Published in final edited form as:

Psychol Assess. 2020 April ; 32(4): 394–406. doi:10.1037/pas0000799.

Optimal Assessment of Protective Behavioral Strategies Among College Drinkers: An Item Response Theory Analysis

Jessica L. Martin,

Department of Educational and Counseling Psychology, University at Albany-State University of New York;

Kimberly F. Colvin,

Department of Educational and Counseling Psychology, University at Albany-State University of New York;

Michael B. Madson,

Department of Psychology, The University of Southern Mississippi;

Byron L. Zamboanga,

Department of Psychology, Smith College;

Rena Pazienza

Department of Educational and Counseling Psychology, University at Albany-State University of New York.

Abstract

College student alcohol use and related consequences continue to warrant significant concern. Extant research demonstrates protective behavioral strategies (PBS; self-regulatory strategies that can be employed before, during, or after drinking to prevent intoxication or negative consequences) have promise for preventative interventions. Variations in conceptualization and measurement of the construct limit generalization of PBS research. To advance generalization of PBS research, there is a need for a brief, comprehensive, content valid, PBS measure that demonstrates equivalence in measurement across demographically diverse college students. The present study aimed to develop a psychometrically robust measure of PBS using item response theory (IRT) to address measurement and methodological issues including identifying optimal response anchors and items that represent the full range of the PBS construct for use with college men and women from different racial backgrounds. Participants were 503 college students enrolled in a midsized university in the Southern United States. IRT and differential item functioning (DIF) analyses of 68 PBS items extracted from 6 existing PBS measures resulted in a 20-item instrument, the Protective Drinking Practices Scale, with items that functioned equivalently for White and Black/African American college men and women. The measure also demonstrated good internal and external validity. Widespread use of this measure will help enhance the comparability of findings in PBS research, allowing for more targeted and impactful research on PBS as a mechanism of change. Suggestions for future research are provided.

Keywords

protective behavioral strategies; alcohol use; alcohol-related problems; college students; item response theory

Alcohol misuse by college students in the U.S. has been a pervasive long-term problem. Recent studies estimate that roughly 76% of college students consumed alcohol in the past year and 62% drank alcohol in the past month (Schulenberg et al., 2018). College students regularly engage in heavy episodic drinking (HED; defined as 4/5 or more drinks in one sitting for women/men, respectively; National Institute of Alcohol Abuse and Alcoholism [NIAAA], 2004) and high intensity drinking (HID; defined as 8 drinks for women and 10 drinks for men; Patrick, 2016; White, Kraus, & Swartzwelder, 2006), which are associated with a plethora of negative consequences for the drinker, nondrinking peers, and the campus community (White & Hingson, 2013). Recent research suggests that those who engage in HED or HID are more likely to be injured because of drinking, be arrested or have legal problems resulting from alcohol use, be the driver in an alcohol-related traffic accident, and have an alcohol use disorder (Hingson, Zha, & White, 2017). Given the pervasive culture of alcohol use among college students resulting in largely stable levels of alcohol consumption and associated negative consequences over time, researchers and university personnel have focused their efforts on harm reduction strategies (Larimer & Cronce, 2007). One approach to harm reduction that has empirical support is protective behavioral strategies (PBS).

PBS have emerged as an evidence supported self-regulatory set of behaviors students can implement to help prevent intoxication and/or alcohol-related problems (for reviews see Pearson, 2013; Prince, Carey, & Maisto, 2013). These strategies include behaviors aimed at directly managing consumption (e.g., avoiding shots, alternate alcoholic and nonalcoholic drink) and indirect strategies focused on reducing the harm associated with drinking (e.g., using a designated driver, leaving the bar or party with a friend). Support for PBS as a valuable component of alcohol interventions with college students (Barnett, Murphy, Colby, & Monti, 2007; Larimer et al., 2007; Murphy et al., 2012) has led to research exploring the mediating effects of PBS between common predictors of alcohol use including, drinking motives (e.g., Martens, Ferrier, & Cimini, 2007), alcohol expectancies (e.g., Madson, Moorer, Zeigler-Hill, Bonnell, & Villarosa, 2013), and alcohol use and negative consequences such that direct effects of the predictors on outcomes are mitigated with more frequent use of PBS. More recently, the value of PBS in the relationship between alcohol-related problems and mental health concerns such as symptoms of depression (Martens et al., 2008; Villarosa, Messer, Madson, & Zeigler-Hill, 2018), anxiety (Linden, Lau-Barraco, & Milletich, 2013), social anxiety (Villarosa, Moorer, Madson, Zeigler-Hill, & Noble, 2014), and posttraumatic stress disorder (Jordan et al., 2019) have emerged. While evidence generally supports the use of PBS as a harm reduction approach, it appears that the use and effects of PBS are influenced by factors such as gender and race (Longo, Martin, Zamboanga, Milroy, & Wyrick, 2018; Madson & Zeigler-Hill, 2013; Martin, Zamboanga, Haase, & Buckner, 2019). Some evidence suggests that among men and women and across racial groups there are differences in the overall use of PBS, including differences in the use of particular strategies (e.g., Clarke et al., 2016; Martin et al., 2019).

Numerous studies have found that college women use more PBS than college men (e.g., LaBrie, Lac, Kenney, & Mirza, 2011; Lewis, Kaysen, Rees, & Woods, 2010; Palmer, McMahon, Rounsaville, & Ball, 2010). Kenney and LaBrie (2013) found that PBS strategies were more protective for White women than students of other races and genders. Importantly, in a test of gender invariance of the Protective Behavioral Strategies Survey (PBSS; Martens et al., 2005), Treloar, Martens, and McCarthy (2014) found systematic difference across genders (i.e., lack of invariance) for several items after controlling for consumption rates. In regard to racial differences in PBS use and racial invariance of PBS items, findings have been mixed. Related to harmful drinking patterns and alcohol quantity, Madson and Zeigler-Hill (2013) found serious harm reduction and limiting/stopping drinking strategies to be more protective for White students than Black students. An examination of alcohol-related problems by race and gender found that limiting/stopping strategies were protective for all groups (e.g., White men, White women, and Black men) except Black women (Madson & Zeigler-Hill, 2013). Martin et al. (2019) found six of 15 PBSS (Martens et al., 2005) item-factor loadings did not demonstrate equivalence across White and Black students suggesting the potential for measurement bias in those items.

Overall, PBS use appears to help reduce alcohol consumption and alcohol-related problems directly as well as indirectly by way of reducing the harm associated with numerous predictors of alcohol-related problems among college student drinkers. Hence, assessment of PBS holds promise as a focus for preventative interventions. However, one limitation of PBS research has been differences in how PBS are conceptualized and assessed (for a review of methodological issues see Prince et al., 2013).

Measurement of PBS

Extant literature notes key inconsistencies in PBS measurement with respect to conceptual and operational definitions and the method by which PBS is assessed. Limitations regarding the design and validation of various measures have also been observed in the literature. Specifically, definitions of PBS vary in scope, such that most researchers define PBS narrowly (e.g., strategies exclusively used during a drinking episode; Martens et al., 2005), yet others use a broad definition, including strategies used before, during, and after drinking episodes as well as those used to avoid drinking entirely (e.g., Sugarman & Carey, 2007). Among measures of PBS that have been subjected to factor analysis, the number of factors that have emerged has ranged from one to 10 (Prince et al., 2013), with the three-factor PBSS (i.e., stopping/limiting drinking, manner of drinking, and serious harm reduction strategies all used during a drinking episode; Martens et al., 2005) being the most commonly used measure (Pearson, 2013). Inconsistencies in the assessment of PBS include variations in response scales used (e.g., *yes/no* dichotomous response, multipoint contingent frequency, absolute frequency, and degree of truth scales; e.g., Kite, Pearson, & Henson, 2013; Luebke, Varvel, & Dude, 2009; Martens et al., 2005; Martens, Pederson, LaBrie, Ferrier, & Cimini, 2007).

