

Research article

Open Access

## Assessing the adequacy of self-reported alcohol abuse measurement across time and ethnicity: cross-cultural equivalence across Hispanics and Caucasians in 1992, non-equivalence in 2001–2002

Adam C Carle

Address: Department of Psychology, University of North Florida, 1 UNF Drive, Jacksonville, FL 32224, USA

Email: Adam C Carle - adam.carle@unf.edu

Published: 19 February 2009

Received: 15 September 2008

*BMC Public Health* 2009, **9**:60 doi:10.1186/1471-2458-9-60

Accepted: 19 February 2009

This article is available from: <http://www.biomedcentral.com/1471-2458/9/60>

© 2009 Carle; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Do estimates of alcohol abuse reflect true levels across United States Hispanics and non-Hispanic Caucasians, or does culturally-based, systematic measurement error (i.e., measurement bias) affect estimates? Likewise, given that recent estimates suggest alcohol abuse has increased among US Hispanics, the field should also ask, "Does cross-ethnic change in alcohol abuse across time reflect true change or does measurement bias influence change estimates?"

**Methods:** To address these questions, I used confirmatory factor analyses for ordered-categorical measures to probe for measurement bias on two large, standardized, nationally representative, US surveys of alcohol abuse conducted in 1992 and 2001–2002. In 2001–2002, analyses investigated whether 10 items operationalizing DSM-IV alcohol abuse provided equivalent measurement across Hispanic ( $n = 4,893$ ) and non-Hispanic Caucasians ( $n = 16,480$ ). In 1992, analyses examined whether a reduced 6 item item-set provided equivalent measurement among 834 Hispanic and 14,8335 non-Hispanic Caucasians.

**Results:** In 1992, findings demonstrated statistically significant measurement bias for two items. However, sensitivity analyses showed that item-level bias did not appreciably bias item-set based alcohol abuse estimates among this cohort. For 2001–2002, results demonstrated statistically significant bias for seven items, suggesting caution regarding the cross-ethnic equivalence of alcohol abuse estimates among the current US Hispanic population. Sensitivity analyses indicated that item-level differences *did* erroneously impact alcohol abuse rates in 2001–2002, underestimating rates among Hispanics relative to Caucasians.

**Conclusion:** 1992's item-level findings suggest that estimates of drinking related social or legal problems may underestimate these specific problems among Hispanics. However, impact analyses indicated no appreciable effect on alcohol abuse estimates resulting from the item-set. Efforts to monitor change in alcohol abuse diagnoses among the Hispanic community can use 1992 estimates as a valid baseline. In 2001–2002, item-level measurement bias on seven items did affect item-set based estimates. Bias underestimated Hispanics' self-reported alcohol abuse levels relative to non-Hispanic Caucasians. Given the cross-ethnic equivalence of 1992 estimates, bias in 2001–2002 speciously minimizes current increases in drinking behavior evidenced among Hispanics. Findings call for increased public health efforts among the Hispanic community and underscore the necessity for cultural sensitivity when generalizing measures developed in the majority to minorities.

## I. Background

Psychological science and its allied disciplines too often stand culturally blind, rarely questioning if concepts and measurements valid in the majority culture demonstrate similar validity among minority communities. Despite the fact that alcohol dependence often leads to greater impairment than alcohol abuse and that some recent arguments call for a single diagnostic category encompassing both, [1] alcohol abuse remains a separate diagnostic category in DSM-IV and embodies a substantial public health problem. For example, recent estimates suggest a DSM-IV alcohol abuse prevalence rate of approximately 4.7% in the US. [2] However, studies note significant differences when comparing the prevalence and comorbidity of alcohol abuse and dependence across Caucasians and cultural minorities. [3,4] Consistent with earlier work, [5,6] these studies show a significantly lower rate of comorbid alcohol disorders among Hispanics as compared to Caucasians. [4] Work also establishes a changing trend of drinking behavior across Caucasians and Hispanics, remaining relatively stable among Caucasians and increasing for Hispanics. [4] These investigations demonstrate etiological and epidemiological differences in alcohol abuse across Caucasians and Hispanics and highlight the need for culturally sensitive public health policy and prevention and intervention efforts, particularly given the presence of health disparities [7-12] and the colossal cost of alcohol abuse to individuals, families, and society. [13,14] Despite this work, research has not adequately explored culture's possible influence on alcohol abuse.

The previous comparisons frequently rest on the *untested* assumption that concepts and measurements reliably and validly estimated among the majority Caucasian culture achieve similar reliability and validity among the Hispanic community. Measurement bias, also labeled differential item functioning (DIF), refers to the possibility that individuals equal in their true levels of alcohol abuse, but from different groups, i.e., Caucasians and Hispanics, do not have identical probabilities of responding to questions concerning their alcohol use. [15] Although studies have established the validity and reliability of standardized alcohol abuse measures and diagnostic criteria in the general population, and provided support for single factor models, [16-26] the role of minority/majority based measurement bias in the instruments used to assess alcohol abuse in the U.S. population goes relatively unexamined.

Modern measurement models, such as confirmatory factor analysis (CFA), offer powerful tools to examine bias. [27,28] Generally, these methods use equations to model item response probabilities and compare the equality of the parameters associated with these models across groups to investigate bias. While investigations of this

type have not examined measurement bias and alcohol abuse in recent data, they have examined cultural differences across a number diagnostic measures generally, e.g., dementia, depression, etc. [27] They have shown that bias can attenuate or accentuate group differences, [28-32] lead to inaccurate diagnoses, [28,33-35] and generally decrease reliability and validity. [36-41] Studies have also uncovered bias so profound as to render cross group comparisons virtually impossible. [42-44] Thus, before validly comparing minority and majority groups, we must ask whether the measurements upon which we base comparisons function similarly across groups. [27,45] The field must consider the extent to which observed differences and change reflect true differences or result from a lack of equivalence in the measures used to assess alcohol abuse across populations.

Theoretical and empirical reasons suggest we should suspect measurement bias across Hispanics and Caucasians. [46] Authors have noted differences in the relation between probabilistic thinking and assignment of numbers, [47] differences in acquiescent responses, [48] and differences in language use [49] across Hispanics, Caucasians, and other minorities. A number of authors have also noted that behavioral exemplars describing a psychological construct for the majority may not be appropriate for a minority group, nor do they necessarily include a set of culturally appropriate indicators for minorities. [30,43,44] Hui and Triandis have described sincerity as a cultural value among Hispanics that may lead to measurement bias, positing 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. [44] suggest that for certain behaviors greater levels of a specific problem may be needed before Hispanics willingly acknowledge a problem. McHorney and Fleishman [50] note that survey questions may trigger differential cultural perceptions regarding socially desirable responses and that question wording may impede symptomatology reporting by Hispanics. In sum, we have strong reason to express concern that measurement bias affects the equivalence of measurement across Hispanics and Caucasians generally, and have no reason to exclude alcohol abuse from these suspicions. Indeed, in a recent reexamination of alcohol dependence among a 1992 cohort, Carle [51] found statistically significant measurement bias across Hispanics and non-Hispanic Caucasians. This bolsters concerns about current estimates, particularly as they relate to earlier assessments.

Woefully, a literature review found no published studies examining the validity of alcohol abuse measures across Caucasians and Hispanics in recent or early data. It remains ambiguous whether measurement bias affects

epidemiological estimates and research across these groups. This leaves unclear whether recent documentation suggesting differential prevalence and comorbidity of alcohol abuse and discrepant drinking pattern changes reflect true self-reported statuses, measurement bias, or both. As a result, in the current study, I had several goals. I used modern measurement models to examine whether measurement bias exists across Hispanic and non-Hispanic Caucasians on two standardized measures of alcohol abuse in a large, nationally representative survey of alcohol use in the United States conducted in 1992 and in 2001–2002, and, if so, to what extent does it impact estimates of alcohol abuse across non-Hispanic Caucasians and Hispanics at these points in time. I used these results to assess whether descriptions noting recent changes in alcohol abuse ought to receive modification. Should we *increase* or *decrease* current estimates as a function of biased measurement?

## 2. Methods

### 2.1 Participants

#### 2.2.1 1992

Participants ( $n = 14,835$ ; 14,001 non-Hispanic Caucasians and 834 Hispanics) were a subset of the larger 1992 National Longitudinal Alcohol Epidemiologic Study (NLAES), [52] designed and sponsored by the National Institute for Alcohol Abuse and Alcoholism (NIAAA), 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 between the ages of 18 and 29, and had household and sample person response rates of 92% and 97% respectively. Sample weights adjust the data to make it representative of the civilian non-institutionalized population of the US [52] non-Hispanic Caucasians and Hispanics with complete data were included in the current study. In the original design, 14,835 individuals who reported consumption of alcohol in the past 12 months were asked the questions studied here.

