Published in final edited form as:
Addict Behav. 2009 January ; 34(1): 43–50. doi:10.1016/j.addbeh.2008.08.004.

Cross-cultural Invalidity of Alcohol Dependence Measurement across Hispanics and Caucasians in 2001 and 2002

Adam C. Carle, Ph.D.
University of North Florida, 1 UNF Drive, Jacksonville, FL, 32224, Phone: 904-620-3573, Fax: 904-620-3814, e-mail: adam.carle@unf.edu

Abstract

Aims—Do assessments of alcohol dependence demonstrate similarly validity across Hispanics and non-Hispanic Caucasians? This investigation examined this question.

Method—It employed confirmatory factor analyses for ordered-categorical measures to search for measurement bias on the AUDADIS, a standardized measure of alcohol dependence, across Hispanic (n = 4,819) and non-Hispanic Caucasians (n = 16, 109) in a nationally representative survey of alcohol use in the United States conducted in 2001 and 2002.

Measurement—Analyses considered whether 27 items operationalizing the DSM-IV alcohol dependence construct provided equivalent measurement.

Findings and Conclusions—Nine items revealed statistically significant bias, suggesting strong caution regarding the cross-ethnic validity of alcohol dependence. Sensitivity analyses established that item level differences erroneously impact alcohol dependence estimates among the 2001 – 2002 US Hispanic population. Biased measurement underestimates differences between Hispanics and non-Hispanic Caucasians, underestimates Hispanics' true use levels, and falsely minimizes current increases in drinking behavior evidenced among Hispanics. Findings urge improved public health efforts among the Hispanic community and underscore the necessity for cultural sensitivity when generalizing measures and constructs developed in the majority to Hispanic individuals.

Keywords

Measurement Bias; Cross-Cultural Differences; Hispanics; Validity; Alcohol Dependence

Cross-cultural Invalidity of Alcohol Dependence Measurement across Hispanics and Caucasians in 2001 and 2002

Psychology and its allied disciplines too often fail to fully consider culture, despite repeated, erudite calls to the contrary (Kazdin, 2003; Sue, 1999), and research often assumes that constructs valid in the majority demonstrate equivalent validity for minorities (Teresi, et al., 2006). Alcohol dependence does not escape this problem. Recent estimates suggest a US DSM-IV alcohol dependence prevalence of 3.81 % (Grant, et al., 2004). However, studies comparing the prevalence and comorbidity of alcohol use disorders across Caucasians and cultural minorities note significant differences, showing significantly lower rates of comorbid alcohol

Correspondence to: Adam C. Carle.

use disorders among Hispanics as compared to non-Hispanic Caucasians and establish a changing drinking behavior trend; it remains relatively stable among Caucasians and increases for Hispanics ((Dawson, et al., 1995; Grant, et al., 1994; Grant, et al., 2004; Hasin & Grant, 2004). These investigations demonstrate noteworthy epidemiological differences. They also underscore a need for culturally sensitive public health policy, especially given health disparities (Bloche, 2004; Fiscella, et al., 2000; Ramírez, et al., 2005; Smedley, et al., 2003; Steinbrook, 2004; Stewart & Nápoles-Springer, 2003) and alcohol dependence's cost to all (Greenfield, 1998; Harwood, 1998).

Unfortunately, comparisons like these frequently rest on the *unassessed* assumption that constructs valid for Caucasians achieve similar status among the Hispanic community. Before validly comparing ethnicities, we must ask whether measurements function similarly across groups (Stahl & Hahn, 2006). The discipline must consider whether observed differences reflect true differences or result from poor measurement. Measurement bias, also labeled differential item functioning (DIF), refers to the possibility that individuals equal in their true levels of alcohol dependence, but from different groups, have dissimilar alcohol use item response probabilities (Mellenbergh, 1989). Distressingly, bias can diminish or amplify differences, lead to inaccurate diagnoses, decrease reliability and validity, and render comparisons impossible (Carle, et al., in press; Carle, in press; McHorney & Fleishman, 2006), a problematic possibility when tracking change. For example, measurement bias may minimize the recent increase in drinking among Hispanics, creating the *appearance* of a smaller increase that what truly occurred. If bias affects estimates, apparent change may not reflect true change. And, although studies have established the validity and reliability of standardized alcohol dependence measures and diagnostic criteria in collectively (Chatterji, et al., 1997; Grant, 1997, Grant, 2000; Grant, et al., 2001; Grant, et al., 2003; Grant, et al., 2007; Hasin, et al., 1994; Harford & Muthén, 2001; Kessler, et al., 1994; Muthén, et al., 1993; Muthén, 1995; Grant, et al., 1995; Hasin, et al., 1997a; Hasin, et al., 1997b; Hasin & Paykin, 1999), the role of minority/majority based measurement bias in instruments used to assess alcohol dependence in the US goes relatively unexamined.

