** We conducted multigroup confirmatory factor analyses on 311 Hispanic and 10,863 non-Hispanic White (NHW) outpatients and their informants diagnosed with dementia or normal cognition at their initial Alzheimer's Disease Center evaluations nationwide. **Results:** We confirmed our hypothesized one-factor FAQ and four-factor NPI-Q models for each ethnic group. We also found evidence for the configural (i.e., number of factors) and factorial (i.e., pattern of factor loadings) invariance of both scales and structural (i.e., factor covariances) invariance of the NPI-Q across groups. However, we did not obtain evidence for ethnic-group scalar (i.e., intercept) invariance for either scale. **Implications:** The FAQ and NPI-Q were operating similarly across Hispanics and NHWs, suggesting that they can be meaningfully used within and across these groups to measure informant-reported dementia symptomatology. However, their scalar noninvariance indicates that

meaningful ethnic-group comparisons of their latent factor mean values cannot be made.

**Key Words:** Factor analysis, Ethnicity, Latino/a, Dementia, Informant reports

---

As the United States elderly population continues to surge, it is projected that Hispanic American older adults will be at a disproportionate risk of dementia through the next 50 years (Valle & Lee, 2002). Accordingly, it is important to assess for possible differences in dementia assessment across Hispanics and non-Hispanic Whites (NHWs), as clinicians will be increasingly likely to work with Hispanic older adults with possible dementia.

Clinicians frequently rely on informant-reported dementia symptomatology, such as declines in functional abilities, during the diagnostic process, as they can be useful indicators of latent dementia. Erzigkeit and colleagues (2001) found that the Bayer-Activities of Daily Living (B-ADL) scale was as effective as or superior to a dementia cognitive screen in identifying individuals with clinically manifest dementia symptoms. Thus, evidence suggests that changes in individuals' ability to function in daily life may represent the earliest stages of a dementing disorder and can be a valuable area to assess through informants. The Functional

Assessment Questionnaire (FAQ; Pfeffer, Kurosaki, Harrah, Chance, & Filos, 1982) is a frequently used informant-report scale that measures patients' ability to conduct 10 instrumental activities of daily living (IADLs) to assist clinicians in dementia diagnosis. Pfeffer and colleagues (1982) validated the FAQ on a sample of 195 older adults aged 61–91 years in a stable retirement community of 22,000 people.

Family members or other informants also often notice changes in patients' personality, behavior, and mood, commonly seen in dementia, even before the assignment of a formal diagnosis (Balsis, Carpenter, & Storandt, 2005). Therefore, behavioral and psychological symptoms of dementia (BPSD), even those present at earlier stages, can also be dependably reported by informants to aid clinicians in diagnosis. One validated, commonly used informant-report measure of BPSD is the Neuropsychiatric Inventory Questionnaire (NPI-Q; Kaufer et al., 2000), a briefer version of the NPI (Cummings et al., 1994), which evaluates changes in BPSD among patients with possible dementia. The NPI-Q was validated on a sample of primarily highly educated NHW participants and their informants from a university-based research clinic, with adequate test–retest reliability and convergent validity reported.

What is lacking in the literature is the examination of whether the FAQ and NPI-Q can be used among diverse ethnic groups to draw meaningful within- and across-group comparisons. The FAQ creators provided no information regarding their sample's ethnic breakdown, thus limiting the ability to extend the validation of this scale to diverse groups. Additionally, it remains unclear how the findings from the NPI-Q validation study would generalize to more diverse populations and settings. Finally, the ethnic-group measurement invariance of these scales has yet to be tested. Accordingly, more research focused on these scales' suitability and measurement invariance across diverse groups is needed.

Though no research to our knowledge has investigated the ethnic-group measurement invariance of either the FAQ or NPI-Q, certain studies have examined the underlying factor structure of similar measures. Regarding functional abilities, Erzigkeit and colleagues (2001) found a one-factor structure that they termed *dementia severity* for the B-ADL for separate samples of individuals with dementia of varying severity in the United Kingdom, Germany, and Spain. Their findings

provided evidence for the factorial invariance of the one-factor B-ADL structure across these three European countries and suggested that the mean differences in scores were likely meaningful and not due to measurement artifact.

Regarding BPSD, Kang, Ahn, Kim, and Kim (2010) examined the factor structure of the 12-item Korean NPI using both exploratory and confirmatory factor analysis (CFA) in their sample of South Koreans with either Alzheimer's disease (AD) or probable AD. They found a 10-item, four-factor model: *hyperactivity* (agitation, disinhibition, and irritability), *affect* (anxiety and depression), *psychosis* (delusions and hallucinations), and *apathy/vegetative symptoms* (apathy, nighttime behavior, and appetite). Aalten and colleagues (2007) found a nearly identical four-factor structure for the 12-item NPI in their sample of dementia patients from 12 European countries: *hyperactivity* (agitation, disinhibition, irritability, and aberrant motor behavior), *affective* (anxiety and depression), *psychosis* (delusions, hallucinations, and nighttime behavior), and *apathy* (apathy and appetite). The only differences between this latter study and the former were the addition of aberrant motor behavior in *hyperactivity* and nighttime behavior in *psychosis* (as opposed to the *apathy/vegetative symptoms* in the former study). Aalten and colleagues noted that these findings were unclear but reported that aberrant motor behavior loaded to a very similar degree on *apathy* and barely met their 0.40 factor loading cutoff for both factors. Additionally, nighttime behavior also loaded strongly on *apathy*, albeit to a lesser degree. Given these results and the strength of Kang and colleagues' combined exploratory and CFA approach, it appears plausible that the NPI-Q may also be best represented by the same 10-item, four-factor structure in other ethnic groups. Although these studies' findings had in common the co-occurrence of similar BPSD syndromes across different ethnicities and nationalities, their slightly divergent findings suggest that the NPI may not necessarily be factorially invariant across ethnicities and that invariance cannot simply be presumed.

