TESTING THE INVARIANCE OF ADOLESCENT SURVEY-BASED ALCOHOL-RELATED BEHAVIORS ACROSS RACE/ETHNICITY AND GENDER

A Thesis
Presented to the
Faculty of
San Diego State University

In Partial Fulfillment
of the Requirements for the Degree
Master of Arts
in
Psychology

by
Loretta Yao Pei Hsueh
Spring 2015
SAN DIEGO STATE UNIVERSITY

The Undersigned Faculty Committee Approves the

Thesis of Loretta Yao Pei Hsueh:

Testing the Invariance of Adolescent Survey-Based Alcohol-Related Behaviors

Across Race/Ethnicity and Gender

______________________________
Elizabeth Klonoff, Chair
Department of Psychology

______________________________
Scott C. Roesch
Department of Psychology

______________________________
Suzanne Lindsay
Graduate School of Public Health

3/16/2015
Approval Date
DEDICATION

This thesis is dedicated to my family, whose support has allowed me to be where I am today, and to my cohort, whose support has confirmed that this is where I want to be.
ABSTRACT OF THE THESIS

Testing the Invariance of Adolescent Survey-Based Alcohol-Related Behaviors Across Race/Ethnicity and Gender

by

Loretta Yao Pei Hsueh
Master of Arts in Psychology
San Diego State University, 2015

Alcohol consumption amongst adolescents continues to be a major public health concern in the United States. State and federal policies have been designed and implemented with the intention of curbing problematic drinking behaviors amongst the adolescents identified as being at the highest risk. These policies are often informed by surveys such as the Youth Risk Behavior Survey (YRBS), sponsored by the Center for Disease Control (CDC). Previous research, however, has found that such surveys lack invariance (or measurement equivalency) across different groups, such as race/ethnicity and gender. A lack of invariance indicates that valid group comparisons cannot be made, since group membership determines perception of the survey question. This project examined the factorial invariance of the 2013 YRBS alcohol-related behaviors across race/ethnicity and gender. Multigroup confirmatory factor analysis statistically tested the psychometric and factorial invariance of the “Alcohol-Related Behaviors” latent variable across four ethnic and two gender groups. Invariance was examined across Hispanic/Latino (n = 1,734), white (n = 5,449), Asian (n = 491), and black/African American (n = 2,991) adolescents, and across males (n = 5,439) and females (n = 5,224). The latent variable was indicated by five questions, representing “drinking and driving,” “lifetime alcohol use,” “30 day alcohol use,” “binge drinking,” and “maximum number of drinks, 30 days.” Invariance was first examined across gender groups to determine if subsequent tests could be done collapsing across gender. While the models achieved configural invariance, metric invariance could not be established. This indicated that associations between measured variables (e.g., binge drinking) and the latent construct (i.e., “Alcohol-Related Behaviors”) differed between males and females. Subsequent tests of invariance were therefore done within gender groups, comparing race/ethnicity groups in pairs. Configural invariance was established across race/ethnicity groups for both genders. However, metric invariance could only be achieved between certain groups. Results suggest that comparisons made between gender groups on alcohol-related behaviors as a construct may be inappropriate. Additionally, even when making comparisons within gender, caution should be taken when making comparisons between race/ethnic groups. Findings highlight the need to tailor alcohol prevention and intervention efforts to particular groups based on group membership.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>viii</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Alcohol Use &amp; Gender</td>
<td>2</td>
</tr>
<tr>
<td>Alcohol Use &amp; Race/Ethnicity</td>
<td>2</td>
</tr>
<tr>
<td>Measurement Equivalency in Alcohol Research</td>
<td>3</td>
</tr>
<tr>
<td>Purpose</td>
<td>4</td>
</tr>
<tr>
<td>METHODS</td>
<td>5</td>
</tr>
<tr>
<td>Procedure</td>
<td>5</td>
</tr>
<tr>
<td>Participants</td>
<td>6</td>
</tr>
<tr>
<td>Measures</td>
<td>7</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>7</td>
</tr>
<tr>
<td>RESULTS</td>
<td>11</td>
</tr>
<tr>
<td>Gender</td>
<td>11</td>
</tr>
<tr>
<td>Race/Ethnicity</td>
<td>15</td>
</tr>
<tr>
<td>Men</td>
<td>15</td>
</tr>
<tr>
<td>Women</td>
<td>18</td>
</tr>
<tr>
<td>DISCUSSION</td>
<td>22</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>25</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1. Responses to Alcohol-Related Behavior Questions as a Percentage of the Sample by Gender and Race/Ethnicity .................................................................12
Table 2. Results of One-Way ANOVA Using White Group as Comparison ..................13
Table 3. Unstandardized Factor Loadings of Baseline Models ....................................14
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my mentor, Dr. Elizabeth A. Klonoff, for your advice and support. More importantly, thank you for humoring me these past years. I would also like to thank Dr. Scott C. Roesch and Dr. Suzanne P. Lindsay for all of your time and advice.
INTRODUCTION

Adolescence is characterized by significant changes in biological, cognitive, and social developments. At the same time, adolescence may also be characterized by increased incidences of alcohol initiation and the development of alcohol-related problems. Despite growing efforts to curb alcohol use during this critical time of development, adolescent drinking continues to be a major public health concern. Adolescent alcohol use has been linked to developmental changes in cognition, as well as risky behaviors such as unprotected sex and drunk driving (Calvert, Bucholz, & Stegar-May, 2010; O’Malley & Johnston, 2007). The Pacific Institute for Research and Evaluation (PIRE) estimated that underage drinking alone cost the US over $25 million in medical care and work loss in 2010 (PIRE, 2011).

According to the Monitoring the Future study, more adolescents drink alcohol than smoke cigarettes or use marijuana. The study estimates that seven out of every ten students (68%) consume alcohol by the end of high school, and three out of ten (28%) consume alcohol by the 8th grade. About half (52%) of 12th graders and one eighth (12%) of 8th graders reported having been drunk at least once in their life (Johnston, O'Malley, Miech, Bachman, & Schulenberg, 2014). The early onset of alcohol use has been shown to increase the chronicity of adult alcohol dependence, and those who begin drinking alcohol before the age of 15 are five times as likely to report alcohol dependence or abuse than those who first drank alcohol at the age of 21 or older (Guttmannova et al., 2011; Substance Abuse and Mental Health Services Administration, 2005). Thus, early identification of groups vulnerable to alcohol-related problems is critical to the current and future health of adolescent-aged populations.

Strategies to curb adolescent alcohol use are often informed by state and national self-report surveys. Surveys such as the California Health Interview Survey (UCLA Center for Health Policy Research, 2013), the Monitoring the Future Survey (Johnston et al., 2014), and the Youth Risk Behavior Survey (Center for Disease Control and Prevention [CDC], 2013), guide prevention and intervention strategies through surveillance of adolescents’ self-reported health-risk behaviors. Data collected from these surveys can aid in comparing rates of incidence and prevalence of alcohol-related behaviors and the identification of at-risk
groups. However, the usefulness of these surveys depends on the reliability and validity of the scale measures. Social factors have been found to influence the differing rates of alcohol behaviors among different race/ethnic groups as well as between genders, and these factors may influence scale responses.

