Published in final edited form as:

J Ethn Cult Divers Soc Work. 2015 ; 24(2): 109–129. doi:10.1080/15313204.2014.977985.

Measurement Invariance of the Short Inventory of Problems-Revised Across African American and Non-Latino White Substance Users

FRANK R. DILLON, KAREN WHITEMAN, and RUI DUAN

Robert Stempel College of Public Health and Social Work, Florida International University, Miami, FL, USA

Abstract

This study investigated measurement invariance properties of the Short Inventory of Problems – Revised (SIP-R) across racial groups. The sample included 195 African American and 194 non-Latino White adult participants in a clinical trial investigating the effectiveness of motivational enhancement therapy in the National Institute on Drug Abuse Clinical Trials Network. The SIP-R demonstrated configural invariance and weak metric invariance, suggesting conceptualizations of adverse consequences of substance use are equivalent across racial groups. The SIP-R also indicated partial strong/scalar and strict metric invariance, suggesting a need for continued research of SIP-R items to ensure valid measurement and outcomes across racial groups.

Keywords

measurement invariance; racial minorities; drug treatment

INTRODUCTION

Many measures used in substance abuse treatment research are susceptible to measurement variance because this type of analysis is frequently overlooked by researchers (Randolph, Gerend, & Miller, 2006). The continued use of measures with different conceptual meaning across racial groups may render invalid analyses comparing such groups (Burlew et al., 2009). Conclusions drawn from invalid findings can lead to ineffective treatments and policy initiatives targeting substance use disorders (Ramirez, Ford, Steward, & Teresi, 2005). Burlew et al. (2009) and Feaster et al. (2010) emphasized that any analysis of comparative effectiveness of substance abuse treatments across racial and ethnic groups should first ensure that measures used to assess treatment effectiveness have the same conceptual meaning or measurement invariance across groups. Thus, the current secondary

Address correspondence to Dr. Frank Dillon, Division of Counseling Psychology, University at Albany – State University of New York, 1400 Washington Avenue, ED 240, Albany, NY 12222; fdillon@albany.edu.
Frank R Dillon, PhD, is now Associate Professor, Division of Counseling Psychology, University at Albany – State University of New York, 1400 Washington Avenue, ED 240, Albany, NY 12222; fdillon@albany.edu
Karen Whiteman, PhD now teaches at the School of Social Work, Florida Atlantic University, 777 Glades Rd., Boca Raton, FL 33431, karen@researchgerontology.com
Rui Duan, MPH, is Senior Research Analyst, Department of Public Health Sciences, University of Miami - Miller School of Medicine, Room 1064, 1120 N.W. 14th Street, Miami, FL33136; rduan@med.miami.edu

analysis study aims to determine measurement invariance properties of a widely-used clinical assessment of the negative consequences of substance use (the Short Inventory of Problems–Revised or SIP-R; Kiluk, Dreifuss, Weiss, Morgenstern, & Carroll, 2012) across African American and non-Latino White participants in a randomized clinical trial in the National Institute on Drug Abuse Clinical Trials Network (NIDA-CTN).

LITERATURE REVIEW

Measurement variance or non-equivalence of an instrument is introduced when groups of participants experience or conceptualize a construct differently (Meredith, 1993; Vandenberg & Lance, 2000; Widaman & Reise, 1997). Determining measurement invariance of an instrument allows researchers to assess whether the construct and scores of a measure are similarly comprehended and measured across salient participant groups (e.g., based on race, ethnicity, gender, age, etc.). Measurement variance across cultural groups is theorized to result from (a) differences in participants' environments to engage in certain behaviors or develop beliefs due to contextual differences, racism, or discrimination; (b) the language of the measure; or (c) cultural differences in norms and relevance of the constructs being assessed (Burlew, Feaster, Brecht, & Hubbard, 2009). In addition, participants from different cultural groups may respond to item content in dissimilar ways causing variable construct validity across groups (Allen & Walsh, 2000). For instance, Crockett, Randall, Shen, Russell, and Driscoll (2005) reported that the factor solution of the *Center for Epidemiologic Studies Depression Scale* (Radloff, 1977) pertained to non-Latino White American adolescents in a national sample, but was not appropriate for Cuban American or Puerto Rican youth. Similarly, Chan Tran, and Nguyen (2012) determined that only some items from the *Williams Everyday Discrimination Scale* (Williams, Yu, Jackson, & Anderson, 1997) were invariant across Vietnamese-American and Chinese-American adults.

Measurement invariance estimates are more contemporary psychometric properties that expand on classical test theory and are important for further validating measures across diverse groups (Burlew et al., 2009; Meredith, 1993; Vandenberg & Lance, 2000; Widaman & Reise, 1997). Researchers often assume that adequate reliability estimates among groups are sufficient to obtain invariant measurement of a construct between groups. However, a measure may appear equally reliable in two populations yet provide a misleading basis for comparing those populations because of measurement variance (Bingenheimer, Raudenbush, Leventhal, & Brooks-Gunn, 2005). If not accounted for, violations of measurement invariance assumptions are as threatening to substantive interpretations as a lack of evidence of reliability or validity of a measure (Vandenberg & Lance, 2000). For instance, when measurement variance occurs, it is unclear whether mean differences in the construct of interest reflect differences in factor structure, differences in levels of item endorsement, or both. As a result, findings from standard statistical analyses are difficult to interpret.

Development of the Short Inventory of Problems–Revised

Drug and alcohol frequency and quantity rates alone are incomplete predictors of substance use disorders (Bender, Griffin, Gallop, & Weiss, 2007). A more comprehensive assessment includes a measurement of adverse consequences of using substances. Assessing negative

consequences of substance use is congruent with diagnostic criteria for substance use disorders according to the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV-TR; American Psychiatric Association, 2000). Measuring problems caused by substance use also is important in social work to target clients' motivation for changing their substance-related behaviors through increasing problem awareness and to facilitate retention in treatment (Kiluk et al., 2012; Miller & Rollnick, 2002). Thus, assessment of substance-related consequences is increasingly regarded as an outcome measure in drug treatment clinical trials (e.g., Ball et al., 2007; Carroll et al., 2006, 2009; Winhusen et al., 2008).