Theoretically and empirically, we should suspect cultural measurement bias (Sue, 1999). Wright, et al. (1978) noted cultural differences in probabilistic thinking and assignment of numbers, Smith (2004) described differences in acquiescent responses, and Bachman and O'Malley (1984) discuss differences in language use across Hispanics, Caucasians, and other minorities. Authors have also noted that the behavioral exemplars describing a psychological construct for the majority may not be appropriate for minorities, nor do exemplars necessarily include the full set of culturally appropriate indicators (Huang, et al., 1997; Prelow, et al., 2000; Prelow et al., 2002). Culturally specific values may also influence responses. Hui and Triandis (1989) hypothesize that the Hispanic culture generally values sincere responses that lead to more ready endorsements of scale end points because the middle of scales often reflect a "don't know", "no opinion", or similar option. Prelow, et al. (2000) suggest that for certain behaviors greater levels of a specific problem may be needed before Hispanics willingly acknowledge a problem. McHorney and Fleishman (2006) note that survey questions may trigger differential cultural perceptions regarding desirable responses and that question wording may impede symptomatology reporting. In sum, we have strong reason to advocate concern that bias affects the validity of measurement across Hispanics and Caucasians generally, and no reason to eliminate alcohol dependence from these qualms. Indeed, in a recent reexamination of alcohol dependence among a 1992 cohort, Carle (in press) found statistically significant measurement bias across Hispanics and non-Hispanic Caucasians.

Disappointingly, a review found no published studies examining the validity of alcohol dependence across Caucasians and Hispanics in recent data, and only one (Carle, in press) examining previous data. Thus, it remains unclear if bias affects current research. This leaves

vague whether findings suggesting differential prevalence and comorbidity and discrepant changes across time reflect true states, measurement bias, or both. Fortunately, modern methods like confirmatory factor analysis (CFA) provide a preferred technique to investigate bias (Stahl & Hahn, 2006; Waller, et al., 2000). As a result, the current study had two goals. It used modern measurement models to examine whether statistically significant measurement bias exists across Hispanic and non-Hispanic Caucasians on a standardized measure of alcohol dependence in a large, nationally representative survey of alcohol use in the United States conducted in 2001 and 2002, and, if so, to what extent it impacts estimates. It used these results to explore whether reports noting recent change need modification. Should the field *enlarge* or *diminish* current estimates as a function of bias measurement?

Methods

Participants—Participants (n = 20,928; 16,109 non-Hispanic Caucasians and 4,819 Hispanics) were a subset of the larger, publicly available 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC: Grant, Kaplan, Shepard, & Moore, 2003) data sponsored by the National Institute for Alcohol Abuse and Alcoholism (NIAAA) and fielded by the US Census Bureau. The original sample represented the non-institutionalized adult US population and consisted of 43,093 adults aged 18 and older. The complex, multistage design incorporated the Census 2000/2001 Supplementary Survey and Census 2000 Group Quarters Inventory sampling frames and oversampled African American, Hispanics, and young adults (18 – 24). Sample weights, specified elsewhere (Grant, et al., 2003), adjust the data to make it representative of the civilian, non-institutionalized US population across several socioeconomic variables using the 2000 Decennial Census and reduce variance resulting from two primary sampling units (PSUs) representing one stratum. They also adjust other issues: the probability of PSUs selection within stratum and housing unit selection within PSUs, the selection of a single individual from each household, young adult oversampling, and household and person nonresponse. The NESARC had household and sample person response rates of 89% and 93% respectively. The current study included participants with complete data reporting on their alcohol consumption in the past 12 months.

Procedures—Skilled Census Bureau interviewers conducted face to face interviews in respondents' homes and recorded information concerning: alcohol consumption and problems, drug use and problems, periods of low mood, utilization of alcohol and drug treatment, alcohol-related physical morbidity, family history of alcoholism, sociodemographic background characteristics, and additional variables.

Measures

Alcohol Dependence: 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: tolerance; withdrawal; drinking in larger amounts or over a longer period than intended; enduring desire or failed efforts to control use; substantial time spent obtaining, using, or recovering from alcohol; reduction of important social, occupational, or recreational activities because of alcohol; or continued use despite knowing a physical or physiological problem is likely to have been caused or exacerbated by alcohol. 27 items selected from the Alcohol Use Disorder and Associated Disabilities Interview Schedule DSM-IV Version (AUDADIS-IV: Grant, et al., 2001), a fully structured diagnostic interview schedule used in the NESARC, mirrored this operationalization. The AUDADIS-IV generates diagnoses consistent with the Fourth Edition of the DSM (DSM-IV; American Psychiatric Association, 1994) among others. Reliabilities established through test-retest generally exceed 0.70 and often surpass 0.90 (Ruan, et al., 2008, Grant, et al., 1995). Additional studies have established several types of validity, e.g., construct, criterion, etc. (Grant, 1997, Grant, 2000; Grant, et al., 2003; Hasin, et al., 1994; Harford & Muthén, 2001; Muthén, et al.,

1993; Muthén, 1995; Grant, et al., 1995; Hasin, et al., 1997a; Hasin, et al., 1997b; Hasin & Paykin, 1999).

Ethnicity: The NESARC used five race codes: American Indian and Alaska Native; Asian; Black or African American; Native Hawaiian and Other Pacific Islander; and White. A single item allowed Hispanic self-identification. The current study considered individuals non-Hispanic Caucasians if they identified themselves as both White and non-Hispanic and regarded anyone who self-identified as Hispanic a Hispanic.

Analytic Strategy

Confirmatory Factor Analysis for Ordered-Categorical Measures: The current study used confirmatory factor analyses for ordered-categorical measures (CFA-OCM) to probe for bias. CFA-OCM appropriately and explicitly models the categorical nature of the items (Muthén & Christoffersson, 1981). I review briefly here this method here, but, for detailed reviews, the reader should consult Byrne (1998), Millsap & Yun-Tien (2004), Muthén (1984), or Muthén & Christoffersson (1981).