**Alcohol Use & Gender**

Research suggests that odds of developing lifetime alcohol abuse and dependence is higher among male adolescents compared to female adolescents (Haberstick et al., 2013). A number of gender differences found in factors influencing alcohol use among adolescents likely explain this discrepancy. Specifically for male drinkers, traditional masculine ideologies are positively related to alcohol consumption (Isenhart, 1993; McCreary, Newcomb, & Sadava, 1999; Uy, Massoth, & Gottdiener, 2014). Conversely, females endorsing stereotypical female characteristics (e.g., nurturance) are less likely to report alcohol involvement (Ricciardelli, Connor, Williams, & Young, 2001). Research has also suggested gender differences in family influence on adolescent alcohol use. Specifically, male adolescents are less likely than their female counterparts to be monitored by parents, which partially explains male adolescents’ increased likelihood of initiating alcohol use and increased alcohol misuse after initiation (Barnes, Reifman, Farrell, & Dintcheff, 2000).

However, it seems that female adolescent drinking rates are converging with adolescent male drinking rates. While older adolescent males still exhibit higher rates of binge drinking than their female counterparts, this gap is closing. In younger adolescents, females now exhibit rates of drinking, binge drinking, and getting drunk at similar rates to their male counterparts. It is expected that this trend will be followed by increases in alcohol-related health consequences such as unintended pregnancies, sexually transmitted infections, and interpersonal violence (Johnston, O’Malley, Bachman, & Schulenberg, 2009).

**Alcohol Use & Race/Ethnicity**

The development and trajectory of adolescent alcohol use has been shown to vary depending on race and ethnicity. Compared to their ethnic minority counterparts, the odds of developing lifetime alcohol abuse and dependence is higher among White adolescents, and they are more likely to report problems associated with use (Blum et al., 2000; Haberstick et al., 2013; Khan, Cleland, Scheidell, & Berger, 2014; Lo, Cheng, & Howell, 2014). While
research suggests that White and Native Americans have a higher probability of developing alcohol dependence, once dependence has been established, Blacks and Hispanics show higher rates of recurring or persistent dependence than White Americans (Chartier & Caetano, 2010). Furthermore, cultural factors such as acculturation have shown to impact adolescent alcohol use. In Asian American adolescents, increased acculturation is associated with increased risk for alcohol use (Wang-Schweig, Kviz, Altfeld, Miller, & Miller, 2014). A study of bicultural stress and alcohol use in immigrant Hispanic adolescents shows a similar link between ethnicity-related stressors and alcohol use (Oshri et al., 2014). Furthermore, higher levels of discrimination increase odds of developing lifetime alcohol use disorders, while family cohesion serves as a protective factor for both Asian and Hispanic populations (Savage & Mezuk, 2014; Unger, Schwartz, Huh, Soto, & Baezconde-Garbanati, 2014).

**measurement equivalency in alcohol research**

Social factors have been found to influence differing rates of alcohol behaviors across race/ethnicities and between genders, and these factors may affect scale responses on measurement surveys. Although race/ethnic experiences can affect beliefs about alcohol that underlie alcohol behaviors, surveys operate under the assumption that race/ethnic differences do not exist in the way alcohol-related behavior questions are perceived. Similarly, while factors such as gender socialization have been found to influence beliefs about alcohol that underlie alcohol behaviors, the validity of survey responses depends on the assumption that males and females are perceiving the questions similarly.

Past research has underscored the importance of examining these assumptions empirically. One such method for examining the equivalency of measurement across groups is multigroup confirmatory factor analysis (MGCFA). The goal of multigroup confirmatory factory analysis is to statistically test for factorial invariance across multiple groups of interest. Factorial invariance tests examine whether members of different groups make similar associations between measured items and their underlying constructs, or latent variables (Cheung & Rensvold, 1999). Factorial invariance was supported for the Short Inventory of Problems among White non-Hispanic English speaking, Hispanic English speaking, and Hispanic Spanish speaking samples, suggesting that the Short Inventory of Problems is a valid measurement of alcohol treatment effectiveness among Hispanic
populations using either language (Marra, Field, Caetano, & von Sternberg, 2014). In a sample of middle school students, Harrington and colleagues (2011) found evidence for factorial invariance for the Temptations to Try Alcohol Scale across gender, across White versus Black students, and across Hispanic versus non-Hispanic students. Furthermore, factorial invariance was supported for the Alcohol Expectancy Questionnaire Tension-Reduction Subscale across gender (Hittner, 1995). However, measurement invariance could not be established across gender groups for the Protective Behavioral Strategies Scale, suggesting the items of the scale had different meanings for each gender group (Treloar, Martens, & McCarthy, 2014).

These studies underscore the importance of examining measurement equivalence across race/ethnicity and gender groups. Establishing measurement equivalency is a critical step in making valid group comparisons on factors of interest. If measurement equivalency cannot be established, then mean differences found between groups may reflect group differences in question perceptions rather than true group differences. Without establishing measurement equivalency, inferences drawn from these surveys may be based on flawed assumptions. Resources are limited, and since intervention and prevention strategies often derive their information from state and national surveys, it is critical to confirm measurement equivalency across these groups of interest.

**Purpose**

The purpose of the proposed study will be to: (a) evaluate whether group differences exist in endorsement of alcohol-related behaviors administered as part of the Youth Risk Behavior Survey (YRBS), (b) evaluate the psychometric properties and factorial invariance of the questions across race/ethnic groups, and (c) evaluate the psychometric properties and factorial invariance of the questions across males and females. Given past research examining the invariance of other risk-behavior-related surveys across groups, it is hypothesized that the alcohol-related behaviors assessed in the YRBS will not be invariant across race/ethnic groups and will not be invariant across gender. The proposed study will highlight the importance of examining the psychometric properties and factorial invariance of behavior scales in order to make valid group comparisons that will aid in the design of effective prevention and intervention strategies.
METHODS

The proposed study will use data from the school-based national 2013 Youth Risk Behavior Survey (YRBS), sponsored by the Center for Disease Control and Prevention (CDC) and conducted by IFC Macro, Inc., an IFC International Company. The YRBS is part of a larger system called the Youth Risk Behavior Surveillance System (YRBSS), which was designed to: (1) describe the prevalence of health-risk behaviors among youths, (2) assess trends in health-risk behaviors over time, and (3) evaluate and improve health-related policies and programs. The YRBSS was developed to focus almost exclusively on health-risk behaviors. The system includes one-time national surveys, special-population surveys, and ongoing surveys including school-based national, state, tribal, and large urban school district surveys. Additionally, the YRBSS was developed to provide comparable data among subpopulation of youths (e.g., race/ethnic subgroups). The survey is sectioned according to groups of questions used to assess behaviors that result in: (1) unintentional injuries and violence, (2) tobacco use, (3) alcohol and other drug use, (4) sexual behaviors that contribute to unintended pregnancy and sexually transmitted disease, including HIV infection, (5) obesity, overweight, and weight control, (6) dietary behaviors, (7) physical activity, and (8) asthma. The national survey has been conducted biennially since 1991 (CDC, 2013).