The Drinker Inventory of Consequences was the first standardized measure of the negative consequences of substance use (DrInC; Miller, Tonigan, & Longabaugh, 1995). The DrInC is a 50-item measure designed to assess alcohol-related consequences based on five domains: (a) impulse control, (b) interpersonal, (c) intrapersonal, (d) physical, and (e) social. The DrInC was later adapted to develop the Inventory of Drug Use Consequences (InDUC; Tonigan & Miller, 2002) by simply modifying the wording of items to include alcohol and illicit drug use (Tonigan & Miller, 2002). The same theorized 5-factor structure of the DrInC was suggested by the InDUC.

To enhance the clinical utility of these measures, a brief 15-item version was developed: the Short Inventory of Problems (SIP; Blanchard, Morgenstern, Morgan, Labouvie, & Bux, 2003). Multiple psychometric studies describe consistent reliability estimates for the SIP across various clinical and non-treatment seeking samples including primary care patients, veterans, college students, persons with co-occurring substance use and mental health disorders, problem drinkers, and men who have sex with men (Allensworth-Davies, Cheng, Smith, Samet, & Saiz, 2012; Alterman, Cacciola, Ivey, Habing, & Lynch, 2009; Bender et al., 2007; Hagman et al., 2009; Feinn, Tennen, & Kranzler, 2003; Gillespie, Holt, & Blackwell, 2007). However, these same studies report equivocal construct validity findings.

Recently, a slightly revised version of the SIP (SIP-R) was implemented as an outcome measure in four multisite randomized clinical trials in the NIDA-CTN (Kiluk et al., 2012). These studies investigated the effectiveness of (a) motivational enhancement therapy (MET; Ball et al., 2007), (b) motivational interviewing (MI) techniques (Carroll et al., 2006), (c) MET for pregnant substance users (Winhusen et al., 2008), and (d) MET for Spanish-speaking clients (Carroll et al., 2009). The SIP-R purports the same 5-factor structure of its predecessors, but includes two more items to broaden assessment of problems with work and legal trouble in the social domain, resulting in 17-items. The two items are: I have gotten into trouble because of drinking or drug use and The quality of my work has suffered because of my drinking or drug use. Another minor modification involved replacing one of the impulse control domain items, I have had an accident while drinking or intoxicated, with Drinking or using one drug has caused me to use other drugs more.

To date, only one study has evaluated the psychometric properties of the SIP-R. Kiluk et al. (2012) described evidence of reliability and validity of the SIP-R in a sample consisting of participants from the previously noted MET (Ball et al., 2007) and MI (Carroll et al., 2006) effectiveness studies. Findings suggested that the hypothesized 5-factor model yielded the best model fit to the data, supporting its construct validity. The SIP-R total score also

evidenced internal consistency, but reliability estimates were not reported for the 5-factors. Convergent and discriminant validity also were evidenced across the 5-factors and total scores. Finally, predictive validity was indicated by strong associations between higher SIP-R scores (i.e., more adverse consequences) and poorer treatment retention.

Although initial psychometric properties support the valid use of the SIP-R as a measure of consequences of substance use among treatment-seeking drug and alcohol users, measurement invariance analyses were omitted from the SIP-R psychometric study. Findings of measurement variance across salient participant groups could potentially threaten the validity of SIP-R outcomes for population subgroups in future substance use treatment outcome research. Measurement variance also may explain equivocal findings concerning the underlying factor structure of the SIP in past studies. Cultural differences may influence African American persons to perceive and experience adverse consequences of substance use differently than their White, European American counterparts (Burlew et al., 2009). Publically available data from the NIDA-CTN MET effectiveness trial (Ball et al., 2007) contains a large and sufficiently diverse sample for the current study to determine measurement equivalence estimates of the SIP-R across African American and non-Latino White participants.

The Present Study

The current secondary analysis study seeks to determine whether African American and non-Latino White adult outpatient participants share a common understanding of the constructs measured by the SIP-R (i.e., evidence measurement invariance). First, measurement models of the SIP-R factor structure will be tested with each racial group to ensure adequate fit of each model for subsequent invariance testing. If the measurement models are found to be appropriate for each group, *configural invariance* will be explored next. Configural invariance indicates that the measure represents similar constructs across the groups (Meredith, 1993; Vandenberg & Lance, 2000; Steinmetz, Schmidt, Tina-Booh, Wiecek, & Schwartz, 2009). That is, the same basic factor structure holds across groups. Configural invariance will be tested by examining the posited measurement model across groups, with no impositions of equality constraints. Second, if evidence of configural invariance is found, *item or metric invariance* will be tested to determine whether SIP-R items have the same meaning and scale across racial groups. Item equivalence will be assessed across *weak*, *strong* (or *scalar*), and *strict* levels as described by Meredith (1993) and Vandenberg and Lance (2000). *Weak metric invariance* will be tested by constraining factor loadings to equality across groups. This type of metric invariance assesses the equality of factor loadings (i.e., the regression weights from item responses to proposed latent factors). This type of metric invariance suggests that the scales of the latent factors are the same across the groups being compared. Therefore, a unit increase in one group implies an equal change in the underlying factor for the other group. (Feaster et al., 2010). Methodologists have noted that weak metric invariance is not sufficient to make comparisons of levels of the factor across groups because it does not indicate that the zero point on the scales are the same across groups (Chen, Sousa, & West, 2005; Feaster et al., 2010). Hence, if weak invariance is established, strong or *scalar invariance* will be tested. It is established by constraining intercepts to be equal across groups. Strong or scalar

invariance indicates that the intercepts of the regression equations for each item on its intended factor is invariant across groups. Since the item intercept is the mean value of an item when the latent factor is equal to zero, when item intercepts are equal then the zero point on the underlying factor is the same across groups (Feaster et al., 2010). If the item intercepts differ across groups, it suggests there are different normative levels of the item across groups. In other words, different groups need to report different levels of the item to yield the same underlying level of the latent factor. Finally, if strong or scalar invariance is established, strict metric invariance will be examined. The test for strict invariance constrains the item residual variances to be equal across groups. An absence of equal item residual variances across groups causes standardized loadings and factor reliability to differ between groups even when other forms of invariance hold (Burlew et al., 2009).

