

Hispanic Journal of Behavioral Sciences

<http://hjb.sagepub.com>

Cross-Cultural Validity of Alcohol Dependence Across Hispanics and Non-Hispanic Caucasians

Adam C. Carle

Hispanic Journal of Behavioral Sciences 2008; 30; 106

DOI: 10.1177/0739986307311618

The online version of this article can be found at:

<http://hjb.sagepub.com/cgi/content/abstract/30/1/106>

Published by:



<http://www.sagepublications.com>

Additional services and information for *Hispanic Journal of Behavioral Sciences* can be found at:

Email Alerts: <http://hjb.sagepub.com/cgi/alerts>

Subscriptions: <http://hjb.sagepub.com/subscriptions>

Reprints: <http://www.sagepub.com/journalsReprints.nav>

Permissions: <http://www.sagepub.com/journalsPermissions.nav>

Citations <http://hjb.sagepub.com/cgi/content/refs/30/1/106>

Cross-Cultural Validity of Alcohol Dependence Across Hispanics and Non-Hispanic Caucasians

Adam C. Carle
University of North Florida

Confirmatory factor analyses for ordered-categorical measures probed for differential item functioning on a standardized measure of alcohol dependence across Hispanics ($n = 834$) and non-Hispanic Caucasians ($n = 14,001$) in a nationally representative survey of alcohol use in the United States conducted in 1992. Analyses investigated whether 30 items operationalizing the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) construct of alcohol dependence provided equivalent measurement. The results demonstrated statistically significant differential item functioning for 7 items, suggesting caution regarding the cross-ethnic validity of alcohol dependence. Sensitivity analyses suggested that item-level differences had a limited impact, lending confidence to previous findings. The findings underscore the necessity of cultural sensitivity when generalizing measures and constructs developed in the majority to Hispanic individuals and demonstrate the need for evaluations of differential item functioning in contemporary data.

Keywords: *differential item functioning; cross-cultural differences; Hispanics; validity; alcohol dependence*

Alcohol-use disorders are among the most prevalent mental disorders in the United States (Grant et al., 2004) and are a considerable public health problem. Estimates indicate a *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.) (DSM-IV; American Psychiatric Association, 1994) alcohol dependence prevalence rate in the general U.S. population of approximately 3.81% (Grant et al., 2004). Yet significant differences exist when comparing the prevalence, incidence, and comorbidity of alcohol-use

Author's Note: This work would not have been possible without the unending support and input of Tara J. Carle, Margaret Carle, and James Rauth.

disorders across Caucasians and Hispanics (Grant et al., 2004; Hasin & Grant, 2004). These studies have noted that Hispanics demonstrate a significantly greater rate of alcohol dependence (3.95%) than Caucasians (3.83%) (Grant et al., 2004), a finding consistent with earlier observations (Dawson, Grant, Chou, & Pickering, 1995; Grant et al., 1994). Patterns of drinking behavior are changing across Caucasians and Hispanics as well (Grant et al., 2004). These investigations have demonstrated important differences in the etiology and epidemiology of alcohol dependence across Caucasians and Hispanics and underscore a need for culturally sensitive public health policies, prevention efforts, and intervention efforts (Grant et al., 2004). Given health disparity concerns (Anderson & Nickerson, 2005) and the immense cost of alcohol-use disorders to individuals, their families, and society, these issues are particularly salient.

However, it is important to consider whether the observed differences in alcohol dependence reflect true differences or resulted from a lack of equivalence in the measures used to assess alcohol dependence across these populations. Measurement bias or differential item functioning (DIF) is the possibility that individuals equal in their true levels of alcohol dependence, but from different groups (e.g., Caucasians and Hispanics), do not have identical probabilities of responding to questions concerning their alcohol use (Mellenbergh, 1989). Thus, it may be that differences described in earlier and current studies may be exacerbated or attenuated by DIF. This possibility is also problematic when charting changes in drinking patterns across time. To the extent that initial estimates were affected by DIF, subsequent changes may not reflect true change across the groups. It is essential to establish the cross-ethnic validity of alcohol-use measures in order to interpret earlier and current work.

Studies have established the validity and reliability of standardized alcohol-dependence measures in the general population (Chatterji et al., 1997; Grant, 1997, 2000; Grant, Dawson, & Hasin, 2001; Grant, Dawson, & Stinson, 2003; Grant, Harford, & Dawson, 1995; Grant et al., 1994; Hasin, Carpenter, & McCloud, 1997; Hasin, Grant, & Cottler, 1997; Hasin, Muthén, Wisnicki, & Grant, 1994; Hasin & Paykin, 1999; Harford & Muthén, 2001; Kessler et al., 1994; Muthén, 1995; Muthén, Hasin, & Wisnicki, 1993; Regier et al., 1988), and factor-analytic studies have addressed construct validity (Harford & Muthén, 2001; Hasin et al., 1994; Muthén, 1995; Muthén et al., 1993). These studies provided reasonable evidence that the diagnostic criteria can be operationalized in a valid and reliable manner for use in the collective population. Lamentably, the role of minority- or majority-based measurement DIF in the instruments used to assess alcohol dependence in the U.S. population

has rarely been examined, and it should not be assumed that findings evidenced in the majority group hold in other groups.

