Published in final edited form as:

Addict Res Theory. 2012 October 1; 45(4): 292–302. doi:10.1177/0748175612449744.

Assessing the Dependability of Drinking Motives via Generalizability Theory

Brooke J. Arterberry¹, Matthew P. Martens¹, Jennifer M. Cadigan¹, and Ashley E. Smith¹

¹University of Missouri, Columbia, MO, USA

Abstract

This study assessed the score reliability of the Drinking Motives Questionnaire–Revised (DMQ-R) via generalizability theory. Participants ($n = 367$ college students) completed the DMQ-R at three time points. Across subscale scores, persons, persons $\times$ occasions, and persons $\times$ items interactions accounted for meaningful variance. Findings illustrate advantages of generalizability theory–based techniques.

Keywords

alcohol; generalizability theory; drinking motives

The college student population tends to engage in excessive alcohol use that contributes to a variety of negative consequences, which has become a growing public health concern. At least 85% of college students have tried alcohol and more than 40% report binge drinking (i.e., +5 drinks for males and +4 drinks for females) in the preceding 2 weeks (Johnston, O'Malley, Bachman, & Schulenberg, 2008; Weschler et al., 2002). Students who engage in such drinking practices are at risk for negative consequences such as assault, sexual assault, injury, and death (Hingson, Zha, & Weitzman, 2009; Wechsler et al., 2002).

The use of reliable assessments in helping students to reduce alcohol consumption and negative consequences related to alcohol use is important for counselors, prevention specialists, and researchers. Understanding the underlying motives for drinking can establish a more holistic conceptualization of the student's drinking and can influence intervention techniques. For example, counseling strategies may differ for a student primarily endorsing drinking motives related to coping (e.g., "To forget my problems") versus a student primarily endorsing drinking motives related to social factors (e.g., "To enjoy a party"). The former would likely benefit from counseling focused on developing alternative coping strategies, whereas the latter would benefit from an approach focusing on ways to enhance social experiences without engaging in excessive drinking. Therefore, identifying and accurately measuring different types of drinking motives has important implications for both clinical and research efforts in the field (Cooper, 1994; Cox & Klinger, 1988; Kuntsche, Knibbe, Gmel, & Engels, 2005; Newcomb, Chou, Bentler, & Huba, 1988).

Influential theoretical models of drinking motives may identify a final pathway mediating the effects of a host of individual and environmental factors on the decision to use alcohol

© The Author(s) 2012

Corresponding Author: Brooke J. Arterberry, University of Missouri, 16 Hill Hall, Columbia, MO 65211, USA, bjaww8@mail.missouri.edu.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

(Cooper, Frone, Russell, & Mudar, 1995; Cox & Klinger, 1988). These models conceptualized motives from the standpoint of positive and negative reinforcement; that is, individuals are motivated to drink in order to either enhance positive affect or reduce negative affect. Several studies indicated both positively and negatively reinforcing drinking motives are associated with alcohol use, alcohol-related problems, and other drinking-related constructs, especially among adolescent and young adult drinkers (e.g., Cooper et al., 2008; Kuntsche et al., 2005, 2006).

The most commonly used measure of positively and negatively reinforcing drinking motives is the Drinking Motives Questionnaire (DMQ; Cooper, 1994; Cooper, Russell, Skinner, & Windle, 1992). The initial version of the DMQ assessed three specific types of motives: Coping (e.g., “To forget about your problems”), Social (e.g., “Because it improves parties and celebrations”), and Enhancement (e.g., “Because it gives you a pleasant feeling”). Later, Cooper (1994) created a four-factor DMQ designed to assess drinking motives specifically among adolescents. This version of the measure (the Drinking Motives Questionnaire–Revised [DMQ-R]) added a Conformity subscale (e.g., “To be liked”) to the three existing subscales; this four-factor DMQ-R has also been used in studies with college students (MacLean & Lecci, 2000; Martens, Rocha, Martin, & Serrao, 2008).