Existing PBS measures have received varying degrees of support for their reliability and validity. While the PBSS (Martens et al., 2005) is the most widely used measure (Pearson, 2013) and one of the few that has received support for its factor structure and convergent and

incremental validity, few authors use the original measure. Variants of the original PBSS have added or removed items, used different time frames (i.e., the period of time across which participants are asked to report engagement in PBS use), and different response anchors (i.e., some use the original 6-point Likert-type scales, while others have used a 5-point scale and others *yes/no*). Such variation between scales and within the use of the same scale may contribute to inconsistent findings across studies and limit researchers' ability to synthesize data. Reviews of the PBS literature reveal that the design of some measures of PBS has not been rigorous, and more psychometric work needs to be done on nearly all existing measures (Prince et al., 2013). To our knowledge, none of the existing measures of PBS has been subjected to advanced item response theory analysis (IRT), which, for several reasons, is a logical next step in this line of research.

Item Response Theory and Protective Behavioral Strategies

Classical test theory (CTT) allows for several important analyses crucial in scale development, including the parceling of items into factors via factor analysis, and tests of construct validity and item-scale and interitem reliability. However, CTT statistics do not allow for the type of item level focus that allows for in depth look at the performance of individual items that IRT analyses do, particularly how items function along the underlying latent trait scale. An important step forward in the assessment of PBS is to examine the degree to which specific strategies map onto the latent continuum of PBS in regard to ease and frequency of use. IRT can provide information on how well individual items place an individual on a specific point along the PBS continuum (e.g., infrequent to frequent use). Therefore, use of IRT will allow us to create, to the best of our ability given existing PBS items, a measure of PBS that taps all levels of PBS use. Another advantage of IRT is that it allows for easy determination of the impact of reducing items on the performance of a measure. Hence, a short, content valid measure can be developed with relative ease (Embretson & Reise, 2000). IRT analyses are frequently followed up with DIF analyses to determine whether an item or items operates differently for different demographic subgroups, such as those based on race or gender.

In sum, using IRT to analyze PBS items will allow us to identify a much smaller set of psychometrically strong items free of gender and racial impact from a large pool of PBS items originating from multiple existing measures. Because IRT parameters are not sample dependent, it can be assumed that the relative positioning of the items in the scale will be consistent across similar samples of college students (Embretson & Reise, 2000).

Present Study

There is a vast number of existing measures of PBS, several of which have demonstrated measurement bias as applied to college men and women and White and Black college students, and many of which are in need of further psychometric evaluation. Existing measures have between six and 62 items, represent varying conceptualizations of the PBS construct, and use different response anchors and time frames. To advance research on PBS there is a need for a relatively brief, content valid measure of PBS that is relatively free of bias, exhibits evidence of validity, and can be readily and consistently used in survey studies

(Prince et al., 2013). The primary goal of the present study was to respond to the limitations associated with the measurement of PBS identified in prior research. Specifically, we used IRT to extract from a large pool of items taken from the most common measures of PBS, a small set of the most optimally functioning PBS items in terms of difficulty (i.e., level of the PBS trait associated with endorsement of each item), discrimination (i.e., relationship between items and the underlying PBS trait), and lack of DIF (i.e., measurement bias). Along with this, we endeavored for the items on the measure to represent the broadest definition of PBS, defined as strategies that can be used before, during, and after alcohol use as well as avoidance of alcohol use altogether (Pearson, 2013). Including on our measure a limited number of broadly defined strategies would address the limitation of most of these broadly defined PBS measures being too lengthy to be practically useful (Prince et al., 2013) and help those completing the measure to consider new strategies that could be used in different settings or situations.

Based on previous investigations of measurement invariance of PBS items across gender and racial groups, we hypothesized that some items would exhibit DIF across men and women and White and Black groups of students. For example, the following items have exhibited significant gender and racial differences: “Make sure you go home with a friend,” “Use a designated driver,” and “Avoid drinking games” (Clarke et al., 2016; Martin et al., 2019; Treloar, Martens, & McCarthy, 2014). No a priori hypotheses were made about which items would function most optimally in regard to item difficulty and discrimination. We aimed to retain 20 items or less to make the measure maximally accessible to researchers and clinicians whose goal is always to minimize burden on potential participants.

A second goal of the current study was to determine the most optimally functioning response anchors (i.e., text vs. numeric). The variety of response anchors used is just one of several methodological critiques of existing PBS measures said to potentially influence comparison of findings across studies (Pearson, 2013; Prince et al., 2013). To our knowledge, no investigations of the performance of various response anchors exists. Due to evidence of recall bias on reports of alcohol use among college student samples (Patrick & Lee, 2010), we hypothesized that the text-based response anchors (e.g., *sometimes*, *always*) would perform better statistically than the numeric response anchors that ask participants to recall a more specific number of times a particular PBS was used within a specific time period.

Finally, we wanted to provide evidence of the construct validity of the new measure. To do so, we correlated scores on the new PBS measure, called the Protective Drinking Practices Scale (PDPS), with measures of alcohol use and consequences commonly used with college student samples. It was expected that the optimal set of PBS items would be negatively correlated with alcohol use and problems as has been exhibited in an abundance of previous research.

Method

Participants

Participants were a convenience sample of 503 traditional age (18–25) undergraduate students ($M = 19.99$, $SD = 1.91$) at a mid-sized university in the Southern U.S. Participants were primarily female ($n = 336$) and White non-Hispanic ($n = 293$; 58.3%). Thirty-five percent of participants identified as Black, 1.8% Hispanic, 0.4% American Indian or Alaska Native, 2.0% Asian, 0.6% Native Hawaiian or Other Pacific Islander, and 1.4% identified as “Other.” The majority of students were classified as freshman ($n = 206$) or sophomores ($n = 131$), though juniors ($n = 88$) and seniors ($n = 76$) were also represented.

Procedure

Participants were recruited using the School of Psychology research participation system (Sona Systems) and received credit that could be used for partial fulfillment of a research requirement. To be eligible for the study participants must have been between the ages of 18–25, report having consumed alcohol at least once in the past 30 days, and provided informed consent. After registering for the study, participants were guided to a University Institutional Review Board informed consent page where they provided electronic consent. Upon providing informed consent, participants completed a demographic form followed by the study measures presented randomly in an effort to minimize order effects. All surveys were presented using a secure online survey system (Qualtrics).

Measures

Demographic questionnaire.—Participants completed a demographic questionnaire that included information about gender, race, age, year in school, and enrollment status.

Typical weekly drinking.—The Daily Drinking Questionnaire (DDQ; Collins, Parks, & Marlatt, 1985) was used to assess typical alcohol consumption. It asks participants to report the number of standard drinks they consumed each day during a typical week. The number of typical drinks per week was calculated by summing the values across all 7 days.

Hazardous and harmful drinking.—The 10-item Alcohol Use Disorders Identification Test (AUDIT; Saunders, Aasland, Babor, de la Fuente, & Grant, 1993) was used to assess hazardous and harmful drinking. Instructions ask that participants respond to questions that assess quantity and frequency of alcohol consumption, negative consequences associated with alcohol use, and symptoms of alcohol dependence using a scale ranging from 0 to 4. Higher AUDIT scores suggest greater severity of alcohol use and/or problems. Internal consistency in this study was adequate ($\alpha = .79$).

Alcohol-related problems.—The Rutgers Alcohol Problem Index (RAPI; White & Labouvie, 1989) is a 23-item self-report instrument that measures frequency of alcohol-related consequences (0 = *never* to 5 = *more than 10 times*) that occurred in the past year. Items assess alcohol abuse and dependence (e.g., “Wanted to stop drinking but you couldn’t”), social/interpersonal consequences (e.g., “Friends or neighbors avoided you”), and personal consequences (e.g., “Passed out or fainted suddenly”). Participants are asked

how often each of the consequences had occurred while drinking, or because of drinking; higher scores indicate more alcohol-related consequences. Internal consistency estimates among college students typically range from .85 to .95; Cronbach's alpha in the present sample was .95 (for a review see Devos-Comby & Lange, 2008).