#### 2.2.2 2001–2002

Participants ( $n = 21,373$ ; 16,480 non-Hispanic Caucasians and 4,893 Hispanics) were a subset of the larger, publicly available 2001–2002 National Epidemiologic Survey on Alcohol and Related Conditions (NESARC [53]) data designed and sponsored by the National Institute for Alcohol Abuse and Alcoholism (NIAAA) and fielded by the US Census Bureau. The original sample consisted of 43,093 US adults aged 18 years and older representing the non-institutionalized adult US population. The complex, multistage design incorporated the Census 2000/2001 Supplementary Survey (C2SS) and Census 2000 Group Quarters Inventory sampling frame and oversampled Afri-

can American, Hispanics, and young adults (18 – 24). Sample weights, described in detail elsewhere [53], adjust the data to make it representative of the civilian non-institutionalized population of the US. The NESARC had household and sample person response rates of 89% and 93% respectively [53]. The current study included participants with complete data who reported consumption of alcohol in the past 12 months.

### 2.2 Procedures

For both surveys, experienced Census Bureau interviewers completed direct 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, and sociodemographic background characteristics.

### 2.3 Measures

#### 2.3.1 Alcohol Abuse

The DSM-IV [54] identifies alcohol abuse as a maladaptive pattern of alcohol use that occurs in the absence of alcohol dependence and leads to significant impairment or distress, and that demonstrates at least one of the following four criteria: 1) continued use despite a social or interpersonal problem caused or exacerbated by the effects of drinking; 2) recurrent drinking in situations in which alcohol use is physically hazardous; 3) recurrent drinking resulting in a failure to fulfill major role obligations; or 4) recurrent alcohol related legal problems.

##### 2.3.1.1 1992

The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS [55]) used in the NLAES uses a total of 6 dichotomous items to operationalize alcohol abuse criteria. The AUDADIS provides 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. It generates diagnoses consistent with the several diagnostic classification systems including the Fourth Edition of the DSM (DSM-IV). I used all 6 items operationalizing alcohol abuse criteria. Reliabilities established through independent test-retest meet acceptable standards. [56] Additional studies have also established several types of validity, e.g., construct, criterion, etc. [17,18,22,23,25,26,56-59]

##### 2.3.1.1: 2001–2002

The Alcohol Use Disorder and Associated Disabilities Interview Schedule DSM-IV Version (AUDADIS-IV [19]) used in the NESARC uses a total of 10 dichotomous items to operationalize alcohol abuse criteria. The AUDADIS-IV is an updated version of the AUDADIS used in the NLAES.

Like its predecessor, it generates diagnoses consistent with DSM-IV and has demonstrated acceptable psychometric standards. [17,18,22,23,25,26,56-59] I used all 10 items operationalizing alcohol abuse criteria.

2.3.2 Ethnicity

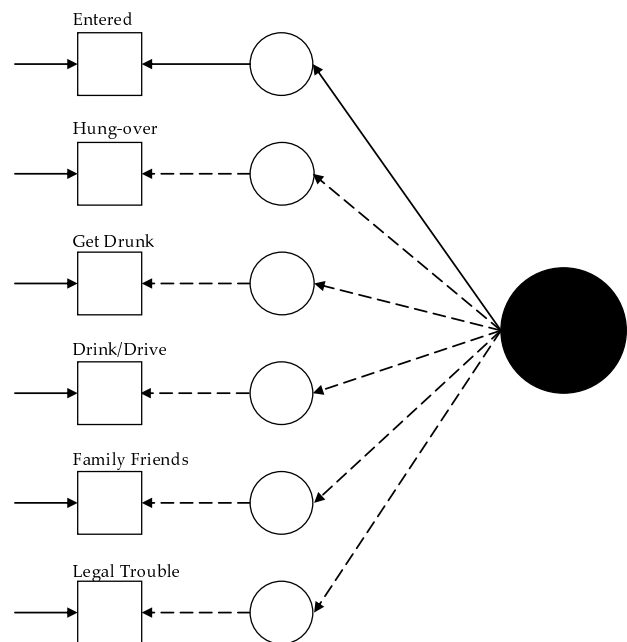
Both the NLAES (1992) and NESARC (2001–2002) coded race using five options: American Indian and Alaska Native; Asian; Black or African American; Native Hawaiian and Other Pacific Islander; and White. A single item allowed Hispanic self-identification. The current study considered individuals non-Hispanic Caucasians if they identified themselves as both White and non-Hispanic and regarded anyone who self-identified as Hispanic a Hispanic.

2.4 Analytic Strategy

The current study used confirmatory factor analyses for ordered-categorical measures (CFA-OCM) to probe for bias. CFA-OCM appropriately models the categorical nature of the items[60] and falls within a larger family of latent variable measurement modeling approaches that includes: CFA for continuous measures, multiple indicator multiple cause (MIMIC) models, and item response theory models. Given its strengths relative to the weaknesses of other approaches, the present study adopted CFA-OCM. For example, MIMIC models: require invariant loadings across the groups, a challenging assumption when working with understudied measures like the one here;[61] require invariant factor variances across the groups; lack formal hierarchical invariance tests; and may miss non-uniform bias. [61,62] Additionally, because a MIMIC model would control for the effects of ethnicity, this approach would not allow analyses to examine whether group differences remain after modeling measurement bias, a specific goal of this study. CFA-OCM does not suffer from these issues. With regard to IRT, Takane and de Leeuw[63] demonstrated the functional equivalence of CFA-OCM and 2 parameter IRT models making the choice between them relatively superficial, although estimation procedures can differ in practice. [64] However, IRT modeling procedures include fewer indices relative to CFA,[61,65] and, as a result, CFA-OCM can provide more informed model fit examinations and one can mathematically derive IRT parameter estimates post hoc. For these reasons, the study used CFA-OCM. Unfortunately, few social scientists receive training in these models. As a result, I review them briefly. However, the interested reader should consult Byrne[66], Millsap & Yun-Tien[67], Muthén[68], or Muthén & Christoffersson[60] for detailed reviews.

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. 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. [69]

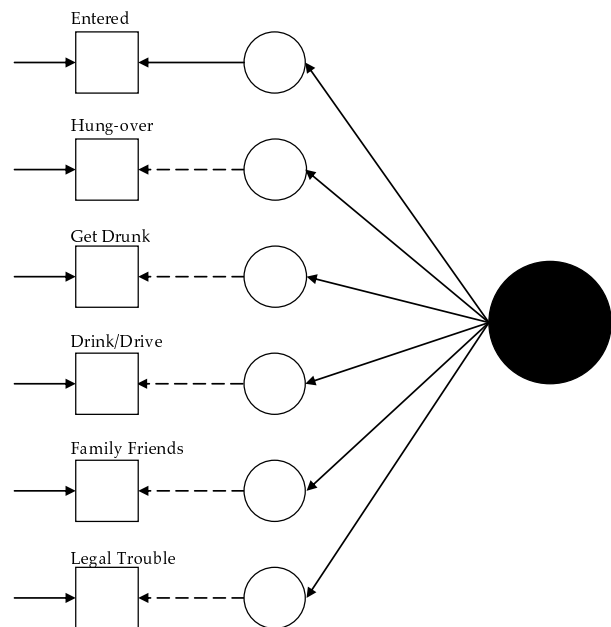


**Figure 1**  
**Path diagram summarizing Model 1, the baseline 1992 alcohol abuse measurement model.** The solid black circle represents the alcohol abuse variable. The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics. Dashed lines represent values freely estimated across non-Hispanic Caucasians and Hispanics.

Figure 1 presents a visual representation of this measurement model. The solid black circle represents the latent variable, here alcohol abuse. The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. Finally, the arrows pointing only to the squares represent the uniquenesses.

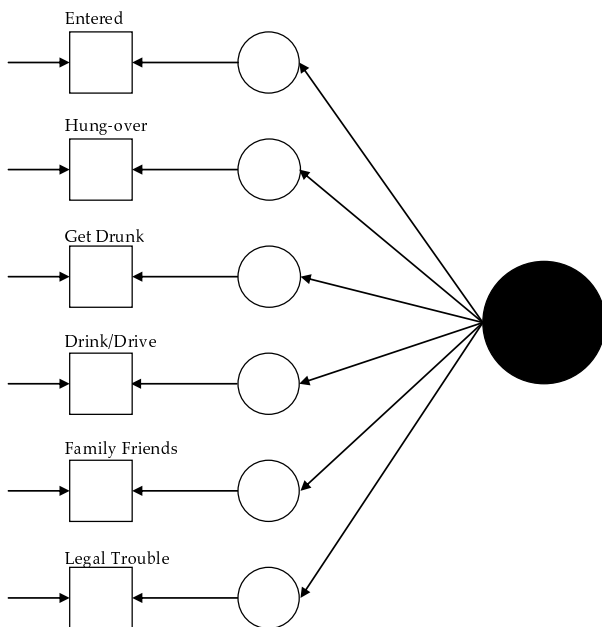
In measurement bias studies, researchers examine the equivalence of the measurement parameters across groups. In practice, a series of hierarchically nested models typically test measurement bias. [66,67,69] The method starts with the least restricted measurement model across groups and adds cross-group 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 achieves 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.