A number of possible reasons exist that may lead to ethnic-group differences in response patterns on these scales, including linguistic barriers, ethnocultural influences on the perception and communication of dementia symptoms, and shame and stigma associated with dementia (Sayegh & Knight, 2013). If the FAQ and NPI-Q do not

demonstrate ethnic-group measurement invariance, then explanations of across-group differences are called into question, as is clinical diagnostic validity. Given that [Erzigkeit and colleagues \(2001\)](#) confirmed the measurement invariance of the B-ADL, a measure similar to the FAQ, among individuals of different nationalities and found a one-factor structure of *dementia severity*, we hypothesized that the FAQ would also have a 10-item, one-factor structure reflecting *dementia severity* and demonstrate measurement invariance across Hispanics and NHWs. Additionally, we predicted that the NPI-Q would have a 10-item, four-factor structure and demonstrate ethnic-group measurement invariance. We hypothesized that these four factors would be similar to those found in the [Kang and colleagues \(2010\)](#) study, as reported previously. Finally, we hypothesized that the FAQ and NPI-Q's latent factor mean values would not differ significantly across groups, as they would not be expected to differ systematically on the basis of ethnicity alone.

## Methods

### *Participants and Procedure*

Participants included existing outpatients and volunteers (who thereby became outpatients) who underwent evaluation and their informants enrolled in the longitudinal National Alzheimer's Coordinating Centers' (NACC) Alzheimer's Disease Center (ADC) study at 33 sites nationwide based on a sample of 23,029 ethnically diverse patients with various diagnoses. This study included data (using the June 2011 data freeze) from 11,174 Hispanic and NHWs (and their informants) with an initial clinical diagnosis of normal cognitive function (NCF) or dementia based on all the information (e.g., neuropsychological data, informant reports, and neurological examinations) available to clinicians as part of this data set. Clinicians, whether individually or through consensus, provided an overall judgment regarding whether patients had NCF (i.e., no mild cognitive impairment [MCI], dementia, or other neurological condition resulting in cognitive impairment) or met criteria for dementia in accordance with standard criteria for either (a) AD (using the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association [NINCDS/ADRDA]; [McKhann et al., 1984](#)), or (b) vascular dementia (using the National

Institute of Neurological Disorders and Stroke and Association Internationale pour la Recherche et l'Enseignement en Neurosciences [NINDS/AIREN]; [Román et al., 1993](#)), or (c) demonstrated sufficient evidence of other non-Alzheimer's or vascular types of dementia (e.g., Lewy body dementia and frontotemporal dementia). Additional inclusion criteria were having an ethnically matched informant deemed reliable to report on patients' symptoms and a completed FAQ and NPI-Q. Patients were excluded if they had diagnoses of stroke or Parkinson's disease without dementia associated with these diagnoses, MCI, and cognitive impairment without MCI or dementia. All patients were assessed in English, though these data do not explicitly indicate the languages in which informants provided information for questionnaire completion.

### *Measures*

*Functional Assessment Questionnaire.*—Informants completed the FAQ, which measures patients' abilities to carry out 10 IADLs with response options including: 0 (*normal*), 1 (*has difficulty, but does by self*), 2 (*requires assistance*), 3 (*dependent*), and 8 (*not applicable* [e.g., never did]). Total scores range from 0 to 30, with higher scores representing more difficulty or requiring assistance with IADLs for more than 4 weeks. Trained clinicians (e.g., physicians and nurses) or other health professionals conducted interview with patients' informants to complete the FAQ, with detailed, standardized administration instructions provided in the study's codebook.

*Neuropsychiatric Inventory Questionnaire.*—Informants completed the NPI-Q, which measures the occurrences of patients' BPSD in the past month. Total scores for the 10 items used in our CFAs range from 0 to 10, with higher scores indicating a greater number of BPSD. NACC required online training to provide certification for clinicians and other health professionals to administer the NPI-Q and provided standardized administration instructions in this study's codebook.

### *Analysis*

We first evaluated the data for multivariate kurtosis, a critically important assumption of structural equation modeling (SEM). To assess the ethnic-group measurement invariance of the FAQ

and NPI-Q, we conducted CFAs on the Hispanic and NHW samples using AMOS 17.0. We tested for measurement invariance in two stages (Byrne, 2010). First, we created baseline models for each scale separately for each group and evaluated the goodness of fit to the data. In the event that two or more factors were found to be highly correlated, alternative models in which these factors' items were combined into one factor were assessed and compared with the hypothesized baseline models. Second, using multigroup CFAs, we examined four assumptions based on the baseline models to assess measurement invariance across groups: (a) configural invariance—the same hypothesized number of factors, (b) factorial invariance—the same factor structure or pattern of factor loadings, (c) structural invariance—the same structural relations (i.e., factorial covariances), which only needed to be tested for the multidimensional NPI-Q, with its four latent factors set to be correlated with each other because they are theoretically all areas of BPSD, and (d) scalar (i.e., intercept) invariance, a prerequisite for assessing for latent mean value differences. In the event of findings that did not support ethnic-group scalar invariance, we conducted post hoc analyses to assess for significant, meaningful item-level group differences to determine whether any specific items may have been accountable for scalar noninvariance. Scalar noninvariance is believed to result from differential additive response bias and suggests that issues unrelated to the pertinent constructs are influencing the presence of systematically different scores on some or all items in one group compared with the other.