Protective behavioral strategies.—Protective behavioral strategies were assessed by combining the nonoverlapping items from six different measures of PBS for a total of 68 nonoverlapping items. Items were extracted from the following measures: the Protective Behavioral Strategies Scale (PBSS; Martens et al., 2005), the Protective Behavioral Strategies Scale—Revised (PBSS-R; Madson, Arnau, & Lambert, 2013), the Strategy Questionnaire (Sugarman & Carey, 2007), the Self-Control Questionnaire (SCQ; Werch, 1990), the Protective Strategies Questionnaire (PSQ; Palmer, 2004), and the Protective Behavioral Strategies Measure (PBSM; Novik & Boekeloo, 2011). Each measure has demonstrated evidence of construct validity via correlations in the expected direction with alcohol use and alcohol-related problems and adequate reliability among college student samples (for reviews of each measure see Pearson, 2013 and Prince et al., 2013).

The entire set of 68 nonoverlapping PBS items were administered twice in the survey with the same instructions (i.e., “How often you have used the following strategies in the past 3 months?”), but with two different response anchors. The first set of anchors were the text from the PBSS (Martens et al., 2005) ranging from 1 (*never*) to 6 (*always*). The second set of anchors were numeric and included 0, 1, 2–3, 4–5, 6–10, 11 or more. These anchors had the same 1–6 values as the text anchors (i.e., 1 = 0, 2 = 1, 3 = 2–3, 4 = 4–5, 5 = 6–10, 6 = 11 or more). To minimize order effects, administration of the two versions of the PBS scale was counterbalanced so that half saw the text-based responses first, and the other half saw the numeric-responses, first. The time frame (three months) and response anchors that correspond to the PBSS were chosen since it is the most frequently used and most extensively researched measure of PBS (Prince et al., 2013).

Data Analysis

Graded response model.—Samejima's graded response model (GRM; Samejima, 1997) is an example of an item response theory (IRT) model for polytomous items (i.e., items with more than two response options). The GRM allows us to estimate simultaneously the relative placement of each item along the latent trait scale and the respondents' locations along the same scale, with an IRT-based trait estimate. Using the mirt package (Chalmers, 2012) for the R environment (R Core Team, 2013), the GRM was fit to two versions of the PBS items (i.e., text and numeric response anchors). The GRM models the probability of a response, u , equal to or greater than a response category, x , for a given location along the latent trait scale, θ_i :

$$P(u \geq x | \theta_i) = \frac{e^{a_j(\theta_i - b_{jx})}}{1 + e^{a_j(\theta_i - b_{jx})}}, \quad (1)$$

where i refers to the person, a_j and b_j are the discrimination and threshold locations, respectively, of the particular response category. In other words, the probability of a person

responding with u of 3 or higher on a scale of 1 to 6, in our case, given the person has a particular θ on the latent trait scale. The probability of a response, u , equal to 3 is found by subtracting adjacent cumulative probabilities: $P(u \geq 3 | \theta_j) - P(u \geq 4 | \theta_j)$. These differences are the individual response categories and depicted in an item's response probability curves.

The PBS items were 6-point Likert-type items. For each item, the graded response model estimates the likelihood of every response, from 1 to 6, along the scale; an example is provided in Figure 1. The estimated probability curves can be used to help assess how respondents are using the response options. Figure 1 displays an ideal response, where the probability curves are spread out and at some point along the scale, each response is most probable, or not completely overlapped by another curve. In this study, the resulting person scores were fixed to have a mean of 0.0, with a standard deviation of 1.0. For the item shown in Figure 1, "Spacing drinks over time," of respondents whose total score is 0.5, that is, one half of a standard deviation above the mean score, 60% are predicted to respond *usually*, while about 20% are expected to respond *occasionally* and 20% *always*.

Because the GRM has an assumption of unidimensionality, a principal components analysis was conducted to ensure that the data were unidimensional enough for IRT, that is, the first component accounted for at least 20% of the variance (Reckase, 1979). Yen's $Q3$ statistic (Yen, 1984) was used to evaluate local item dependence, where average values should be close to 0. To further assess model fit, standardized residuals should be unimodal with a mean close to 0. Finally, item-level fit was evaluated with Yen's $Q1$ statistic, distributed χ^2 , where the degrees of freedom equal number of groupings less the number of model parameters (Yen, 1981).

Comparison of two sets of items/two PBS scales.—The first step in our analysis was to determine whether to use responses from the items when presented with text-based anchors or numeric-based anchors. The results of fitting the GRM to both sets of response options were compared; specifically, we considered the relationship of the IRT-based scores with observed total scores on each scale, the item-level fit, standard error measurement along the entire scale, and external validity based on the correlations of the IRT-based scores with alcohol use and alcohol-related problems as measured by the DDQ, RAPI, and AUDIT. Item-level fit was evaluated with both Yen's $Q1$ and the relative locations of the probability curves for a given item, where distinct probability curves along the trait scale would be most desirable and indicative of good separation of the response options.

Differential item functioning.—DIF occurs when subjects who have the same level of the underlying trait and should be predicted to have the same responses to a given item, actually differ in their predicted responses, due to an external factor, such as gender (Clauser & Mazor, 1998). DIF is a violation of unidimensionality in that the probability of a response to a given item is affected by not just the latent trait, but also by some other factor. Two DIF analyses, one for gender ($N = 503$) and one for race ($N = 471$; White, $n = 293$ vs. Black, $n = 178$), were conducted using ordinal logistic regression (Crane, Gibbons, Jolley, & van Belle, 2006; Zumbo, 1999) to determine whether the demographic variable affected the response probabilities.

For each item, three logistic regression models were fit; Model 1: the null model with only the latent trait (as estimated by the IRT-based score) as a coefficient; Model 2: in addition to the latent trait, group membership was added to the model; and Model 3: included the latent trait, group membership, and the interaction of the latent trait by group membership. An item was considered to have exhibited DIF if the coefficient for group membership in Model 2 or the coefficient for the interaction between group membership and the latent trait in Model 3 was statistically significant at an alpha of .01. DIF effect sizes were computed using McFadden's likelihood ratio R^2 (Menard, 2000):

$$R^2 = 1 - \frac{\log \text{likelihood}(\text{full model})}{\log \text{likelihood}(\text{null model})}, \quad (2)$$

where three pairs of nested models were used (Choi, Gibbons, & Crane, 2011). Using Cohen (1988) effect size cutoffs, items where the maximum of the three computed McFadden's R^2 s is greater than .02 are considered to exhibit small DIF, greater than .13 medium DIF, and greater than .26 are considered large DIF (Kim, Cohen, Alagoz, & Kim, 2007).

Creation of the Protective Drinking Practices Scale.—To create a short form that functioned equally as well as the 68-item original measure of PBS, the next step was to fit the GRM to the non-DIF items. From these remaining items we looked for items that provided the most information (i.e., were highly discriminating), had good spread of response probability curves across the scale, and had good item-level fit. Items were not considered if they offered little information, or had large, patterned residuals. In addition to the psychometric properties, the content of the items was considered. For example, items with similar locations along the scale and similar item information were compared based on the content of the items; the item that best reflected the construct was selected. This process yielded the set of items that would be the first proposed short form.

Next, the GRM was fit to the proposed short form. The total test information, standard errors of the IRT-based scores, and correlations between these scores and the external variables (i.e., alcohol use and alcohol-related problems) were compared for the proposed short form and the larger set of all non-DIF items. As evidence that the short form was functioning well, a similar pattern in total test information and standard errors was desired across the two forms, as well as a strong correlation between the IRT-based scores resulting from both forms. A scatterplot of the two sets of scores was used to identify any possible floor or ceiling effects created by eliminating items. Additional non-DIF items, selected based on content, item information, item location, and small, random residuals plots, were then added to the proposed short form. The same comparisons were made between the new short forms and the original set of non-DIF items until a satisfactory short form was obtained.