Visually, for the least restricted model in the 1992 data, this would mean fitting a model like that presented in Figure 1 for each group and allowing the measurement parameters to vary across the groups (excepting those constrained to equality for statistical identification, see below). Thus, in Figure 1, dashed lines represent measurement parameters allowed to vary across groups and the solid lines represent measurement parameters constrained to equality across the groups. As Figure 2 shows, the model constraining the loadings across the groups has solid black lines from the latent variable (solid circle) to the latent response variates, indicating that the loadings have been constrained to equality. Figures 3 and 4 continue the visual representation for the 1992 data. Figures 5, 6, 7, and 8 illustrate the models for the 2001–2002 data.

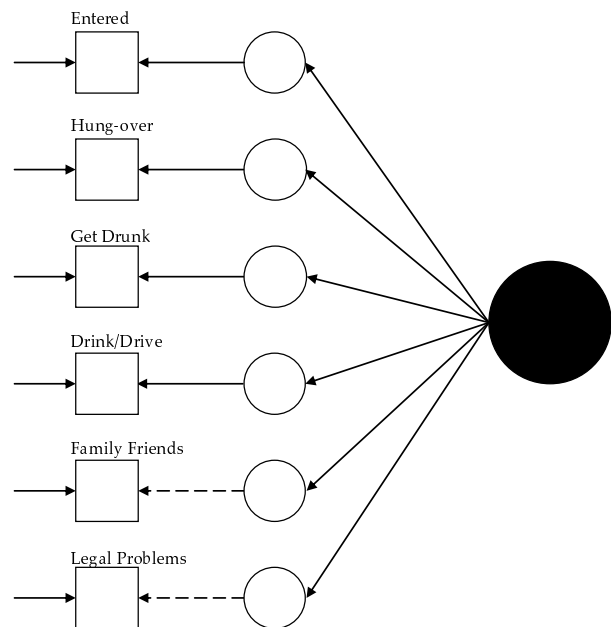


**Figure 2**  
**Path diagram summarizing Model 2, the 1992 alcohol abuse measurement model constraining the loadings to equality across groups.** The solid black circle represents the alcohol abuse variable. The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics. Dashed lines represent values freely estimated across non-Hispanic Caucasians and Hispanics.

The current investigation adopted this approach and conducted all analyses using Mplus, a program capable of appropriately handling complex survey data, its theta parameterization and robust weighted least squares (WLSMV) estimator (Mplus 4.2)[64]. Analyses examined measurement invariance following the hierarchical method described above and in detail by Millsap and Yun-Tien[67]. A priori, the studied adopted preferred fit index levels suggested by Hu and Bentler[70,71], Muthén and Muthén[64], Steiger[72], and Cheung and Rensvold:[73] root mean square error of approximation (RMSEA) values less than 0.05; comparative fit index (CFI), Tucker-Lewis Index (TLI), and Gamma Hat values greater than 0.95; and McDonald's noncentrality index (NCI) values greater than 0.90. Fit evaluation focused on the index set. Models included means and covariances at each step and statistical identification conformed to Millsap and Yun-Tien's[67] description. Consistent with argu-



**Figure 3**  
**Path diagram summarizing Model 3a, the 1992 alcohol abuse measurement model constraining the thresholds to equality across groups.** The solid black circle represents the alcohol abuse variable. The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics.



**Figure 4**  
**Path diagram summarizing Model 4, the final 1992 alcohol abuse measurement model.** The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics. Dashed lines represent values freely estimated across non-Hispanic Caucasians and Hispanics.

ments for more stringent error control in modeling.[74,75] an  $\alpha$  of 0.01 was adopted a priori.

### 3. Results

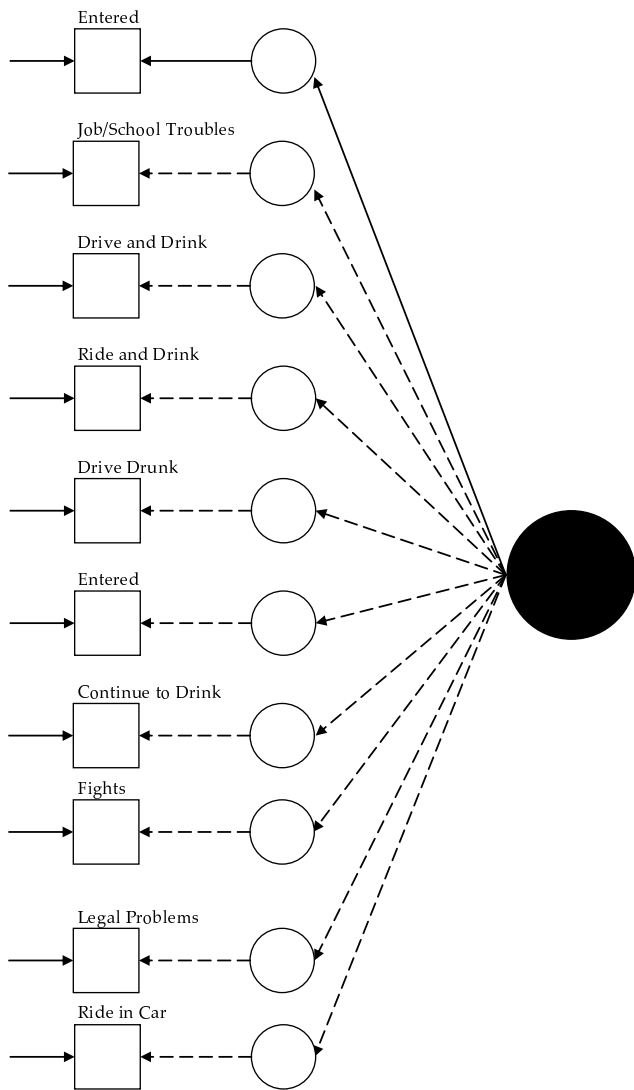
#### 3.1 Measurement Bias

##### 3.1.1 1992

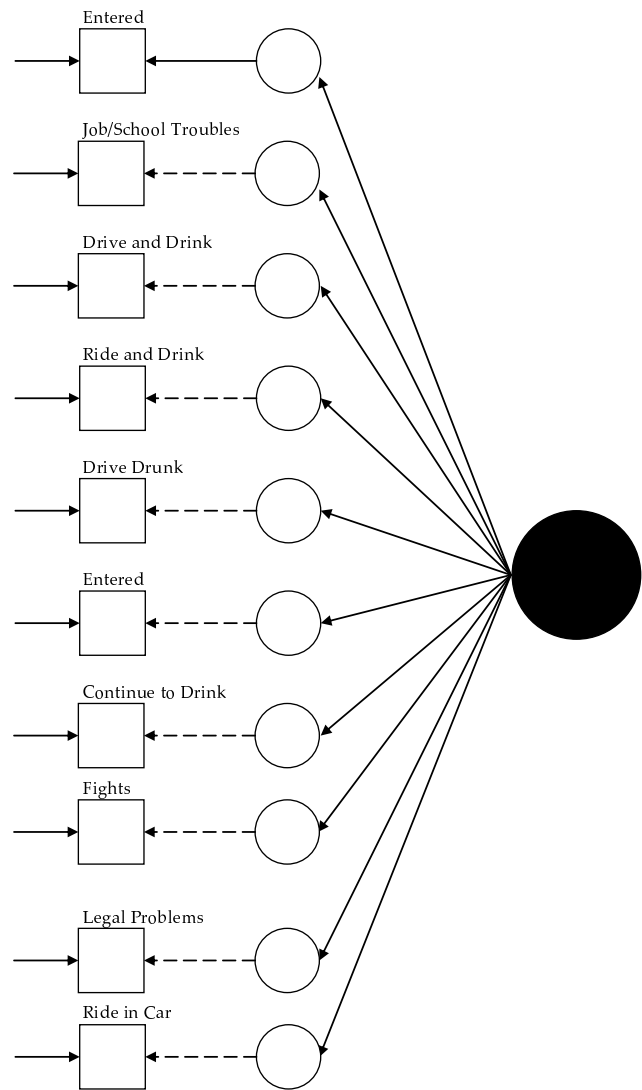
Previous work guided initial model selection. [25,26] Consistent with Muthén and colleague's work suggesting the adequacy of a single factor alcohol abuse model[25,26], analyses tested the fit of a baseline single factor model across the Hispanic and non-Hispanic Caucasian groups. Second, preceding work [76-78] recommended using the "entered into dangerous situation after drinking" item as an item free of bias across these groups. Thus, for statistical identification, the baseline model: fixed the factor mean at zero and variance at one for the non-Hispanic Caucasian group, constrained item intercepts to zero in each group, constrained the loading for the "entered " item to equality across the groups, con-

strained the threshold for the "entered" item to equality across the groups, and fixed the uniquenesses to a value of one in each group. It included no covariates.