Several studies were conducted with the express purpose of examining the reliability and validity scores from the DMQ/DMQ-R among college students (e.g., Cooper, 1994; Cooper et al., 1992; MacLean & Lecci, 2000; Martens, Cox, Beck, & Heppner, 2003; Martens et al., 2008; Simons, Correia, Carey, & Borsari, 1998; Stewart, Zeitlin, & Samoluk, 1996). In general, factor analytic findings from these studies supported the validity of scores from either the three- or four-factor model (depending on the version of the DMQ being used), as well as the convergent validity evidence of scores from the DMQ/DMQ-R with relevant constructs (e.g., alcohol use and alcohol-related problems). These studies also examined the reliability of scores from individual DMQ subscales, but these analyses were largely limited to tests of internal consistency. Although, in general, internal consistency estimates regarding scores on the separate subscales were found to be satisfactory (i.e., $\alpha > .80$ usually reported for each subscale), relying only on internal consistency to examine score reliability is problematic because internal consistency only accounts for one source of error (described in more detail below). We are aware of only one study that examined the test–retest reliability of scores on the DMQ. Grant, Stewart, O’Connor, Blackwell, and Conrod (2007) assessed test–retest reliability of scores on a modified five–factor version of the measure, with results indicating satisfactory score reliability on the subscales (intraclass correlation coefficient = .65–.78). However, the general lack of thorough reliability studies on the DMQ warrants additional examination, considering the popularity of the measure and the importance of the construct it measures. In particular, using generalizability theory (Brennan, 2001; Hoyt & Melby, 1999) to assess the reliability of DMQ-R scores among college students may be especially beneficial.

Generalizability Theory

Generalizability theory (GT) provides a more thorough explanation of measurement error than classical test theory (CTT; Webb, Shavelson, & Haertel, 2007). Many forms of reliability analyses exist within the framework of CTT, such as test–retest reliability, internal consistency, and parallel/equivalent forms reliability. Each of these reliability estimates focuses on different areas of measurement error and is assessed independently from each other. This can produce variability in reliability estimates due to the examination of distinctive sources of error, resulting in misleading interpretations. For example, coefficient alpha accounts for error associated with inconsistency across items (Helms, Henze, Sass, & Mifsud, 2006) but does not account for error associated with inconsistency

in scores across measurement occasions. Similarly, a test–retest reliability coefficient provides information on error associated with the stability of the measure across separate time points (Webb et al., 2007) but provides no information on the possible impact of other sources of error on the accuracy of test scores. In sum, reliability analyses derived from CTT cannot identify more than one specific source of measurement error within a set of scores (Hoyt & Melby, 1999).

In contrast to CTT-based approaches, GT can not only identify multiple levels and multiple sources of measurement error but also assesses how these sources of error may be interacting with one another. GT provides a more holistic view of reliability by examining scores through interaction effects such as how a person responds differently to items depending on the time point, whereas CTT can only identify consistency in scores across persons or time points or items specifically. Furthermore, GT methods allow for analyses used to identify possible areas in which error could be reduced, such as increasing the number of items on a scale or adding additional raters. In terms of GT, reliability is assessed by examining measurement error in observed scores to establish how well the observed scores estimate the dependability of the hypothetical scores (Brennan, 2001; Hoyt & Melby, 1999; Webb et al., 2007). The universe of admissible observations includes all possible observations on the object of measurement that could be obtained where one or more specific facets (with one or more levels of error) affect the object of measurement. GT employs two types of studies to examine the dependability of scores on a measure: the generalizability study (G-study) and the decision study (D-study; Brennan, 2001; Webb et al., 2007).

Through the G-study, the researcher examines the variance components associated with the object of measurement and the universe of admissible observations. The G-study estimates variance components that reflect both main effects and interaction effects associated with multiple sources of error (Hoyt & Melby, 1999). In addition, the G-study estimates a G-coefficient (i.e., defines role of relative error with an analogous representation of the reliability coefficient of CTT) and Phi coefficient (i.e., defines role of absolute error with no analogous analysis in CTT) for each of the D-study designs (Brennan, 2001; Webb et al., 2007). The Phi coefficient in GT is called the index of dependability and is not analogous to the Phi coefficient used to describe the correlation between dichotomous variables (Brennan, 2001; Thompson, 2003). In the D-study, the information provided by the G-study enables the researcher to specify designs he or she may wish to examine in order to determine how score reliability would differ (Webb et al., 2007). The D-study uses estimated variances from the G-study to introduce possible strategies to reduce error. For example, in a G-Study where the researcher calculated a G-coefficient based on scores from a 10-item measure assessed on three occasions, the D-study could assess how the G-coefficient would change if the measure's scores were examined on two through four occasions. Furthermore, the D-study could show how changing the number of items on the measure would impact score reliability (e.g., decrease to 5 items, increase to 15 items).