To evaluate the criteria used to accept models, we examined various goodness of fit statistics, including the goodness of fit indices (GFI), comparative fit indices (CFI), Tucker–Lewis Indices (TLI), and the root mean square error of approximation (RMSEA). Though rules for establishing goodness of fit for models vary widely, adequate fit is often indicated by GFI, CFI, and TLI values greater than .95 and RMSEA values less than .06 (Schreiber, Nora, Stage, Barlow, & King, 2006). To assess for significant changes in model fit, we examined the  $\Delta$ CFI values, with changes of less than .01 indicative of invariance (Cheung & Rensvold, 2002). To evaluate the adequacy of our sample sizes, we examined Hoelter's (1983) critical *N* (CN) for the .05 level, with values greater than 200 suggestive of a large enough sample and values

less than 75 deemed unacceptably low. Although an arbitrary 0.40 standardized factor loading cutoff is often used in the literature to determine whether items meaningfully load onto factors (Aaltonen et al., 2007), we based our decisions regarding the acceptability of factor loadings both on our hypothesized factor structures and whether specific item loadings were noticeably lower than those of other items. Finally, given the discrepancy in sample sizes across ethnic groups (311 Hispanics and 10,863 NHWs), we randomly selected a sample of 311 NHWs and reconducted our CFAs to assess whether sample size influenced our multigroup invariance analyses.

## Results

The data showed evidence of multivariate kurtosis for both Hispanics (Mardia's coefficient:  $FAQ = 61.51$ ;  $NPI-Q = 26.12$ ) and NHWs (Mardia's coefficient:  $FAQ = 381.69$ ;  $NPI-Q = 239.55$ ). Therefore, CFAs were based on asymptotically distribution-free estimation (Browne, 1984). There was no evidence of multivariate outliers for the Hispanic or NHW data for either scale based on examination of Mahalanobis distance values.

The sample was composed of 311 Hispanic and 10,863 NHW outpatients and their informants, matched by ethnicity. A total of 6,151 (55.0%, standard error of percentage [ $SE_{\text{percentage}}$ ] = 0.47) patients were diagnosed with dementia and 5,023 patients with NCF. Of the 311 Hispanic patients, 187 (60.13%) were Mexican/Chicano/Mexican American, 58 (18.65%) were Puerto Rican, 17 (5.47%) were South American, 14 (4.50%) were Central American, 11 (3.54%) were Cuban, 7 (2.25%) were Dominican, 14 (4.50%) were classified as other, and 3 (0.96%) were coded as unknown. The origins of the Hispanic informants in this study were similar to those of the patients in terms of frequencies. Of the Hispanic patients, 181 (58.20%,  $SE_{\text{percentage}} = 2.80$ ) were diagnosed with dementia and 130 with NCF. Among the NHW patients, there were similar proportions of diagnoses, with 5,970 (54.96%,  $SE_{\text{percentage}} = 0.48$ ) diagnosed with dementia and 4,893 with NCF. Table 1 provides additional descriptive information on demographic and other key variables separated by ethnicity and diagnosis. Table 2 provides information on the number and percentages of the specific types of and contributing factors to dementia for the entire sample and separate ethnicities.

Table 1. Descriptive Statistics and *p* Values for *t* Tests and Chi-Square Statistics by Ethnicity and Diagnosis

	Overall ( <i>N</i> = 11,174), <i>M</i> ( <i>SD</i> )	NCF ( <i>N</i> = 5,023), <i>M</i> ( <i>SD</i> )	Dementia ( <i>N</i> = 6,151), <i>M</i> ( <i>SD</i> )	NCF vs. Dementia, <i>p</i> Values
Age (years)				
Hispanic ( <i>N</i> = 311)	72.59 (9.88)	70.09 (9.42)	74.39 (9.84)	<.01
NHW ( <i>N</i> = 10,862)	72.67 (10.74)	71.56 (10.92)	73.59 (10.51)	<.01
Hispanic vs. NHW, <i>p</i> value	.08	.13	.31	—
Patient education (years)				
Hispanic ( <i>N</i> = 310)	12.61 (3.81)	13.33 (3.55)	12.09 (3.92)	.01
NHW ( <i>N</i> = 10,782)	15.25 (2.98)	15.82 (2.73)	14.78 (3.10)	<.01
Hispanic vs. NHW, <i>p</i> value	<.01	<.01	<.01	—
Informant education (years)				
Hispanic ( <i>N</i> = 286)	14.07 (2.93)	14.06 (2.90)	14.07 (2.96)	.96
NHW ( <i>N</i> = 10,336)	15.59 (2.67)	15.81 (2.68)	15.41 (2.66)	<.01
Hispanic vs. NHW, <i>p</i> value	<.01	<.01	<.01	—
	Overall, <i>N</i> (%)	NCF, <i>N</i> (%)	Dementia, <i>N</i> (%)	NCF vs. Dementia, <i>p</i> value
Women				
Hispanic ( <i>N</i> = 311)	189 (60.77)	92 (70.77)	97 (53.59)	.00
NHW ( <i>N</i> = 10,863)	5,998 (55.21)	3,026 (61.84)	2,972 (49.78)	<.01
Hispanic vs. NHW, <i>p</i> value	.05	.04	.31	—
Coresidency				
Hispanic ( <i>N</i> = 311)	161 (51.77)	59 (45.38)	102 (56.35)	.06
NHW ( <i>N</i> = 10,863)	7,071 (65.09)	2,782 (56.86)	4,289 (71.84)	<.01
Hispanic vs. NHW, <i>p</i> value	.00	.01	<.01	—
	Overall, <i>M</i> ( <i>SD</i> )	NCF, <i>M</i> ( <i>SD</i> )	Dementia, <i>M</i> ( <i>SD</i> )	NCF vs. Dementia, <i>p</i> value
Total FAQ				
Hispanic ( <i>N</i> = 311)	10.58 (11.02)	0.50 (1.34)	17.81 (9.05)	<.01
NHW ( <i>N</i> = 10,863)	9.22 (10.38)	0.48 (1.96)	16.37 (8.90)	<.01
Hispanic vs. NHW, <i>p</i> value	.03	.93	.03	—
Total NPI-Q				
Hispanic ( <i>N</i> = 311)	2.50 (2.58)	0.91 (1.51)	3.64 (2.59)	<.01
NHW ( <i>N</i> = 10,863)	2.18 (2.46)	0.69 (1.28)	3.40 (2.52)	<.01
Hispanic vs. NHW, <i>p</i> value	.03	.10	.22	—

Notes: NCF = normal cognitive functioning; NHW = non-Hispanic White; FAQ = Functional Assessment Questionnaire; NPI-Q = Neuropsychiatric Inventory Questionnaire.