METHODS

Procedures

This secondary analysis study used baseline data from a clinical trial conducted in the NIDA-CTN via the Clinical Trials Network Data Share web site (www.ctndatashare.org). The clinical trial from which the data were obtained investigated the effect of three sessions of motivational enhancement therapy (MET) on retention and substance use among individuals seeking treatment at five community-based outpatient treatment programs (Ball et al., 2007). Specific clinical trial procedures have been described in past studies (Ball et al., 2007; Field, Adinoff, Harris, Ball, & Carroll, 2009). This secondary analysis study was reviewed and approved by the Institutional Review Board of a large, southeastern university.

Sample

Baseline SIP-R data are available for 464 participants outpatients who were initially evaluated in the MET study (Ball et al., 2007). The current study examines data from 195 African American participants and 194 non-Latino White participants. Data from 59 Latino participants and 16 participants coded as *other* were not included in analyses because these subgroups were too small for multigroup confirmatory factor analyses necessary to determine measurement invariance. The average age of the total sample of African American and non-Latino White participants ($n = 389$) was 35.93 years ($SD = 10.22$). Approximately 71% of the total sample was male. Significant age differences were detected between racial groups [$F(1,388) = 15.22, p < .001, \eta^2 = .20$]. African American participants tended to be older ($M = 37.90, SD = 10.01$) than non-Latino White participants ($M = 33.93, SD = 10.07$). No differences were detected in participant gender between racial groups.

Measure

The Short Inventory of Problems-Revised (SIP-R; Kiluk et al., 2012)—The SIP-R is a 17-item instrument used to assess participants' perceptions of the adverse consequences of their substance use during 3 months prior to assessment. Each item is rated by the respondent on a four-point Likert-type scale. Anchors for Items 1-9 are 1 = *never* to 4 = *daily or almost daily*; while anchors for 10-17 are 1 = *not at all* to 4 = *very much*. The SIP-R has five factors: impulse control (3 items), interpersonal (3 items), intrapersonal (3 items), physical (3 items), and social (5 items). Sample items include: I have taken foolish risks

when I have been drinking or using drugs, (impulse control); *My family has been hurt by my drinking or drug use*, (interpersonal); I have been unhappy because of my drinking or drug use, (intrapersonal); Because of my drinking or drug use, I have lost weight or not eaten properly, (physical); and I have failed to do what is expected of me because of my drinking or drug use, (social). The SIP-R has yielded evidence of internal reliability ($\alpha = .95$) and several forms of validity (Kiluk et al., 2012). In the current study, the impulse control ($\alpha = .79$), interpersonal ($\alpha = .84$), intrapersonal ($\alpha = .82$), physical ($\alpha = .82$), and social ($\alpha = .82$) factors yielded an adequate estimate of internal consistency across African American participants. Adequate Cronbach's alpha reliability estimates also were found with non-Latino White sample: impulse control ($\alpha = .82$), interpersonal ($\alpha = .85$), intrapersonal ($\alpha = .92$), physical ($\alpha = .85$), and social ($\alpha = .89$) factors.

Analytic Methods

Data analyses proceeded in several steps. First, SIP-R items were examined initially for univariate and multivariate normality. Lei and Lomax (2005) suggests a cutoff of absolute value of 2.30 for univariate skewness and kurtosis estimates. Mahalanobis distances were calculated to detect multivariate outliers (Tabachnick & Fidell, 2007). Second, a series of confirmatory factor analyses (CFA) via structural equation modeling were conducted using Mplus 6.12 statistical software (Muthén & Muthén, 1998–2011) to assess the fit of the hypothesized five factors of the SIP-R measurement model. Measurement model fit was assessed using the confirmatory fit index (CFI) and the root mean square error of approximation (RMSEA). A measurement or structural model with excellent fit to the data have a CFI .95 and RMSEA .06; while, an adequate fit has CFI .90 and RMSEA .08 (Hancock & Freeman, 2001; Hu & Bentler, 1999; McDonald & Ho, 2002; Tomarken & Waller, 2005).

The first CFA tested the SIP-R measurement model separately across African American and non-Latino White participants. Next, the SIP-R measurement model was fit as a multigroup CFA model with parameters free to vary across groups. This model was used to establish configural invariance. If the configural invariance model yielded acceptable fit indices, item loadings were constrained across groups to test metric or weak invariance (Meredith, 1993; Vandenberg & Lance, 2000). The next level of invariance to be tested was equal item intercepts or scalar/strong invariance (Meredith, 1993; Vandenberg & Lance, 2000). The final level of invariance to be tested was strict invariance (Meredith, 1993; Vandenberg & Lance, 2000), which additionally requires equal item residual variances.

When testing differences in model fit between the types of invariance, two standard indices were used: the difference in chi-square values (χ^2) and the difference in CFI values (CFI) and RMSEA (RMSEA) values. In each model comparison conducted, for the null hypothesis of invariance across groups to be statistically rejected, any of the following three criteria had to be met: (a) χ^2 significant at $p < .05$ (Byrne, 2001), (b) CFI > .01, or (c) RMSEA > .015 (Cheung & Rensvold, 2002).

FINDINGS

The data met guidelines for univariate normality. Univariate skewness and kurtosis values were less than 2.3 for all items (Lei & Lomax, 2005). Mahalanobis distances for 6 cases were significant ($p < .001$), suggesting multivariate nonnormality (Tabachnick & Fidell, 2007). Therefore, these 6 cases were omitted from subsequent analyses resulting in a sample of 192 African American participants and 191 non-Latino, White participants.