CFA-OCM indicates a set of equations to describe the relations among a set of ordered-categorical items, suggesting that individuals' item responses are determined by their value on an underlying factor or factors and several measurement parameters, e.g., intercepts, loadings, thresholds, and uniquenesses. In the CFA-OCM model, loadings, similar to correlations, represent the degree to which an item relates to the factor(s); the greater the value of the factor loading, the greater the relation between the item and the latent variable of interest. The threshold parameters reflect the ordered-categorical nature of the items. The model assumes that a continuous latent response variate underlies discrete item response categories. If an individual's value on the latent response variate is less than the threshold, they will respond in one category, but, if their value is greater than the threshold, they will respond in at least the next highest category. Intercept parameters give the expected value of an item when the value of the underlying factor(s) is zero, and uniquenesses include sources of variance not attributable to the factor(s), including measurement error (Bollen, 1989).

In measurement bias studies, researchers examine the equivalence of the measurement parameters across groups. In practice, a series of hierarchically nested models typically test measurement bias Bollen, 1989; Byrne, 1998; Millsap and Yun-Tein, 2004). The method starts with the least restricted measurement model across groups and adds cross-group equivalence constraints in the measurement parameters in a stepwise fashion in later models. Fit indices describe the tenability of the equivalence constraints in a given set of measurement parameters at each step. When these indices suggest untenable constraints, analyses have identified statistically significant measurement bias. Finally, work of this type distinguishes between full and partial measurement invariance. Full measurement invariance implies that an entire set of item parameters achieve equality across the groups, e.g., all of the loadings, thresholds, intercepts, and uniqueness demonstrate equivalence. However, statistically significant measurement bias may result from a limited number of parameters rather than bias across the entire set of item, e.g., a small number of loadings. To investigate this, analysts test a partial measurement invariance hypothesis. This hypothesis constrains some measurement parameters to equality across the groups and allows inequivalence in others. In this way, researchers can fully model cross-cultural differences in measurement bias and examine whether some or all items demonstrate bias. The current investigation adopted this approach.

In the current study. I conducted all analyses using Mplus, its theta parameterization, and robust weighted least squares (WLSMV) estimator (Mplus 5.1; Muthén & Muthén, 2007). Analyses examined measurement bias following the method described above and in detail by Millsap and Yun-Tein (2004). A priori, the studied adopted preferred levels of fit for indices of global

and local model fit suggested by Hu and Bentler (1998), Muthén and Muthén (2007), Steiger (1998), and Cheung and Rensvold (2002). Fit assessment focused on the set and preferred root mean square error of approximation (RMSEA) values less than 0.05; comparative fit index (CFI), Tucker-Lewis Index (TLI), Gamma Hat values greater than 0.95; and McDonald's noncentrality index (NCI) values greater than 0.90. Models included means and covariances at each step. Statistical identification conformed to Millsap and Yun-Tein's (2004) description. Consistent with calls for more stringent error control in modeling (Green & Babyak, 1997; Thissen, Steinberg, & Wainer, 1993), analyses used an α of 0.01.

Results

Across the Hispanic and non-Hispanic Caucasian groups, a baseline single factor model tested configural invariance. This essentially examined whether the same general measurement model held in both groups. It asked, did a single factor model adequately describe the relations among the items for both groups? For statistical identification, this model: fixed the factor mean at zero for the Caucasian group, fixed the factor variance at one for Caucasians, constrained item intercepts to zero across group, constrained the loading for the "*usual number of drinks had less effect*" item to equality across groups, constrained the threshold for the "*usual number*" item to equality across groups, and fixed the uniquenesses at one across groups. The index set indicated a well fitting model (RMSEA = 0.035, CFI = 0.93, TLI, 0.97, McDonald's NCI = 0.93, Gamma Hat = 0.990, and χ^2 = 3204.01, 228, n = 20,928, $p < 0.01$).

Analyses turned to equivalence in the loadings. These analyses essentially asked whether each item related to the underlying alcohol dependence construct to a similar degree across Caucasians and Hispanics. This model retained the previous specification and additionally constrained the loadings to equality across the groups. This model fit the data well ($\Delta\chi^2$ = 25.082, 17, n = 20,928, p = 0.09, RMSEA = 0.032, CFI = 0.96, TLI = 0.98, McDonald's NCI = 0.96, Gamma Hat = 0.994, χ^2 = 1817.812, 158, n = 20,928, $p < 0.01$).

To investigate whether differential reporting tendencies at similar levels of alcohol existed across ethnicity, analyses assessed the fit of a model constraining the threshold parameters to equivalence across Caucasians and Hispanics. This model retained the previous constraints and added thresholds equality constraints. The $\Delta\chi^2$ test noted statistically significant measurement bias: $\Delta\chi$ 2 = 244.34 (22, n = 20,928, $p < 0.01$). Modification indices (MIs) and expected parameter change indices (EPCs) suggested that constraining nine thresholds predominantly accounted for the misfit. For the: 1) "*drink less to get effect wanted*", 2) "*trouble falling/staying asleep when effects end*", 3) "*kept drinking longer than intended*", and 4) "*drinking more than intended*" items, the equality constraint underestimated Hispanics' thresholds. For the: 5) "*give up pleasurable activities*", 6) "*multiple unsuccessful attempts to quit*", 7) "*see, feel, hear things when effects end*", 8) "*drink more to get effect*", and 9) "*more than once want to stop*" items, the equality constraint overestimated Hispanics' thresholds. A model allowing partial measurement invariance for these nine thresholds fit the data well: RMSEA = 0.031, CFI = 0.96, TLI = 0.98, McDonald's NCI = 0.96, Gamma Hat = 0.994, χ^2 = 1838.82 (163, n = 20,928, $p < 0.01$), and $\Delta\chi^2$ = 29.081 (16, n = 20,928, p = 0.02). The partially invariant thresholds hypothesis was not rejected and analyses moved to the uniquenesses.