Measurement models such as confirmatory factor analysis (CFA) and item response theory offer tools to examine DIF (Waller, Thompson, & Wenk, 2000). A set of mathematical equations modeling item responses ask whether model parameters (e.g., loadings, thresholds, uniquenesses) are equivalent for each group. Studies have shown that DIF can attenuate or accentuate group differences (Cole, 1999; Huang, Church, & Katigbak, 1997; Pentz & Chou, 1994; Smith & Reise, 1998; Waller et al., 2000), lead to inaccurate diagnoses (Cole, Martin, Peeke, Henderson, & Harwell, 1998; Gallo, Anthony, & Muthén, 1994; Reid et al., 1998, Waller et al., 2000), and generally decrease reliability and validity (Byrne, Baron, & Baley, 1996, 1998; Byrne, Baron, & Campbell, 1993, 1994; Byrne & Campbell, 1999; Knight & Hill, 1998; Schafer & Caetano, 1996). Studies have uncovered DIF so severe that cross-group comparisons were impossible (Knight, Tein, Shell, & Roosa, 1992; Prelow, Tein, Roosa, & Wood, 2000).

These findings are problematic when coupled with theoretical and empirical reasons to suspect the presence of DIF across majority and minority groups generally and Hispanics and Caucasians specifically (Sue, 1999). Compared with Caucasians, minorities, Hispanics included, have shown differences in the relation between probabilistic thinking and the assignment of numbers (Wright et al., 1978), differences in acquiescent responses (Smith, 2004), and differences in language use (Bachman & O'Malley, 1984). Hui and Triandis (1989) noted that Hispanics described sincerity as a cultural value. They suggest that this may lead to DIF, positing that Hispanic culture values sincere responses that in turn lead to more ready endorsements of scale end points, because the middles of scales often reflect options of no opinion, "don't know," or similar responses. Prelow et al. (2000) suggested that cultural differences affect the salience of behaviors and the degree to which they are considered problematic; thus, for certain behaviors, greater levels of a specific problem may be needed before Hispanics are willing to acknowledge the problem. Finally, the behavioral exemplars that describe a psychological construct for Caucasians may not be appropriate for minorities and may not include the set of indicators that would be appropriate (Huang, et al., 1997; Prelow, Michaels, Reyes, Knight, & Barrera, 2002; Prelow et al., 2000). Despite these possibilities, a literature review found no published studies examining the validity of alcohol-dependence measures across Caucasians and Hispanics, leaving unclear to what extent DIF affects epidemiological estimates, research, or clinical conclusions across these groups.

The current study addressed these issues by using CFA for ordered-categorical measures (CFA-OCM) to probe for bias across self-identified minority status in a large, nationally representative sample of noninstitutionalized individuals in the U.S. population surveyed in 1992. Analyses examined whether DIF was exhibited in a standardized survey of alcohol dependence across non-Hispanic Caucasians and Hispanics. Given previous evidence to suspect DIF, it was hypothesized that statistically significant DIF would be found. However, given the dearth of studies addressing alcohol dependence between these groups, no item-level predictions were made.

Methods

Participants

Participants ($n = 14,835$; 14,001 non-Hispanic Caucasians and 834 Hispanics) were a subset of the larger 1992 National Longitudinal Alcohol Epidemiologic Study, designed and sponsored by the National Institute on Alcohol Abuse and Alcoholism and fielded by the U.S. Census Bureau. The original sample consisted of 42,862 U.S. adults aged 18 years and older, selected at random from a sample representative of U.S. households nationwide. The complex, multistage design oversampled both the African American population and young adults aged 18 to 29 years and had household and sample person response rates of 92% and 97%, respectively (Grant et al., 1994). The current study included participants with complete data on the items under study. Individuals who reported the consumption of alcohol in the past 12 months were asked the studied questions.

Procedures

Direct face-to-face interviews were completed in English or bilingually in respondents' homes by experienced Census Bureau interviewers, lasted approximately 1 hour, and recorded information concerning alcohol consumption and problems, drug use and problems, periods of low mood, the use of alcohol and drug treatment, alcohol-related physical morbidity, family history of alcoholism, and sociodemographic background characteristics. Although interviewers may have conducted parts of the interview in Spanish, they adopted no formal procedure, did not record whether these particular items were asked in English or Spanish, and did not record participants' language preferences.