Next, SIP-R measurement models were tested separately across African American and non-Latino White participants. Fit statistics are presented in Table 1.

For the African-American base model, fit indices fell within acceptable ranges. However, for the non-Latino White base model, fit indices fell slightly outside acceptable ranges due to a RMSEA value of .09. To improve base model fit, we consulted modification indices provided by Mplus. We allowed the error terms for two items on the Social subscale (items 11 and 17) to be correlated, as well as two items from the Intrapersonal subscale (items 1 and 4). We included these modifications in both baseline models and further equivalence testing analyses. Table 2 lists SIP-R items and the corresponding item-level descriptive statistics and standardized parameter estimates for the 5-factor measurement model for each racial group.

Table 3 is an inter-item correlation matrix for the two samples.

Configural invariance was the first level of equivalence examined. Results of this model are reported in Table 1. Analyses indicated that fit indices were acceptable for the configural invariance model. The SIP-R appeared to reflect similar constructs across the samples of African American and non-Latino White participants.

To test weak metric invariance, equality constraints for items loading onto their intended latent factors were included across the two groups. The model passed this level of testing when compared to the configural invariance model (see Table 1). Thus, factor loadings of items in the measure are invariant across the racial groups.

Strong or scalar equivalence (i.e., equivalence of intercepts) was assessed next. This model was created by imposing equality constraints on the intercepts of items across the two racial groups. The model did not pass this level of equivalence testing, indicating that scale items do not reliably assess the same level of the construct in both groups. Therefore, we examined modification indices for each item and iteratively freed equality constraints for the items with the highest modification indices until the chi-square difference between the models was no longer significant at $p < .05$. Freeing the constraints on the following four items established a viable partial strong/scalar invariance measure: Item 3 (I have failed to do what is expected of me because of my drinking or drug use—social), Item 17 (I have spent too much or lost a lot of money because of my drinking or drug use—social), Item 11 (I have had money problems because of my drinking or drug use—social), and Item 6 (When drinking or using drugs, I have done impulsive things that I regretted later—impulse control). This finding suggests that item intercepts were significantly different between the two

groups. African American participants had reliably higher intercepts across Item 17 (2.27 vs. 1.87), Item 11 (1.97 vs. 1.59), and Item 6 (1.54 vs. 1.35); and Item 3 (1.64 vs. 1.60).

Finally, strict invariance was tested by constraining residual variances of items in the partial strong/scalar invariance model. The equal residual variance restrictions as a group were rejected by the chi-square difference test. This implies that reliabilities of the latent factors vary by racial group. Modification indices were examined to identify potentially disparate residual variances. Two items demonstrated high modification indices. They were freed iteratively, beginning with the highest modification index. African American participants yielded larger residual variability on Item 1 (I have been unhappy because of my drinking or drug use–intrapersonal; 0.73 vs. 0.46) and Item 4 (I have felt guilty or ashamed because of my drinking or drug use–intrapersonal; 0.60 vs. 0.47). When both items were unconstrained, the resultant model demonstrated acceptable invariance (see Table 1).

DISCUSSION

The study of racial/ethnic variation in response to drug abuse treatment is critically needed to provide effective treatments to all patients (Carroll et al., 2007, 2009). Historically, racial and ethnic minority adults tend to be over-represented in drug treatment programs (Carroll et al., 2007) and under-represented in clinical trial studies (Magruder, Ouyang, Miller, & Tilley, 2009). Few studies have addressed potential differential effectiveness of treatments across racial and ethnic groups (Amaro, Arévalo, Gonzalez, Szapocznik, & Iguchi, 2006; Strada, Donohue, & Lefforge, 2006). Nevertheless, drug treatment researchers are developing culturally-appropriate methods to better engage and retain more racial and ethnic minority adults in studies (e.g., Barry, Sullivan, & Petry, 2009; Carroll et al., 2009; Field & Caetano, 2010). Conceptually and structurally equivalent outcome measures across racial/ethnic groups are needed to validly test potential differential treatment outcomes and mechanisms of change. The current study facilitates such studies by reporting measurement invariance estimates of the SIP-R across African American and non-Latino White adult outpatients.

The initial psychometric assessment of the SIP-R reported evidence of five domains with adequate reliability and validity in substance abusing patients (Kiluk et al., 2012). Thus, it appeared as if the five-factor structure representing domains of adverse consequences of substance use could be supported. However, despite reports of biased SIP items across racial groups of men who have sex with men (Hagman et al., 2009), the assumption of measurement invariance had not yet been tested across African American and non-Latino White outpatient groups for the SIP-R. Hence, we examined the measurement invariance properties of the SIP-R in the current study. Our findings supported the comparable understanding of the factors and most of the items of the SIP-R in African American and non-Latino White participants.

Our findings indicated configural and weak and metric invariance between racial groups, demonstrating that the SIP-R effectively functions as a measure of its intended construct for both groups. This finding informs past and future social work research conducted with and comparing these groups using the SIP-R. It suggests that the five constructs are similarly

understood by both racial groups, which allows for confident assessment of mean differences in adverse consequences of substance use. However, only partial strong/scalar invariance was demonstrated. That is, four items suggested differences in the normative levels of the items across racial groups. While evidence of partial invariance is sufficient to support equivalence of the measure between groups (Steinmetz et al., 2009), African American and non-Latino White outpatients may need to report different levels of these items, particularly in the social and impulsive domains, to yield the same underlying level of these latent factors. Therefore, future social work research and clinical use of the SIP-R may need to consider differences in levels of awareness of adverse consequences of substance use among African American and Non-Latino White adults as measured by the identified items, especially in the social domain.

The SIP-R also only indicated partial strict metric invariance. Two items from the intrapersonal subscale were identified as contributing to the nonequivalence in the residual variances of items. The residual variances of the two items were largest in African American participants' reports. The larger the variance exhibited by an item, the smaller the standardized loading and subsequent reliability of the factor the item represents (Feaster et al., 2010). Thus, strict metric nonequivalence suggests that African American participants' report of these two indicators are less reliable in comparison to non-Latino White participants.