To examine uniqueness invariance, analyses developed a new hierarchy given that a free uniquenesses model did not nest within earlier models. Dichotomous models cannot statistically identify a model simultaneously allowing variation in all parameters (Millsap & Yun-Tein, 2004). By incorporating the constraints described for the partially invariant model, analyses established a new "baseline" model that allowed variation in the uniquenesses and compared this model's fit to that of an equivalent uniquenesses model. Thus, this set of analyses examined whether similar amounts of measurement error presented cross-culturally.

The variant uniquenesses model retained the partial invariance constraints in the loadings and thresholds but did not include cross-group uniqueness constraints. This model fit well: RMSEA = 0.034, CFI = 0.96, TLI = 0.98, McDonald's NCI = 0.96, Gamma Hat = 0.994, and $\Delta\chi^2$ = 2692.71 (208, n = 20,928, $p < 0.01$), analyses compared this model's fit to an invariant model specifying equivalent uniquenesses, and found no evidence for biased uniquenesses: RMSEA = 0.031, CFI = 0.96, TLI = 0.98, McDonald's NCI = 0.96, Gamma Hat = 0.994, 1838.82 (163, n = 20,928, $p < 0.01$), and $\Delta\chi^2$ = 22.66 (16, n = 20,928, p = 0.12). Given the final indices, analyses rejected the fully invariant measurement model, identified measurement bias, and specified a model with partially invariant thresholds. Table 1 describes the final estimates.

Statistically significant item differences do not necessarily lead to consequential scale differences (Byrne & Campbell, 1999; Cole & Maxwell, 1985). As such, a sensitivity analysis examined whether statistically significant item bias impacted mean estimates. Analyses compared the mean differences resulting from a fully invariant model ignoring observed measurement bias to those resulting from the model incorporating measurement bias. This comparison addresses the extent to which analyses that fail to acknowledge measurement bias suffer (Carle, in press). For fully invariant model, non-Hispanic Caucasians had a mean of zero (a function of statistical identification) and Hispanics had a mean of −0.11 ($M_{Caucasians}$ = 0.00, $SD_{Caucasians}$ = 1, $M_{Hispanics}$ = −0.10, $SD_{Hispanics}$ = 0.98, z = −2.47[1], $p < 0.01$, d = 0.10). For these items, higher scores reflect less use (1 = "yes", 2 = "no"). Under the partially invariant model, Hispanics' mean use *increased* ($M_{Caucasians}$= 0.00, $SD_{Caucasians}$ = 1, $M_{Hispanics}$= −0.14, $SD_{Hispanics}$ = 0.97, z = −3.14[1], $p < 0.01$, d = 0.14). Thus, failing to acknowledge statistically significant measurement bias: 1) meaningfully impacts mean estimates and comparisons, 2) underestimates differences between the groups, and 3) underestimates Hispanics' true use.

Discussion

Do assessments of alcohol dependence provide cross-culturally valid measurement among Hispanics in the US? This study sought an answer. It investigated whether statistically significant measurement bias existed across Hispanic and non-Hispanic Caucasians on a standardized, 27 item measure of DSM-IV alcohol dependence in a 2001–2002, large, nationally representative survey of US alcohol use and addressed whether statistically significant bias impacted the validity of this cohort's alcohol dependence estimates. Thus, it asked; are contemporary estimates of alcohol dependence valid across Hispanics and non-Hispanic Caucasians? It answered, no. CFA-OCM found statistically significant, impactful measurement bias for nine of twenty-seven items. These items addressed: drinking more to get desired the effects; drinking less to get the desired effects; drinking longer than intended; drinking more than intended; giving up pleasurable activities to drink; trouble sleeping when alcohol's effects ended; seeing, feeling, or hearing things when alcohol's effects wore off; multiple instances of wanting to stop drinking; and unsuccessful attempts to stop drinking. Differences in responses to these items underestimated alcohol dependence among Hispanics.

Biased thresholds revealed differential reporting tendencies at similar levels of alcohol dependence for nine items. CFA-OCM assumes a continuous latent variate underlies observed responses and a threshold determines responses. If an individual's level of the variate is less than the threshold, they answer yes. If not, they answer no. In this study, Hispanics found it *more* "difficult" to endorse several items. Compared to non-Hispanic Caucasians, they needed to experience more "*drinking less to get desired effect*" and more "*alcohol related sleeping troubles*" in order to endorse these items. Likewise, they needed to "*keep drinking longer than intended*" and have more "*periods of drinking more than intended*" before saying yes. Five

[1]The ratio of the mean divided to its standard error approximately follows the z distribution, and this ratio can be used to generate the probability the estimate differs from zero (Muthén & Muthén, 2007).

items showed a reverse pattern. At the same level of alcohol dependence, Hispanics more readily endorsed items that described "*giving up pleasurable activities*", "*seeing, feeling, or hearing things effects wore off*", "*drinking more to get effect*", "*multiple instances of wanting to quit*", and "*multiple unsuccessful attempts to stop*".