PROCEDURE

The national school-based YRBS is conducted under contract with IFC Macro, Inc., an IFC International Company, with CDC oversight. To obtain a nationally representative sample of US students in grades 9-12 in the 50 states and the District of Columbia, the school-based national YRBS uses a three-stage, cluster sample design. The first stage of sampling includes primary sampling units (PSUs) consisting of large-sized counties or groups of smaller, adjacent counties. The PSUs are then divided into sub-PSU units. Schools are then sorted by size and assigned to sub-PSU units. The second stage involves selecting schools from PSUs and classifying them by size (large vs. small), which aids selection decision. At this stage, oversampling techniques are employed to enable separate analysis of data for black and Hispanic students. The final stage of sampling consists of randomly
selecting classes from each chosen school and in each grade 9-12. Weights based on student sex, race/ethnicity, and school grade are applied to adjust for oversampling and student nonresponse.

Local procedures are followed to obtain parental permission before administering the YRBS. Trained data collectors administer the questionnaires to students at the participating schools and record information about schools and classrooms in order to verify sample selection and to weight data. Students complete the self-administered questionnaire directly onto a computer-scannable answer sheet or booklet. Further information regarding study design can be found on the YRBS website (http://www.cdc.gov/yrbs).

**PARTICIPANTS**

For the 2013 national YRBS, 13,633 participants completed the questionnaire. Of these, 50 failed quality control (less than 20 remaining responses after editing for inconsistencies or has the same answer to 15 or more consecutive questions), leaving 13,583 questionnaires. YRBS identified eight race/ethnic groups, including American Indian or Alaska Native (n = 121), Asian (n = 491), black or African American (n = 2,993), Native Hawaiian or other Pacific Islander (n = 135), white (n = 5,449), Hispanic/Latino (n = 1,734), Multiple – Hispanic (n = 1,661), and Multiple – Non-Hispanic (n = 681). Participants that did not report their race/ethnicity (n = 318) will be excluded from analyses. Participants were deemed Multiple – Hispanic if they selected “Yes” to the question “Are you Hispanic or Latino?” and then selected one or more of the race categories (American Indian/Alaska Native, Asian, black/African American, Native Hawaiian/other PI, white). Participants were deemed Multiple – Other if they selected “No” to the question “Are you Hispanic or Latino?” and then selected more than one of the race categories. Due to small sample size, American Indian/Alaska Native, Native Hawaiian or other Pacific Islander, and Multiple – Other will be excluded from analyses. The Multiple – Hispanic category will also excluded due to the possibility of selecting multiple race groups, which would leave race/ethnic categorization for analyses ambiguous. Participants that did not indicate either sex (n = 4) would also be excluded, which will affect the size of the black/African American group (n = 2,991).

After proposed exclusions, the final sample size will consist of 10,663 participants. Of these participants, about half will be male (n = 5,439), and half will be female (n = 5,224).
More than half of the participants \((n = 6,007)\) will 16 years of age or younger, with the rest \((n = 4,648)\) 17 years of age or older.

**Measures**

While the YRBS asks a number of questions that related to alcohol in general, only five questions will be of interest to the proposed study. In order to form the latent variable of “Alcohol-Related Behaviors,” participants were asked the following questions: “During the past 30 days, how many times did you drive a car or other vehicle when you had been drinking alcohol?” “During your life, on how many days have you had at least one drink of alcohol?” “During the past 30 days, on how many days did you have at least one drink of alcohol?” “During the past 30 days, on how many days did you have 5 or more drinks of alcohol in a row, that is, within a couple of hours?” “During the past 30 days, what is the largest number of alcoholic drinks you had in a row, that is, within a couple of hours?”

Participants could respond by selecting one of several answer options that ranged in wording from *I did not/have not (done the behavior in question)* or *0 days* to *All (days in question)*, depending on the wording of the question.

**Data Analysis**

A multigroup confirmatory factor analysis (MGCFA) examined the factorial invariance of the five items across the factor of interest. Invariance was tested through a multi-step process that compared two groups of interest at a time (e.g., Male vs. Female, White vs. Hispanic/Latino). All statistical procedures were done through MPlus statistical software (version 7), using the MLR estimator to estimate all models (Muthén & Muthén, 2012). The MLR estimator is robust to non-normality, uses all available data from participants, and assumes data are missing at random. The MLR estimator also produces the Satorra-Bentler scaled \(\chi^2\) \((S-B\chi^2; Satorra & Bentler, 2001)\), which differs from the usual \(\chi^2\) statistic in that it is a test of model fit when data is deemed nonnormal. Statistical significance of all model parameters was set at a significance level of .05.

The first step established overall model fit and tested the configural invariance model. A one-factor, five item model representing the alcohol-related behaviors latent variable was first specified. The factor was indicated by the five scale measures in the survey that were intended to measured alcohol-related behaviors. Because the alcohol-related behavior latent
variable is comprised of the five scale measures and is inherently not measured itself, the latent variable does not have its own metric. In order to identify the model and give the latent variable “meaning,” the factor variance was fixed to 1 and the factor mean was fixed to 0 at all steps of testing.

Overall model fit will be determined through a S-B$\chi^2$ value for statistical fit and descriptive fit indices for descriptive fit. While the S-B$\chi^2$ value indicates statistical fit, this value is strongly influenced by sample size. Model fit with large sample sizes (as is the case with the YRBS data) are therefore better interpreted through descriptive fit indices. While the S-B$\chi^2$ value will be reported for completeness, a model was instead determined a good fit to the data if at least one of the cut-offs for descriptive fit indices are met. The study used two descriptive fix indices: the (a) comparative fit index (CFI) (Bentler, 1990), and the (b) standardized root mean squared residuals (SRMR) (Bentler, 1995; Jöreskog & Sörbom, 1981). CFI compares the estimated model to an independence (null) model and is normed to the 0 to 1 range. Thus, higher CFI values reflect the relative improvement of the estimated model from the independence (null) model. The SRMR can be considered a measure of “lack of fit.” Conceptually, the SRMR measures the discrepancy between the observed covariance matrix and the model-implied covariance matrix, and therefore smaller SRMR values indicate a closer fitting model (Hu & Bentler, 1998). In this study, the CFI indicated good model fit with values greater than .95 and acceptable model fit with values greater than .90; the SRMR indicated good model fit with values less than .05 and acceptable model fit with values less than .08. These indices were examined at each level of analysis.

In order to be considered configurally invariant, patterns of relations between observed variables and the corresponding “alcohol-related behaviors” latent variable should be similar in valence and statistical significance across groups. This is examined by establishing baseline models for each group of interest. A baseline model allows all parameters to be estimated freely, thus establishing the “best” fitting model, independently, for each group. These parameters include factor loadings of the five scale measures to the “alcohol-related behaviors” latent variable. These factor loadings represent the strength of the relationship between the individual scale measures and their underlying construct (the “alcohol-related behaviors” latent variable). These factor loadings were then examined for significance and valence. If the significance and valence of the factor loadings for each
observed variable on the latent variable are the same across groups, the model can be considered configurally invariant. If the configural invariance can be established, then the next step of the analysis is to test the metric invariance model.