The difference in reliability is apparent in the large differences in reliability estimates between racial groups for the Intrapersonal factor. The difference may be due to the few items assessing the complex Intrapersonal construct. This concern could be remedied with further item development. Heterogeneity of item content, and conceptualization of the content, also may lead to the inconsistencies across groups. One Intrapersonal scale item assesses effects of drug use on personal growth while the other two items assess negative affectivity about drug use. Alternatively, other differences between the racial groups on factors such as age or primary drug of choice could influence inconsistent responding to items about unhappiness and guilt due to drug use as it has influenced results in relevant past studies (Carroll et al., 2006; Field et al., 2009). In the current study, non-Latino, white participants reported more primary alcohol, opioid, and methamphetamine use while African American participants reported more cocaine, marijuana, and polydrug use. Future studies are needed to investigate potential differences in emotional consequences related to different drugs of abuse which may be contributing to the strict metric variance. Finally, although the subscale evidenced adequate internal consistency across groups, the difference in reliability caused by the two identified items is worthy of verification in future social work research studies to ensure use of equally reliable latent factors.

The present findings should be interpreted in light of limitations. First, the current assessment was limited to African American and non-Latino White participants. Due to sample size restrictions, the current study was unable to determine measurement equivalence properties across other racial and ethnic groups as well as other key variables potentially influencing measurement equivalence between demographic groups. Future researchers are encouraged to conduct similar analytic procedures to ensure valid measurement across racial, ethnic, and other salient groups. Second, because this study involved a clinical

sample, the subsamples were not representative of populations of African American and non-Latino White adults. Future substance abuse treatment researchers need to continue to increase enrollment of racial and ethnic minorities to improve external validity and allow for secondary studies such as measurement equivalence studies to ensure valid results and to better understand the sociocultural aspects of race/ethnicity and substance abuse.

CONCLUSION

Despite these limitations, the present study is the first to report measurement equivalence properties of the SIP-R across racial groups. This study contributes to the growing literature on how substance abuse treatment researchers can ensure unbiased assessment when studying treatment outcomes across racial/ethnic minority groups (e.g., Burlew et al., 2009; Feaster et al., 2010; Field et al., 2009). Since MI and MET (Miller & Rollnick, 2002; Miller, Zweben, DiClemente, & Rychtarik, 1992) techniques have become standard social work practice in substance-abuse treatment settings (Ondersma, Winhusen, Erickson, Stine, & Wang, 2009), the SIP-R is expected to continue to be one of the most prominent outcome measures in social work practice involving motivational techniques. The current findings encourage the valid use of the SIP-R with samples of African American and non-Latino White adult outpatients to validly examine treatment outcomes and test potential differential effectiveness of treatments across racial groups.

Acknowledgments

The information reported here results from secondary analyses of data from clinical trials conducted as part of the National Drug Abuse Treatment Clinical Trials Network (CTN) sponsored by National Institute on Drug Abuse (NIDA). Specifically, data from CTN-0004 (MET to Improve Treatment Engagement and Outcome in Subjects Seeking Treatment for Substance Abuse) were included. CTN databases and information are available at www.ctndatashare.org. This study also was partially supported by National Institute on Minority Health and Health Disparities (NIMHD) Award P20MD002288. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Minority Health and Health Disparities or the National Institutes of Health.

REFERENCES

- Allen, J.; Walsh, J. A construct-based approach to equivalence: Methodologies for cross-cultural/multicultural personality assessment research.. In: Dana, R., editor. Handbook of cross-cultural and multicultural personality assessments. Lawrence Erlbaum; Mahwah, NJ: 2000. p. 63-86.
- Allensworth-Davies D, Cheng DM, Smith PC, Samet JH, Saitz R. The Short Inventory of Problems—Modified for Drug Use (SIP-DU): Validity in a primary care sample. *The American Journal on Addictions*. 21:257–262. doi: 10.1111/j.1521-0391.2012.00223.x. [PubMed: 22494228]
- Alterman AI, Cacciola JS, Ivey MA, Habing B, Lynch KG. Reliability and validity of the alcohol short index of problems and a newly constructed drug short index of problems. *Journal of Studies on Alcohol and Drugs*. 2009; 70:34–307. PMID: PMC2653616.
- Amaro H, Arévalo S, Gonzalez G, Szapocznik J, Iguchi MY. Needs and scientific opportunities for research on substance abuse treatment among Hispanic adults. *Drug and Alcohol Dependence*. 2006; 84:S64–S75. doi: 10.1016/j.drugalcdep.2006.05.008. [PubMed: 16766137]
- American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 4th ed.. APA; Washington, DC: 2000. text rev.
- Ball SA, van Horn D, Crits-Christoph P, Woody GE, Martino S, Nich C, Carroll KM. Site matters: multisite randomized trial of motivational enhancement therapy in community drug abuse clinics. *Journal of Consulting and Clinical Psychology*. 2007; 75:556–567. doi: 10.1037/0022-006X.75.4.556. [PubMed: 17663610]