In sum, GT methods provide a more comprehensive estimate of a measure's score reliability than those based on CTT (Hoyt & Melby, 1999; Webb et al., 2007). To date, however, there have been very few studies in the counseling literature that have used GT to examine the reliability of scores on relevant measures. The DMQ-R is widely used to measure drinking motives, and a more robust assessment of reliability will help provide additional information regarding the psychometric properties of scores from the measure among college students. Thus, the purpose of this study was to use GT analyses to examine the reliability of scores on the DMQ-R among a sample of at-risk college drinkers.

Method

Participants

Participants were 367 college students from a Northeastern state university who received a university judicial sanction for an alcohol-related offense and were participating in a larger study examining the efficacy of three group-based alcohol interventions (see Cimini et al., 2009). Participants in the present study were those who provided complete data on the DMQ-R on three separate measurement occasions (53% of total sample). Although we did not track the specific percentage of eligible students who participated in the study, very few of those eligible to participate chose not to. A majority of the sample were male (55.6%) and most participants were White (83.4%), with other ethnic representations as follows: 3.8% Asian/Asian American, 3.8% multiracial, 1.9% African American, 0.3% American Indian, 0.3% Native Hawaiian/Pacific Islander, and 6.5% “other.” Ten percent of the sample also self-identified as Hispanic or Latino(a). The mean age of the sample was 18.90 years ($SD = 0.81$), and the bulk of the participants were either first- or second-year students (82.0%). Since eligibility for the study involved a university judicial sanction for alcohol use, almost all the participants reported living in an on-campus residence hall (97.8%).

Measures

Drinking Motives Questionnaire—Revised—The DMQ-R (Cooper, 1994) was used to assess motives for alcohol use among adolescents and adults. The 20-item questionnaire is composed of four subscales: Coping (e.g., “To forget your worries”), Social (e.g., “To celebrate special occasions with friends”), Enhancement (e.g., “Because it’s fun”), and Conformity (e.g., “Because your friends pressure you to drink”). Each subscale contains five items, and participants are asked to indicate how frequently they drink alcohol for each reason listed. Responses are scored on a 5-point scale ranging from *Almost Never/Never* (scored as 1) to *Almost Always/Always* (scored as 5). Responses are then averaged for each sub-scale. Factor analytic studies among college students have indicated good fit to a four-factor model (e.g., comparative fit index = .92–.95, Tucker–Lewis index = .94, incremental fit index = .92) and shown that the four-factor model provided a better fit than one-, two-, and three-factor models (MacLean & Lecci 2000; Martens et al., 2008). Furthermore, internal consistency estimates for the DMQ-R scores have been adequate for college populations. For example, MacLean and Lecci (2000) reported coefficient alphas ranging from .81 (Conformity) to .91 (Social), whereas Martens et al. (2008) reported alphas ranging from .78 (Coping) to .87 (Enhancement).

Demographics—Participants completed a measure that obtained information such as gender, ethnicity, age, and year in school.

Procedures

Details of this study’s procedures have been reported elsewhere (see Cimini et al., 2009) and will only be summarized. Institutional review board approval was obtained prior to initiating data collection for this project. All students who committed an on-campus judicial infraction involving alcohol (e.g., having alcohol in a dorm room) were eligible to participate in the study. Students who committed an infraction were mandated by the university to attend an alcohol intervention program. They were given the option of participating in an intervention associated with the research project or an alternative intervention provided by the university counseling center, all of which required roughly the same time commitment. Those who were interested in participating in the study completed a series of questionnaires, including the DMQ-R, at baseline, 6-month follow-up, and 12-month follow-up, and received a \$25 gift card each time they completed the battery. Participants attended one of three 90-minute group intervention sessions after completing their baseline measures. A prior study indicated

the interventions had no effect on alcohol use or alcohol-related problems (Cimini et al., 2009) with no between-group differences on any of the DMQ-R subscales at any follow-up point. Therefore, participants were collapsed across conditions for this study.