Measures

Alcohol dependence. A set of 30 items operationalizing the *DSM-IV* construct of alcohol dependence were selected from the larger Alcohol Use Disorders and Associated Disorders Interview Schedule (AUDADIS; Grant & Hasin, 1992; Grant & Towle, 1990). The *DSM-IV* describes alcohol dependence as a set of cognitive, behavioral, and physiological symptoms continued despite significant alcohol-related problems demonstrated by at least three of the following: (a) tolerance; (b) withdrawal; (c) drinking in larger amounts or over a longer period than intended; (d) enduring desire or failed efforts to control use; (e) substantial time spent obtaining, using, or recovering from alcohol; (f) the reduction of important social, occupational, or recreational activities because of alcohol; or (g) continued use despite knowing that a physical or physiological problem is likely to have been caused or exacerbated by alcohol. Items reflected this operationalization.

The AUDADIS is a fully structured diagnostic interview schedule that includes modules to measure alcohol and drug use, major mood disorder, substance-related medical conditions, and family history of alcohol and drug use disorders (Grant & Towle, 1990), and it can be used to generate diagnoses consistent with the *Diagnostic and Statistical Manual of Mental Disorders* (3rd ed., rev.) (American Psychiatric Association, 1987), the *DSM-IV*, and the *International Classification of Diseases* (10th rev.) (World Health Organization, 1992). Reliabilities, established through an independent test-retest study, generally exceed .70 and often exceed .90 (Grant et al., 1995). Additional studies have also established several types of validity (e.g., construct validity, criterion validity; Grant, 1997, 2000; Grant et al., 1995, 2003; Hasin et al., 1994; Hasin, Carpenter, et al., 1997; Hasin, Grant, et al., 1997; Hasin & Paykin, 1999; Harford & Muthén, 2001; Muthén, 1995; Muthén et al., 1993).

Ethnicity. Minority status was calculated using two items. The first asked participants to choose their race from among five categories: White, Black, American Indian, Asian, or other. A following question asked whether participants were Hispanic or non-Hispanic. In the current study, individuals were considered non-Hispanic Caucasians if they identified themselves as both White and non-Hispanic. Anyone who self-identified as Hispanic was considered Hispanic. In the current sample, this included approximately 96% White Hispanics, 4% Black Hispanics, and a small number of Hispanics identifying other categories. The interview collected no information regarding Hispanics' generational level, nor did it specifically ask for information on country of origin.

Results

Differential Item Functioning

CFA-OCM was used to probe for DIF. All analyses were conducted using Mplus 4.0 (Muthén & Muthén, 1998-2006). DIF was examined following the methods described by Millsap and Tein (2004), Byrne (1998), and Cheung and Rensvold (2002). Preferred levels of fit for indices of global and local model fit were adopted a priori and followed those suggested by Hu and Bentler (1998), Muthén and Muthén (1998-2006), Steiger (1998), and Cheung and Rensvold (2002). Given the functional dependence of the χ^2 statistic on sample size and its sensitivity to trivial deviations of fit (Cheung & Rensvold, 2002), model assessments were conducted using the set of indices. Rejection occurred when the majority of indices indicated inadequate fit. Models were identified as described by Millsap and Tein (2004), the θ parameterization and robust weighted least squares estimator were used in all analyses, and means and covariances were included at each step. Given arguments for more stringent error control in model tests of this type (Green & Babyak, 1997; Thissen, Steinberg, & Wainer, 1993), an α value of .01 was adopted.

A baseline single-factor model examined configural invariance across the Hispanic and Caucasian groups. The factor mean was fixed at zero for the Caucasian group, the factor variance was fixed at one for Caucasians, item intercepts were constrained to zero in each group, the loading for the "cut interests" item was constrained to equality across the groups, the threshold for the "cut interests" item was constrained to equality across the groups, and the uniquenesses were fixed to a value of one in each group for statistical identification.

The fit indices suggested the data were well fit by the configural invariance model: root mean square error of approximation (RMSEA) = .041, comparative fit index (CFI) = .95, McDonald's noncentrality index (NCI) = .94, γ hat = .992, $\chi^2(138, N = 14,835) = 1,895.23, p < .01$; configural invariance was not rejected, and metric invariance was examined. Retaining the constraints in the configural model, this model constrained loadings to equality across the groups and allowed variation in the remaining parameters. The $\Delta\chi^2$ test suggested the presence of significant DIF, $\Delta\chi^2(15, N = 14,835) = 32.57, p < .01$, and the hypothesis of metric invariance was rejected. Modification indices and expected parameter change indices identified the loading for the "used drugs to get over effects of drinking" item as predominantly responsible for the increase in misfit observed in the $\Delta\chi^2$ statistic. The model overestimated

the extent to which this item was related to alcohol dependence for Hispanics. This constraint was relaxed, and the model allowing partial measurement invariance for the “used drugs to get over effects of drinking” item’s loading fit the data well: RMSEA = .038, CFI = .97, McDonald’s NCI = .97, γ hat = .995, $\chi^2(99, N = 14,835) = 1,136.80, p < .01$, and $\Delta\chi^2(15, N = 14,835) = 22.86, p = .09$. Partial measurement invariance was not rejected, and analyses moved to examining invariance in the thresholds.