As in the first step, the metric invariance model was examined for overall fit. A metric invariance model constrains factor loadings to equivalency across groups. The metric invariance model simultaneously estimates parameters of both groups under this constraint, and thus is a more “restrictive” version of the previously estimated configural invariance model. This process will produce a null model to which the previously estimated configural invariance model can be compared. Since the metric invariance model can be thought of as a more restrictive version of the configural invariance model, the metric invariance model is considered “nested” within the configural invariance model. In the case of two good-fitting models, a S-B$\chi^2$ difference test ($\Delta$ S-B$\chi^2$) (Satorra, 2000) and an examination of the differences in CFI values ($\Delta$CFI) were consulted to determine whether the configural invariance or the metric invariance model was a better fit to the data. With $\Delta$S-B$\chi^2$, if no significant differences are found, then the more parsimonious model (the more restrictive model, i.e., the metric invariance model) was deemed a better fit to the data. If the metric invariance model is found to be a better fit to the data, it would indicate that the relationship between the measured items and the latent variable do not differ depending on group membership. Furthermore, because $\Delta$S-B$\chi^2$ can be influenced by sample size, $\Delta$CFI values were also examined to determine model fit (Kelloway, 1995), with $\Delta$CFI < .01 indicating no difference between the two models (Chen, 2007; Cheung & Rensvold, 2002). In some cases, the $\Delta$S-B$\chi^2$ could not be computed because doing so resulted in a negative chi-square value. This type of finding could be the result of a number of reasons (Satorra, 2000; Muthén, 2005). While the implication is that there is no significant difference between models, a maximum likelihood chi-square difference test ($\Delta \chi^2$) was computed to confirm that the models were indeed properly specified.

In the case that the metric invariance model was found to not have good fit, the modification indices were consulted to determine which factor loading(s) differed between groups. Freeing the offending parameter(s) and re-estimating the model would result in a revised metric invariance model. A revised metric invariance model would then be tested in which only the identified loading(s) are freely estimated with the rest remaining restrained to
equivalency. The freely estimated loadings would then be reported and interpreted. Additionally, the revised metric invariance model is subjected to a $\Delta S-B\chi^2$ and $\Delta CFI$ values would be examined to determine which model was a better fit to the data.

Normally, if a metric invariance model can be established and found to fit well to the data, then a factor variance covariance invariance model would be tested. A factor variance covariance invariance model constrains factor variance and covariance to equivalency and tests whether or not the scaling of the items is equivalent across groups. However, since the factor variance will be constrained to a value of 1 for identification purposes during the configural and metric invariance testing stages, the factor variance covariance invariance model was not tested in this study. Furthermore, the consensus is that tests of factor loading equivalency (i.e., metric invariance) are the most important prerequisites to measurement equivalency (Meade & Kroustalis, 2006), thus these tests became the focus of the project.
RESULTS

Descriptive statistics and responses of each race/ethnic group and each gender group can be seen in Table 1. Differences in behaviors related to drinking and driving, lifetime alcohol use, 30 day alcohol use, binge drinking, and 30 day maximum number of drinks across the two genders and four race/ethnicity groups were explored through a one-way ANOVA using simple planned contrast comparisons. The White group was used as a comparison group for the three minority groups. Results indicated that all minority groups reported engaging in fewer alcohol-related behaviors compared to the White group, and females reported engaging in fewer alcohol-related behaviors compared to males (all \( p < .05 \)). Across all alcohol-related behaviors, Asians reported the least amount of engagement. Results are presented in Table 2.

The one factor, five-item model representing “Alcohol-Related Behaviors” was specified for all subsequent tests of invariance across gender groups and race/ethnic groups. All parameter estimates reported are unstandardized values. The one-factor alcohol-related behavior model was tested across the entire sample using CFA. The one-factor model did not fit well statistically (S-By\( \chi^2 \)[5, \( N = 10,643 \)] = 491.569, \( p < .001 \)), but did fit well descriptively (CFI = .967; SRMR = .018), indicating a good overall fit to the data. All unstandardized factor loadings were generally large and statistically significant (values ranged from .344 to 1.841). Specific factor loadings are presented in Table 3.

GENDER

The “Alcohol-Related Behaviors” latent variable was first tested across the two gender groups. Baseline models for each group were established in the first step. For males, this one-factor model did not fit well statistically (S-By\( \chi^2 \)[5, \( n = 5,426 \)] = 282.537, \( p < .001 \)) but did fit well descriptively (CFI = .966, SRMR = .017). All unstandardized factor loadings were generally large and statistically significant (values ranged from .433 to 2.006). For the female group, the one-factor model did not fit well statistically (S-By\( \chi^2 \)[5, \( n = 5,217 \)] = 242.558, \( p < .001 \)) but did fit well descriptively (CFI = .962, SRMR = .022). All unstandardized factor loadings were generally large and statistically significant (values
Table 1. Responses to Alcohol-Related Behavior Questions as a Percentage of the Sample by Gender and Race/Ethnicity