Results

Dimensionality and Model Fit

Based on a principal components analysis, the full set of the original 68 items was found to be unidimensional enough for the IRT analysis with the first eigenvalue explaining 55.5% of the variance. The ratio of the first to second eigenvalues was 8.18. No issues with local item

dependence were found (Yen's $Q3: M = .01, SD = .17$). However, an inspection of the standardized residuals plots and estimated probability curves revealed that two items had severe misfit: "Leave a drink unattended" and "Punished self for drinking alcohol," where the residuals indicated a pattern of overestimation of the predicted response, rather than a random pattern, and that only the extreme response categories were most probable (see Figure 2). These two items were removed from further analysis. For the final short form of the Protective Drinking Practices Scale, the first eigenvalue explained 62.3% of the variance and ratio of the first to second eigenvalues was 10.5. Coupled with a check of local dependence (Yen's $Q3: M = -.03, SD = .13$) and the standardized residuals, no issues with model fit were found.

Comparing the Two Versions of the PBS Scale

Both versions of the scale (i.e., text and numeric response anchors) had some items where the item-level fit was better with one version than the other. In fact, the IRT-based scores and their standard errors were quite similar across both versions, although they differed at the extremes of the scale. For most items, the probability response curves were better defined with text-based responses, as seen in Figure 3. With the numeric response anchors, many items displayed overlapping of categories, which indicated that the responses were not separated along the scale and that respondents who were otherwise matched on total IRT-score were predicted to give different responses. Scores from the numeric-based scale had almost no correlation with the three external measures of alcohol use and problems (absolute values ranged from .03 to .07), while the text-based version had some relationship with alcohol use and problems (absolute values ranged from .15 to .23). For these reasons, the text-based version of PBS was used for the remainder of the analyses and to construct the Protective Drinking Practices Scale.

Differential Item Functioning

After checking for model and item fit, the 66 items were subjected to two DIF analyses: one based on gender, the other race. Due to limitations of sample size within ethnic groups, only a comparison of White and Black participants was conducted. An item is considered to exhibit DIF when males and females, for example, who are matched on the same IRT-based score, are predicted to have different responses to an item. The DIF analysis comparing the probability of response for males and females, matched on their latent trait score, identified seven items as exhibiting DIF, and 24 items were identified in the DIF analysis based on race (one item exhibited DIF for both gender and race; see Table 1). Because the goal of this study was to develop a brief, optimal measure of PBS that functions equivalently for men and women of diverse racial identities, all items found to exhibit DIF were removed, leaving 36 non-DIF items for further consideration. After fitting the GRM to the remaining 36 items, the IRT-based scores correlated almost perfectly with scores based on the larger set of 66 items ($r = .99$), providing evidence for the validity of the 36 remaining PBS items.

Creation of Protective Drinking Practices Scale and Tests of Validity

After removing the DIF items, a set of strong items was identified; these items provided the most statistical information, had strong item-level fit and residuals plots, had distinct response probability curves, and minimized content duplication with other items. At this

point items were evaluated for both their psychometric properties and based on their content. We tested different sets of items, removing items that were almost identical in content and statistical properties. Items were selected to maximize information along the entire continuum of the PBS latent construct and maintain a small conditional standard error of measurement. Ultimately, a set of 20 items was selected to comprise the PDPS. The resulting IRT location parameters for the final set of 20 items are depicted in Table 2. Items on the PDPS had a coefficient alpha of .95 and an IRT-based marginal reliability of .96. The PDPS had a dominant first component explaining 62.3% of the variance and the ratio of the first to second component was about 10.5:1. Fit statistics also demonstrated adequate fit with respect to dimensionality ($M2_{df=90}^* = 265.80$, RMSEA = .065, 90% CI [.056, .074]), there is adequate unidimensionality for an IRT analysis when RMSEA is less than .089 Maydeu-Olivares & Joe (2014).

The total test information curve for the PDPS had a similar shape to that of the 36 non-DIF items; the curve was also wide rather than pointed, indicating that the items in the scale tapped a wide range of the latent trait (see Figure 4). Also shown in Figure 4 is the distribution of IRT-based scores, demonstrating that the short form items were targeting the current sample. The IRT-based scores from the 20-item PDPS correlate nearly perfectly ($r = .99$) with scores from the 36 non-DIF items and were moderately correlated in the expected direction with the DDQ, $r = -.21$, $p < .001$; RAPI, $r = -.15$, $p < .01$; and AUDIT, $r = -.24$, $p < .001$, providing support for the validity of test score interpretations.

Discussion

The main goal of the study was to use IRT analytic techniques to create a brief measure of PBS wherein the items represent the broadest definition of the construct (i.e., strategies that can be used before, during, and after drinking as well as alcohol avoidance strategies) and function equivalently across diverse groups of college students so as to enhance the generalizability of future survey research. A second goal was to determine whether text or numeric anchors were more valid from a psychometric standpoint, because existing measures of PBS use both. Finally, we aimed to provide evidence of the validity of the set of items we chose to comprise the Protective Drinking Practices Scale (PDPS) by demonstrating the relationship between scores on the measure and various assessments of alcohol use and alcohol-related problems commonly used with college student samples.

The first decision to be made was whether to use items utilizing the 6-point Likert type text response anchors ranging from 1 (*never*) to 6 (*always*) or a 6-point Likert-type scale with numeric response anchors ranging from 1 (*0 times*) to 6 (*11 or more times*). Based on statistical and psychometric considerations, the numeric response anchors did not hold up well. As shown in Figure 2, the probability curve for each response was not distinct for many items and the IRT-based scores for the numeric version did not correlate with measures of drinking behaviors. It appears that the numeric anchors did not allow the respondent to adequately reflect their behavior and use of such anchors may lead to misleading information for researchers. For example, if a person drank with friends twice in the past 3 months and used a designated driver on both occasions, on the version of the scale with numeric response options the person would respond with a 3 (*2 to 3 times*), but on the

version with text-based response options the response would be 6 (*always*). This inconsistency likely led to the poor item-level fit for the numeric version. Having determined the superiority of the text-based response anchor when used within the 3-month time-frame provided to participants, analyses were next focused on determining DIF in regard to gender and racial/ethnic identity.

Deciding which items exhibited DIF based on the significance of the resulting coefficients from ordinal logistic regression, rather than the effect sizes, meant that items were liberally discarded, removing 30 of 66 items, rather than none as would have been the case had we chosen to use effect size as the criterion. The rationale for our decision to use the significance of the resulting coefficients was that the goal of the study was to establish a PBS measure that optimally measured the construct (i.e., provided as much information as possible with as few items as possible) equivalently across White and Black men and women. Hence, we wanted to err on the side of caution by removing items that could possibly function differentially for different groups to ensure that the final set of items were particularly strong.

Differential performance of several items was expected based on findings of previous studies. For example, Treloar et al. (2014) found that women were more likely than men to endorse a higher response category (indicating more frequent use) to the items “Make sure you go home with a friend” and were more likely than men to endorse a lower response category (indicating less frequent use) to the item “Use a designated driver.” Martin et al. (2019) found a lack of measurement invariance for both of the aforementioned items and examination of item means indicated that White students reported more frequent use of both PBS items. Martin et al. (2019) and Clarke et al. (2016) also found significant racial differences for “Avoid drinking games” such that White respondents utilized this strategy less often than Black respondents did. Treloar et al. (2014) found that women reported avoiding drinking games more often than men. Our results indicated that, as expected, “Use a designated driver” and “Avoid drinking games” exhibited DIF in regard to race/ethnicity. In contrast to previous findings, neither item exhibited DIF based on gender. The individual item response curves for Black and White students on the “Avoid drinking games” item are shown in Figure 5. The figure shows that Black and White students who were both about one standard deviation below the mean on PBS have different probabilities of responding *never* to the item, with Black students at that trait level estimated to respond *never* 50% of the time, while White students were estimated to respond with *never* 80% of the time. This finding may reflect research that suggests that Black students engage in drinking games less frequently than White students (Zamboanga et al., 2015).