This model retained the restraints in the previous model, constrained the thresholds to equality across the groups, and allowed variation in the remaining parameters. The $\Delta\chi^2$ statistic demonstrated significant DIF: $\Delta\chi^2(22, N = 14,835) = 121.138, p < .01$. The modification indices and expected parameter change indices suggested that constraining the thresholds for the items “used drugs to get over effects of drinking,” “used drinking to get over drinking,” “usual number of drinks have less effect,” “drink more to get same effect,” “wanted to stop drinking,” “couldn’t stop drinking,” and “started drinking despite not wanting to” was predominantly responsible for the increase in misfit, overestimating the thresholds for Hispanics in each case. A model allowing partial measurement invariance for these thresholds fit the data well: RMSEA = .038, CFI = .97, McDonald’s NCI = .97, γ hat = .995, $\chi^2(100, N = 14,835) = 1,146.46, p < .01$, and $\Delta\chi^2(18, N = 14,835) = 33.601, p = .014$. The hypothesis of partial measurement invariance in the loadings and thresholds was not rejected, and analyses moved to examining invariance in the uniquenesses.

A new model hierarchy was motivated to examine invariance in the uniquenesses, given that a model with free uniquenesses was not nested within the previous models. It is not possible to statistically identify a model that simultaneously allows variation in the loadings, thresholds, and uniquenesses (Millsap & Tein, 2004). By incorporating the constraints just described, it was possible to establish a new free uniquenesses “baseline” model that constrained the loadings and thresholds across groups and allowed variation in the uniquenesses. The fit of this model was then compared with the fit of a model that constrained the loadings, thresholds, and uniquenesses across the groups.

The variant uniquenesses model retained the partial invariance constraints in the loadings and thresholds but relaxed the cross-group uniqueness constraints. This model fit the data well: RMSEA = .040, CFI = .96, McDonald’s NCI = .95, γ hat = .995, $\chi^2(123, N = 14,835) = 1,588.42, p < .01$. No evidence for DIF was found in the uniquenesses when the fit of this model was compared with that of a model that specified equivalent uniquenesses: RMSEA = .038, CFI = .98, McDonald’s NCI = .97, γ hat =

.995, $\chi^2(34, N = 14,835) = 1,146.46, p < .001$, and $\Delta\chi^2(16, N = 14,835) = 29.909, p = .018$. Given the final set of fit indices, a fully invariant measurement model was rejected and a partially invariant measurement model specifying partial invariance in the loadings and thresholds was specified in its place. The final completely standardized estimates for this model are summarized in Table 1.

Impact

Statistical significance does not always translate into practical differences, and a sensitivity analysis examined impact. The means from a fully invariant model ignoring observed DIF (i.e., the analysis one would perform if one were not aware of DIF) were compared with those from the model incorporating measurement differences. If results from the models differed, this would suggest impact. In both the fully and partially invariant models, the group means were not significantly different from each other (fully invariant model: $M_{\text{Caucasians}} = 0.00, SD_{\text{Caucasians}} = 1, M_{\text{Hispanics}} = -0.09, SD_{\text{Hispanics}} = 1, z = -1.56$; partially invariant model: $M_{\text{Caucasians}} = 0.00, SD_{\text{Caucasians}} = 1, M_{\text{Hispanics}} = 0.01, SD_{\text{Hispanics}} = 1, z = 0.2$). Given that significant mean differences were not observed in either model, the hypothesis that DIF affected scores was rejected.

Discussion

The current study used CFA-OCM to examine DIF on a standardized measure of alcohol dependence across Hispanic Americans and non-Hispanic Caucasian Americans in a large survey representative of U.S. adults in 1992. It addressed the validity of alcohol dependence cross-culturally and whether disparities and levels of use described in this cohort reflected true differences or were influenced by DIF. Analyses examined whether 30 items operationalizing the *DSM-IV* construct of alcohol dependence provided equivalent measurement. The results demonstrated the presence of statistically significant DIF for 7 items. These items addressed whether individuals used drugs to get over the effects of drinking, whether they used drinking to get over the effects of drinking, whether the usual number of drinks had less effect, whether they drank more to get the same effect, whether they wanted to stop drinking, whether they couldn't stop drinking, and whether they started drinking despite not wanting to. This suggests caution regarding the cross-ethnic validity of the alcohol

Table 1
Final Partial Alcohol-Dependence Measurement Invariance Model
Across Hispanics and Non-Hispanic Caucasians