<table>
<thead>
<tr>
<th></th>
<th>Male (n = 5,426)</th>
<th>Female (n = 5,217)</th>
<th>Asian (n = 489)</th>
<th>Black/African American (n = 2,978)</th>
<th>White (n = 5,444)</th>
<th>Hispanic/Latino (n = 1,732)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Drinking &amp; driving, 30 days</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I did not drive the past 30 days/ 0 times</td>
<td>85.8</td>
<td>88.5</td>
<td>88.1</td>
<td>86.1</td>
<td>87.7</td>
<td>86.7</td>
</tr>
<tr>
<td>1 to 5 times</td>
<td>35.7</td>
<td>3.3</td>
<td>2.5</td>
<td>2.8</td>
<td>6.0</td>
<td>3.8</td>
</tr>
<tr>
<td>6 or more times</td>
<td>1.3</td>
<td>0.3</td>
<td>1.8</td>
<td>0.7</td>
<td>0.8</td>
<td>0.6</td>
</tr>
<tr>
<td><strong>Lifetime drinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 days</td>
<td>33.5</td>
<td>30.3</td>
<td>47.2</td>
<td>34.2</td>
<td>31.3</td>
<td>25.5</td>
</tr>
<tr>
<td>1 to 9 days</td>
<td>30.2</td>
<td>36.6</td>
<td>34.1</td>
<td>36.3</td>
<td>29.7</td>
<td>39.6</td>
</tr>
<tr>
<td>10 to 99 days</td>
<td>24</td>
<td>24.5</td>
<td>12.2</td>
<td>19.3</td>
<td>27.8</td>
<td>25.1</td>
</tr>
<tr>
<td>100 or more days</td>
<td>8.7</td>
<td>5.6</td>
<td>13.4</td>
<td>5.5</td>
<td>8.6</td>
<td>6.8</td>
</tr>
<tr>
<td><strong>30 days drinking</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 days</td>
<td>59.4</td>
<td>59.1</td>
<td>73</td>
<td>60.1</td>
<td>58.5</td>
<td>56.7</td>
</tr>
<tr>
<td>1 to 19 days</td>
<td>30.0</td>
<td>31.0</td>
<td>19.6</td>
<td>25.6</td>
<td>33.9</td>
<td>32.5</td>
</tr>
<tr>
<td>20 to all 30 days</td>
<td>1.5</td>
<td>0.9</td>
<td>1.0</td>
<td>0.8</td>
<td>1.2</td>
<td>0.9</td>
</tr>
<tr>
<td><strong>Binge drinking, 30 days</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0 days</td>
<td>75.5</td>
<td>80.4</td>
<td>87.5</td>
<td>83.7</td>
<td>74.6</td>
<td>75.5</td>
</tr>
<tr>
<td>1 to 19 days</td>
<td>19.5</td>
<td>17.0</td>
<td>10</td>
<td>11.6</td>
<td>22.3</td>
<td>19.5</td>
</tr>
<tr>
<td>20 or more days</td>
<td>0.9</td>
<td>0.1</td>
<td>0.4</td>
<td>0.3</td>
<td>0.7</td>
<td>0.4</td>
</tr>
<tr>
<td><strong>Maximum # of drinks, 30 days</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I did not drink alcohol during past 30 days</td>
<td>59.9</td>
<td>60.3</td>
<td>74.2</td>
<td>61.2</td>
<td>59</td>
<td>57.6</td>
</tr>
<tr>
<td>1 to 7 drinks</td>
<td>21.7</td>
<td>27.2</td>
<td>16.3</td>
<td>24.4</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>8 to 10 or more drinks</td>
<td>9.4</td>
<td>4.5</td>
<td>3.6</td>
<td>1.9</td>
<td>9.8</td>
<td>7.7</td>
</tr>
</tbody>
</table>
Table 2. Results of One-Way ANOVA Using White Group as Comparison

<table>
<thead>
<tr>
<th>Ethnicity/Race</th>
<th>Drinking and Driving</th>
<th>Lifetime Alcohol Use</th>
<th>30 Day Alcohol Use</th>
<th>Binge Drinking</th>
<th>Max # of Drinks, 30 Days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian vs. White</td>
<td>42.09*</td>
<td>91.60*</td>
<td>44.24*</td>
<td>43.17*</td>
<td>51.52*</td>
</tr>
<tr>
<td>Black/African-American vs. White</td>
<td>53.11*</td>
<td>103.76*</td>
<td>44.64*</td>
<td>146.20*</td>
<td>187.03*</td>
</tr>
<tr>
<td>Hispanic/Latino vs. White</td>
<td>53.044*</td>
<td>4.25**</td>
<td>3.98**</td>
<td>18.24*</td>
<td>8.18**</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td><strong>115.34</strong></td>
<td><strong>8.09</strong>**</td>
<td><strong>15.61</strong></td>
<td><strong>47.00</strong></td>
<td><strong>48.84</strong></td>
</tr>
</tbody>
</table>

Note. All values are F.

*df = 1, 9850, **df = 1, 9852, ***df = 1, 10310, ****df = 1, 10312, *****df = 1, 9697, ******df = 1, 9699, *******df = 1, 10301, ********df = 1, 10305, *********df = 1, 9748, **********df = 1, 9750.

*p < .001; **p < .05

ranged from .219 to 1.638). Because factor loadings for both the male and female groups were all similar in significance and valence, the models were considered configurally invariant.

As the models were found to be configurally invariant, a metric invariance model was tested. The metric invariance model constrained the factor loadings to equivalency across males and females. The metric invariance model did not fit well statistically (S-Bχ²[19, N = 10,643] = 759.891, p < .001) but was an acceptable fit descriptively (CFI = .949, SRMR = .115). While the metric invariance model and configural invariance model could not be compared statistically (ΔS-Bχ²[9] = -28,782.845), the result of a maximum likelihood chi-square difference test (Δχ²) confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, Δχ²(9) = 488.807, p < .001, and descriptively, ΔCFI = .015. An examination of the modification indices revealed that “drinking and driving,” “binge drinking,” and “30 day alcohol use” factor loadings should be freely estimated by gender group. A revised metric invariance model was then reestimated with those three parameters freed. The revised metric invariance model was a poor fit statistically (S-Bχ²[16, N = 10,643] = 788.552, p < .001), but an acceptable fit descriptively
Table 3. Unstandardized Factor Loadings of Baseline Models

<table>
<thead>
<tr>
<th>Observed Variable</th>
<th>Drinking and driving</th>
<th>Lifetime alcohol use</th>
<th>30 day alcohol use</th>
<th>Binge drinking</th>
<th>Maximum # of drinks, 30 days</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Entire Sample</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(N = 10,643)</td>
<td>.344</td>
<td>1.409</td>
<td>1.035</td>
<td>.989</td>
<td>1.841</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males ((n = 5,426))</td>
<td>.433</td>
<td>1.513</td>
<td>1.134</td>
<td>1.107</td>
<td>2.006</td>
</tr>
<tr>
<td>Females ((n = 5,217))</td>
<td>.219</td>
<td>1.286</td>
<td>.921</td>
<td>.847</td>
<td>1.638</td>
</tr>
<tr>
<td><strong>Female Race/Ethnic Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White ((n = 2,602))</td>
<td>.260</td>
<td>1.398</td>
<td>.995</td>
<td>.976</td>
<td>1.802</td>
</tr>
<tr>
<td>Asian ((n = 256))</td>
<td>.267</td>
<td>1.087</td>
<td>.738</td>
<td>.591</td>
<td>1.339</td>
</tr>
<tr>
<td>Black/AA ((n = 1,492))</td>
<td>.140</td>
<td>1.096</td>
<td>.855</td>
<td>.617</td>
<td>1.238</td>
</tr>
<tr>
<td>Hispanic/Latino ((n = 867))</td>
<td>.143</td>
<td>1.185</td>
<td>.863</td>
<td>.793</td>
<td>1.702</td>
</tr>
<tr>
<td><strong>Male Race/Ethnic Group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White ((n = 2,842))</td>
<td>.471</td>
<td>1.597</td>
<td>1.209</td>
<td>1.229</td>
<td>2.239</td>
</tr>
<tr>
<td>Asian ((n = 233))</td>
<td>.522</td>
<td>1.264</td>
<td>.951</td>
<td>.901</td>
<td>1.665</td>
</tr>
<tr>
<td>Black/AA ((n = 1,486))</td>
<td>.326</td>
<td>1.331</td>
<td>1.026</td>
<td>.815</td>
<td>1.349</td>
</tr>
<tr>
<td>Hispanic/Latino ((n = 865))</td>
<td>.420</td>
<td>1.473</td>
<td>1.111</td>
<td>1.081</td>
<td>2.020</td>
</tr>
</tbody>
</table>