Also as expected the DIF analyses by race and by gender revealed a similar, nonuniform DIF pattern for “Make sure you go home with a friend.” Black students with PBS latent trait scores about one-and-a-half standard deviations below average were predicted to respond *never* to the item when White students, with the same score, were predicted to respond *rarely*. However, for White and Black students with latent trait scores around the mean or higher, the response patterns of Black and White students were the same. When comparing males and females matched on the same score for the item “Make sure you go home with a friend,” females were predicted to be more likely to respond positively than males with the

same latent trait score. Another way to interpret this is that females moved from a weaker to a stronger response (e.g., *sometimes* to *always*) at a lower latent trait score than males. The reasons for observed differences in responses to various PBS items across genders and races is difficult to know because no research has investigated this. However, we can speculate that cultural differences across college men and women and White and Black students having to do with drinking norms such as where, when, and how alcohol is consumed, may impact the relevance or utility of particular PBS for members of certain demographic groups.

It is important to consider that while differences in the psychometric performance of items could make it difficult to properly compare PBS use across men and women or White and Black students, those items exhibiting differences in performance may be most useful in the development of a personalized measure of PBS to be used in experimental investigations of interventions that contain PBS (Prince et al., 2013). Based on our results and the findings of others, it seems that a personalized approach to using PBS as an intervention tool would be focused on different PBS for men versus women and White versus Black students. Within those groups, researchers and clinicians should consider the utility of specific strategies for unique individuals. Future research should investigate why we observe racial and gender differences for various PBS items and how the use of personalized approaches to the measurement and discussion of PBS may impact the effectiveness of interventions.

The final set of 20 items chosen for the PDPS represent strategies that can be used before, during, and after drinking episodes as well as alcohol avoidance strategies. Creation of a unidimensional, relatively brief measure of broadly defined PBS strategies addresses critiques of existing measures that have complicated factor structures and scoring criteria and excessive length that can limit their practical use (Prince et al., 2013). Further, we intentionally chose a small set of items that represented a broad range of item difficulty to ensure that the measure represented strategies that may be applicable to a broad range of people and contexts. Each item, by itself, covered a wide range of item difficulty from about a standard deviation below the mean latent trait score to anywhere from about .3 to 1.5 standard deviations above the mean. Almost four times as many people responded with *always* for the easiest to endorse item (i.e., “Have a plan with a friend to watch out for each other”) compared with the hardest (i.e., “Thought about problems drinking can cause”), while about half as many people responded *never* to the easiest item compared to the hardest item. Future research focused on understanding what makes particular strategies more or less difficult to use, considering such things as conformity to gender or peer norms (Whitley, Madson, & Zeigler-Hill, 2018) and drinking context (Villarosa-Hurlocker, Madson, Mohn, Zeigler-Hill, & Nicholson, 2018), may help clinicians talk to students about how to increase their use of particular strategies in the face of obstacles.

Validity evidence indicated that scores on the PDPS correlated in the expected direction with all measures of alcohol use and alcohol-related problems. The strongest negative relationship was between scores on the PDPS and hazardous and harmful alcohol use as measured by the AUDIT. These findings provide support for the construct validity of the measure, as increased use of a broad range of strategies are intended to reduce not only alcohol use, but also the harms associated with use.

Because some researchers and clinicians are interested in obtaining a quick snapshot of a person's use of PBS but have minimal time to do so, such as when utilizing an ecological momentary assessment approach to data collection or as part of an intake assessment in a clinical setting, we attempted to identify an extremely short form of the PDPS by selecting three of the stronger short-form items that are spread across the latent trait scale. The IRT scores from this three-item version correlated .94 with the IRT scores from the 20-item version of the PDPS. The three items were: "Spacing drinks over time," "Only drink in safe environments," and "Avoid hard liquor." Because so many of the 20-item version were strong, other sets of three items would also be useful. Future research could investigate the psychometric properties and comparability of various three-item scales.

Limitations and Future Directions

Several limitations of the study should be considered and addressed in future research. The retrospective self-report nature of measures of alcohol use, alcohol-problems, and PBS allow for the possibility that recall bias impacted findings. Future research should aim to replicate and extend the current study by using ecological momentary assessments of alcohol outcomes and PBS. A limitation of our sample was that it was a convenience sample obtained from one university in the southeastern U.S. It is important for future research to examine the performance of the 20 PBS items and relationships between PBS and alcohol outcomes on diverse college student samples obtained from various regions of the U.S. to increase generalizability of our findings. Especially important will be future research that includes a larger, more diverse college student sample that will allow for additional DIF analyses on more racial and ethnic groups. Future research should also investigate the relevance and utility of various PBS for male and female college students and students who identify as White and Black to better understand DIF results obtained with the present study sample. Strengths of the associations we obtained between PBS, alcohol use, and problems, while moderate to small, are consistent with findings of other studies (e.g., Borden et al., 2011; LaBrie, Lac, Kenney, & Mirza, 2011; Madson, Moorer et al., 2013). Nonetheless, this limitation should be considered in future studies of PBS as clearly there is a great deal of variance in alcohol use accounted for by other factors.

Beyond confirming that the 20 items comprising the PDPS were essentially unidimensional to allow for the IRT analysis, we did not examine the factor structure of the PDPS as that was beyond the scope of the present study. However, it will be important for researchers to do so because most studies tend to find measures of PBS strategies are multidimensional (i.e., typically at least three subscales). We examined the most common forms of response anchors used on existing PBS measures (i.e., numeric and text-based Likert type scales) and found that the text-based anchors outperformed the numeric anchors from a psychometric standpoint. Some measures of PBS use a *yes/no* response format, which would be important to examine against the text-based Likert-type response anchors in a future IRT study. It may be that a simple *yes/no* format would perform equally as well as the more complex Likert-type scale.

Summary

Findings of the current study extends literature on the measurement of PBS by addressing identified limitations using IRT analysis to examine individual PBS items representing a host of different PBS measures. The Protective Drinking Practices Scale (PDPS) improves upon existing measures that did not have psychometric support, or that used narrow definition of PBS, or that contained items that functioned differently for individuals of different genders or races. This study contributes to the literature a brief measure of PBS comprised of items representing the broadest definition of PBS (i.e., can be used before, during, and after drinking episodes, as well as avoidance of drinking strategies) and a range of difficulty from relatively easy to endorse through difficult to endorse. It also assesses PBS equivalently across college men and women identifying as White and Black. We anticipate that the brevity and psychometric support for this measure will increase the likelihood of its use in future research. Widespread use of this measure will enhance the comparability of findings from survey studies of PBS making it possible to conduct a meta-analysis of PBS that will help researchers develop more targeted and impactful research on PBS as a mechanism of change.