Data Analysis

A random effects design for both the G-study and the D-study was implemented for each set of scores on each subscale to provide a robust examination of main effects for each facet as well as all possible interaction effects among facets. We let p represent the object of measurement, i represent items, and o represent occasions in the universe of admissible observations (i.e., facets)— $p \times i \times o$. The random effects design for the G-study examined all main effects, p , i , o , and all interaction effects, $p \times i$, $p \times o$, $i \times o$, $p \times i \times o$. Each effect then had a corresponding variance component. G- and Phi-coefficients were calculated as global indices of score reliability for each subscale. We then identified a design ($p \times I \times O$) for the D-study where G-coefficients were calculated for various combinations of each facet: 1, 2, 3, 4, and 5 measurement occasions and 1, 3, 7, 10, 12, and 15 items. We focused on G-coefficients as opposed to the Phi-coefficient since the G-coefficient best reflects score reliability when the researcher is interested in the relative standing of a variable (e.g., how individuals compare with other individuals) rather than absolute standing (e.g., how an individual compares with an established criterion; Brennan, 2001; Hoyt & Melby, 1999; Webb et al., 2007). Drinking motives are almost always interpreted in terms of relative rather than absolute standing. Generalizability analyses on scores from each DMQ-R subscale were conducted using SPSS with syntax developed by Mushquash and O'Connor (2006). For illustrative purposes, we also calculated internal consistency estimates and test–retest reliability of scores for each DMQ-R subscale.

Results

Internal Consistency/Test–Retest Reliability

Internal consistency and test–retest reliability analyses are presented in Table 1. Cronbach's alpha indicated acceptable internal consistency at all three time points for scores from each subscale: $\alpha = .83$ –.89. Intraclass correlation was used to examine test–retest reliability of scores as it is more effective in detecting systematic error and allows for one analysis when assessments are administered more than twice, whereas Pearson's product–moment correlation (r) does not allow for these distinctions (e.g., Cicchetti, 1994, Shrout & Fleiss, 1979). Intraclass correlation coefficients also indicated adequate test–retest reliability of scores, with values ranging from .70 (Conformity) to .80 (Enhancement).

Generalizability Theory Tests

G- and Phi-coefficients—The G- and Phi-coefficients for scores from each subscale are presented in Table 2. G-coefficients ranged from .66 (Conformity) to .74 (Enhancement); Phi-coefficients ranged from .66 (Conformity/Coping) to .73 (Social). None of the coefficients achieved the .80 threshold, which is often used as a rule-of-thumb for acceptable overall score reliability in a GT study (e.g., Mushquash & O'Connor, 2006). This threshold, though, has not been validated as an acceptable cutoff score in any empirical analysis.

G-study findings—The variance components for each subscale score are presented in Table 3. Several general themes should be noted from these findings. First, variance attributed to persons, which indicates the degree to which individual differences in level of drinking motives accounts for subscale scores and is therefore not considered a source of error, was relatively high in comparison to the other variance components for each subscale's score. Differences attributed to persons accounted for either the most (Social Motives, 33.0%; Enhancement Motives, 26.6%), second-most (Coping Motives, 26.2%), or

third-most (Conformity Motives, 26.2%) variability in scores. Second, the persons $\times$ items $\times$ occasions three-way interaction accounted for a large portion of variance on each subscale score (25.7% to 35.8%), but this source of error is very difficult to interpret because it is the residual term and therefore confounded with random error (Brennan, 2001; Hoyt & Melby, 1999; Mushquash & O'Connor, 2006). Third, the persons $\times$ occasions interaction accounted for a greater percentage of the variability in scores on each subscale (15.7% to 28.2%) than the persons $\times$ items interaction (7.7% to 12.3%), which indicates inconsistency in the degree to which individuals responded across measurement occasions was a greater source of error than the inconsistency in which individuals responded across items. Finally, with the exception of items for Enhancement Motives (20.1%), and to a lesser degree items for Coping Motives (6.3%), very little variance in scores was attributable to differences among items themselves or occasions themselves.

D-Study findings—Results from the D-studies are presented in Tables 4 and 5. In general, results for reliability of scores on each sub-scale indicated fairly large improvements when increasing the number of items from one to five, but more modest improvements at greater numbers of items. For example, the estimated G-coefficient for scores on three measurement occasions increased from .55 to .74 on the Enhancement subscale and from .46 to .69 on the Coping subscale when the number of items increased from one to five. Yet the G-coefficient only increased an additional .06 (Enhancement) and .06 (Coping) percentage points when the number of items increased from five to fifteen. Similarly, there were relatively large increases in the estimated G-coefficient when the number of measurement occasions increased from one to three, but more modest increases with additional measurement occasions. For example, the estimated G-coefficient with three items increased from .47 to .69 on the Social sub-scale and from .38 to .62 on the Conformity subscale between one and three measurement occasions, but only an additional .07 (Social) to .09 (Conformity) percentage points between three and five measurement occasions.