Abbreviated Item Label	Loading		Threshold	
	Caucasians	Hispanics	Caucasians	Hispanics
Cut interests	.94	.94	-1.30	-1.09
Usual number less effect	.69	.70	-1.66	-1.35
Drank more to get effect	.80	.81	-1.03	-0.77
Wanted to stop drinking	.70	.72	-2.03	-1.63
Couldn't stop drinking	.79	.80	-1.58	-1.34
Started drinking didn't want to	.77	.78	-0.82	-0.79
Drank more than wanted	.84	.85	-1.18	-1.15
Drank longer than intended	.80	.81	-1.49	-1.47
Trouble sleeping	.60	.61	-1.85	-1.80
Shakes when alcohol wears off	.77	.78	-1.46	-1.43
Feel irritable when alcohol wears off	.79	.80	-1.08	-1.05
Feel sick when alcohol wears off	.71	.72	-0.69	-0.67
Headache after drinking	.71	.72	-1.70	-1.66
Alcohol sweats	.71	.73	-2.48	-2.44
Alcohol hallucinations	.68	.69	-1.74	-1.49
Drank to get over drinking	.92	.93	-2.14	-2.02
Used drugs to get over drinking	.66	.38	-1.80	-1.73
Drink to avoid hangover	.93	.94	-2.25	-2.18
Too much time drinking	.90	.91	-2.21	-2.15
Hung over a lot	.80	.81	-2.21	-2.16
Time making alcohol available	.82	.83	-2.30	-2.22
Cut important activities	.90	.91	-2.40	-2.32
Think only of drinking	.89	.89	-2.29	-2.22
Strong desire to drink	.77	.78	-1.38	-1.34
Heavy drinking same way all the time	.88	.89	-2.10	-2.03
Regular drinking no matter what or where	.83	.84	-2.16	-2.10
Not able to remember after drinking	.80	.81	-1.40	-1.36
Kept drinking without sobriety	.83	.84	-2.13	-2.07
Drank fifth of liquor one day	.75	.76	-1.63	-1.60
Couldn't handle daily life without drinking	.92	.93		
	Caucasians		Hispanics	
Dependence factor mean	0.00		0.01	
Dependence factor variance	1.00		1.00	

dependence construct and items across Hispanics and non-Hispanic Caucasians in the 1992 cohort. Impact analyses indicated that the practical effect on aggregate scores may be minimal.

Analyses demonstrated that a partially invariant model allowing non-equivalence in one of the loadings and seven of the thresholds provided the best fit to the data. This has several interpretive implications. First, DIF in the loadings demonstrated that responses to the item assessing whether Hispanics used drugs to get over the effects of drinking were not as well predicted from the underlying construct of alcohol dependence compared with Caucasians. Loadings can be interpreted similarly to correlations; more variance in the item is accounted for by the underlying factor (i.e., alcohol dependence) when values are large. Substantively, endorsements of this item were not as closely tied to the underlying construct of alcohol dependence for Hispanics as they were for non-Hispanic Caucasians. One would have less faith that Hispanics' responses to this item were a function of alcohol dependence. Alcohol dependence does predict responses to this item for both groups, but less so for Hispanics.

Evidence was also found for DIF in the thresholds. CFA-OCM assumes that observed yes or no responses are a function of a threshold along a continuous latent response variate. For the items studied here, if the individual's level of the latent response variate is less than the threshold, the individual will answer yes to the item. If not, the individual will answer no. Seven items demonstrated DIF in the thresholds. For each of the items, Hispanics needed to experience less of the concept the item was measuring before endorsing the item. For example, consider the item assessing whether individuals wanted to stop drinking. The difference in the threshold suggested that on average, it was easier for Hispanics to endorse this item compared with non-Hispanic Caucasians (i.e., the desire to stop drinking did not need to be as strong in order for Hispanics to answer yes to the question of whether they had wished to stop drinking). Hispanics more readily endorsed this item compared with non-Hispanic Caucasians. The pattern of differences and interpretation are the same for the remaining items. Taken together, these results demonstrate the presence of statistically significant DIF in the measure of alcohol dependence across Hispanics and non-Hispanic Caucasians and that caution is warranted when estimating rates and levels of alcohol dependence across these groups.

Given that many psychological assessments use aggregate, total scores resulting from measures such as these, it is worthwhile to examine whether observed DIF affected conclusions drawn from the instrument. Partial measurement invariance need not lead to biased observed scores (Byrne, 1988, Byrne & Campbell, 1999; Cole & Maxwell, 1985), nor do statistically significant criteria necessarily translate into meaningful differences (Byrne & Campbell, 1999). A sensitivity analysis compared the pattern, size, and direction

of mean differences across a model proceeding as if DIF were not present and a model incorporating DIF. This comparison asked whether analyses conducted ignoring DIF would diverge from those incorporating bias. Under both models, the pattern was similar and significant. Although measuring alcohol dependence is affected by differential item responses across Hispanics and non-Hispanic Caucasians and estimates comparing these groups will more accurately reflect true levels of alcohol dependence when incorporating DIF, these differences likely do not affect the observed scores using aggregated item scores. This finding is particularly important; it supports the validity of previous research and clinical work using the alcohol dependence construct among the Hispanic population and supports the continued use of the construct.