- Barry D, Sullivan B, Petry NM. Comparable efficacy of contingency management for cocaine dependence among African American, Hispanic, and White methadone maintenance clients. *Psychology of Addictive Behaviors*. 2009; 23:168–174. doi: 10.1037/a0014575. [PubMed: 19290703]
- Bender RE, Griffin ML, Gallop BG, Weiss RD. Assessing negative consequences in patients with substance use and bipolar disorders: Psychometric properties of the Short Inventory of Problems (SIP). *American Journal on Addictions*. 2007; 16:503–509. doi: 10.1080/10550490701641058. [PubMed: 18058418]
- Bingenheimer JB, Raudenbush SW, Leventhal T, Brooks-Gunn J. Measurement equivalence and differential item functioning in family psychology. *Journal of Family Psychology*. 2005; 19:441–455. doi: 10.1037/0893-3200.19.3.441. [PubMed: 16221024]
- Blanchard KA, Morgenstern J, Morgan TJ, Labouvie EW, Bux DA. Assessing consequences of substance use: Psychometric properties of the Inventory of Drug Use Consequences. *Psychology of Addictive Behaviors*. 2003; 17:328–331. doi:10.1037/0893-164X.17.4.328. [PubMed: 14640829]
- Burlew AK, Feaster DJ, Brecht M, Hubbard R. Measurement and data analysis in research addressing health disparities in substance abuse. *Journal of Substance Abuse Treatment*. 2009; 36:25–43. doi: 10.1016/j.jsat.2008.04.003. [PubMed: 18550320]
- Byrne, BM. *Structural Equation Modeling with AMOS*. Erlbaum; New Jersey: 2001.
- Carroll KM, Ball SA, Nich C, Martino S, Frankforter TL, Farentinos C, Woody GE, the NIDA-CTN. Motivational interviewing to improve treatment engagement and outcome in individuals seeking treatment for substance abuse: a multisite effectiveness study. *Drug and Alcohol Dependence*. 2006; 81:301–312. doi: 10.1016/j.drugalcdep.2005.08.002. [PubMed: 16169159]
- Carroll KM, Martino S, Ball S, Nich C, Frankforter TL, Anez LM, Farentinos C. A multisite randomized effectiveness trial of motivational enhancement therapy for Spanish-speaking substance users. *Journal of Consulting and Clinical Psychology*. 2009; 77:993–999. doi: 10.1037/a0016489. [PubMed: 19803579]
- Carroll KM, Rosa CL, Brown LS, Daw R, Magruder KM, Beatty L. Addressing ethnic disparities in drug abuse treatment in the clinical trial networks. *Drug and Alcohol Dependence*. 2007; 90:101–106. doi: 10.1016/j.drugalcdep.2006.12.033.
- Chan KT, Tran TV, Nguyen T. Cross cultural equivalence of a measure of perceived discrimination between Chinese-Americans and Vietnamese-Americans. *Journal of Ethnic & Cultural Diversity in Social Work*. 2012; 21:20–36. doi: 10.1080/15313204.2011.647348.
- Chen FF, Sousa KH, West SG. Testing measurement equivalence of second-order factor models. *Structural Equation Modeling*. 2005; 12:471–492.
- Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*. 2002; 9:233–255. doi: 10.1207/S15328007SEM0902_5.
- Crockett L, Randall B, Shen Y, Russell S, Driscoll A. Measurement equivalence of the Center for Epidemiological Studies Depression Scale for Latino and Anglo Adolescents: A national study. *Journal of Consulting and Clinical Psychology*. 2005; 73:47–58. doi: 10.1037/0022-006X.73.1.47. [PubMed: 15709831]
- Feaster DJ, Robbins MS, Henderson C, Horigan V, Puccinelli MJ, Burlew AK, Szapocznik J. Equivalence of family functioning and externalizing behaviors in adolescent substance users of different race/ethnicity. *Journal of Substance Abuse Treatment*. 2010; 38:S113–S124. doi: 10.1016/j.jsat.2010.01.010. [PubMed: 20307791]
- Feinn R, Tennen H, Kranzler HR. Psychometric properties of the Short Index of Problems as a measure of recent alcohol-related problems. *Alcoholism: Clinical and Experimental Research*. 2003; 27:1436–1441. doi:10.1097/01.ALC.0000087582.44674.AF.
- Field CA, Adinoff B, Harris TR, Ball SA, Carroll KM. Construct, concurrent and predictive validity of the URICA: Data from two multi-site clinical trials. *Drug and Alcohol Dependence*. 2009; 101:115–123. doi: 10.1016/j.drugalcdep.2008.12.003. [PubMed: 19157723]
- Field C, Caetano R. The role of ethnic matching between patient and provider on the effectiveness of brief alcohol interventions with Hispanics. *Alcoholism: Clinical and Experimental Research*. 2010; 34:262–271. doi: 10.1111/j.1530-0277.2009.01089.x.