Post Hoc Tests on Enhancement Motives

Because the variance attributable to items themselves was so much higher on the Enhancement Motives subscale than any other subscale, we conducted additional analyses in an attempt to determine if any particular item was disproportionately contributing to this finding. We conducted five additional G-studies on the Enhancement subscale with one of the items removed on each analysis and examined the amount of variance attributable to items on each test. Presumably, if an item were disproportionately contributing to the variance attributable to items, the percentage would decrease when the item was removed from the analyses. In four of the five analyses, the amount of variance attributable to items remained relatively high when an item was removed (17.9% to 24.2%). However, when the item “To get high” was removed from the GT analyses, the amount of variance attributable to items reduced considerably (10.7%). Furthermore, the amount of variance attributable to persons increased from 26.7% to 33.3%.

A high variance associated with items indicates inconsistency in terms of item difficulty (Webb et al., 2007), which on the DMQ-R subscales would reflect an item with a mean score that differed considerably from the other DMQ-R subscales. Mean scores on the “To get high” item were 1.95, 2.17, and 2.13 at baseline, 6-month, and 12-month follow-up, whereas mean scores on the item closest in score (“Because it’s exciting”) were 2.60, 2.75, and 2.56, respectively. This spread in scores between the “To get high” item and the other items on its subscale was the largest for any item across all four subscales.

Discussion

The purpose of this study was to use GT analyses to examine the reliability of scores on the measure used most frequently to assess drinking motives—the DMQ-R. Results from this study allow one to more precisely examine reliability of scores on the DMQ-R subscales. In many instances, findings were relatively consistent across the four subscales, but in others there were important distinctions.

Relative to traditional tests of reliability, one of the most important aspects of GT analyses involves the ability to determine the amount of variability in scores attributable to different sources. In the present study, the variance in scores attributable to persons (interpreted as being due to individual differences in drinking motives rather than due to error) remained relatively consistent across the four subscales. Variance in scores attributed to persons also accounted for the highest proportion of variance for Social and Enhancement subscales. On the other hand, in absolute terms, the amount of variance in scores attributable to persons seems relatively low. On a perfectly reliable scale, persons would account for 100% of the variance, but this is unrealistic in social sciences research. Evaluating how good the amount of variance in scores accounted for by persons is difficult, as guidelines for evaluating this percentage presently do not exist. One instructive option would be to compare this percentage to other published examples in the psychological literature. In GT analyses involving other constructs, such as self-reported schizotypy symptoms (Winterstein, Willse, Kwapił, & Silvia, 2010), self-reported self-esteem (Mushquash & O'Connor, 2006), teacher ratings of attention deficit hyperactive disorder symptoms (Volpe, Briesch, & Gadow, 2011), and single-item direct behavior ratings (Christ, Riley-Tillman, Chafouleas, & Boice, 2010), variance in scores attributable to persons ranged from 6.2% to 32.6%. In general, findings from the present study suggest variability in DMQ-R scores due to actual differences in level of drinking motives is similar to or exceeds comparable estimates from other types of psychological assessment measures.

Another finding relatively consistent across subscale scores was the proportion of variance attributable to the persons $\times$ items ($p \times i$) and persons $\times$ occasions ($p \times o$) interactions. The variance in scores attributable to $p \times i$ ranged from 7% to 12%, whereas the score variance attributable to $p \times o$ ranged from 15% to 28%. These findings indicate individual scores on the DMQ-R items were relatively consistent. For example, on the Coping subscale, the item with the highest mean score on all three measurement occasions was “Because you feel more self-confident and sure of yourself,” followed by “To cheer you up when you are in a bad mood.” A relatively low percentage of variance attributable to the $p \times i$ interaction indicates generally little variation exists across subjects in terms of scoring higher on those two Coping items relative to the other three items. Or, put another way, it is likely that both those with relatively high, medium, and low Coping Motives scored higher on those two particular items than the other three Coping items. One might expect this to be the case given the overall mean scores for the items. But it is important to note that a diverse array of patterns among individual scores, including both consistent and inconsistent between-person responding among items, could result in the same overall mean. In contrast, the variance in scores accounted for by the $p \times o$ interaction indicated DMQ-R subscale scores generally varied among individuals over time. If the variance accounted for by the $p \times o$ interaction was relatively low, it would suggest that the participants with the highest scores on a particular DMQ-R subscale at Time 1 would also have the highest scores on that subscale on subsequent occasions (Mushquash & O'Connor, 2006). In contrast, the relatively high proportion of variance attributable to the $p \times o$ interaction in the present study indicates that participants' relative rankings on the DMQ-R subscale scores differed across time (e.g., those with the highest Coping Motives subscale scores at Time 1 may not have been the same individuals with the highest scores at Time 2). These findings suggest individual

DMQ-R subscale scores may ebb and flow over time due to a host of intrapersonal and environmental factors unique to a specific individual. For example, one individual might report an increase in Social Motives scores between Times 1 and 2 due to a change in his or her social environment (e.g., increased socialization with fraternity/sorority members), whereas a second individual might report a lower score on the Social Motives subscale due to environmental changes that promote less alcohol use (e.g., engaging in less fraternity/sorority system activity).