\(\text{CFI} = .947, \text{SRMR} = .075\). The revised metric invariance model differed from the configural invariance model statistically, \(\Delta S - B \chi^2(6) = 355.140, p < .001\), and descriptively, \(\Delta \text{CFI} = .017\). Modification indices were consulted again and indicated that the “maximum # of drinks, 30 days” item should be freely estimated. The resulting revised metric invariance
model with the four items freed was a poor fit statistically \((S-B\chi^2; 15, N = 10,643) = 729.113, p < .001\), but an acceptable fit descriptively \((\text{CFI} = .951, \text{SRMR} = .045)\). However, the revised metric invariance model still differed from the configural invariance model statistically, \(\Delta S-B\chi^2; (5) = 165.5625, p < .001\), and descriptively, \(\Delta \text{CFI} = .013\). While the difference in CFI values was marginal, freeing four out of five factor loadings reduces the practical utility of the construct. Thus, metric invariance was not achieved across males and females. Males loaded more strongly on the “drinking and driving” (.431), “binge drinking” (1.082), “30 day alcohol use” (1.102), and “maximum # of drinks, 30 days” (1.960) items compared to females (.224; .870; .949; 1.683). Since metric invariance could not be established across males and females, all subsequent tests of invariance across race/ethnicity were conducted within each gender group.

**Race/Ethnicity**

Since metric invariance could not be established across males and females, all subsequent tests of invariance across race/ethnicity were conducted within each gender group.

**Men**

The “Alcohol-Related Behaviors” latent variable was then tested for invariance across the race/ethnic groups within the male group. Baseline models were first established for each race/ethnic group, independently. For the White male group, the one-factor model did not fit well statistically \((S-B\chi^2; 5, n = 5,426) = 177.515, p < .001\) but did fit well descriptively \((\text{CFI} = .966, \text{SRMR} = .018)\). All unstandardized factor loadings were generally large and statistically significant (values ranged from .471 to 2.239). For the Asian male group, the one-factor model did not fit well statistically \((S-B\chi^2; 5, n = 233) = 22.488, p < .001\) but did fit well descriptively \((\text{CFI} = .953, \text{SRMR} = .020)\). All unstandardized factor loadings were generally large and statistically significant (values ranged from .522 to 1.665). For the Black/African American male group, the one-factor model did not fit well statistically \((S-B\chi^2; 5, n = 1,486) = 54.364, p < .001\) but did fit well descriptively \((\text{CFI} = .961, \text{SRMR} = .023)\). All unstandardized factor loadings were generally large and statistically significant (values ranged from .326 to 1.349). Finally, for the Hispanic/Latino male group, the one-factor
model also did not fit well statistically ($S-B\chi^2 [5, n = 865] = 35.937, p < .001$) but did fit well descriptively (CFI = .979, SRMR = .014). All unstandardized factor loadings were generally large and statistically significant (values ranged from .420 to 2.020). Specific factor loadings are presented in Table 2. Because factor loadings across race/ethnic groups were all similar in significance and valence, the models were considered configurally invariant.

As the models were found to be configurally invariant, metric invariance was then tested across pairs of male race/ethnic groups. Invariance was first tested across the White and Asian groups. The metric invariance model did not fit well statistically ($S-B\chi^2 [19, n = 3,075] = 177.803, p < .001$) but was an acceptable fit descriptively (CFI = .969, SRMR = .080). While the revised metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2 [9] = 204.7241$), the result of a maximum likelihood chi-square difference test ($\Delta\chi^2$) confirmed correct model specification. While the metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2 (9) = 59.593, p < .001$, it did not differ descriptively, $\Delta$CFI = .009. Thus, the metric invariance model was deemed a better fit to the data than the configural invariance model.

Invariance was then tested across the White and Black/African American male groups. The metric invariance model did not fit well statistically ($S-B\chi^2 [19, n = 4,328] = 431.222, p < .001$) but was an acceptable fit descriptively (CFI = .929, SRMR = .171). The metric invariance model differed from the configural invariance model statistically, $\Delta S-B\chi^2 [9] = 163.240, p < .001$, and descriptively, $\Delta$CFI = .034. An examination of the modification indices revealed that the “maximum # of drinks, 30 days,” “30 day alcohol use,” and “binge drinking” factor loadings should be freely estimated. The revised metric invariance model was a poor fit statistically ($S-B\chi^2 [16, n = 4,328] = 298.618, p < .001$), but an acceptable fit descriptively (CFI = .951, SRMR = .074). The revised metric invariance model differed from the configural invariance model statistically, $\Delta S-B\chi^2 (6) = 222.067, p < .001$, and descriptively, $\Delta$CFI = .012. Although the descriptive difference was marginal, freely estimating three out of five factor loadings decreases the practical utility of the construct. Thus, metric invariance could not be established across the White and Black/African American male groups. The White male group loaded more strongly on the “maximum # of drinks, 30 days” (2.207), “30 day alcohol use” (1.160), and “binge drinking” (1.193) items compared to the Black/African American male group (1.437; 1.098; .871).
Next, invariance was tested across the White and Hispanic/Latino male groups. The metric invariance model did not fit well statistically ($S$-$B\chi^2 [19, n = 3,707] = 211.872, p < .001) but was a good fit descriptively (CFI = .971, SRMR = .044). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S$-$B\chi^2[9] = -2667.2440$), the $\Delta\chi^2$ results confirmed correct model specification. While the metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(9) = 167.999, p < .001$, it did not differ descriptively, $\Delta$CFI = .001. Therefore, the metric invariance model was deemed a better fit to the data than the configural invariance model.

Invariance was then tested across the Asian and Black/African American male groups. The metric invariance model did not fit well statistically ($S$-$B\chi^2 [19, n = 1,719] = 104.184, p < .001) but was a good fit descriptively (CFI = .952, SRMR = .044). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S$-$B\chi^2[9] = -75.3865$), the $\Delta\chi^2$ confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(9) = 162.486, p < .001$, but did not differ descriptively, $\Delta$CFI = .005. Thus, the metric invariance model was deemed a better fit to the data than the configural invariance model.

Next, invariance was tested across the Asian and Hispanic/Latino male groups. The metric invariance model did not fit well statistically ($S$-$B\chi^2 [19, n = 1,098] = 59.475, p < .001) but was a good fit descriptively (CFI = .976, SRMR = .071). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S$-$B\chi^2[9] = -62.4601$), the $\Delta\chi^2$ confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(9) = 64.356, p < .001$, and descriptively, $\Delta$CFI = .010, although the difference was marginal. Thus, the metric invariance model was deemed a better fit to the data than the configural invariance model.