Although summary scores may not be affected, caution is warranted at the item level. Item-level DIF suggests that some ethnic differences do exist in the measurement of alcohol dependence for Hispanics. Several possible explanations exist. Hispanics may share a common cultural value or experience that influences their responses (Prelow et al., 2000, 2002). Hui and Triandis (1989) described sincerity as a cultural value among Hispanics. Perhaps a desire to be honest and sincere makes it "easier" to endorse items that describe problematic drinking patterns. Likewise, it may be that less stigma is associated with divulging a desire to drink less, easing endorsements. That is, the cultural experience of dependence may be different for Hispanics. Alternatively, methodological issues alone may have affected Hispanics' responses to these items. It may be the survey setting itself that resulted in the differences (Tourangeau, Rips, & Rasinski, 2000). Perhaps a large-scale survey fielded by the federal government systematically affects individuals' perceptions of these items and their subsequent desire to answer them. It may also be that the construct validity of alcohol dependence as defined by these behaviors is poor for this group. These possibilities need exploration.

Before concluding, it is worthwhile to note some of the study's limits. First, invariance was examined in a representative sample of the U.S. population. It remains unclear whether results would persist in clinical samples. Additionally, the analysis was unable to explore DIF across more specifically defined Hispanic groups (Mexican Americans, Puerto Ricans), given their smaller sample sizes with the included Hispanic group. Perhaps the results do not hold for all Hispanics. Third, the study was unable to address whether differences resulted from methodological problems, differential perceptions of alcohol dependence, cultural differences, or some combination. Fourth, the study was unable to empirically examine possible explanations for the observed DIF. Future research would extend the field by addressing underlying sources of DIF. Finally, the study's goal was to

examine whether earlier findings were influenced by DIF and so examined a cohort from the early 1990s. It did not assess DIF in current population data. It remains unclear whether DIF affects current population assessments. Fortunately, this work is under way at the University of North Florida, and the results of this study will be available soon.

In conclusion, the current study demonstrated statistically significant DIF across Hispanics and non-Hispanic Caucasians in the United States for seven items that were part of a set operationalizing *DSM-IV* alcohol dependence. In a representative sample of the 1992 United States, a partial measurement invariance model best fit the data. Researchers and clinicians should note some concern when diagnosing Hispanic individuals on the basis of these items, and some caution may be judicious when estimating rates for Hispanic individuals. However, given minimal impact, results continue to support findings evidenced among Hispanics in the early 1990s.

References

- American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed. rev.). Washington, DC: Author.
- American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders* (4th ed.). Washington, DC: Author.
- Anderson, N. B., & Nickerson, K. J. (2005). Genes, race, and psychology in the genome era: An introduction. *American Psychologist, 60*, 5-8.
- Bachman, J. G., & O'Malley, P. M. (1984). Yes-saying, nay-saying, and going to extremes: Black-White differences in response styles. *Public Opinion Quarterly, 48*, 491-509.
- Byrne, B. (1988). Measuring adolescent self-concept: Factorial validity and equivalency of the SDQ-III across gender. *Multivariate Behavioral Research, 24*, 361-375.
- Byrne, B. (1998). *Structural equation modeling with LISREL, PRELIS, and SIMPLIS: Basic applications and programs*. Hillsdale, NJ: Lawrence Erlbaum.
- Byrne, B. M., Baron, P., & Baley, J. (1996). The Beck Depression Inventory: Testing for its factorial validity and invariance across gender for Bulgarian non-clinical adolescents. *Personality and Individual Differences, 21*, 641-651.
- Byrne, B. M., Baron, P., & Baley, J. (1998). The Beck Depression Inventory: A cross-validated test of second-order factorial structure for Bulgarian adolescents. *Educational and Psychological Measurement, 58*, 241-251.
- Byrne, B. M., Baron, P., & Campbell, T. L. (1993). Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender. *Journal of Research on Adolescence, 3*, 127-143.
- Byrne, B. M., Baron, P., & Campbell, T. L. (1994). The Beck Depression Inventory (French version): Testing for gender-invariant factorial structure for nonclinical adolescents. *Journal of Adolescent Research, 9*, 166-179.