References

- Barnett NP, Murphy JG, Colby SM, & Monti PM (2007). Efficacy of counselor vs. computer-delivered intervention with mandated college students. *Addictive Behaviors*, 32, 2529–2548. 10.1016/j.addbeh.2007.06.017 [PubMed: 17707594]
- Borden LA, Martens MP, McBride MA, Sheline KT, Bloch KK, & Dude K (2011). The role of college students' use of protective behavioral strategies in the relation between binge drinking and alcohol-related problems. *Psychology of Addictive Behaviors*, 25, 346–351. 10.1037/a0022678 [PubMed: 21443303]
- Chalmers RP (2012). mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, 48, 1–29. 10.18637/jss.v048.i06
- Choi SW, Gibbons LE, & Crane PK (2011). lordif: An R Package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software*, 39, 1–30. 10.18637/jss.v039.i08
- Clarke N, Kim SY, Ray AE, White HR, Jiao Y, & Mun EY (2016). The association between protective behavioral strategies and alcohol-related problems: An examination of race and gender differences among college drinkers. *Journal of Ethnicity in Substance Abuse*, 15, 25–45. 10.1080/15332640.2014.1002877 [PubMed: 26114577]
- Clauser BE, & Mazor KM (1998). Using statistical procedures to identify differentially functioning test items. *An NCME Instructional Module. Educational Measurement: Issues and Practice*, 17, 31–44. 10.1111/j.1745-3992.1998.tb00619.x
- Cohen J (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Collins RL, Parks GA, & Marlatt GA (1985). Social determinants of alcohol consumption: The effects of social interaction and model status on the self-administration of alcohol. *Journal of Consulting and Clinical Psychology*, 53, 189–200. 10.1037/0022-006X.53.2.189 [PubMed: 3998247]
- Crane PK, Gibbons LE, Jolley L, & van Belle G (2006). Differential item functioning analysis with ordinal logistic regression techniques. DIFdetect and difwithpar. *Medical Care*, 44, S115–S123. 10.1097/01.mlr.0000245183.28384.ed [PubMed: 17060818]
- Devos-Comby L, & Lange JE (2008). Standardized measures of alcohol-related problems: A review of their use among college students. *Psychology of Addictive Behaviors*, 22, 349–361. 10.1037/0893-164X.22.3.349 [PubMed: 18778128]

- Embretson SE, & Reise SP (2000). Item response theory for psychologists. Mahwah, NJ: Erlbaum, Inc.
- Hingson RW, Zha W, & White AM (2017). Drinking beyond the binge threshold: Predictors, consequences, and changes in the U.S. *American Journal of Preventive Medicine*, 52, 717–727. [PubMed: 28526355]
- Jordan HR, Madson MB, Nicholson BC, Bravo AJ, Pearson MR, & Protective Strategies Study Team. (2019). Posttraumatic stress disorder symptoms and problematic alcohol use in college students: The moderating role of alcohol protective behavioral strategies and gender. *Psychological Trauma: Theory, Research, Practice, and Policy*, 11, 247–255.
- Kenney SR, & LaBrie JW (2013). Use of protective behavioral strategies and reduced alcohol risk: Examining the moderating effects of mental health, gender, and race. *Psychology of Addictive Behaviors*, 27, 997–1009. 10.1037/a0033262 [PubMed: 24079648]
- Kim SH, Cohen AS, Alagoz C, & Kim S (2007). DIF detection effect size measures for polytomously scored items. *Journal of Educational Measurement*, 44, 93–116. 10.1111/j.1745-3984.2007.00029.x
- Kite BA, Pearson MR, & Henson JM (2013). The assessment of protective behavioral strategies: Comparing the absolute frequency and contingent frequency response scales. *Psychology of Addictive Behaviors*, 27, 1010–1018. 10.1037/a0031366 [PubMed: 23438243]
- LaBrie JW, Lac A, Kenney SR, & Mirza T (2011). Protective behavioral strategies mediate the effect of drinking motives on alcohol use among heavy drinking college students: Gender and race differences. *Addictive Behaviors*, 36, 354–361. 10.1016/j.addbeh.2010.12.013 [PubMed: 21215529]
- Larimer ME, & Cronce JM (2007). Identification, prevention, and treatment revisited: Individual-focused college drinking prevention strategies 1999–2006. *Addictive Behaviors*, 32, 2439–2468. 10.1016/j.addbeh.2007.05.006 [PubMed: 17604915]
- Larimer ME, Lee CM, Kilmer JR, Fabiano PM, Stark CB, Geisner IM, ... Neighbors C (2007). Personalized mailed feedback for college drinking prevention: A randomized clinical trial. *Journal of Consulting and Clinical Psychology*, 75, 285–293. 10.1037/0022-006X.75.2.285 [PubMed: 17469886]
- Lewis MA, Kaysen DL, Rees M, & Woods BA (2010). The relationship between condom-related protective behavioral strategies and condom use among college students: Global- and event-level evaluations. *Journal of Sex Research*, 47, 471–478. 10.1080/00224490903132069 [PubMed: 19657943]
- Linden AN, Lau-Barraco C, & Millettich RJ (2013). The role of protective behavioral strategies and anxiety in problematic drinking among college students. *Journal of Studies on Alcohol and Drugs*, 74, 413–422. 10.15288/jsad.2013.74.413 [PubMed: 23490570]
- Longo LM, Martin JL, Zamboanga BL, Milroy JJ, & Wyrlic DL (2018). Ethnic, gender and seasonal differences in heavy drinking and protective behavioral strategies among student-athletes. *Journal of Ethnicity in Substance Abuse*, 12, 1–18. 10.1080/15332640.2018.1520172
- Luebke AM, Varvel S, & Dude K (2009). Reciprocal relations of protective behavioral strategies and risk-amplifying behaviors with alcohol-related consequences: Targets for intervention with female college students. *Journal of Alcohol and Drug Education*, 53, 7–30.
- Madson MB, Arnau RC, & Lambert SJ (2013). Development and psychometric evaluation of the revised Protective Behavioral Strategies Scale. *Psychological Assessment*, 25, 556–567. 10.1037/a0031788 [PubMed: 23397929]
- Madson MB, Moorer KD, Zeigler-Hill V, Bonnell MA, & Villarosa M (2013). Alcohol expectancies, protective behavioral strategies, and alcohol-related outcomes: A moderated mediation study. *Drugs Education Prevention & Policy*, 20, 286–296. 10.3109/09687637.2013.766788
- Madson MB, & Zeigler-Hill V (2013). Protective behavioral strategies, alcohol consumption, and negative alcohol-related consequences: Do race and gender moderate these associations? *Journal of Ethnicity in Substance Abuse*, 12, 242–258. 10.1080/15332640.2013.798848 [PubMed: 23967885]