- Gillespie W, Holt JL, Blackwell RL. Measuring outcomes of alcohol, marijuana, and cocaine use among college students: A preliminary test of the Shortened Inventory of Problems–Alcohol and Drugs (SIP-AD). *Journal of Drug Issues*. 2007; 37:549–567. doi:10.1177/002204260703700304.
- Hagman BT, Kuerbis AN, Morgenstern J, Bux DA, Parsons JT, Heidinger BE. An item response theory (IRT) analysis of the Short Inventory of Problems-Alcohol and Drugs (SIP-AD) among nontreatment seeking men-who-have-sex-with-men: Evidence for a shortened 10-item SIP-AD. *Addictive Behaviors*. 2009; 34:948–954. doi:10.1016/j.addbeh.2009.06.004. [PubMed: 19564078]
- Hancock GR, Freeman MJ. Power and sample size for the root mean square error of approximation test of not close fit in structural equation modeling. *Educational and Psychological Measurement*. 2001; 61:741–758. doi: 10.1177/00131640121971491.
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6:1–55. doi: 10.1080/10705519909540118.
- Kiluk BD, Dreifuss JA, Weiss RD, Morgenstern J, Carroll KM. The Short Inventory of Problems – Revised (SIP-R): Psychometric properties within a large, diverse sample of substance use disorder treatment seekers. *Psychology of Addictive Behaviors*. May 28.2012 Advance online publication. doi: 10.1037/a0028445.
- Magruder KM, Ouyang B, Miller S, Tilley BC. Retention of under-represented minorities in drug abuse treatment studies. *Clinical Trials*. 2009; 6(3):252–260. doi: 10.1177/1740774509105224. [PubMed: 19528134]
- McDonald RP, Ho M. Principles and practice in reporting structural equation analyses. *Psychological Methods*. 2002; 7:64–82. PMID: 11928891. [PubMed: 11928891]
- Meredith W. Measurement invariance, factor analysis and factorial invariance. *Psychometrika*. 1993; 58:525–543. doi: 10.1007/BF02294825.
- Miller, WR.; Rollnick, S. *Motivational interviewing: Preparing people for change*. Guilford Press; New York: 2002.
- Miller, WR.; Tonigan, JS.; Longabaugh, R. Vol. 4. National Institute on Alcohol Abuse and Alcoholism (NIAAA); Rockville, MD: 1995. *The Drinker Inventory of Consequences (DRrINC): An instrument for assessing adverse consequences of alcohol abuse*. Test manual..
- Miller, WR.; Zweben, A.; DiClemente, CC.; Rychtarik, RG. *Motivational Enhancement Therapy manual: A clinical research guide for therapists treating individuals with alcohol abuse and dependence*. National Institute on Alcohol Abuse and Alcoholism; Rockville, MD: 1992.
- Muthén, B.; Muthén, L. *Mplus user's guide*. Muthén and Muthén; Los Angeles, CA: 1998-2011.
- Ondersma SJ, Winhusen TM, Erickson SJ, Stine SM, Wang Y. Motivation enhancement therapy with pregnant substance-abusing women: Does baseline motivation moderate efficacy? *Drug and Alcohol Dependence*. 2009; 101:74–79. PMID: PMC2792933. [PubMed: 19101099]
- Radloff LS. The CES-D Scale: A self report depression scale for research in the general population. *Applied Psychological Measurement*. 1977; 1:385–401. doi: 10.1177/014662167700100306.
- Ramirez M, Ford M, Steward A, Teresi J. Measurement issues in health disparities research. *Health Research and Educational Trust*. 2005; 40:1640–1657. doi: 10.1111/j.1475-6773.2005.00450.x.
- Randolph K, Gerend M, Miller B. Measuring alcohol expectancies in youth. *Journal of Youth and Adolescence*. 2006; 35:939–948. doi: 10.1007/s10964-006-9072-3.
- Steinmetz H, Schmidt P, Tina-Booh A, Wiecek S, Schwartz SH. Testing measurement equivalence using multigroup CFA: Differences between groups in human values measurement. *Quality and Quantity*. 2009; 43:599–616. doi:10.1007/s11135-007-9143-x.
- Strada M, Donohue B, Lefforge N. Examination of ethnicity in controlled treatment outcome studies involving adolescent substance abusers: A comprehensive literature review. *Psychology of Addictive Behaviors*. 2006; 20:11–27. PMID: 16536661. [PubMed: 16536661]
- Tonigan JS, Miller WR. The Inventory of Drug Use Consequences (InDUC): Test–retest stability and sensitivity to detect change. *Psychology of Addictive Behaviors*. 2002; 16:165–168. doi: 10.1037/0893-164X.16.2.165. [PubMed: 12079257]
- Tomarken A, Waller NG. Structural equation modeling as a data-analytic framework for clinical science: Strengths, limitations, and misconceptions. *Annual Review of Clinical Psychology*. 2005; 1:31–65. doi: 10.1146/annurev.clinpsy.1.102803.144239.

- Vandenberg R, Lance C. A review and synthesis of the measurement equivalence literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*. 2000; 3:4–70. doi: 10.1177/109442810031002.
- Widaman, KF.; Reise, SP. Exploring the measurement equivalence of psychological instruments: applications in the substance use domain.. In: Bryant, KJ.; Windle, M.; West, SG., editors. *The science of prevention: methodological advances from alcohol and substance abuse research*. American Psychological Association; Washington, DC: 1997. p. 281-324.
- Williams DR, Yu Y, Jackson JS, Anderson NB. Racial differences in physical and mental health: Socioeconomic status, stress and discrimination. *Journal of Health Psychology*. 1997; 2:335–351. doi:10.1177/135910539700200305. [PubMed: 22013026]
- Winhusen TM, Kropp F, Babcock D, Hague D, Erickson SJ, Renz C, Somoza EC. Motivational enhancement therapy to improve treatment utilization and outcome in pregnant substance users. *Journal of Substance Abuse Treatment*. 2008; 35:161–173. PMID: PMC2546520. [PubMed: 18083322]

TABLE 1

Fit Indices by Measurement Invariance Model Test

	X ²	df	X ²	df	X ² difference test p	CFI	CFI	RMSEA (90% CI)	RMSEA
Base Model: African American	186.41	107				0.96		0.06 (0.05, 0.08)	n/a
Base Model: White, Non-Latino	242.14	107				0.94		0.08 (0.07, 0.09)	n/a
Configural Invariance	428.55	214				0.95		0.07 (0.06, 0.08)	n/a
Metric Invariance	448.55	226	19.98	12	.07	0.95	0.00	0.07 (0.06, 0.08)	0.0
Strong/Scalar Invariance	483.56	238	35.03	12	<.001	0.94	−0.01	0.07 (0.06, 0.08)	0.0
Partial strong/scalar invariance model by freeing Items 3, 6, 11, 17	455.23	234	6.69	8	.57	0.95	0.00	0.07 (0.06, 0.08)	0.0
Strict invariance with partial strong/scalar invariance model	493.19	251	37.96	17	<.001	0.94	−0.01	0.07 (0.06, 0.08)	0.0
Partial strict invariance model by freeing Items 1 and 4	476.05	249	20.82	15	.14	0.94	−0.01	0.07 (0.06, 0.08)	0.0

Note. df = degrees of freedom; CFI = comparative fit index; RMSEA = root mean square error of approximation; CI = confidence interval.