Model 1 (see Figure 1) tested the cross-group fit of a single factor model. Fit indices suggested Model 1 fit well (RMSEA = 0.034, CFI = 0.99, TLI = 0.98,  $\chi^2 = 141.33$ , 15,  $n = 14,001$ ,  $p < 0.01$ ). Model 2 (see Figure 2) retained Model 1's constraints, constrained the loadings to equality, and uncovered no biased loadings (RMSEA = 0.028; CFI = 0.99; TLI = 0.99;  $\chi^2 = 118.20$ , 17,  $n = 14,001$ ,  $p < 0.01$ ;  $\Delta\chi^2 = 7.31$ , 4,  $n = 14,001$ ,  $p = 0.12$ ). Model 3a (see Figure 3) retained Model 2's restraints and constrained the thresholds. A statistically significant  $\Delta\chi^2$  demonstrated bias:  $\Delta\chi^2 = 16.78$  (5,  $n = 14,001$ ,  $p < 0.01$ ). For "...causing trouble with family/friends", and "...legal problems", the constraint overestimated Hispanics' thresholds. Model 3b, allowing partial invariance for these two thresholds, fit well (RMSEA = 0.027; CFI = 0.99; TLI = 0.99;  $\chi^2 = 129.85$ , 20,  $n = 14,001$ ,  $p < 0.01$ ;  $\Delta\chi^2 = 7.16$ , 3,  $n = 14,001$ ,  $p =$



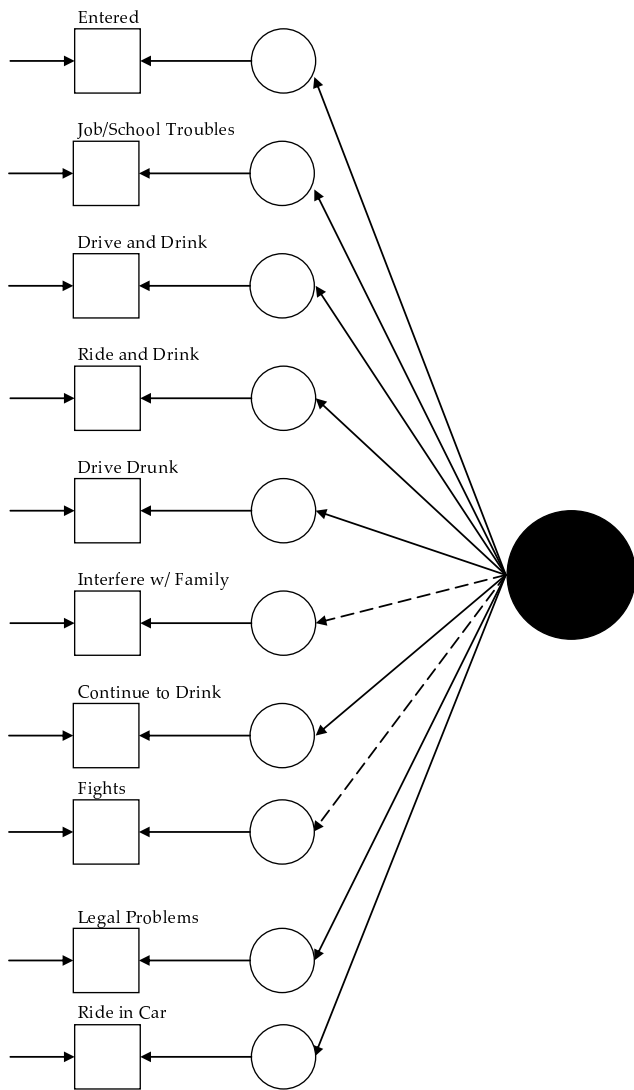
**Figure 5**  
**Path diagram summarizing Model 5, the baseline 2001–2002 alcohol abuse measurement model.** The solid black circle represents the alcohol abuse variable. The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics. Dashed lines represent values freely estimated across non-Hispanic Caucasians and Hispanics.



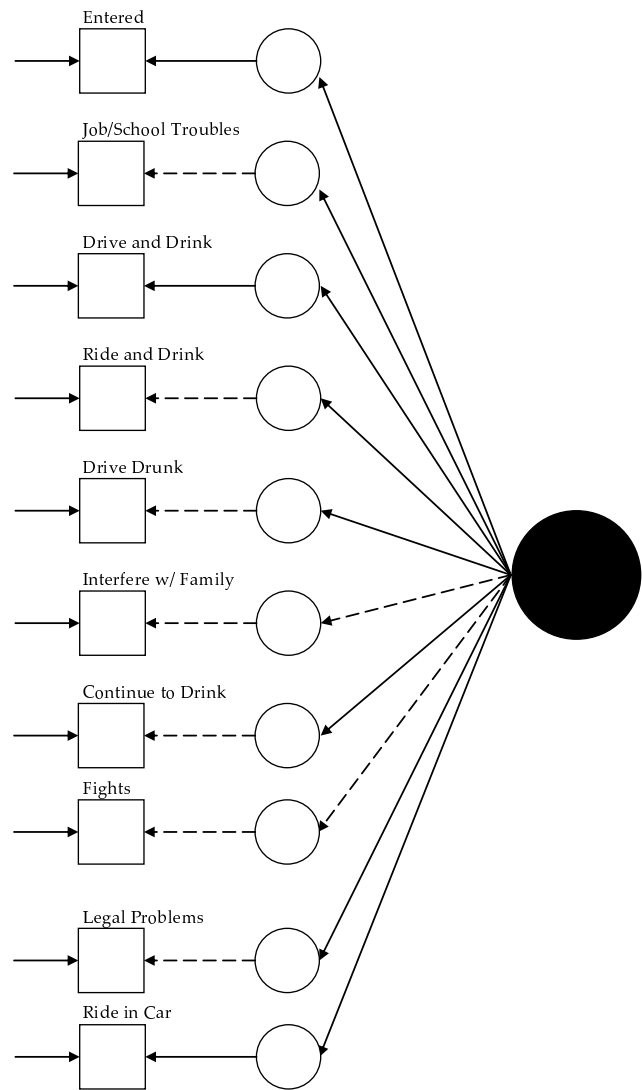
**Figure 6**  
**Path diagram summarizing Model 6a, the 2001–2002 alcohol abuse measurement model constraining the loadings to equality across groups.** The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics. Dashed lines represent values freely estimated across non-Hispanic Caucasians and Hispanics.

0.07). Analyses next compared the fit of a model (4) with free uniquenesses to the equivalent uniquenesses model. Model 4 (see Figure 4) fit well (RMSEA = 0.032; CFI = 0.99; TLI = 0.99;  $\chi^2 = 134.43$ , 16,  $n = 14,001$ ,  $p < 0.01$ ).

Constraining the uniquenesses uncovered no biased uniquenesses (RMSEA = 0.027; CFI = 0.99; TLI = 0.99;  $\chi^2 = 129.85$ , 20,  $n = 14,001$ ,  $p < 0.01$ ;  $\Delta\chi^2 = 11.2$ , 5,  $n = 14,001$ ,  $p = 0.05$ ). Given the final set of fit indices, analyses rejected the fully invariant measurement model and



**Figure 7**  
**Path diagram summarizing Model 7a, the 2001–2002 alcohol abuse measurement model constraining the thresholds to equality across groups.** The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics. Dashed lines represent values freely estimated across non-Hispanic Caucasians and Hispanics.



**Figure 8**  
**Path diagram summarizing Model 8, the final 2001–2002 alcohol abuse measurement model.** The small circles represent the continuous latent response variates underlying the dichotomous items (represented by the squares). The arrows from the solid black circle to the smaller circles represent the loadings. The arrows from the small circles to the squares represent the thresholds. The arrows pointing only to the squares represent the uniquenesses. Solid lines represent values constrained to equality across non-Hispanic Caucasians and Hispanics. Dashed lines represent values freely estimated across non-Hispanic Caucasians and Hispanics.

specified a 1992 model with partial invariance in the loadings and thresholds in its place. Table 1 summarizes the final estimates across Hispanics and non-Hispanic Caucasians in 1992 and Figure 4 presents the model visually.