- Martens MP, Ferrier AG, & Cimini MD (2007). Do protective behavioral strategies mediate the relationship between drinking motives and alcohol use in college students? *Journal of Studies on Alcohol and Drugs*, 68, 106–114. 10.15288/jsad.2007.68.106 [PubMed: 17149524]
- Martens MP, Ferrier AG, Sheehy MJ, Corbett K, Anderson DA, & Simmons A (2005). Development of the Protective Behavioral Strategies Survey. *Journal of Studies on Alcohol*, 66, 698–705. 10.15288/jsa.2005.66.698 [PubMed: 16329461]
- Martens MP, Martin JL, Hatchett ES, Fowler RM, Fleming KM, Karakashian MA, & Cimini MD (2008). Protective behavioral strategies and the relationship between depressive symptoms and alcohol-related negative consequences among college students. *Journal of Counseling Psychology*, 55, 535–541. 10.1037/a0013588 [PubMed: 22017560]
- Martens MP, Pederson ER, LaBrie JW, Ferrier AG, & Cimini MD (2007). Measuring alcohol-related protective behavioral strategies among college students: Further examination of the Protective Behavioral Strategies Scale. *Psychology of Addictive Behaviors*, 21, 307–315. 10.1037/0893-164X.21.3.307 [PubMed: 17874881]
- Martin JL, Zamboanga BL, Haase RF, & Buckner L (2019). Measurement invariance of the Protective Behavioral Strategies Scale across racial groups. *Measurement and Evaluation in Counseling and Development*, 53, 17–33. 10.1080/07481756.2019.1640615
- Maydeu-Olivares A, & Joe H (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49, 305–328. 10.1080/00273171.2014.911075 [PubMed: 26765800]
- Menard S (2000). Coefficients of determination for multiple logistic regression analysis. *The American Statistician*, 54, 17–24.
- Murphy JG, Dennhardt AA, Skidmore JR, Borsari B, Barnett NP, Colby SM, & Martens MP (2012). A randomized controlled trial of a behavioral economic supplement to brief motivational interventions for college drinking. *Journal of Consulting and Clinical Psychology*, 80, 876–886. 10.1037/a0028763 [PubMed: 22663899]
- National Institute on Alcohol Abuse and Alcoholism. (2004). NIAAA council approves definition of binge drinking. *NIAAA Newsletter*, 3, 3.
- Novik MG, & Boekeloo BO (2011). Dimensionality and psychometric analysis of an alcohol protective behavioral strategies scale. *Journal of Drug Education*, 41, 65–78. 10.2190/DE.41.1.d [PubMed: 21675325]
- Palmer RS (2004). Efficacy of the Alcohol Skills Training Program in mandated and non-mandated heavy drinking college students (Unpublished doctoral dissertation). University of Washington, Seattle, Washington.
- Palmer RS, McMahon TJ, Rounsaville BJ, & Ball SA (2010). Coercive sexual experiences, protective behavioral strategies, alcohol expectancies and consumption among male and female college students. *Journal of Interpersonal Violence*, 25, 1563–1578. 10.1177/0886260509354581 [PubMed: 20040711]
- Patrick ME (2016). A call for research on high-intensity alcohol use. *Alcoholism, Clinical and Experimental Research*, 40, 256–259. 10.1111/acer.12945
- Patrick ME, & Lee CM (2010). Comparing numbers of drinks: College students' reports from retrospective summary, followback, and prospective daily diary measures. *Journal of Studies on Alcohol and Drugs*, 71, 554–561. 10.15288/jsad.2010.71.554 [PubMed: 20553664]
- Pearson MR (2013). Use of alcohol protective behavioral strategies among college students: A critical review. *Clinical Psychology Review*, 33, 1025–1040. 10.1016/j.cpr.2013.08.006 [PubMed: 24036089]
- Prince MA, Carey KB, & Maisto SA (2013). Protective behavioral strategies for reducing alcohol involvement: A review of the methodological issues. *Addictive Behaviors*, 38, 2343–2351. 10.1016/j.addbeh.2013.03.010 [PubMed: 23584196]
- R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing Retrieved from <http://www.R-project.org/>
- Reckase MD (1979). Unifactor latent trait models applied to multifactor tests: Results and implications. *Journal of Educational Statistics*, 4, 207–230. 10.3102/10769986004003207

- Samejima F (1997). Graded response model In van der Linden WJ & Hambleton RK (Eds.), *Handbook of modern item response theory* (pp. 85–100). New York, NY: Springer
10.1007/978-1-4757-2691-6_5
- Saunders JB, Aasland OG, Babor TF, de la Fuente JR, & Grant M (1993). Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol Consumption-II. *Addiction*, 88, 791–804. 10.1111/j.1360-0443.1993.tb02093.x [PubMed: 8329970]
- Schulenberg JE, Johnston LD, O'Malley PM, Bachman JG, Miech RA, & Patrick ME (2018). *Monitoring the Future national survey results on drug use, 1975–2017: Volume II, college students and adults ages 19–55*. Ann Arbor: Institute for Social Research, The University of Michigan.
- Sugarman DE, & Carey KB (2007). The relationship between drinking control strategies and college student alcohol use. *Psychology of Addictive Behaviors*, 21, 338–345.
10.1037/0893-164X.21.3.338 [PubMed: 17874884]
- Treloar HR, Martens MP, & McCarthy DM (2014). Testing measurement invariance of the protective behavioral strategies scale in college men and women. *Psychological Assessment*, 26, 307–313.
10.1037/a0034471 [PubMed: 24059477]
- Villarosa MC, Messer MA, Madson MB, & Zeigler-Hill V (2018). Depressive symptoms and drinking outcomes: The mediating role of drinking motives and protective behavioral strategies among college students. *Substance Use & Misuse*, 53, 143–153. 10.1080/10826084.2017.1327974 [PubMed: 28813174]
- Villarosa MC, Moorer KD, Madson MB, Zeigler-Hill V, & Noble JJ (2014). Social anxiety and negative alcohol-related consequences among college drinkers: Do protective behavioral strategies mediate the association? *Psychology of Addictive Behaviors*, 28, 887–892. 10.1037/a0037628 [PubMed: 25222176]
- Villarosa-Hurlocker MC, Madson MB, Mohn RS, Zeigler-Hill V, & Nicholson BC (2018). Social anxiety and alcohol-related outcomes: The mediating role of drinking context and protective strategies. *Addiction Research & Theory*, 26, 396–404.
- Werch CE (1990). Behavioral self-control strategies for deliberately limiting drinking among college students. *Addictive Behaviors*, 15, 119–128. 10.1016/0306-4603(90)90015-P [PubMed: 2343785]
- White A, & Hingson R (2013). The burden of alcohol use: Excessive alcohol consumption and related consequences among college students. *Alcohol Research*, 35, 201–218. [PubMed: 24881329]
- White AM, Kraus CL, & Swartzwelder H (2006). Many college freshmen drink at levels far beyond the binge threshold. *Alcoholism, Clinical and Experimental Research*, 30, 1006–1010. 10.1111/j.1530-0277.2006.00122.x
- White HR, & Labouvie EW (1989). Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol*, 50, 30–37. 10.15288/jsa.1989.50.30 [PubMed: 2927120]
- Whitley RB, Madson MB, & Zeigler-Hill V (2018). Protective behavioral strategies and harmful alcohol use among male college students: Conformity to male norms as a moderator. *Psychology of Men & Masculinity*, 19, 477–483.
- Yen W (1981). Using simulation results to choose a latent trait model. *Applied Psychological Measurement*, 5, 245–262. 10.1177/014662168100500212
- Yen W (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8, 125–145.
10.1177/014662168400800201
- Zamboanga BL, Pesigan IJA, Tomaso CC, Schwartz SJ, Ham LS, Bersamin M, & Hurley EA (2015). Frequency of drinking games participation and negative drinking consequences in a multiethnic sample of college students: Do gender and ethnicity matter? *Addictive Behaviors*, 41, 112–116. 10.1016/j.addbeh.2014.10.002 [PubMed: 25452053]
- Zumbo BD (1999). *A handbook on the theory and methods of differential item functioning (DIF): Logistic regression modeling as a unitary framework for binary and Likert-type (ordinal) item scores*. Ottawa, Ontario, Canada: Directorate of Human Resources Research and Evaluation, Department of National Defense.

Public Significance Statement

A 20-item psychometrically robust measure of protective behavioral strategies (PBS) called the Protective Drinking Practices Scale was developed using item response theory to address measurement and methodological issues of existing measures. The new measure includes optimal response anchors and items, represents the full range of the PBS construct, and functions equivalently for college men and women from different racial backgrounds.

Figure 1.
Probability curves for each response option for protective behavioral strategies (PBS) item
“Spacing drinks over time.”

Figure 2.
Protective behavioral strategies (PBS) item, “Leave a drink unattended,” with overlapping probability response curves and poor residuals plot. Item removed from scale.

Figure 3.

Response probability curves for the same item “Refuse drinks,” when administered with text- (left) and numeric-based response options (right). PBS = protective behavioral strategies.

Figure 4. Distribution of IRT-based scores and total test information of the Protective Drinking Practices Scale (PDPS) with 20 items. PBS = protective behavioral strategies.

Figure 5.
Example of item exhibiting DIF: “Avoid drinking games.” PBS = protective behavioral strategies.