Collectively, these findings give compelling evidence that US Hispanics respond to criteria operationalizing alcohol dependence differently than non-Hispanic Caucasians. Coupled with similar findings among a 1992 cohort (Carle, in press) they express substantial doubt about the cross-cultural validity of alcohol dependence measurement. Given this skepticism, the study also investigated bias' impact. A sensitivity analysis compared the size and direction of mean differences across a model proceeding as if bias didn't exist and a model incorporating it. This comparison examined whether analyses conducted ignoring measurement would diverge from those incorporating bias. Importantly, this analysis suggested that current estimates of alcohol dependence among Hispanics *underestimate* the true change in alcohol dependence relative to non-Hispanic Caucasians. Acknowledging and incorporating measurement bias in the model *increased* reporting levels. Observed scores incorrectly estimate alcohol dependence and fail to provide cross-culturally valid measurement. These findings suggest greater levels of alcohol dependence among Hispanics than previously conveyed. Alcohol use has increased alarmingly among Hispanics (Grant, et al., 2004) and more readily than suspected.

This inquiry provides evidence that measurement bias impairs alcohol dependence measurement across Hispanics and non-Hispanic Caucasians. A number of causal mechanisms may result in this bias. Research notes cultural differences in social desirability and the extent to which Hispanics see psychiatric symptoms as undesirable (McHorney & Fleischman, 2006). Cultural differences in sincerity may influence Hispanic responses (Hui & Triandis, 1989). Language barriers and socioeconomic variability may also affect responses (McHorney & Fleischman, 2006). Each of these influences may lead to bias. And, regardless of its cause, the findings have implications for public health, clinical practice, and psychological theory.

The US should reflect upon public health efforts among the Hispanic community. Previous work shows alcohol use disparities for minorities (Grant, et al., 2004) and the US has devoted resources specifically to address these unique health concerns (Department of Health and Human Services, 2007). Nevertheless, results demonstrate that alcohol dependence has increased more disproportionately among this community, deserves more strenuous elimination efforts, and highlights behaviors especially worth targeting (those Hispanics found more difficult to endorse). Clinicians should also pay particular attention when diagnosing Hispanic individuals based on endorsements of these nine items. Finally, the findings also call into question the cross-cultural validity of alcohol dependence generally. The discovery of bias here generally mirrors Carle's (in press) findings among a 1992 cohort. In that study, seven items resulted in biased alcohol dependence measurement across Hispanics and non-Hispanic Caucasians. Those items reflected similar concepts covered here: "*drinking more to get the desired effect*", "*drinking less to get the desired effect*", "*drinking longer than intended*", "*periods of drinking more than intended*", "*multiple instances of wanting to quit*", and "*multiple unsuccessful attempts to stop drinking*", and the bias occurred in the same direction in both studies. Thus, these differences appear stable and likely reflect underlying, reliable differences in the experience and expression of alcohol dependence among different ethnicities. Psychological science should seek bias' source and carefully examine the appropriateness of diagnosing and describing cultural minorities using these criteria. Likewise, psychological science should consider how much the current definition of alcohol dependence reflects cultural myopia and whether we need alternative criteria.

Before concluding, the study's fortes and boundaries deserve review. First, the study did not describe analyses examining sex differences cross-culturally. Dividing the sample across

culture and sex resulted in sparse data with a large number of bivariate empty item comparisons, this limits the reliability and validity of findings based on these groupings (Muthén & Muthén, 2007). Thus, while analyses addressing culture and sex simultaneously mirrored the cultural differences described above, demonstrated no sex by culture interactions, and found a single main effect for sex (the loading associated with "seizure" item differed across sexes but not culture), the reader should consider these findings probative. Second, consistent with other work (Grant, et al., 2004), the study treated Hispanics homogenously despite within group heterogeneity (Dawson, 1998). Unfortunately, this may miss notable measurement heterogeneity. Nevertheless, many in the field argue that single group classifications remain meaningful (Burchard, et al., 2003) and the study followed this advice. Finally, the study used a representative sample; it remains unclear whether results persist clinically. These weaknesses leave some issues unaddressed. Future studies should collect larger samples and examine the replicablility of the finding that culture and sex did *not* interact. Research should also examine the validity of alcohol dependence within the Hispanic community; for example, do assessments validly measure alcohol dependence across Hispanics from disparate US regions or different countries of origin? Finally, clinical samples can differ from community samples (Kazdin, 2003) and research should examine whether findings hold clinically.

Despite limits, the study has numerous assets. It makes a distinctive contribution. A review found no studies examining the cross-cultural measurement equivalence of alcohol dependence in recent US data. It fills this gap using a well designed, large, nationally representative sample, which alleviates sampling bias and methodological concerns. Moreover, by addressing measurement bias across Hispanics and Caucasians generally, it allows readers to evaluate the validity of the large body of previous work that used these same groupings and it maintains consistency with recent arguments to continue general cross-cultural comparisons (Carle, 2008; Fiscella, et al., 2000; Flores, et al., 2002). Finally, the study uses modern techniques that allow rigorous investigation (Teresi, et al., 2006). These strengths allow it to unequivocally call upon science to reconsider its cultural nescience.