### Functional Assessment Questionnaire

*Establishing Baseline Models.*—SEM-based congeneric estimate of score reliability (Graham, 2006) for this scale was 0.99 for Hispanics. For the Hispanic group, our CFA yielded a GFI value of .96, a CFI value of .94, a TLI value of .92, a RMSEA value of .09, 90% confidence interval (CI) [0.07, 0.11], and a CN value of 125. All of these values represent indicators of reasonably adequate to good model fit and acceptable, albeit somewhat low, sample size. SEM-based congeneric estimate of score reliability for this scale was 0.99 for NHWs. When we tested the model for the NHW group, results yielded a GFI value of .95, a CFI value of .90, a TLI value of .87, a RMSEA value of .08, 90% CI [0.08, 0.08], and

a CN value of 218. Overall, this model appeared to adequately fit the data and sample size was deemed sufficient. The factor loadings for both models are listed in Table 3.

*Multigroup Measurement Invariance.*—First, we tested whether the one-factor structure was equivalent across groups (Model 1), and the obtained fit indices were overall indicative of adequate fit (Table 4). These results suggested that the FAQ structure is indeed most appropriately described by a one-factor (*dementia severity*) model for both the Hispanic and NHWs samples. Second, we assessed whether the pattern of factor loadings was equivalent across groups. Results from this

**Table 2. Number and Percentages of Types of and Contributing Factors to Dementia for the Overall Sample and Separate Ethnicities**

	Overall (N = 6,151), N (%)	Hispanic (N = 181), N (%)	NHW (N = 5,970), N (%)
Probable Alzheimer’s disease	3,991 (64.88)	130 (71.82)	3,861 (64.67)
Possible Alzheimer’s disease	780 (12.68)	27 (14.92)	753 (12.61)
Dementia with Lewy bodies	485 (7.88)	10 (5.52)	475 (7.96)
Vascular dementia	186 (3.02)	11 (6.08)	175 (2.93)
Alcohol-related dementia	31 (0.50)	1 (0.55)	30 (0.50)
Dementia of undetermined etiology	138 (2.24)	3 (1.66)	135 (2.26)
Frontotemporal dementia (behavioral/ executive dementia)	559 (9.09)	4 (2.21)	555 (9.30)
Primary progressive aphasia (aphasic dementia)	337 (5.48)	2 (1.10)	335 (5.61)
Progressive supranuclear palsy	55 (0.89)	0 (0.00)	55 (0.92)
Corticobasal degeneration	124 (2.02)	2 (1.10)	122 (2.04)
Huntington’s disease	1 (0.02)	0 (0.00)	1 (0.02)
Prion disease	40 (0.65)	1 (0.55)	39 (0.65)
Cognitive dysfunction from medications	54 (0.88)	0 (0.00)	54 (0.90)
Cognitive dysfunction from medical illnesses	83 (1.35)	2 (1.10)	81 (1.36)
Depression	1,315 (21.38)	55 (30.39)	1,260 (21.11)
Other major psychiatric illness	74 (1.20)	2 (1.10)	72 (1.21)
Down’s syndrome	6 (0.19)	0 (0.00)	6 (0.10)
Parkinson’s disease	219 (3.56)	5 (2.76)	214 (3.58)
Stroke	265 (4.31)	11 (6.08)	254 (4.25)
Hydrocephalus	51 (0.83)	2 (1.10)	49 (0.82)
Traumatic brain injury	81 (1.32)	3 (1.66)	78 (1.31)
Central nervous system neoplasm	16 (0.26)	0 (0.00)	16 (0.27)
Other	482 (7.84)	10 (5.52)	472 (7.91)

Note. NHW = non-Hispanic White.

**Table 3. Standardized Factor Loadings from the Functional Assessment Questionnaire (FAQ) Baseline Models**

	Hispanics	Non-Hispanic White
	Dementia severity	Dementia severity
Bills/finances	.99	.95
Taxes/papers	.97	.92
Shopping	.95	.94
Games/hobbies	.84	.85
Stove	.87	.83
Meal preparation	.92	.89
Tracking events	.95	.92
Paying attention	.89	.88
Remembering dates	.94	.92
Travel	.93	.91

Notes: All factor loadings are significant at the <.01 level. Unstandardized factor loadings with standard errors can be provided upon request by contacting the corresponding author.

test of Model 2 (Table 4) determined that the postulated equality of factor loadings across groups was tenable,  $\Delta CFI = 0.00$ , suggesting that pattern of factor loadings was invariant across groups.

*Scalar Invariance.*—Third, we tested the FAQ for ethnic-group scalar invariance. The goodness of fit statistics from this test is presented in Table 4 (Model 3) and were mainly indicative of adequate fit though suggestive of a less than adequate sample size. However, the increases in the CFI values for this model compared with both the configural model (Model 1,  $\Delta CFI = 0.01$ ) and the measurement model (Model 2,  $\Delta CFI = 0.01$ ) were not less than 0.01. Thus, we did not find evidence for full scalar invariance across groups.