To examine the possibility that the choice of the "entered" item as an anchor might influence the final model's results and interpretation, I iterated a set of analyses using the "riding in car" item that exhibited no bias in these analyses as an anchor. These analyses arrived at the exact



**Table 1: Partially Invariant Alcohol Abuse Measurement Model Across Hispanics and Non-Hispanic**

Abbreviated Item Labels	Loadings		Thresholds		Proportion Endorsing Symptom	
	Caucasians	Hispanics	Caucasians	Hispanics	Caucasians	Hispanics
Hung-over During Important Activities	2.00	2.00	-3.96	-3.96	0.038	0.043
Get Drunk During Important Activities	1.67	1.67	-3.62	-3.62	0.031	0.034
Drink And Drive	1.30	1.30	-1.94	-1.94	0.119	0.112
Enter Hurtful/Harmful Situations While Drinking	1.23	1.23	-2.69	-2.69	0.046	0.034
Continue To Drink Causing Trouble with Family and Friends	1.27	1.27	-3.13	-2.79	0.026	0.041
Legal Problems Because of Drinking	0.97	0.97	-3.05	-2.80	0.014	0.022

Caucasians in 1992 (Uniquenesses Constrained to 1 for both groups)

same partial measurement invariance model as those using the "entered" item and minimize concerns that analyses using a different item anchor diverge from these.

3.1.2 2001–2002

As above, previous work guided initial model selection, [25,26] and analyses tested the fit of a baseline single factor model across the Hispanic and non-Hispanic Caucasian groups. For statistical identification, the baseline model: fixed the factor mean at zero and variance at one for the non-Hispanic Caucasian group, constrained item intercepts to zero in each group, constrained the loading for the "entered" item to equality across the groups, constrained the threshold for the "entered" item to equality across the groups, and fixed the uniquenesses to a value of one in each group. It included no covariates. The set of fit indices suggested this model (Model 5) fit the data well (RMSEA = 0.039, CFI = 0.98, TLI = 0.98, McDonald's NCI = 0.98, Gamma Hat = 0.996, and  $\chi^2 = 733.91$ , 43,  $n = 21,373$ ,  $p < 0.01$ ). Given the well fitting model, analyses moved to metric invariance, i.e., equivalence in the loadings. This model (6a) retained restraints in the previous model, constrained the loadings to equality across the groups, and allowed variation in the remaining parameters. The  $\Delta\chi^2$  test suggested the presence of statistically significant measurement bias:  $\Delta\chi^2 = 37.88$  (8,  $n = 21,373$ ,  $p < 0.01$ ), and the hypothesis of metric invariance was rejected. Modification indices (MIs) and expected parameter change indices (EPCs) suggested that constraining the loadings for the "drinking interfered with taking care of home or family" and "get into physical fight while or after drinking" items predominantly accounted for the increased misfit. These constraints underestimated the extent to which the items related to alcohol abuse for Hispanics. A partially invariant model (6b) relaxing the equivalence constraint for the items' loadings fit the data well: RMSEA = 0.035,

CFI = 0.98, TLI = 0.98. McDonald's NCI = 0.99, Gamma Hat = 0.996,  $\chi^2 = 620.69$  (44,  $n = 21,373$ ,  $p < 0.01$ ), and  $\Delta\chi^2 = 10.62$  (6,  $n = 21,373$ ,  $p = 0.10$ ). The hypothesis of partial measurement invariance was not rejected, and analyses moved to examining invariance in the thresholds.

This model (7a) retained the partially invariant restraints in the previous model, constrained the thresholds to equality across the groups, and allowed variation in the remaining parameters. Again, the  $\Delta\chi^2$  demonstrated statistically significant measurement bias:  $\Delta\chi^2 = 57.91$  (7,  $n = 21,373$ ,  $p < 0.01$ ). The MIs and EPCs suggested that the increased misfit principally resulted from seven items. For the "drinking interfered with taking care of home or family", "more than once drive vehicle after drinking", and "get into physical fight while or after drinking" items, the equality constraint underestimated Hispanics' thresholds. For the "job or school troubles because of drinking", "more than once ride in vehicle while drinking", "continue to drink despite causing trouble with family or friends", and "get arrested or have legal problems because of drinking" items, the equality constraint overestimated Hispanics' thresholds. A model allowing partial measurement invariance (7b) for these seven thresholds fit the data well: RMSEA = 0.035, CFI = 0.98, TLI = 0.98, McDonald's NCI = 0.99, Gamma Hat = 0.996,  $\chi^2 = 622.93$  (45,  $n = 21,373$ ,  $p < 0.01$ ), and  $\Delta\chi^2 = 4.87$  (2,  $n = 21,373$ ,  $p = 0.09$ ). The hypothesis of partial measurement invariance in the loadings and thresholds was not rejected and analyses moved to the uniquenesses. Analyses next compared the fit of a model (8) with free uniquenesses to the equivalent uniquenesses model. Model 8 fit well (RMSEA = 0.038, CFI = 0.98, TLI = 0.98, McDonald's NCI = 0.98, Gamma Hat = 0.996;  $\chi^2 = 740.34$ , 44,  $n = 21,373$ ,  $p < 0.01$ ). Constraining the uniquenesses uncovered no biased uniquenesses (RMSEA = 0.035, CFI = 0.98,

TLI = 0.98, McDonald's NCI = 0.99, Gamma Hat = 0.996;  $\chi^2 = 622.93$ , 45,  $n = 21,373$ ,  $p < 0.001$ ;  $\Delta\chi^2 = 9.738$ , 6,  $n = 21,373$ ,  $p = 0.14$ ). Given the final set of fit indices, analyses rejected the fully invariant measurement model and specified a 2001–2002 model with partial invariance in the loadings and thresholds in its place. Table 2 summarizes the final estimates across Hispanics and non-Hispanic Caucasians in 2001–2002.

As with the 1992, I reiterated a set of analyses using the "riding in car" item as an anchor. These analyses arrived at the exact same partial measurement invariance model as those using the "entered" item and minimize concerns that analyses using a different item anchor diverge from these.

**3.2 Impact**

Because statistically significant item level criteria do not always translate into meaningful or practical differences on scale scores, a sensitivity analysis examined the extent to which the statistically significant bias impacted estimates. No gold standard exists for evaluating DIF's impact, especially with ordered-categorical models. [79] In light of this, a number of authors argue for and have shown the utility of conducting analyses that compare the direction and size of mean differences resulting from a fully invariant model ignoring observed measurement bias to those resulting from the model incorporating

measurement bias to evaluate impact. [80-82] Changes in mean differences reflect impactful bias. Analyses adopted this approach.

**3.2.1 1992**

In the invariant model, non-Hispanic Caucasians' mean equaled zero (a function of statistical identification) and Hispanics' mean equaled -0.09 ( $z = -0.82$ ,  $p = 0.41$ ). Under the partially invariant model, this pattern persisted ( $M_{Caucasians} = 0.00$ ,  $M_{Hispanics} = -0.05$ ,  $z = -0.44$ ,  $p = 0.66$ ), suggesting little impact. [62,83] Failing to incorporate measurement bias did not affect mean estimates and cross-group comparisons.