- Byrne, B. M., & Campbell, T. L. (1999). Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. *Journal of Cross Cultural Psychology, 30*, 555-574.
- Chatterji, S., Saunders, J. B., Vrsti, R., Grant, B. F., Hasin, D., & Mager, D. (1997). Reliability of the alcohol and drug modules of the Alcohol Use Disorder and Associated Disabilities Interview Schedule-Alcohol/Drug-Revised (AUDADIS-ADR): An international comparison. *Drug and Alcohol Dependence, 47*, 171-185.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233-255.
- Cole, D. A., Martin, J. M., Peeke, L., Henderson, A., & Harwell, J. (1998). Validation of depression and anxiety measures in White and Black youths: Multitrait-multimethod analyses. *Psychological Assessment, 10*, 261-276.
- Cole, D. A., & Maxwell, S. E. (1985). Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. *Multivariate Behavioral Research, 20*, 389-417.
- Cole, S. R. (1999). Assessment of differential item functioning in the Perceived Stress Scale-10. *Journal of Epidemiology and Community Health, 53*, 319-320.
- Dawson, D. A., Grant, B. F., Chou, P. S., & Pickering, R. P. (1995). Subgroup variation in U.S. drinking patterns: Results of the 1992 National Longitudinal Alcohol Epidemiologic Study. *Journal of Substance Abuse, 7*, 331-344.
- Gallo, J. J., Anthony, J. C., & Muthén, B. O. (1994). Age differences in the symptoms of depression: A latent trait analysis. *Journal of Gerontology, 49*, 251-264.
- Grant, B. F. (1997). Convergent validity of DSM-III-R and DSM-IV alcohol dependence: Results from the National Longitudinal Alcohol Epidemiologic Survey. *Journal of Substance Abuse, 9*, 89-102.
- Grant, B. F. (2000). Theoretical and observed subtypes of DSM-IV alcohol abuse and dependence in a general population sample. *Drug and Alcohol Dependence, 60*(3), 287-293.
- Grant, B. F., Dawson, D. A., & Hasin, D. S. (2001). *The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version (AUDADIS-IV)*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
- Grant, B. F., Dawson, D. A., & Stinson, F. S. (2003). The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of alcohol consumption, tobacco use, family history of depression and psychiatric diagnostic modules in a general population sample. *Drug and Alcohol Dependence, 71*, 7-16.
- Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., Dufour, M. C., & Pickering, R. P. (2004). The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991-1992 and 2001-2002. *Drug and Alcohol Dependence, 74*, 223-234.
- Grant, B. F., Harford, T. C., & Dawson, D. (1995). The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): Reliability of alcohol and drug modules in a general population sample. *Drug and Alcohol Dependence, 39*, 37-44.
- Grant, B. F., Harford, T. C., Dawson, D. A., Chou, P., Dufour, M., & Pickering, R. (1994). Prevalence of DSM-IV alcohol abuse and dependence. *Alcohol Health and Research World, 18*, 243-248.
- Grant, B. F., & Hasin, D. S. (1992). *The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS)*. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism.
- Grant, B. F., & Towle, L. H. (1990). Standardized diagnostic interviews for alcohol research. *Alcohol Health and Research World, 14*, 340-348.

- Green, S. B., & Babyak, M. A. (1997). Control of Type I errors with multiple tests of constraints in structural equation modeling. *Multivariate Behavioral Research, 32*, 39-51.
- Harford, T. C., & Muthén, B. O. (2001). Alcohol use among college students: The effects of prior problem behaviors and change of residence. *Journal of Studies on Alcohol, 62*, 306-312.
- Hasin, D., Carpenter, K. M., & McCloud, S. (1997). The alcohol use disorder and associated disabilities interview schedule (AUDADIS): Reliability of alcohol and drug modules in a clinical sample. *Drug and Alcohol Dependence, 44*, 133-141.
- Hasin, D. S., & Grant, B. F. (2004). The co-occurrence of DSM-IV alcohol abuse in DSM-IV alcohol dependence: Results of the National Epidemiologic Survey on Alcohol and Related Conditions on heterogeneity that differ by population subgroup. *Archives of General Psychiatry, 61*, 891-896.
- Hasin, D. S., Grant, B., & Cottler, L. (1997). Nosological comparisons of alcohol and drug diagnoses: A multisite, multi-instrument international study. *Drug and Alcohol Dependence, 47*, 217-226.
- Hasin, D. S., Muthén, B., Wisnicki, K. S., & Grant, B. F. (1994). Validity of the bi-axial dependence concept: A test in the US general population. *Addiction, 89*, 573-579.
- Hasin, D., & Paykin, A. (1999). Alcohol dependence and abuse diagnoses: Concurrent validity in a nationally representative sample. *Alcoholism: Clinical and Experimental Research, 23*, 144-150.
- Hu, L., & Bentler, P. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424-453.
- Huang, C. D., Church, A. T., & Katigbak, M. S. (1997). Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. *Journal of Cross Cultural Psychology, 28*, 192-218.
- Hui, C. H., & Triandis, H. C. (1989). Effects of culture and response format on extreme response style. *Journal of Cross-Cultural Psychology, 20*, 296-309.
- Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H. U., & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: Results from the National Comorbidity Survey. *Archives of General Psychiatry, 51*, 8-19.
- Knight, G. P., & Hill, N. E. (1998). Measurement equivalence in research involving adolescents. In V. C. McLoyd & L. Steinberg (Eds.), *Studying minority adolescents: Conceptual, methodological, and health issues* (pp. 183-210). Hillsdale, NJ: Lawrence Erlbaum.
- Knight, G., Tein, J.-Y., Shell, R., & Roosa, M. (1992). The cross-ethnic equivalence of parenting and family interaction measures among Hispanic and Anglo-American families. *Child Development, 63*, 1392-1403.
- Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research, 13*, 127-143.
- Millsap, R. E., & Tein, J.-Y. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*, 479-515.
- Muthén, B. O. (1995). Factor analysis of alcohol abuse and dependence symptom items in the 1988 National Health Interview Survey. *Addiction, 90*, 637-645.
- Muthén, B. O. (1998-2006). *Mplus user's guide* (4th ed.). Los Angeles: Muthén & Muthén.
- Muthén, B. O., Hasin, D., & Wisnicki, K. S. (1993). Factor analysis of ICD-10 symptom items in the 1988 National Health Interview Survey on Alcohol Dependence. *Addiction, 88*, 1071-1077.
- Pentz, M. A., & Chou, C. P. (1994). Measurement invariance in longitudinal clinical research assuming change from development and intervention. *Journal of Consulting and Clinical Psychology, 62*, 450-462.