Because we did not find evidence for full scalar invariance, our next step was to conduct analyses to assess for partial scalar invariance, which is suggested as a compromise between full and a lack of measurement invariance. A conservative method of implementing partial invariance constraints is to use modification indices (MI) and degree of expected parameter change values to determine which constraints to relax (Steenkamp & Baumgartner, 1998). We considered MI values that stood out as excessively large in comparison to other values (Byrne, 2010) or values greater than 20 to be indicative of a significant scalar invariance problem, though we

Table 4. Summary of Tests for Ethnic-Group Invariance of the Functional Assessment Questionnaire (FAQ) Measurement Model

	GFI	CFI	TLI	RMSEA	90% CI of RMSEA	CN
Model 1: Number of factors (i.e., 1) invariant	.95	.90	.87	.06	0.06, 0.06	389
Model 2: Model 1 with pattern of factor-loading invariant	.95	.90	.88	.05	0.05, 0.06	430
Model 3: Model 1 with intercepts invariant and latent mean freely estimated in one group	—	.91	.91	.11	0.11, 0.11	110

Notes: GFI = goodness of fit indices; CFI = comparative fit indices; TLI = Tucker–Lewis indices; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval; CN = Hoelter’s Critical N (.05 level).

did not find any values that met these criteria. An alternative approach is to detect noninvariant item intercepts by sequentially removing equality constraints on each item’s measurement intercept and examining the  $\Delta$ CFI value until one identifies noninvariant items (Van Lieshout, Cleverley, Jenkins, & Georgiades, 2011). However, we still did not find evidence for partial scalar invariance using this approach, as the CFI values were identical or very similar in all cases. Therefore, we could not meaningfully test for significant differences on this scale’s latent factor mean values.

To assess whether there were specific items accounting for the lack of ethnic-group scalar invariance, we compared the FAQ’s item mean values across groups and found that Hispanics had significantly higher scores than NHWs on shopping,  $z = -2.03, p = .04$ ; remembering events,  $z = -2.42, p = .02$ ; remembering dates,  $z = -3.67, p < .01$ ; and traveling,  $z = -2.05, p = .04$ . Nonetheless, these group differences generally had small effect sizes ( $r = -.02, -.02, -.03, \text{ and } -.02$ , respectively), which means that they may not be meaningfully interpretable differences in light of our large sample size.

### Neuropsychiatric Inventory Questionnaire

**Establishing Baseline Models.**—Hancock and Mueller’s (2001) weighted construct reliability coefficient for Hispanics was .87. Subscale reliability coefficients using the Kuder–Richardson Formula 20 (KR-20) were as follows: *hyperactivity*, .68; *affect*, .52; *psychosis*, .50; and *apathy/vegetative symptoms*, .48. For the Hispanic group, our CFA yielded a GFI value of .98, a CFI value of .98, a TLI value of .97, an RMSEA value of .02, 90% CI [0.00, 0.05], and a CN value of 399. All of these values represent indicators of good model fit and sufficient sample size. Hancock and Mueller’s weighted construct reliability coefficient for NHWs

was .85. Subscale reliability coefficients using the KR-20 were as follows: *hyperactivity*, .65; *affect*, .53; *psychosis*, .50; and *apathy/vegetative symptoms*, .54. When we tested the model for the NHW group, results yielded a GFI value of .98, a CFI value of .92, a TLI value of .87, an RMSEA value of .03, 90% CI [0.03, 0.04], and a CN value of 1,155. Overall, this four-factor model also appeared to fit the data reasonably well and sample size was deemed sufficient. The factor loadings and correlations for both models are listed in Tables 5 and 6, respectively. Given the high correlations between the *affect* and *apathy/vegetative symptoms* factors among both groups, we tested an alternative three-factor model in which these two factors were combined. However, the goodness of fit indices for this model were nearly identical to the hypothesized four-factor model and were not significantly different based on  $\Delta$ CFI values. As the focus of these analyses was confirmatory and aimed at testing multigroup invariance rather than factor structure per se, we conducted our multigroup analyses using our originally proposed four-factor model.

**Multigroup Measurement Invariance.**—First, we tested whether the four-factor structure was equivalent across groups (Model 1), and our results (Table 7) suggested that the NPI-Q structure is indeed most appropriately described by our hypothesized four-factor model for both groups. Second, we assessed whether the pattern of factor loadings was equivalent across ethnic groups. Results from this test of Model 2 provided evidence for invariance,  $\Delta$ CFI = .00. Third, we examined whether the factor covariances were equivalent across groups. Goodness of fit statistics from the estimation of Model 3 were indicative of adequate to good model fit and the  $\Delta$ CFI of .00 indicated a nominal change in fit, suggestive of invariant factor covariances.

**Table 5. Standardized Factor Loadings from the Neuropsychiatric Inventory Questionnaire (NPI-Q) Baseline Models**

	Hispanic				Non-Hispanic White			
	Hyperactivity	Affect	Psychosis	Apathy/ vegetative	Hyperactivity	Affect	Psychosis	Apathy/ vegetative
Agitation	.72				.71			
Disinhibition	.60				.53			
Irritability	.68				.67			
Anxiety		.65				.64		
Depression		.61				.58		
Delusions			.77				.68	
Hallucinations			.50				.48	
Apathy				.69				.64
Sleep				.31				.46
Appetite				.48				.46

Notes: All factor loadings are significant at the <.01 level. Unstandardized factor loadings with standard errors can be provided upon request by contacting the corresponding author.