**3.2.2 2001–2002**

In the fully invariant model, non-Hispanic Caucasians had a group mean of zero (a function of statistical identification) and Hispanics had a group mean of 0.49 significantly greater than zero ( $M_{Caucasians} = 0.00$ ,  $SD_{Caucasians} = 1$ ,  $M_{Hispanics} = 0.49$ ,  $SD_{Hispanics} = 1.26$ ,  $z = 3.39$ ,  $p < 0.01$ ). For these items, higher values reflect less use (1 = "Yes", 2 = "No"). Under the partially invariant model, non-Hispanic Caucasians' and Hispanics' means did not differ significantly ( $M_{Caucasians} = 0.00$ ,  $SD_{Caucasians} = 1$ ,  $M_{Hispanics} = 0.07$ ,  $SD_{Hispanics} = 0.98$ ,  $z = 1.29$ ,  $p = 0.10$ ). Thus, failing to incorporate statistically significant measurement bias: 1) meaningfully impacts mean estimates and cross-group comparisons, 2) overestimates differences between the

**Table 2: Partially Invariant Alcohol Abuse Measurement Model across Hispanics and Non-Hispanic Caucasians in 2001–2002 (Uniquenesses Constrained to 1 for both groups)**

Abbreviated Item Labels	Loadings		Thresholds		Proportion Endorsing Symptom	
	Caucasians	Hispanics	Caucasians	Hispanics	Caucasians	Hispanics
Drinking Interfered with Taking Care of Home or Family	1.27	2.31	-3.97	-5.78	0.007	0.008
Job or School Troubles Because of Drinking	1.43	1.43	-4.58	-4.25	0.004	0.006
More Than Once Drive Vehicle While Drinking	1.91	1.91	-3.18	-3.18	0.070	0.056
More Than Once Ride In Vehicle While Driver was Drinking	2.29	2.29	-3.25	-2.99	0.097	0.099
More Than Once Drive Vehicle After Too Much to Drink	1.60	1.60	-3.15	-3.25	0.048	0.036
Entered Into Dangerous Situation After Drinking	1.32	1.32	-3.17	-3.17	0.028	0.021
Continue to Drink Despite Causing Trouble With Family or Friends	1.47	1.47	-4.00	-3.58	0.012	0.017
Get Into Physical Fight While or After Drinking	1.02	1.54	-3.20	-3.66	0.013	0.019
Get Arrested or Have Legal Problems Because of Drinking	1.01	1.01	-3.41	-3.13	0.008	0.011
Ride in Car as Passenger While Drinking	1.96	1.96	-2.69	-2.69	0.110	0.101

groups, and 3) underestimates Hispanics' true use levels. Table 3 completely summarizes these estimates.

#### 4. Discussion

How well does the field measure and estimate change in alcohol abuse among Hispanics in the US? In this study, I sought an answer. First, I examined whether statistically significant, impactful measurement bias presented across Hispanic and non-Hispanic Caucasians on a standardized, six-item measure of DSM-IV alcohol abuse in a nationally representative 1992 US sample. Confirmatory factor analysis for ordered-categorical measures (CFA-OCM) uncovered two biased items. These items addressed drinking related troubles with family and friends and whether individuals experienced drinking related legal problems. Bias resulted in differential reporting tendencies at similar levels of alcohol abuse. Relative to non-Hispanic Caucasians, Hispanics needed to experience fewer "trouble[s] with family and friends" and fewer "legal problems" to say yes. However, given that partial measurement invariance does not by default lead to biased observed scores,[38,84] nor do statistically significant criteria necessarily translate into meaningful differences[38] I also investigated whether item-level bias affected 1992 item-set alcohol abuse estimates. A sensitivity analysis compared the size and direction of mean differences across a model proceeding as if bias didn't present and a model incorporating measurement bias. This comparison examined whether analyses conducted ignoring measurement bias would diverge from those incorporating bias. These analyses revealed that item-level differences minimally affected item-set alcohol abuse estimates. In other words, do 1992 cross-ethnic alcohol abuse estimates provide valid baseline estimates? Yes.

Second, to better evaluate recent alcohol abuse estimates and differential change in alcohol abuse across time, I examined whether statistically significant measurement bias existed across Hispanic and non-Hispanic Caucasians on a standardized, 10 item measure of DSM-IV alcohol abuse in a recent (2001–2002), large, nationally representative survey of US alcohol use. This addressed

whether statistically significant bias impacted the validity of *current* alcohol abuse estimates across Hispanics and non-Hispanic Caucasians. CFA-OCM demonstrated the presence of statistically significant, impactful measurement bias for seven of ten items. These items addressed drinking related legal problems, physical fights, job or school troubles, troubles with family and friends, and drinking related interference with taking care of the home or family, as well as whether individuals drove vehicles after drinking too much and whether they rode in vehicles as passengers while drinking. Differences in responses to these items underestimated rates of alcohol abuse among 2001–2002 Hispanics as compared to non-Hispanic Caucasians. In other words, how valid are current estimates of alcohol abuse across Hispanics and non-Hispanic Caucasians? Not valid enough.

Bias in the loadings showed that two problems better predicted alcohol abuse for Hispanics than they did for non-Hispanic Caucasians; "*drinking related interference with taking care of the home*" and "*physical fights while or after drinking*". Endorsements of these items more closely tied to alcohol abuse for Hispanics than non-Hispanic Caucasians; one would have more faith that Hispanics' item responses reflected alcohol abuse as opposed to some other influence. Bias in the thresholds demonstrated differential reporting tendencies at similar levels of alcohol abuse for seven items. CFA-OCM assumes a continuous latent variate underlies observed yes/no responses and that a threshold determines responses, thus, 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 were less likely to endorse several items. Compared to non-Hispanic Caucasians, they needed to experience more "*drinking related interference with taking care of the home or family*", more "*physical fights while or after drinking*", or "*drive [a] vehicle after drinking*" more to say yes. Four items saw a reversed pattern. As a function of alcohol abuse, Hispanics more readily endorsed "*drinking related legal problems*", "*drinking related trouble with family or friends*" and "*drinking related job or school troubles*". Also,

**Table 3: Sensitivity Analyses Presenting Changes in Means across Models Ignoring Measurement Bias (Fully Invariant) and Allowing Measurement Bias (Partially Invariant)**

	1992				2001–2002			
	Mean (SD)		z	p	Mean (SD)		z	p
	Caucasian	Hispanic			Caucasian	Hispanic		
Fully Invariant	0.00 (1.00)	-0.09 (0.96)	-0.82	0.41	0.00 (1.00)	0.49 (1.26)	3.39	< 0.01
Partially Invariant	0.00 (1.00)	-0.05 (0.95)	-0.44	0.66	0.00 (1.00)	0.07 (0.98)	1.29	0.10

Hispanics needed to "ride in a vehicle as passengers while drinking" less frequently before upholding the item.

Taken together, these findings present strong evidence that Hispanics in the US currently respond to several items operationalizing alcohol abuse criteria differently than non-Hispanic Caucasians and they call into doubt the cross-cultural equivalence of alcohol abuse measurement across these groups, especially given that the 1992 analyses with a different cohort and reduced item set also identified problems with the "legal problems" and "trouble with family and friends" items. Moreover, unlike 1992, acknowledging and incorporating measurement bias in the 2001–2002 model lead to *increased* mean reporting levels. Observed scores incorrectly estimate alcohol abuse and fail to provide cross-culturally valid measurement in 2001–2002. Relative to non-Hispanic Caucasians, these findings suggest greater levels of alcohol abuse among Hispanics than previously reported. Given that 1992 estimates do provide a valid baseline, not only is alcohol abuse increasing alarmingly among Hispanics,[3,4] it increases at a greater rate than suspected.

This investigation provides strong evidence that measurement bias presents across Hispanics and non-Hispanic Caucasians when measuring alcohol abuse. A number of mechanisms may *simultaneously* result in this bias, particularly given bias' non-uniform distribution (e.g., some items were more difficult to endorse, others were easier). For example, research notes cultural differences in social desirability and the extent to which Hispanics see psychiatric symptoms as undesirable. [50] Hui and Triandis[85] note that cultural differences in sincerity may influence Hispanic responses. Language skills and socioeconomic variability may also differentially affect responses across these groups. [50] Additionally, the findings may not represent error, but rather accurately reflect fundamental differences in alcohol abuse patterns across non-Hispanic Caucasians and Hispanics. Each of these influences may lead to measurement bias. Future research should seek to elucidate what leads to these differences.

Regardless of bias' source, the findings have implications for public health and clinical practice. First, the US should increase public health alcohol abuse prevention and intervention efforts among Hispanics. Previous work shows that drinking behaviors have increased among this group[4] and the US has devoted resources specifically to address the unique health concerns of minorities[86]. Nevertheless, this study's results demonstrate that alcohol abuse has increased more than suspected and this community deserves more concerted efforts to stem this disease's increase. Second, item level findings suggest that health care interventions aimed at any of the seven criteria that demonstrated bias in 2001–2002 need to consider

that responses about these behaviors among Hispanics likely do not reflect problems as they do among non-Hispanic Caucasians. Given similar item-level differences for the "trouble[s] with family and friends" and "legal problems" items at 1992 and 2001–2002, health care interventions and clinicians should pay particular attention to these two criteria. Ethnic differences exist in the experience and expression of alcohol abuse for Hispanics as compared to non-Hispanic Caucasians. Third, the findings call into question the cross-cultural equivalence of alcohol abuse and highlight the need for culturally sensitive research and prevention and intervention efforts generally. Psychological science should seek the source of this bias and carefully examine the appropriateness of diagnosing and describing cultural minorities using biased items.