TABLE 2
SIP-R Confirmatory Factor Analysis Results by Racial Group: Item Means, Standard Deviations, and Standardized Parameter Estimates

Latent Construct	Item	Item M	African American	White, Non-Latino	African American	White, Non-Latino	Item Factor Loading	African American	White, Non-Latino	Item Standard Error
Intrapersonal	1. I have been unhappy because of my drinking or drug use.	1.83	2.09	1.16	1.05	0.70	0.88	0.04	0.02	0.02
	4. I have felt guilty or ashamed because of my drinking or drug use.	1.88	2.02	1.15	1.07	0.77	0.89	0.03	0.02	0.02
	15. My drinking or drug use has gotten in the way of my growth as a person.	1.94	2.08	1.17	1.14	0.82	0.81	0.03	0.03	0.03
Physical	2. Because of my drinking or drug use, I have lost weight or not eaten properly.	1.50	1.62	1.22	1.17	0.67	0.80	0.04	0.03	0.03
	10. My physical health has been harmed by my drinking or drug use.	1.31	1.29	1.08	1.05	0.61	0.64	0.05	0.05	0.05
Social	12. My physical appearance has been harmed by my drinking or drug use.	1.35	1.28	1.19	1.08	0.78	0.72	0.04	0.04	0.04
	3. I have failed to do what is expected of me because of my drinking or drug use.	1.64	1.74	1.10	1.05	0.82	0.80	0.03	0.03	0.03
	8. I have gotten into trouble because of drinking or drug use.	1.15	1.21	1.01	0.93	0.57	0.44	0.05	0.06	0.06
	9. The quality of my work has suffered because of my drinking or drug use.	1.02	1.14	1.13	1.12	0.59	0.63	0.05	0.05	0.05
Impulse Control	11. I have had money problems because of my drinking or drug use.	1.97	1.72	1.21	1.19	0.76	0.63	0.04	0.05	0.05
	17. I have spent too much or lost a lot of money because of my drinking or drug use.	2.27	1.99	1.06	1.19	0.67	0.71	0.04	0.04	0.04
	5. I have taken foolish risks when I have been drinking or using drugs.	1.61	1.75	1.10	1.04	0.79	0.80	0.03	0.03	0.03
	6. When drinking or using drugs, I have done impulsive things that I regretted later.	1.54	1.49	1.10	1.00	0.85	0.86	0.03	0.03	0.03
	7. Drinking or using one drug has caused me to use other drugs more.	1.09	1.20	1.17	1.16	0.61	0.67	0.05	0.05	0.05
Interpersonal	13. My family has been hurt by my drinking or drug use.	1.96	1.90	1.17	1.15	0.74	0.71	0.04	0.04	0.04
	14. A friendship or close relationship has been damaged by my drinking or drug use.	1.50	1.60	1.27	1.17	0.79	0.76	0.03	0.04	0.04
	16. My drinking or drug use has damaged my social life, popularity, or reputation.	1.52	1.54	1.22	1.17	0.86	0.76	0.02	0.04	0.04

Note. Each SIP-R item is rated on a four-point Likert-type scale. Anchors for Items 1-9 are 1 = *never* to 4 = *daily or almost daily*; while anchors for 10-17 are 1 = *not at all* to 4 = *very much*.

TABLE 3

Inter-item Correlations by African American and Non-Latino White Samples

Item	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1. I have been unhappy because of my drinking or drug use.	1	.63	.57	.83	.42	.45	.39	.16	.44	.51	.50	.51	.45	.43	.68	.58	.54
2. Because of my drinking or drug use, I have lost weight or not eaten properly.	.55	1	.64	.65	.54	.58	.44	.33	.46	.51	.50	.52	.52	.50	.65	.48	.52
3. I have failed to do what is expected of me because of my drinking or drug use.	.59	.64	1	.69	.62	.66	.47	.32	.58	.42	.49	.53	.51	.51	.60	.57	.53
4. I have felt guilty or ashamed because of my drinking or drug use.	.66	.55	.65	1	.47	.52	.38	.20	.52	.43	.43	.47	.46	.45	.68	.60	.46
5. I have taken foolish risks when I have been drinking or using drugs.	.56	.49	.65	.59	1	.69	.55	.41	.39	.42	.38	.48	.45	.49	.46	.43	.48
6. When drinking or using drugs, I have done impulsive things that I regretted later.	.57	.53	.68	.68	.68	1	.56	.40	.49	.41	.46	.49	.47	.50	.52	.53	.56
7. Drinking or using one drug has caused me to use other drugs more.	.42	.35	.51	.47	.45	.53	1	.41	.35	.28	.42	.43	.34	.45	.36	.42	.48
8. I have gotten into trouble because of drinking or drug use.	.38	.28	.45	.38	.47	.58	.44	1	.42	.26	.25	.25	.34	.33	.26	.27	.30
9. The quality of my work has suffered because of my drinking or drug use.	.37	.40	.49	.41	.43	.46	.35	.44	1	.39	.37	.47	.37	.43	.55	.50	.41
10. My physical health has been harmed by my drinking or drug use.	.39	.36	.43	.42	.43	.43	.29	.36	.41	1	.50	.58	.42	.38	.51	.39	.42
11. I have had money problems because of my drinking or drug use.	.54	.51	.64	.56	.56	.57	.43	.37	.46	.51	1	.57	.45	.46	.50	.41	.76
12. My physical appearance has been harmed by my drinking or drug use.	.45	.52	.60	.56	.58	.54	.39	.34	.48	.49	.61	1	.52	.49	.57	.56	.61
13. My family has been hurt by my drinking or drug use.	.44	.39	.51	.52	.49	.51	.34	.33	.36	.43	.59	.60	1	.61	.58	.48	.47
14. A friendship or close relationship has been damaged by my drinking or drug use.	.44	.48	.48	.55	.57	.60	.38	.38	.46	.43	.60	.60	.64	1	.63	.57	.53
15. My drinking or drug use has gotten in the way of my growth as a person.	.58	.55	.61	.56	.60	.65	.48	.48	.45	.53	.65	.62	.67	.64	1	.67	.58
16. My drinking or drug use has damaged my social life, popularity, or reputation.	.55	.51	.55	.61	.57	.61	.45	.38	.47	.53	.64	.71	.60	.67	.72	1	.53
17. I have spent too much or lost a lot of money because of my drinking or drug use.	.41	.49	.53	.46	.53	.53	.37	.33	.36	.42	.72	.55	.51	.55	.60	.60	1

Note.

* $p < .05$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Correlations for African American participants are presented below the diagonal, and correlations for Non-Latino White participants are presented above the diagonal.

**
 $p < .01$.