In conclusion, results found statistically significant, impactful measurement bias across Hispanics and non-Hispanic Caucasians in a representative sample of the 2001–2002 US for nine items operationalizing DSM-IV alcohol dependence. Results suggest caution when estimating rates and levels of alcohol dependence across these groups. Current findings *underestimate* the rate of alcohol dependence among Hispanics and alcohol dependence appears to be increasing at a rate greater than formerly understood. Results underscore the need for culturally sensitive research, prevention, and intervention efforts and support empirically questioning the generalization of psychological findings from majorities to minorities. Summarily, do DSM-IV based measures of alcohol dependence provide equivalent measurement across Hispanics and non-Hispanics? Lamentably, no.

ACKNOWLEDGMENTS

I would like to thank the National Institute on Alcohol Abuse and Alcoholism and Dr. Bridget Grant for generously making this data publicly available. I would also like to thank Tara J. Carle whose unending support and thoughtful comments made my work possible. A University of North Florida 2007 Summer Research Grant partially supported this manuscript's preparation. UNF played no role: in study design; in the collection, analysis and interpretation of data; in the writing of the report; or in the decision to submit the paper for publication. I have no conflicts of interest.

REFERENCES

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders Fourth Edition. Washington, DC: American Psychiatric Association; 1994.

Bachman JG, O'Malley PM. Yes-Saying Nay-Saying, and Going to Extremes: Black-White Differences in Response Styles. The Public Opinion Quarterly 1984;48:491–509.

Bloche MG. Health care disparities—science, politics, and race. New England Journal of Medicine 2004;350:1568–1570. [PubMed: 15071133]

Burchard EG, Ziv E, Coyle N, Gomez SL, Tang H, Karter AJ, Mountain JL, Perez-Stable EJ, Sheppard D, Risch N. The importance of race and ethnic background in biomedical research and clinical practice. New England Journal of Medicine 2003;348:1170–1176. [PubMed: 12646676]

Byrne, B. Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic applications and programs. New Jersey: Lawrence Erlbaum; 1998.

Byrne BM, Baron P, Baley J. The Beck Depression Inventory A cross-validated test of second-order factorial structure for Bulgarian adolescents. Educational and Psychological Measurement 1998;58:241–251.

Byrne BM, Campbell TL. Cross-cultural comparisons and the presumption of equivalent measurement and theoretical structure: A look beneath the surface. Journal of Cross Cultural Psychology 1999;30:555–574.

Carle AC, Millsap RE, Cole DA. Measurement bias across gender on the Children's Depression Inventory: Evidence for Invariance from Two Latent Variable Models. Educational and Psychological Measurement. in press

Carle AC. Cross-Cultural Validity of Alcohol Dependence across Hispanics and Non-Hispanic Caucasians. Hispanic Journal of Behavioral Sciences 2008;30:106–120.

Chatterji S, Saunders JB, Vrasti. 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 1997;47:171–185. [PubMed: 9306043]

Cheung GW, Rensvold RB. Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. Structural Equation Modeling 2002;9:233–255.

Cole DA, Maxwell SE. Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach. Multivariate Behavioral 1985;20:389–417.

Dawson DA. Beyond black, white and Hispanic: race, ethnic origin and drinking patterns in the United States. Journal of Substance Abuse 1998;10:321–339. [PubMed: 10897287]

Dawson DA, Grant BF, Chou PS, Pickering RP. Subgroup variation in U.S. drinking patterns: Results of the 1992 National Longitudinal Alcohol Epidemiologic Study. Journal of Substance Abuse 1995;7:331–344. [PubMed: 8749792]

Department of Health and Human Services. Healthy People 2010. 2007 Dec 21. Retrieved, February 21, 2007, from http://www.health.gov/healthypeople/document/html/uih/uih_2.htm.

Fiscella K, Franks P, Gold M, Clancy C. Inequality in quality. Addressing socioeconomic, racial, and ethnic disparities in health care. Journal of the American Medical Association 2000;283:2579–2584. [PubMed: 10815125]

Flores G, Fuentes-Afflick E, Barbot O, Carter-Pokras O, Claudio L, Lara M, McLaurin J, Pachter L, Gomez FR, Mendoza F, Valdez RB, Villaruel AM, Zambrana RE, Greenberg R, Weitzman M. The health of Latino children: urgent priorities, unanswered questions, and a research agenda. Journal of the American Medical Association 2002;288:82–90. [PubMed: 12090866]

Grant BF. Convergent Validity of DSM-III-R and DSM-IV Alcohol Dependence: Results from the National Longitudinal Alcohol Epidemiologic Survey. Journal of Substance Abuse 1997;9:89–102. [PubMed: 9494941]

Grant BF. Theoretical and observed subtypes of DSM-IV alcohol abuse and dependence in a general population sample. Grant, B. Drug and Alcohol Dependence 2000 Nov 2000;60(3):287–293. [PubMed: 11053763]

Grant, BF.; Dawson, DA.; Hasin, DS. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version (AUDADIS-IV). Bethesda, MD: National Institute of on Alcohol Abuse and Alcoholism; 2001.

Grant BF, Dawson DA, Stinson FS. 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 2003;71:7–16. [PubMed: 12821201]

Grant BF, Dawson DA, Stinson FS, Chou SP, Dufour MC, Pickering RP. The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. Drug and Alcohol Dependence 2004;74:223–234. [PubMed: 15194200]

Grant BF, Harford TC, Dawson DA, Chou P, Dufour M, Pickering R. Prevalence of DSM-IV Alcohol Abuse and Dependence. Alcohol Health and Research World 1994;18:243–248.