Before concluding, the study's strengths and limits deserve review. First, the study focused on ethnic differences. The study did this because a vast body of work examines cross-ethnic differences in alcohol use without regard to other sociodemographic differences and this study intentionally adopted this approach as well to address the validity of these considerable and similarly oriented studies. Probative analyses in the 2001–2002 data examining whether the exact pattern of measurement differences described above present across ethnicity and sex suggested that a relatively similar pattern of measurement bias results when incorporating sex and culture simultaneously, although some minor sex differences present uniquely within culture and some sex differences exhibit exclusively across cultures. However, sample size restrictions resulted in frequent bivariate empty cells, limiting the reliability and interpretability of these analyses. Thus, I report them cautiously here. Second, consistent with other work,[4] the study treated Hispanics as a homogenous cultural group, despite their heterogeneity in America. [87] Analyses could not explore more specifically defined Hispanic groups given their smaller sample sizes within the included Hispanic group, e.g., South American Hispanics  $n = 28$ , even using Mplus' robust WLSMV estimator. [64] This inability to estimate models for groups this small may miss measurement heterogeneity among Hispanic Americans. Third, the study used a representative sample; it remains unclear whether results would persist in clinical samples.

Finally, these data represent self-reports and may not reflect actual experiences. Without an external gold standard criterion, it remains unclear the extent to which self-reports differ from actual experiences. Additionally, this leaves open the possibility that these questions provide more accurate measurement for Hispanics and poorer measurement for non-Hispanic Caucasians. In other words, without a gold standard, it is possible estimates over-report non-Hispanic Caucasians' alcohol abuse lev-

els rather than under-reporting Hispanics' alcohol abuse. However, given the development of alcohol abuse among the majority non-Hispanic Caucasian population, it seemed reasonable to use non-Hispanic Caucasians as the reference group, as much of the measurement research does. Nevertheless, readers should interpret the findings here with this caution in mind.

These limits leave some issues unaddressed. First, the results of exploratory analyses investigating ethnicity and sex simultaneously highlight the need for future studies with larger sample sizes that could address sociodemographic variability simultaneous to ethnicity. Likewise, by collecting data from larger number of Hispanic individuals, research could examine the equivalence of alcohol abuse measurement within the Hispanic community. A new, larger sample could address this issue. Because clinical samples can differ from community samples,[88] research should examine whether these findings hold in clinical samples. Finally, future research should collect additional data and use an external criterion and examine the extent to which self-reports correspond to the external criterion across non-Hispanic Caucasians and Hispanics. This would clarify whether these findings reflect under-reporting for Hispanics or over-reporting for non-Hispanic Caucasians.

Despite limits, the study has numerous assets. First, it makes an important and unique contribution. A literature review found no studies examining the cross-ethnic measurement equivalence of alcohol abuse in previous or recent US data. Second, it fills this substantial gap using well designed, large, nationally representative samples, alleviating sampling bias and other methodological concerns. Third, it uses modern measurement modeling techniques that allow a sophisticated, precise, and preferred examination of the bias[27]. Finally, it explicitly calls awareness to social science's oft displayed ignorance of cultural variability in measurement.

## 5. Conclusion

In conclusion, results demonstrated the presence of statistically significant measurement bias across Hispanics and non-Hispanic Caucasians for two of six items assessing DSM-IV alcohol abuse in a representative sample of the US in 1992. These item-level differences did not affect alcohol abuse estimates based on the set, though. However, analyses did reveal impactful measurement bias across Hispanics and non-Hispanic Caucasians in a representative sample of the 2001–2002 US for a set of seven items operationalizing DSM-IV alcohol abuse. Results currently suggest caution when diagnosing and estimating rates and levels of alcohol abuse across these groups. Moreover, the study notes that current descriptions *underestimate* the rate of alcohol abuse among Hispanics rela-

tive to non-Hispanic Caucasians and that alcohol abuse may be increasing at a greater rate than previously suspected. Finally, these results underscore the need for culturally sensitive research, prevention, and intervention efforts and support the need to empirically question the generalization of psychological findings from the majority group to minority groups in current population data. Summarily, how well does the field currently measure and estimate alcohol abuse across non-Hispanic Caucasians and Hispanics? Not well enough.

## Competing interests

The author declares that they have no competing interests.

## Authors' contributions

Using publicly available data, I worked individually, conducted the literature searches and summaries of previous related work, undertook the statistical analyses, wrote the manuscript, conducted all revisions, and read and approved the final manuscript.

## Acknowledgements

I would like to thank the National Institute on Alcohol Abuse and Alcoholism and Dr. Bridget Grant for making the data publicly available. I would also like to thank Tara J. Carle and Margaret Carle whose unending support and thoughtful comments made my work possible. I am also grateful for Manuel Barrera, David Jaffee, and Jacob Vigil's comments as I developed the manuscript.

## References

1. Saha T, Chou SP, Grant BF: **Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions.** *Psychological Medicine* 2006, **36**:931-941.
2. 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.
3. 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.
4. 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.
5. 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**:331-334.
6. 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.
7. Bloche MG: **Health care disparities – science, politics, and race.** *New England Journal of Medicine* 2004, **350**:1568-1570.
8. 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.
9. Ramirez M, Ford ME, Stewart AL, Teresi JA: **Measurement issues in health disparities research.** *Health Services Research* 2005, **40**:1640.

10. Smedley BD, Stith AY, Neslon AR: **Unequal treatment: Confronting racial and ethnic disparities in health care.** Academic Press: Washington DC; 2003.
11. Steinbrook R: **Disparities in health care – from politics to policy.** *New England Journal of Medicine* 2004, **350**:1486-1488.
12. Stewart AL, Nápoles-Springer AM: **Advancing health disparities research: can we afford to ignore measurement issues?** *Medical care* 2003, **41(11)**:1207-1220.
13. 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.
14. 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.
15. Mellenbergh GJ: **Item bias and item response theory.** *International Journal of Educational Research* 1989, **13**:127-143.
16. 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.
17. Grant BF: **Convergent validity of DSM-III-R and DSM-IV alcohol dependence: Results from the national longitudinal alcohol epidemiologic survey.** *J Subst Abuse* 1997, **9**:89-102.
18. Grant BF: **Theoretical and observed subtypes of DSM-IV alcohol abuse and dependence in a general population sample.** *Drug Alcohol Depend* 2000, **60(3)**:287-293.
19. 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.
20. Grant BF, Dawson DA, Stinson FS, Chou PS, Kay W, Pickering R: **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 Alcohol Depend* 2003, **71(1)**:7-16.
21. 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.
22. Hasin DS, Muthuen B, Wisnicki KS, Grant BF: **Validity of the bi-axial dependence concept: A test in the US general population.** *Addiction* 1994, **89(5)**:573-579.
23. Harford , Muthén BO: **The dimensionality of alcohol abuse and dependence: A multivariate analysis of DSM-IV symptom items in the National Longitudinal Survey of Youth.** *J Stud Alcohol* 2001, **62(2)**:150-157.
24. 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.
25. Muthén BO: **Factor analysis of alcohol abuse and dependence symptom items in the 1988 National Health Interview Survey.** *Addiction* 1995, **90(5)**:637-645.
26. 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(8)**:1071-1077.
27. Stahl SM, Hahn AA: **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(11 Suppl 3)**:S1-2.
28. 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.
29. Cole SR: **Assessment of differential item functioning in the Perceived Stress Scale-10.** *Epidemiology and Community Health* 1999, **53**:319-320.
30. 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(2)**:192-218.
31. Pentz MA, Chou C: **Measurement invariance in longitudinal clinical research assuming change from development and intervention.** *J Consult Clin Psychol* 1994, **62(3)**:450-462.
32. Smith L, Reise SP: **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* 1998, **75**:1350-1362.
33. Gallo JJ, Anthony JC, Muthén BO: **Age differences in the symptoms of depression: A latent trait analysis.** *J Gerontol* 1994, **49(6)**:P251-P264.
34. Reid R, DuPaul GJ, Power TJ, Anastopoulos AD, Rogers-Adkinson D, Noll MB, Riccio C: **Assessing culturally different students for attention deficit hyperactivity disorder using behavior rating scales.** *Journal of Abnormal Child Psychology* 1998, **26**:187-198.
35. Cole DA, Martin JM, Peeke L, Henderson A, Harwell J: **Validation of depression and anxiety measures in White and Black youths: Multitrait-multimethod analyses.** *Psychol Assess* 1998, **10(3)**:261-276.
36. Byrne BM, Baron P: **The Beck Depression Inventory: Testing and cross-validating a hierarchical factor structure for non-clinical adolescents.** *Measurement and Evaluation in Counseling and Development* 1993, **26**:164-178.
37. 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.
38. 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.
39. Byrne BM, Baron P, Campbell TL: **Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender.** *J Res Adolesc* 1993, **3(2)**:127-143.
40. Knight GP, Hill NE: **Studying minority adolescents: Conceptual, methodological, and health issues.** In *Measurement equivalence in research involving adolescents* Edited by: McLoyd VC. New Jersey: Laurence Erlbaum; 1998.
41. Schafer J, Caetano R: **The DSM-IV construct of cocaine dependence in a treatment sample of Black, Mexican American, and White men.** *Psychological Assessment* 1996, **8**:304-311.
42. Knight GP, Tein JY, Shell R, Roosa M: **The cross-ethnic equivalence of parenting and family interaction measures among Hispanic and Anglo-American families.** *Child Dev* 1992, **63(6)**:1392-1403.
43. Prelow HM, Michaels ML, Reyes L, Knight GP, Barrera JM: **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.
44. Prelow H, Tien JY, Roosa MW, 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.
45. Meredith W, Teresi JA: **An essay on measurement and factorial invariance.** *Medical care* 2006, **44(11 Suppl 3)**:S69-77.
46. Sue S: **Science, ethnicity, and bias: Where have we gone wrong?** *American Psychologist* 1999, **54**:1070-1077.
47. Wright GN, Phillips LD, Whalley PC, Gerry TC, Kee-Ong N, Tan I: **Cultural differences in probabilistic thinking.** *Journal of Cross-Cultural Psychology* 1978, **9**:285-299.
48. Smith PB: **Acquiescent response bias as an aspect of cultural communication style.** *Journal Of Cross-Cultural Psychology* 2004, **35**:50-61.
49. 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.
50. McHorney CA, Fleishman JA: **Assessing and understanding measurement equivalence in health outcome measures. Issues for further quantitative and qualitative inquiry.** *Medical care* 2006, **44(11 Suppl 3)**:S205-10.
51. Carle AC: **Assessing the Cross-Cultural Validity of Alcohol Dependence across Hispanics and Non-Hispanic Caucasians.** *Hispanic Journal of Behavioral Sciences* 2008, **30**:106-120.
52. Grant BF, Peterson A, Dawson DA, Chou SP: *Source and Accuracy Statement for the National Longitudinal Alcohol Epidemiologic Survey*