**Table 6. Factor Correlations for the Neuropsychiatric Inventory Questionnaire (NPI-Q) Baseline Models**

	Hispanic				Non-Hispanic White			
	Hyperactivity	Affect	Psychosis	Apathy/ vegetative	Hyperactivity	Affect	Psychosis	Apathy/ vegetative
Hyperactivity	—				—			
Affect	.74	—			.70	—		
Psychosis	.56	.39	—		.55	.51	—	
Apathy/vegetative	.69	.93	.31	—	.75	.82	.54	—

Note: All correlations are significant at the .05 level ( $t > 1.96$ ).

**Table 7. Summary of Tests for Ethnic-Group Invariance of the Neuropsychiatric Inventory Questionnaire (NPI-Q) Measurement and Structural Models**

	GFI	CFI	TLI	RMSEA	90% CI of RMSEA	CN
Model 1: Number of factors (i.e., four) invariant	.98	.92	.87	.02	0.02, 0.03	1,981
Model 2: Model 1 with pattern of factor-loading invariant	.98	.92	.89	.02	0.02, 0.03	2,122
Model 3: Model 1 with pattern of factor-loading invariant and factor variances and covariances invariant	.98	.92	.89	.02	0.02, 0.02	2,252
Model 4: Model 3 with intercepts invariant and latent mean freely estimated in one group	—	.96	.95	.03	0.03, 0.03	1,339

Notes: GFI = goodness of fit indices; CFI = comparative fit indices; TLI = TuckerLewis Indices; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval; CN = Hoelter's Critical N (.05 level).

*Scalar Invariance.*—Fourth, we tested the NPI-Q for scalar invariance across ethnic groups. Table 7 provides the goodness of fit statistics from this test (Model 4). All fit indices were suggestive of a well-fitting model. However, the  $\Delta$ CFI value exceeded .01 for this model compared with both the configural model (Model 1,  $\Delta$ CFI = .04), the measurement

model (Model 2,  $\Delta$ CFI = .05), and the structural model (Model 3,  $\Delta$ CFI = .05). Therefore, we did not find evidence for full scalar invariance across groups.

Given the lack of evidence obtained for full scalar invariance, we conducted analyses to assess for partial scalar invariance using the previously described techniques used with the FAQ. However,



we failed to find evidence for partial scalar invariance using these approaches, as the CFI values remained identical or very similar in all cases. Thus, we were unable to meaningfully test for significant group differences on this scale's latent mean values. Consequently, we compared NPI-Q item-level mean differences across groups. Hispanics had significantly higher scores than NHWs on delusions,  $\chi^2(1, N = 11,174) = 17.13, p < .01$ ; hallucinations,  $\chi^2(1, N = 11,174) = 5.14, p = .02$ ; agitation,  $\chi^2(1, N = 11,174) = 4.30, p = .04$ ; depression,  $\chi^2(1, N = 11,174) = 9.06, p = .00$ ; and appetite,  $\chi^2(1, N = 11,174) = 7.31, p = .01$ . However, in light of our large sample, the effect sizes associated with these differences were rather small ( $\Phi = 0.04, 0.02, 0.02, 0.03, \text{ and } 0.03$ , respectively), suggesting that they may not be meaningful differences.

### Results Using Matched Sample Sizes

To assess whether discrepant sample sizes may have affected our findings, we randomly selected a sample of 311 (out of 10,863) NHWs to match the Hispanic sample size. Results revealed an identical overall pattern of results, suggesting that sample size was not affecting our key findings regarding ethnic-group invariance.

## Discussion

This study examined the ethnic-group measurement invariance of the FAQ and NPI-Q among Hispanic and NHW outpatients. This study is the first to examine the psychometric properties of these scales to determine whether they are operating equivalently and whether meaningful ethnic-group latent mean comparisons of these scales' factors can be made, which bears relevance in terms of dementia diagnostic validity across groups.

Before we discuss our results, we duly note that in some cases, including our baseline models, goodness of fit statistics did not meet the strict cutoff criteria suggested by some researchers (Schreiber et al., 2006). However, Byrne (2010) strongly advised against modifying baseline models when testing for multigroup invariance, even if fit statistics are modest at best and modifications result in good fit. The more an originally hypothesized model is modified at this stage of analysis, the more difficult it is to determine measurement and structural equivalence, particularly when MI values are small, as was the case for our baseline models. Byrne (2010) also noted that there is a

risk of "overfitting" a model by making post hoc modifications in that the addition of extra parameters can (a) represent weak effects that are not likely replicable, (b) lead to significant increases in standard errors, and (c) affect the model's primary parameters. Model alterations may simply be fitting the sample's small, distinctive characteristics. Furthermore, other researchers have used less stringent fit index cutoffs, noting that although cutoffs suggested by Schreiber and colleagues (2006) are now widely used, these higher thresholds may not reflect a significant improvement and may not be applicable to all models (Marsh, Hau, & Wen, 2004). Taken together, although we believe that our model fit statistics generally support our conclusions, we nonetheless suggest that caution be used in the interpretation of our results without further study involving other samples of Hispanics and NHWs.

### Functional Assessment Questionnaire

As hypothesized, results from our multigroup CFAs revealed that the FAQ demonstrated most aspects of ethnic-group measurement (i.e., configural and factorial) invariance, suggesting that this scale has the same number (i.e., one) of factors and a similar pattern of factor loadings across groups. The one-factor (*dementia severity*) structure found for the FAQ is consistent with the findings of a prior study that used a similar measure of functional abilities that also found a comparable one-factor structure that was invariant across participants from three European countries (Erzigkeit et al., 2001). Given that we found evidence for the ethnic-group configural and factorial invariance of the FAQ, we can conclude that this scale is likely operating similarly across groups and is measuring the same latent construct of *dementia severity* with regard to IADLs.