- Prelow, H. M., Michaels, M. L., Reyes, L., Knight, G. P., Barrera, M., Jr. (2002). Measuring coping in low-income European American, African American, and Mexican American adolescents: An examination of measurement equivalence. *Anxiety, Stress and Coping*, 15, 135-147.
- Prelow, H., Tein, J.-Y., Roosa, M. W., & Wood, J. (2000). Do coping styles differ across socio-cultural groups? The role of measurement equivalence in making this judgment. *American Journal of Community Psychology*, 28, 225-244.
- Regier, D. A., Boyd, J. H., Burke, J. D., Jr., Rae, D. S., Myers, J. K., Kramer, M., Robins, L. N., George, L. K., Karno, M., & Locke, B. Z. (1988). One-month prevalence of mental disorders in the United States. Based on five epidemiologic catchment area sites. *Archives of General Psychiatry*, 45, 977-986.
- Reid, R., DuPaul, G. J., Power, T. J., Anastopoulos, A. D., Rogers-Adkinson, D., Noll, M.-B., & Riccio, C. (1998). Assessing culturally different students for attention deficit hyperactivity disorder using behavior rating scales. *Journal of Abnormal Child Psychology*, 26, 187-198.
- Schafer, J., & Caetano, R. (1996). The DSM-IV construct of cocaine dependence in a treatment sample of Black, Mexican American, and White men. *Psychological Assessment*, 8, 304-311.
- Smith, L., & Reise, S. P. (1998). Gender differences on negative affectivity: An IRT study of differential item functioning on the Multidimensional Personality Questionnaire Stress Reaction Scale. *Journal of Personality and Social Psychology*, 75, 1350-1362.
- Smith, P. B. (2004). Acquiescent response bias as an aspect of cultural communication style. *Journal of Cross-Cultural Psychology*, 35, 50-61.
- Steiger, J. H. (1998). A note on multiple sample extensions of the RMSEA fit index. *Structural Equation Modeling*, 5, 411-419.
- Sue, S. (1999). Science, ethnicity, and bias: Where have we gone wrong? *American Psychologist*, 54, 1070-1077.
- Thissen, D., Steinberg, L., & Wainer, H. (1993). Detection of differential item functioning using the parameters of item response models. In P. W. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 67-113). Hillsdale, NJ: Lawrence Erlbaum.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge, UK: Cambridge University Press.
- Waller, N. G., Thompson, J. S., & Wenk, E. (2000). Using IRT to separate measurement bias from true group differences on homogeneous and heterogeneous scales: An illustration with the MMPI. *Psychological Methods*, 5, 125-146.
- World Health Organization. (1992). *Tenth revision of the International Classification of Diseases—Clinical descriptions and diagnostic guidelines version (ICD-10 CDDG)*. Geneva, Switzerland: Author.
- Wright, G. N., Phillips, L. D., Whalley, P. C., Gerry, T. C., Kee-Ong, N., Tan, I., & Wisudha, A. (1978). Cultural differences in probabilistic thinking. *Journal of Cross-Cultural Psychology*, 9, 285-299.

Adam C. Carle, PhD, is an assistant professor of psychology at the University of North Florida teaching undergraduate and graduate courses in research methods. He pursues two streams of programmatic research. One, exemplified here, examines the cross-cultural validity of psychological measures and the effects of measurement bias on public health efforts to understand and eliminate health disparities, particularly among the Hispanic community. The second applies experimental theories of psychology to the science of learning in young adulthood. Outside of academia, he enjoys baking with his wife Tara.