- (NIAES) Bethesda, MD: National Institute on Alcohol Abuse and Alcoholism; 1994.
53. 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.
  54. American Psychiatric Association: *Diagnostic and Statistical Manual of Mental Disorders* Fourth edition. Washington, DC: American Psychiatric Association; 1994.
  55. Grant BF, Hasin DS: *The Alcohol Use Disorder and Associated Disabilities Interview Schedule-DSM-IV Version (AUDADIS)* Bethesda, MD: National Institute of on Alcohol Abuse and Alcoholism; 1992.
  56. Grant BF, Harford TC, Dawson DD, Chou PS: **The Alcohol Use Disorder and Associated Disabilities Interview Schedule (AUDADIS): Reliability of alcohol and drug modules in a general population sample.** *Drug Alcohol Depend* 1995, **39(1)**:37-44.
  57. Hasin DS, Grant B, Cottler L: **Nosological comparisons of alcohol and drug diagnoses: A multisite, multi-instrument international study.** *Drug and Alcohol Dependence* 1997, **47**:217-226.
  58. Hasin D, Carpenter KM, McCloud S, Smith M, Grant BF: **The alcohol use disorder and associated disabilities interview schedule (AUDADIS): Reliability of alcohol and drug modules in a clinical sample.** *Drug Alcohol Depend* 1997, **44(2–3)**:133-141.
  59. Hasin D, Paykin A: **Alcohol dependence and abuse diagnoses: Concurrent validity in a nationally representative sample.** *Alcohol Clin Exp Res* 1999, **23(1)**:144-150.
  60. Muthén BO, Christoffersson A: **Simultaneous factor analysis of dichotomous variables in several groups.** *Psychometrika* 1981, **46**:407-419.
  61. Millsap RE: **Comments on methods for the investigation of measurement bias in the Mini-Mental State Examination.** *Medical care* 2006, **44(11 Suppl 3)**:S171-5.
  62. Teresi JA: **Different approaches to differential item functioning in health applications. Advantages, disadvantages and some neglected topics.** *Medical care* 2006, **44(11 Suppl 3)**:S152-70.
  63. Takane Y, de Leeuw J: **On the relationship between item response theory and factor analysis of discretized variables.** *Psychometrika* 1987, **52**:393-408.
  64. Muthén LK, Muthén BO: *Mplus User's Guide* Fourth edition. Los Angeles, CA: Muthén & Muthén; 1998.
  65. Millsap RE, Everson HT: **Methodology review: Statistical approaches for assessing measurement bias.** *Applied Psychological Measurement* 1993, **17(4)**:297-334.
  66. Byrne B: **Structural Equation Modeling with LISREL, PRELIS, and SIMPLIS: Basic applications and programs.** Mahwah, New Jersey: Lawrence Erlbaum; 1998.
  67. Millsap RE, Yun-Tein J: **Assessing factorial invariance in ordered-categorical measures.** *Journal of Multivariate Behavioral Research* 2004, **39**:479-515.
  68. Muthén B: **A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators.** *Psychometrika* 1984, **49(1)**:115-132.
  69. Bollen KA: **Structural Equations with Latent Variables.** New York, New York: Wiley; 1989.
  70. Hu L, Bentler PM: **Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification.** *Psychol Methods* 1998, **3(4)**:424-453.
  71. Hu L, Bentler PM: **Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives.** *Structural Equation Modeling* 1999, **6(1)**:1-55.
  72. Steiger JH: **A note on multiple sample extensions of the RMSEA fit index.** *Structural Equation Modeling* 1998, **5**:411-419.
  73. Cheung GW, Rensvold RB: **Evaluating goodness-of-fit indexes for testing measurement invariance.** *Structural Equation Modeling* 2002, **9**:233-255.
  74. Green SB, Babyak MA: **Control of Type I errors with multiple tests of constraints in structural equation modeling.** *Multivariate Behavioral Research* 1997, **32(1)**:39-51.
  75. Thissen D, Steinberg L, Wainer H: **Detection of differential item functioning using the parameters of item response models.** In *Differential item functioning* Edited by: Holland PW. Hillsdale, NJ: Lawrence Erlbaum Associates; 1993:67-113.
  76. Carle AC: **Assessing cross-cultural differences in the DSM-IV alcohol use disorder constructs across non-Hispanic Caucasians, non-Hispanic African-Americans, and Hispanics.** Paper presented at the 133rd Meeting of the American Public Health Association; 2005.
  77. Carle AC: **Cross-cultural differences in the constructs of alcohol dependence and abuse: Assessing measurement bias across Hispanic, non-Hispanic African-Americans, and non-Hispanic Caucasian adults.** Poster presented at the 2005 annual APA Convention; 2005.
  78. Carle AC: **Measurement bias across Hispanics and non-Hispanic Caucasians on a standardized measure of alcohol abuse: Empirically evaluating the impact of observed measurement bias.** Paper presented at the 13th SPR Meeting; 2005.
  79. Millsap RE, Kwok O: **Evaluating the Impact of Partial Factorial Invariance on Selection in Two Populations.** *Psychol Methods* 2004, **9(1)**:93-115.
  80. Carle AC: **Cross-cultural validity of alcohol dependence across Hispanics and non-Hispanic Caucasians.** *Hispanic Journal of Behavioral Sciences* 2008, **30(1)**:106-120.
  81. Millsap RE, Kwok O: **Evaluating the Impact of Partial Factorial Invariance on Selection in Two Populations.** *Psychol Methods* 2004, **9(1)**:93-115.
  82. Teresi JA, Stewart AL, Morales LS, Stahl SM: **Measurement in a multi-ethnic society. Overview to the special issue.** *Medical care* 2006, **44(11 Suppl 3)**:S3-4.
  83. Borsboom D: **When does measurement invariance matter?** *Medical care* 2006, **44(11 Suppl 3)**:S176-81.
  84. Cole DA, Maxwell SE: **Multitrait-multimethod comparisons across populations: A confirmatory factor analytic approach.** *Multivariate Behavioral* 1985, **20(389)**:417.
  85. Hui CH, Triandis HC: **Effects of culture and response format on extreme response style.** *Journal of Cross-Cultural Psychology* 1989, **20(3)**:296-309.
  86. **Healthy People 2010** [[http://www.health.gov/healthypeople/document/html/aih/aih\\_2.htm](http://www.health.gov/healthypeople/document/html/aih/aih_2.htm)]
  87. 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.
  88. Kazdin AE: *Research Design in Clinical Psychology* Fourth edition. Boston, MA: Allyn and Bacon; 2003.

### Pre-publication history

The pre-publication history for this paper can be accessed here:

<http://www.biomedcentral.com/1471-2458/9/60/prepub>

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