Finally, invariance was tested across the Black/African American and Hispanic/Latino male groups. The metric invariance model did not fit well statistically ($S$-$B\chi^2 [19, n = 2,351] = 187.086, p < .001) but had acceptable fit descriptively (CFI = .939, SRMR = .099). The metric invariance model differed from the configural invariance model statistically, $\Delta S$-$B\chi^2(9) = 907.324, p < .001$, and descriptively, $\Delta$CFI = .031. An examination of the modification indices revealed that the “maximum # of drinks, 30 days,” “30 day
alcohol use,” and “binge drinking” factor loadings should be freely estimated. The revised metric invariance model was a poor fit statistically ($S-B\chi^2 [16, n = 2,351] = 130.650, p < .001$), but a good fit descriptively (CFI = .958, SRMR = .049). The revised metric invariance model differed from the configural invariance model statistically, $\Delta S-B\chi^2(6) = 67.4058, p < .001$, and descriptively, $\Delta CFI = .012$. While the descriptive difference was marginal, freeing three out of five factor loadings decreases the practical utility of the construct. Thus, the configural invariance model was deemed a better fit to the data than the metric invariance model. The Hispanic/Latino male group loaded more strongly on the “maximum # of drinks, 30 days” (1.988), “30 day alcohol use” (1.065), and “binge drinking” (1.049) items compared to the Black/African American male group (1.375; 1.046; .831).

Women

As with men, the “Alcohol-Related Behaviors” latent variable was tested for invariance across the race/ethnic groups within the female group. Again, baseline models were first established for each race/ethnic group, independently. For the White female group, the one-factor model did not fit well statistically ($S-B\chi^2 [5, n = 2,602] = 116.768, p < .001$) but did fit well descriptively (CFI = .972, SRMR = .017). All unstandardized factor loadings were generally large and statistically significant (values ranged from .260 to 1.802). For the Asian female group, the one-factor model did not fit well statistically ($S-B\chi^2 [5, n = 256] = 54.408, p < .001$) but was an acceptable fit descriptively (CFI = .836, SRMR = .034). All unstandardized factor loadings were generally large and statistically significant (values ranged from .267 to 1.339). For the Black/African American female group, the one-factor model did not fit well statistically ($S-B\chi^2 [5, n = 1,492] = 75.564, p < .001$) but did fit well descriptively (CFI = .943, SRMR = .029). All unstandardized factor loadings were generally large and statistically significant (values ranged from .140 to 1.238). Finally, for the Hispanic/Latina female group, the one-factor model also did not fit well statistically ($S-B\chi^2 [5, n = 867] = 47.367, p < .001$) but did fit well descriptively (CFI = .949, SRMR = .029). All unstandardized factor loadings were generally large and statistically significant (values ranged from .143 to 1.702). Specific factor loadings are presented in Table 2. Because factor loadings across race/ethnic groups were all similar in significance and valence, the models were considered configurally invariant.
As the models were found to be configurally invariant, metric invariance was then tested across pairs of female race/ethnic groups. Invariance was first tested across the White and Asian groups. The metric invariance model did not fit well statistically ($S-B\chi^2[19, n = 2,858] = 125.212, p < .001$) but was an acceptable fit descriptively (CFI = .975, SRMR = .121). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2[9] = -15.7708$), $\Delta\chi^2$ results confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(9) = 127.327, p < .001$, and descriptively, $\Delta$CFI = .071. The modification indices of the metric invariance model were consulted, and indicated that the “binge drinking” factor loading should be freely estimated. The revised metric invariance model was a poor fit statistically ($S-B\chi^2[18, n = 2,858] = 129.397, p < .001$) but was an acceptable fit descriptively (CFI = .974, SRMR = .084). The revised metric model still could not be compared statistically to the configural invariance model ($\Delta S-B\chi^2[8] = -37.3867$), but $\Delta\chi^2$ results again confirmed correct model specification. The revised metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(8) = 129.397, p < .001$, and still differed descriptively, $\Delta$CFI = .070. As the modification indices could not detect any additional parameters to freely estimate in order to improve model fit, the configural invariance model was deemed a better fit to the data than the revised metric invariance model. The White female group loaded more strongly on the “binge drinking” item (.957) compared to the Asian female group (.742).

Invariance was then tested across the White and Black/African American female groups. The metric invariance model did not fit well statistically ($S-B\chi^2[19, n = 4,094] = 320.629, p < .001$) but was an acceptable fit descriptively (CFI = .940, SRMR = .155). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2[9] = -325.6007$), $\Delta\chi^2$ results confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(9) = 338.267, p < .001$, and descriptively, $\Delta$CFI = .017. An examination of the modification indices revealed that the “binge drinking” and “maximum # of drinks, 30 days” factor loadings should be freely estimated. The revised metric invariance model was a poor fit statistically ($S-B\chi^2[16, n = 4,094] = 248.266, p < .001$), but a good fit descriptively (CFI =
The revised metric invariance model differed from the configural invariance model statistically, $\Delta S-B\chi^2(7) = 368.3523, p < .001$, but did not differ descriptively, $\Delta CFI = .003$. The revised metric invariance model was deemed a better fit to the data than the configural invariance model. The White female group loaded more strongly on the “binge drinking” (.943) and “maximum # of drinks, 30 days” (1.745) items compared to the Black/African American female group (.663; 1.328).

Next, invariance was tested across the White and Hispanic/Latino female groups. The metric invariance model did not fit well statistically ($S-B\chi^2[19, n = 3,469] = 212.927, p < .001$) but was a good fit descriptively (CFI = .959, SRMR = .068). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2[9] = -770.4759$), the $\Delta \chi^2$ results confirmed correct model specification. While the metric invariance model differed from the configural invariance model statistically, $\Delta \chi^2(9) = 114.225, p < .001$, it did not differ descriptively, $\Delta CFI = .001$. Therefore, the metric invariance model was deemed a better fit to the data than the configural invariance model.

Invariance was then tested across the Asian and Black/African American female groups. The metric invariance model did not fit well statistically ($S-B\chi^2[19, n = 1,748] = 97.955, p < .001$) but was a good fit descriptively (CFI = .952, SRMR = .059). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2[9] = -35.0929$), the $\Delta \chi^2$ confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, $\Delta \chi^2(9) = 161.731, p < .001$, and descriptively, $\Delta CFI = .062$. The modification indices of the metric invariance model indicated that the “30 day alcohol use” item should be freed. The revised metric invariance model did not fit well statistically ($S-B\chi^2[18, n = 1,748] = 96.559, p < .001$) but was a good fit descriptively (CFI = .953, SRMR = .054). While the revised metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2[8] = -46.1744$), the $\Delta \chi^2$ results confirmed correct model specification. The revised metric invariance model differed from the configural invariance model statistically, $\Delta \chi^2(8) = 56.881, p < .001$, and descriptively, $\Delta CFI = .063$. As with the White and Asian female group comparison, the modification indices did not detect any additional parameters that should be freed to improve model fit. Therefore, the configural invariance model was deemed a better
fit to the data than the revised metric invariance model. The Black/African American female group loaded more strongly on the “30 day alcohol use” item (.864) compared to the Asian female group (.726).