Grant BF, Harford TC, Dawson D. 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 1995;39:37–44. [PubMed: 7587973]

Grant BF, Harford TC, Muthén BO, Yi H, Hasin DS, Stinson FS. DSM-IV alcohol dependence and abuse: Further evidence of validity in the general population. Drug and Alcohol Dependence 2007;86:154–166. [PubMed: 16814489]

Grant, BF.; Kaplan, K.; Shepard, J.; Moore, T. Source and Accuracy Statement for Wave 1 of the 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions. Bethesda MD: National Institute on Alcohol Abuse and Alcoholism; 2003.

Green SB, Babyak MA. Control of Type I errors with multiple tests of constraints in structural equation modeling. Multivariate Behavioral Research 1997;32:39–51.

Greenfield, LA. Alcohol and Crime: an Analysis of National Data on the Prevalence of Alcohol in Crime. Washington, DC: U.S. Department of Justice; 1998.

Harford TC, Muthén BO. Alcohol use among college students: The effects of prior problem behaviors and change of residence. Journal of Studies on Alcohol 2001;62:306–312. [PubMed: 11414340]

Harwood, H. Updating Estimates of the Economic Costs of Alcohol Abuse in the United States: Estimates, Update Methods and Data. Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1998.

Hasin D, Carpenter KM, McCloud S. The alcohol use disorder and associated disabilities interview schedule (AUDADIS): Reliability of alcohol and drug modules in a clinical sample. Drug and Alcohol Dependence 1997a;44:133–141. [PubMed: 9088785]

Hasin DS, Grant B, Cottler L. Nosological comparisons of alcohol and drug diagnoses: A multisite, multi-instrument international study. Drug and Alcohol Dependence 1997b;47:217–226. [PubMed: 9306047]

Hasin DS, Grant BF. 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 2004;61:891–896. [PubMed: 15351767]

Hasin DS, Muthén B, Wisnicki KS, Grant BF. Validity of the bi-axial dependence concept: A test in the US general population. Addiction 1994;89:573–579. [PubMed: 8044123]

Hasin D, Paykin A. Alcohol dependence and abuse diagnoses: Concurrent validity in a nationally representative sample. Alcoholism: Clinical and Experimental Research 1999;23:144–150.

Hasin DS, Stinson FS, Ogburn E, Grant BF. Prevalence, correlates, disability, and comorbidity of DSM-IV alcohol abuse and dependence in the United States: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Archives of General Psychiatry 2007;64:830–842. [PubMed: 17606817]

Hu L, Bentler P. Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. Psychological Methods 1998;3:424–453.

Huang CD, Church AT, Katigbak MS. Identifying cultural differences in items and traits: Differential item functioning in the NEO Personality Inventory. Journal of Cross Cultural Psychology 1997;28:192–218.

Hui CH, Triandis HC. Effects of Culture and Response Format on Extreme Response Style. Journal of Cross-Cultural Psychology 1989;20:296–309.

Kazdin, AE. Research Design in Clinical Psychology. Vol. Fourth Edition. Boston, MA: Allyn and Bacon; 2003.

Kessler RC, McGonagle KA, Zhao S, Nelson CB, Hughes M, Eshleman S, Wittchen HU, Kendler KS. 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 1994;51:8–19. [PubMed: 8279933]

Knight, GP.; Hill, NE. Measurement equivalence in research involving adolescents. In: McLoyd, VC.; Steinberg, L., editors. Studying minority adolescents: Conceptual, methodological, and health issues. New Jersey: Laurence Erlbaum; 1998.

McHorney C, Fleishman J. Assessing and understanding measurement equivalence in health outcome measures: Issues for further quantitative and qualitative inquiry. Medical Care 2006;44:S205–S210. [PubMed: 17060829]

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

Meredith W, Teresi J. An essay on measurement and factorial invariance. Medical Care 2006;44:S69–S77. [PubMed: 17060838]

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

Muthén BO, Christoffersson A. Simultaneous factor analysis of dichotomous variables in several groups. Psychometrika 1981;46:407–419.

Muthén BO. Factor analysis of alcohol abuse and dependence symptom items in the 1988 National Health Interview Survey. Addiction 1995;90:637–645. [PubMed: 7795500]

Muthén BO, Hasin D, Wisnicki KS. Factor analysis of ICD-10 symptom items in the 1988 National Health Interview Survey on Alcohol Dependence. Addiction 1993;88:1071–1077. [PubMed: 8401161]

Muthén, LK.; Muthén, BO. Mplus User's Guide. Los Angeles: Muthén & Muthén; 2007.

Prelow HM, Michaels ML, Reyes L, Knight GP, Barrera M Jr. Measuring Coping in Low-Income European American, African American, and Mexican American Adolescents: An Examination of Measurement Equivalence. Anxiety, Stress and Coping 2002;15:135–147.

Prelow H, Tien JY, Roosa, Mark W, Wood J. Do coping styles differ across sociocultural groups? The role of measurement equivalence in making this judgment. American Journal of Community Psychology 2000;28:225–244. [PubMed: 10836092]

Ramírez M, Ford ME, Stewart AL, Teresi JA. Measurement issues in health disparities research. Health Services Research 2005;40:1640. [PubMed: 16179000]

Ruan JW, Goldstein RB, Chou SP, Smith SM, Saha TD, Pickering RP, Dawson DA, Huang B, Stinson FS, Grant BF. The Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-IV): Reliability of new psychiatric diagnostic modules and risk factors in a general population sample. Drug and Alcohol Dependence 2008;92:27–36. [PubMed: 17706375]

Smedley, BD.; Stith, AY.; Neslon, AR. Unequal treatment: Confronting racial and ethnic disparities in health care. Washington, DC: National Academies Press; 2003.