Our analyses failed to provide evidence for the ethnic-group scalar invariance of the FAQ, which may have in part been due to a less than adequate sample size. The finding of scalar noninvariance suggests that item mean scores were not similar across ethnic groups. Therefore, meaningful comparisons of levels of the latent *dementia severity* factor could not be made across groups. Item intercept invariance is a requirement for the comparison of latent means, as it implies identical intervals and zero points of the scale across groups. When scalar invariance is not tenable, the comparison of latent means becomes equivocal as the between-group

differences in latent means are confounded with the scale and origin of the latent variable (Cheung & Rensvold, 2002). It is possible that linguistic barriers may have been present among some Hispanic informants that may have made it more difficult for them to accurately convey patients' functional abilities to the primarily English-speaking clinicians in this study. Additionally, Hispanic and NHW informants differed significantly in terms of education level, which could also be driving scalar noninvariance. Informants with higher education levels may be better able to recognize and report on dementia-related symptoms and be more likely to obtain a timely dementia evaluation for their patients, which could influence scores on these measures. Our post hoc analyses did not reveal any specific item(s) that caused the scalar noninvariance. Therefore, the most probable reason for the scalar noninvariance is that these item intercept-level disparities were distributed across some or all items but were not significant on an item-level basis, consistent with our post hoc analyses that failed to find meaningful item-level differences across ethnicities.

Our findings suggest that the FAQ can likely continue to be used as a useful measure of patients' functional abilities among both Hispanics and NHWs. Indeed, it proved to be invariant across ethnic groups in terms of both its factor structure and pattern of factor loadings. However, because we did not find evidence for the scalar invariance of this scale, meaningful contrasts of the *dementia severity* factor derived from this scale cannot be made across these ethnic groups. This finding is a limitation of which researchers should be mindful when employing this scale to compare latent mean values of this factor across Hispanics and NHWs. Despite this limitation, future research should examine whether this widely used, validated, practical informant-report scale of patients' functional abilities demonstrates measurement invariance across other ethnic groups.

### *Neuropsychiatric Inventory Questionnaire*

We also found evidence to confirm our hypothesis regarding most aspects of the NPI-Q's ethnic-group measurement invariance. These findings suggest that this scale has the same number (i.e., four) of factors and a similar pattern of both factor loadings and covariances across Hispanics and NHWs. The hypothesized four-factor (*hyperactivity, affect, psychosis, and apathy/vegetative*

*symptoms*) structure of the NPI-Q that we confirmed for both groups is consistent with the findings of prior studies using comparable versions of this BPSD measure that also found a similar invariant four-factor structure among individuals from 12 European countries (Aalten et al., 2007) and South Korea (Kang et al., 2010). However, as the correlations between the *affect* and *apathy/vegetative symptoms* factors for both groups in this study were high, it should be noted that these two factors are essentially indistinguishable. Given the evidence found for the NPI-Q's ethnic-group factorial invariance, we can conclude that this scale was operating similarly across both groups and was measuring the same four latent constructs of BPSD.

However, the NPI-Q did not demonstrate ethnic-group scalar invariance, suggesting that mean differences in this scale's four latent factors cannot be meaningfully interpreted and compared across Hispanics and NHWs. It is possible that education-level differences, language barriers, and ethnocultural differences in how BPSD are interpreted and communicated may have influenced ethnic-group differences on responses to the NPI-Q items. Additionally, the shame and stigma associated with dementia and its BPSD may be particularly salient among Hispanics (Sayegh & Knight, 2013), which could have systematically influenced response patterns across ethnic groups. Similarly, informant-reporting styles on patients' BPSD may have differed across groups given that Hispanic caregivers may be more sensitive to BPSD than NHW caregivers (Valle, 1994). Because our post hoc analyses failed to reveal any particular item(s) contributing to the scalar noninvariance, it is likely that these differences were spread out across some or all items but may not have been significant for any individual item(s), as may be the case with the FAQ and as was supported by our post hoc analyses assessing for meaningful ethnic-group item-level differences on this measure.

Our multigroup CFA revealed that the NPI-Q can continue to be used as a useful measure of patients' BPSD among both Hispanics and NHWs, as it demonstrated ethnic-group configural, factorial, and structural invariance. These findings confirm that this scale is operating similarly across these groups, supporting the measurement validity of the NPI-Q in clinical and research contexts. Additionally, our results suggested that this scale can be readily and validly used among both groups to measure specific dimensions of BPSD beyond a more general operationalization of these symptoms.

However, the NPI-Q cannot be used to make meaningful comparisons across ethnic groups in terms of its four latent factors' mean values. Researchers should be mindful of this limitation when considering comparing Hispanics and NHWs on latent mean values of these factors. Aside from this limitation, both the overall usefulness of this scale and its validity and specificity suggest that future research should examine whether the NPI-Q demonstrates measurement invariance across other ethnic groups.

### Limitations and Strengths

This study has certain limitations that suggest that caution should be used in the interpretation of these results. The generalizability of these findings may be somewhat limited, as ADC participants essentially represent a convenience sample composed of patients and informants who presented to academic AD clinics. Additionally, this sample lacked enough statistical power to examine differences across subgroups of the Hispanic outpatients, which could provide richer findings regarding differences across more specific ethnic groups. We also may have lacked adequate statistical power for both the FAQ's baseline model for Hispanics and the ethnic-group scalar invariance model. Furthermore, these data do not explicitly state the language in which informants completed the questionnaires, though it is presumable that the vast majority were completed in English rather than Spanish (or another language) as these data derived from an English language module. Despite these limitations, this study also has a number of strengths, including its use of a nationwide, multi-site data set characterized by standardized methods, which bolsters both the external and internal validity of this study. Additionally, the inclusion of a relatively large and diverse Hispanic sample and a large NHW sample is another strength. Finally, this study contributes to the literature by being the first to examine the ethnic-group measurement properties of the FAQ and NPI-Q.