Next, invariance was tested across the Asian and Hispanic/Latino female groups. The metric invariance model did not fit well statistically ($S-B\chi^2 [19, n = 3,469] = 212.927, p < .001$) but was a good fit descriptively (CFI = .959, SRMR = .068). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2 [9] = -21.4400$), the $\Delta\chi^2$ confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(9) = 67.492, p < .001$, and descriptively, $\Delta\text{CFI} = .060$. The modification indices of the metric invariance model did not detect any parameters that should be freed to improve model fit. Thus, the configural invariance model was deemed a better fit to the data than the metric invariance model.

Finally, invariance was tested across the Black/African American and Hispanic/Latina female groups. The metric invariance model did not fit well statistically ($S-B\chi^2 [19, n = 2,359] = 177.404, p < .001$) but had good fit descriptively (CFI = .926, SRMR = .078). While the metric invariance model and configural invariance model could not be compared statistically ($\Delta S-B\chi^2 [9] = -308.5984$), the $\Delta\chi^2$ confirmed correct model specification. The metric invariance model differed from the configural invariance model statistically, $\Delta\chi^2(9) = 129.413, p < .001$, and descriptively, $\Delta\text{CFI} = .020$. The modification indices of the metric invariance model indicated that the “maximum # of drinks, 30 days” item should be freely estimated. The revised metric invariance did not fit well statistically ($S-B\chi^2 [18, n = 2,359] = 161.242, p < .001$) but had good fit descriptively (CFI = .933, SRMR = .062). The revised metric invariance model differed from the configural invariance model statistically, $\Delta S-B\chi^2 [8] = 568.6297$, and descriptively, $\Delta\text{CFI} = .013$, although the difference was marginal. Therefore, the revised metric invariance model was deemed a better fit to the data than the configural invariance model. The Hispanic/Latina female group loaded more strongly on the “maximum # of drinks, 30 days” item (1.629) compared to the Black/African American female group (1.282).
DISCUSSION

This study examined the factorial invariance of the 2013 Youth Risk Behavior Survey across gender and race/ethnicity groups. Results indicated that factorial invariance could not be achieved across the two factors of interest. Across gender groups, the one-factor five-item latent variable of “Alcohol-Related Behaviors” achieved configural invariance, indicating that the same latent construct existed for both groups. However, four out of five parameters had to be freed in order to achieve a revised metric invariance model that only marginally differed (descriptively) from the configural invariance model. Freeing the majority of parameters from equivalency reduces the practical utility of the latent variable. In the case of male versus female comparisons, only “lifetime alcohol use” emerged as an indicator that did not differ between groups. The four other items were stronger indicators of the latent variable for males than for females, which suggests that these items are more representative of engaging in “Alcohol-Related Behaviors” for males than for females. While configural invariance could be established across race/ethnic groups when looking within gender, metric invariance could only be established in a few cases.

For males, metric invariance was achieved between the White and Asian groups, White and Hispanic/Latino groups, Asian and Black/African American groups, and finally, the Asian and Hispanic/Latino groups. The same three items were problematic in the two pairings that did not achieve metric invariance. This indicates that the “maximum # of drinks, 30 days,” “30 day alcohol use,” and “binge drinking” items were more representative of “alcohol-related behaviors” for White and Hispanic/Latino adolescent males compared to their Black/African American counterparts.

For females, the results were more varied. Metric invariance was achieved between White and Hispanic/Latina groups, White and Black/African American groups, and Black/African American and Hispanic/Latina groups. While metric invariance was established in these groups, it is worth noting that for the latter two, the metric invariance model had to be revised before the metric invariance model could be deemed a better fit to the data than the configural invariance model. The “binge drinking” and “maximum # of drinks, 30 days” items were more representative for White adolescent females compared to
their Black/African American counterparts, and “maximum # of drinks, 30 days” was more representative for Hispanic/Latina adolescent females compared to their Black/African American peers.

Particularly problematic within the female group was the Asian group, as metric invariance could not be achieved in any tests. One possibility for this finding is that estimating the baseline model for the Asian female group resulted in a relatively lower obtained CFI value. Since ΔCFI was used in nested model comparisons, the Asian female group was at a “disadvantage.” Further, modification indices were only able to identify one parameter to be freed in order to improve model fit in comparisons with White and Black/African American females. When the metric invariance model was tested with Hispanic/Latina females, no parameters were indicated. As this finding could be due to a number of other factors (e.g., correlated error terms, sample size), interpretations of the results are difficult.

These findings have implications for how researchers use data on latent variable analyses. The lack of metric invariance across multiple groups suggests that when using the 2013 YRBS, caution should be taken when making comparisons on latent variable group means. While lack of metric invariance is problematic for all affected cases, of particular concern is the lack of metric invariance across gender groups. This suggests that when using YRBS alcohol-related behaviors data, male and female data should not be collapsed, as measurement equivalency is lacking. Essentially, while the same latent variable exists for males and females, the degree to which individual measured items represent that latent variable significantly differ.

Within gender groups, of concern is the lack of metric invariance for tests done with the Black/African American adolescent male and Asian adolescent female groups. As mentioned before, the lack of metric invariance in tests with the Asian adolescent female group could be attributed to a number of reasons. One possible reason may be that the one-factor, five-item model may is not the best-fitting model for this group. When considering the Black/African American male adolescent group, it is worth noting that items related to quantity of heavy drinking (“maximum # of drinks, 30 days,” “binge drinking”) and consistency of drinking (“30 day alcohol use”) were less representative of the latent variable compared to White and Hispanic/Latino adolescent males. Indeed, this general pattern
emerged in cases that the metric invariance model was revised to improve model fit (e.g.,
between White and Black/African American females).

This pattern of findings has implications for intervention and prevention efforts aimed
at reducing adolescent alcohol use. Specifically, alcohol-related behaviors should be
examined at the within-group level, that is, within the race/ethnic group of a specific gender.
Without doing so, prevention and intervention efforts may be misguided in both (1) selecting
the groups that are deemed the “highest risk” and (2) selecting the most “problematic”
alcohol-related behaviors. As these efforts are often costly in time and money, it is justified
to tailor efforts to specific groups—that is, to narrow the scope of intervention and
prevention efforts for the sake of efficiency and efficacy. The one case that may be an
exception is of White and Hispanic/Latino groups, although metric invariance could only be
established at the level of each gender. Further research is needed to examine the consistency
of measurement equivalency in other alcohol-related measurements.

Although the results of this project contribute to the literature on the importance of
empirically examining measurement equivalency, this project also had several limitations.
Although the data came from a large nationally representative sample of adolescents, the
sample sizes of certain race/ethnic groups precluded them from inclusion in the analyses.
Therefore, results may not generalize to the population at large. This is particularly important
when considering the groups that were excluded from analyses, such as Native Alaskan and
American Indians. Previous research has indicated that adolescents from these groups are at
high risk for developing alcohol-related problems, yet they remain underserved (Walls,
Whitbeck, Hoyt, & Johnson, 2007). Future research should increase efforts in including these
groups in analyses. In addition, even within the included race/ethnic groups, differences in
sample size may have influenced the power of the tests to detect changes in factorial
invariance.
REFERENCES