Smith PB. Acquiescent Response Bias as an Aspect Of Cultural Communication Style. Journal Of Cross-Cultural Psychology 2004;Vol. 35:50–61.

Stahl S, Hahn A. The National Institute on Aging's Resource Centers for Minority Aging Research: Contributions to measurement in research on ethnically and racially diverse populations. Medical Care 2006;44:S1–S2. [PubMed: 17060815]

Steiger JH. A note on multiple sample extensions of the RMSEA fit index. Structural Equation Modeling 1998;5:411–419.

Steinbrook R. Disparities in health care—from politics to policy. New England Journal of Medicine 2004;350:1486–1488. [PubMed: 15071122]

Stewart AL, Nápoles-Springer AM. Advancing health disparities research: Can we afford to ignore measurement issues? Medical Care 2003;41:1207–1220. [PubMed: 14583684]

Sue S. Science, Ethnicity, and Bias: Where Have We Gone Wrong? American Psychologist 1999;54:1070–1077. [PubMed: 15332528]

Teresi JA, Stewart AL, Morales MS, Stahl SM. Measurement in a Multi-Ethnic Society. Medical Care 2006;44:S3–S4. [PubMed: 17060831]

Thissen, D.; Steinberg, L.; Wainer, H. Detection of differential item functioning using the parameters of item response models. In: Holland, PW.; Wainer, H., editors. Differential item functioning. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993. p. 67-113.

Waller NG, Thompson JS, Wenk E. Using IRT to separate measurement bias from true group differences on homogeneous and heterogeneous scales: An illustration with the MMPI. Psychological Methods 2000;5:125–146. [PubMed: 10937326]

Wright GN, Phillips LD, Whalley PC, Gerry TC, Kee-Ong N, Tan I, Wisudha A. Cultural Differences in Probabilistic Thinking. Journal of Cross-Cultural Psychology 1978;9:285–299.

Table 1

Partially Invariant Alcohol Dependence Measurement Model across Hispanics and Non-Hispanic Caucasians (Uniquenesses Constrained to 1 for both groups)

	Loadings		Thresholds	
Abbreviated Item Labels	**Caucasians**	**Hispanics**	**Caucasians**	**Hispanics**
Find Usual Number of Drinks Had Less Effect Than Before	1.27	1.27	−2.73	−2.73
Drink More to Get Wanted Effect	1.93	1.93	−4.25	−4.02
Drink Equivalent of Fifth of Liquor in 1 Day	1.12	1.12	−2.96	−2.96
Increase Because Amount No Longer Gave Desired Effect	1.95	1.95	−4.62	−4.62
More Than Once Want to Stop or Cut Down on Drinking	1.12	1.12	−2.12	−1.97
More Than One Unsuccessful Stop/Cut Down Attempt	1.35	1.35	−3.74	−3.31
Have Period When Ended Up Drinking More Than Intended	2.09	2.09	−3.06	−3.47
Kept Drinking Longer than Intended	2.05	2.05	−3.25	−3.62
Trouble Falling/Staying Asleep When Effects Wearing Off	1.00	1.00	−2.38	−2.75
Shake When Effects Wearing Off	1.39	1.39	−3.73	−3.73
Anxious or Nervous When Effects Wearing Off	1.49	1.49	−3.78	−3.78
Nausea When Effects Wearing Off	1.36	1.36	−2.35	−2.35
Unusually Restless When Effects Wearing Off	1.36	1.36	−3.16	−3.16
Sweat or Heart Beat Fast When Effects Wearing Off	1.23	1.23	−3.05	−3.05
See, Feel, Hear Things When Effects Wearing Off	1.12	1.12	−2.12	−1.97
Fits/Seizures When Effects Wearing Off	0.94	0.94	−4.28	−4.28
Headaches When Effects Wearing Off	1.21	1.21	−1.95	−1.95
Drink/Use Medicine to Get Over Aftereffects	1.09	1.09	−2.99	−2.99
Drink/Use Medicine to Avoid Aftereffects	1.14	1.14	−3.37	−3.37
Period When Spent Lot of Time Drinking	1.60	1.60	−3.84	−3.84
Spent Lot of Time Sick/Getting Over Aftereffects	1.43	1.43	−4.13	−4.13
Give Up/Cut Down Important Activities to Drink	2.49	2.49	−6.64	−6.64
Give Up/Cut Down Pleasurable Activities to Drink	2.78	2.78	−7.68	−7.20
Continue Even Though Made Depressed, Etc.	1.75	1.75	−4.47	−4.47
Continue Even Though Causing Health Problem	1.31	1.31	−3.37	−3.37
Continue Despite Prior Blackout	1.44	1.44	−3.62	−3.62
Could Drink Less Than Before To Get Effect Wanted	0.56	0.56	−1.42	−1.58

	Loadings		Thresholds	
Abbreviated Item Labels	**Caucasians**	**Hispanics**	**Caucasians**	**Hispanics**
Dependence Factor Mean	0.00	−0.14		
Dependence Factor Variance	1.00	0.95		