Table 1

DIF Items Based on Analyses by Race (Whites vs. Blacks) and Gender (Males vs. Females)

Item	Race			Gender			DIF Analysis
	group	group* Φ	R ²	group	group* Φ	R ²	
Know what you are drinking	-0.50	1.44*	0.02	-0.20	0.55	0.00	Race
Avoid mixing alcohol with prescription drugs (whether prescribed for you or not)	-0.67*	0.94	0.02	-0.27	0.27	0.00	Race
Avoid getting in a car with someone who has been drinking	-0.60*	0.24	0.01	0.10	-0.36	0.00	Race
Know where your drink has been at all times	-0.37	1.04*	0.01	0.37	0.18	0.00	Race
Use a designated driver	-0.87*	0.92*	0.03	0.13	0.83	0.01	Race
Make sure you go home with a friend	-0.33	0.97*	0.01	0.54*	0.61	0.01	Race & Gender
Have a reliable designated driver	-0.67*	0.46	0.01	0.18	0.80	0.01	Race
Preplan transportation to get home	-0.70*	0.48	0.01	0.12	0.10	0.00	Race
Finding other ways besides drinking to reduce stress	-0.53*	-0.42	0.01	0.21	-0.21	0.00	Race
Used nonalcohol, healthy ways to deal with stress or nerves	-0.79*	-0.29	0.02	0.13	-0.73	0.01	Race
Used nonalcohol, healthy ways to feel good or high	-0.56*	-0.42	0.01	0.27	-0.53	0.00	Race
Used nonalcohol, healthy ways to feel at ease with people	-0.53*	-0.24	0.01	0.13	-0.72	0.01	Race
Make your own drinks	-0.28	0.93*	0.01	-0.22	0.35	0.00	Race
Practicing ways to be more comfortable in social settings without using alcohol	0.14	-0.10	0.00	0.30	-0.89*	0.01	Gender
Choose nonalcoholic drink	-0.17	0.25	0.00	0.20	-0.73*	0.01	Gender
Determine not to exceed a set number of drinks	0.52	-0.11	0.01	0.14	1.14*	0.01	Gender
Avoid mixing different types of alcohol	0.51*	0.12	0.01	-0.07	-0.17	0.00	Race
Stop drinking at least 1–2 hr before going home	0.33	-0.02	0.00	-0.63*	-0.21	0.01	Gender
Determine in advance not to exceed a set number of drinks	0.49	-0.25	0.01	-0.15	1.05*	0.01	Gender
Have a friend let you know when you've had enough to drink	0.05	-0.88*	0.01	0.11	0.04	0.00	Race
Choose not to drink	-0.06	-0.18	0.00	0.03	-0.90*	0.01	Gender
Choose not to do shots when available	0.12	-0.80*	0.01	0.05	0.02	0.00	Race
Choose to avoid situations where heavy drinking is likely	0.74*	-0.17	0.01	-0.10	-0.41	0.00	Race

Item	Race			Gender			DIF Analysis
	group	group* Φ	R ²	group	group* Φ	R ²	
Avoid drinking games	0.54 *	-0.79 *	0.01	-0.07	-0.48	0.00	Race
Choose not to participate in drinking games when given the opportunity	0.21	-0.76 *	0.01	-0.11	0.09	0.00	Race
Refuse drinks	-0.03	-0.59 *	0.01	-0.13	-0.59	0.00	Race
Pace your drinks to 1 or fewer per hour	0.34	-0.95 *	0.01	0.03	0.22	0.00	Race
Stayed away from people who drink	0.57 *	-0.14	0.01	-0.02	-0.46	0.00	Race
Rewarded self for not drinking	0.54 *	0.02	0.01	-0.21	-0.22	0.00	Race
Looking for more info about alcohol and health	0.72 *	0.12	0.01	-0.19	0.16	0.00	Race

Note. Items are ordered in ease of endorsement from easier to harder. *group* = ordinal logistic coefficient for group membership in DIF Model 2; *group* Φ* = ordinal logistic coefficient for group membership by latent trait in DIF Model 3. DIF = differential item functioning; R² = maximum McFadden's likelihood ratio R² of nested model pairs: Model 1: Model 2, Model 2: Model 3, and Model 2: Model 3.

* $p < .01$.

Table 2
Protective Drinking Practices Scale Items: Descriptive Statistics, IRT Parameters, and Fit Statistics

Item	<i>M</i>	<i>SD</i>	<i>a</i>	<i>b1</i>	<i>b2</i>	<i>b3</i>	<i>b4</i>	<i>b5</i>	Average location	$\chi^2(12)$	<i>p</i> -value
Have a plan with a friend to watch out for each other	4.40	1.76	2.82	-1.37	-0.89	-0.48	-0.17	0.29	-0.62	15.85	0.32
Only drink in safe environments	4.39	1.72	3.29	-1.33	-0.92	-0.48	-0.18	0.37	-0.61	15.70	0.33
Eating before and while you are drinking	4.27	1.69	3.06	-1.40	-0.89	-0.44	-0.11	0.50	-0.59	29.54	0.01
Choosing to participate in enjoyable activities that do not include alcohol consumption	4.20	1.61	2.40	-1.43	-1.02	-0.47	0.03	0.79	-0.53	30.51	0.01
Being aware of internal bodily sensations that indicate you are getting intoxicated	4.13	1.77	3.58	-1.14	-0.76	-0.31	-0.01	0.50	-0.47	38.93	0.00
Limiting cash before going out drinking	4.02	1.82	3.01	-1.09	-0.67	-0.33	0.02	0.57	-0.43	10.17	0.75
Avoid trying to "keep up" or "out drink" others	4.01	1.79	3.34	-1.07	-0.69	-0.26	0.05	0.59	-0.40	11.24	0.67
Keeping track of how many drinks you have	4.00	1.76	4.22	-1.10	-0.66	-0.22	0.06	0.58	-0.40	22.57	0.07
Drinking slowly	3.96	1.63	4.83	-1.19	-0.71	-0.27	0.13	0.73	-0.39	13.06	0.52
Being prepared with effective coping strategies in situations where you think heavy drinking is likely	3.94	1.72	3.29	-1.13	-0.69	-0.25	0.14	0.74	-0.37	34.25	0.00
Drink for quality not quantity	3.86	1.74	3.28	-1.12	-0.69	-0.23	0.21	0.68	-0.37	34.58	0.00
Spacing drinks over time	3.85	1.68	4.87	-1.05	-0.63	-0.21	0.19	0.73	-0.33	26.46	0.02
Leave the bar/party at a predetermined time	3.74	1.79	3.32	-0.96	-0.55	-0.16	0.24	0.73	-0.29	14.77	0.39
Stop drinking at a predetermined time	3.65	1.75	3.52	-0.97	-0.53	-0.14	0.30	0.85	-0.25	11.66	0.63
Have a set number of drinks you will have for the social occasion	3.60	1.77	3.52	-0.91	-0.51	-0.04	0.31	0.83	-0.22	22.91	0.06
Choose not to "pregame" or "prebar" (i.e., drinking before going out)	3.61	1.73	1.99	-1.26	-0.59	0.03	0.44	1.09	-0.21	66.13	0.00
Avoid hard liquor or spirits	3.59	1.75	2.82	-1.02	-0.47	-0.09	0.36	0.94	-0.20	15.46	0.35
Alternating alcoholic and nonalcoholic beverages when you are drinking	3.49	1.75	2.45	-0.99	-0.50	-0.06	0.39	1.11	-0.16	37.00	0.00
Avoid situations where there was alcohol	3.39	1.62	2.14	-1.23	-0.43	0.15	0.76	1.33	-0.05	91.11	0.00
Thought about problems drinking can cause	3.36	1.69	1.65	-1.15	-0.53	0.10	0.75	1.49	-0.03	41.37	0.00

Note. Items are ordered in ease of endorsement from easier to harder. IRT = item response theory. *a* = IRT discrimination parameter; *b1* to *b5* = IRT threshold parameters; Average location = mean of all five threshold parameters; $\chi^2(12)$ = is the goodness of fit statistic for standardized residuals, with accompanying *p*-value.