### Conclusion

In sum, our findings regarding the ethnic-group measurement invariance of the FAQ and NPI-Q bear relevance in terms of diagnostic validity among Hispanics and NHWs. We found that both scales demonstrated all aspects of ethnic-group measurement invariance except scalar invariance,

suggesting that they are measuring the same underlying constructs and are operating similarly across groups. However, because we did not find evidence for ethnic-group scalar invariance for the FAQ and NPI-Q, they may not be used to make meaningful comparisons based on latent mean estimates of these measures' factors.

### Funding

This work was supported by the National Alzheimer's Coordinating Center Grant (#U01 AG016976).

### Acknowledgment

This content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or Aging.

### References

- Aalten, P., Verhey, F. R. J., Boziki, M., Bullock, R., Byrne, E. J., Camus, V., et al. (2007). Neuropsychiatric syndromes in dementia. Results from the European Alzheimer Disease Consortium: Part I. *Dementia and Geriatric Cognitive Disorders*, *24*, 457–463. doi:10.1159/000111082
- Balsis, S., Carpenter, B. D., & Storandt, M. (2005). Personality change precedes clinical diagnosis of dementia of the Alzheimer type. *Journal of Gerontology: Psychological Sciences*, *60*, P98–P101. doi:10.1093/geronb/60.2.P98
- Browne, M. W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, *37*, 62–83. doi:10.1111/j.2044-8317.1984.tb00789.x
- Byrne, B. M. (2010). *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.). New York: Taylor & Francis Group.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*, 233–255. doi:10.1207/S15328007SEM0902\_5
- Cummings, J. L., Mega, M., Gray, K., Rosenberg-Thompson, S., Carusi, D. A., & Gornbein, J. (1994). The Neuropsychiatric Inventory: Comprehensive assessment of psychopathology in dementia. *Neurology*, *44*, 2308–2314. doi:10.1212/WNL.44.12.2308
- Erzigkeit, H., Lehfeld, H., Peña-Casanova, J., Bieber, F., Yekrang-Hartmann, C., Rupp, M., et al. (2001). The Bayer-Activities of Daily Living Scale (B-ADL): Results from a validation study in three European countries. *Dementia and Geriatric Cognitive Disorders*, *12*, 348–358. doi:10.1159/000051280
- Graham, J. M. (2006). Congeneric and (essentially) tau-equivalent estimates of score reliability: What they are and how to use them. *Educational and Psychological Measurement*, *66*, 930–944. doi:10.1177/0013164406288165
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In S. D. T. Cudeck & D. Sorbom (Eds.), *A festschrift in honor of Karl Jöreskog* (pp. 195–216). Lincolnwood, IL: Scientific Software International.
- Hoelter, J. W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods & Research*, *11*, 325–344. doi:10.1177/0049124183011003003
- Kang, H. S., Ahn, I. S., Kim, J. H., & Kim, D. K. (2010). Neuropsychiatric symptoms in Korean patients with Alzheimer's disease: Exploratory factor analysis and confirmatory factor analysis of the Neuropsychiatric Inventory. *Dementia and Geriatric Cognitive Disorders*, *29*, 82–87. doi:10.1159/000264629
- Kaufner, D. I., Cummings, J. L., Ketchel, P., Smith, V., MacMillan, A., Shelley, T., et al. (2000). Validation of the NPI-Q, a brief clinical form of the Neuropsychiatric Inventory. *Journal of Neuropsychiatry and Clinical Neurosciences*, *12*, 233–239. doi:10.1176/appi.neuropsych.12.2.233
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for

- fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling: An Multidisciplinary Journal*, 11, 320–341. doi:10.1207/s15328007sem1103\_2
- McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., & Stadlan, E. M. (1984). Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology*, 34, 939–944. doi:10.1212/WNL.34.7.939
- Pfeffer, R. I., Kurosaki, T. T., Harrah, C. H., Chance, J. M., & Filos, S. (1982). Measurement of functional activities in older adults in the community. *Journal of Gerontology*, 37, 323–329. doi:10.1093/geronj/37.3.323
- Román, G. C., Tatemichi, T. K., Erkinjuntti, T., Cummings, J. L., Masdeu, J. C., Garcia, J. H., et al. (1993). Vascular dementia: Diagnostic criteria for research studies: Report of the NINDS-AIREN International Workshop. *Neurology*, 43, 250–260. doi:10.1212/WNL.43.2.250
- Sayegh, P., & Knight, B. G. (2013). Cross-cultural differences in dementia: The Sociocultural Health Belief Model. *International Psychogeriatrics*, 25, 517–530. doi:10.1017/S104161021200213X
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *Journal of Educational Research*, 99, 323–337. doi:10.3200/JOER.99.6.323-338
- Steenkamp, J. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–107. doi:10.1086/209528
- Valle, R. (1994). Culture-fair behavioral symptom differential assessment and intervention in dementing illness. *Alzheimer Disease & Associated Disorders*, 8, 21–45. doi:10.1097/00002093-199404000-00003
- Valle, R., & Lee, B. (2002). Research priorities in the evolving demographic landscape of Alzheimer's Disease and associated dementias. *Alzheimer Disease and Associated Disorders*, 16, S64–S76. doi:10.1097/00002093-200200002-00006
- Van Lieshout, R. J., Cleverley, K., Jenkins, J. M., & Georgiades, K. (2011). Assessing the measurement invariance of the Center for Epidemiologic Studies Depression Scale across immigrant and non-immigrant women in the postpartum period. *Archives of Women's Mental Health*, 14, 413–423. doi:10.1007/s00737-011-0236-0