The most distinctive finding from the study in terms of differences among subscale scores involved the variance attributable to items themselves on the Enhancement subscale. It appears a large portion of this effect can be attributable to the item assessing “To get high” as a motivation for drinking, with an overall mean score considerably lower than other items on the subscale. One possible explanation for this finding is that for many college students the term *high* is generally applied to illicit drugs such as marijuana and not alcohol, thus the item may be underendorsed. These findings should be taken with some degree of caution because the analyses were conducted post hoc and may represent sample-specific score variability, although the explanation of the finding has clinical relevance for those who engage in applied work with college students (e.g., college students rarely refer to themselves as getting “high” from drinking). Future researchers should consider rewording the item, for example, “To get high/drunk.” These findings also demonstrate how GT analyses provide more refined information than traditional CTT-based tests, as internal consistency estimates for the four DMQ-R subscale scores were relatively similar across all three occasions and would therefore not have revealed this important subscale distinction.

Perhaps the most important finding from the D-study was that the current length of each DMQ-R subscale (5 items) appeared reasonably adequate for score reliability purposes. Of course, increasing the number of items (assuming adequate sampling occurred from the relevant item universe) would always increase score reliability estimates, but researchers must consistently be aware of logistical considerations such as response burden. For all DMQ-R subscale scores, the increase in G-coefficient estimates between one and five items dwarfed the increases that would occur with the addition of subsequent items. Researchers interested in drinking motives need to consider if the relatively modest increases in score reliability gained by adding five or ten additional items to each subscale warrant an increase in scale length.

Several limitations to this study should be noted. First, the majority of students were White and in their first and second years of school. The generalizability of the findings to a more diverse array of students remains unknown. Second, all students in the study committed some type of judicial infraction resulting in a mandate to receive an alcohol intervention. Although one can argue the importance in examining drinking motives among “at-risk” college student drinkers, the possibility remains that our findings would not generalize to nonjudicially mandated heavy drinking students. Finally, all the students in this study received some type of alcohol-related intervention. It would have been preferable to use a sample that did not receive an alcohol intervention, but no evidence suggests any of the interventions meaningfully affected drinking motives (or any alcohol-related variable).

Despite these limitations, we believe this study achieved our major aim of providing additional reliability data on scores from the DMQ-R. We encourage additional researchers to continue to examine the psychometric properties of scores from this measure, particularly among more diverse racial and ethnic samples. In addition, clinicians should consider using the DMQ-R when working with college students who have alcohol-related issues in an effort to tailor their intervention efforts. Our science is only as good as our measures, and, given the importance of drinking motives in understanding alcohol use among both college

students and other populations, continued examination of the DMQ-R and other related measures is warranted.

Acknowledgments

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This research project was in part supported by National Institute on Alcohol Abuse and Alcoholism Grant # U18AA015039.

References

- Brennan, RL. Generalizability theory. New York, NY: Springer-Verlag; 2001.
- Christ TJ, Riley-Tillman TC, Chafouleas SM, Boice CH. Direct behavior rating (DBR): Generalizability and dependability across raters and observations. *Educational and Psychological Measurement*. 2010; 70:825–843.
- Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*. 1994; 6:284–290.
- Cimini MD, Martens MP, Larimer ME, Kilmer JR, Neighbors C, Monserrat JM. Assessing the effectiveness of peer-facilitated interventions addressing high-risk drinking among judicially mandated college students. *Journal of Studies on Alcohol and Drugs*. 2009; 70:57–66. [PubMed: 19538913]
- Cooper ML. Motivations for alcohol use among adolescents: Development and validation of a four-factor model. *Psychological Assessment*. 1994; 6:117–128.
- Cooper ML, Frone MR, Russell M, Mudar P. Drinking to regulate positive and negative emotions: A motivational model of alcohol use. *Journal of Personality and Social Psychology*. 1995; 69:990–1005. [PubMed: 7473043]
- Cooper ML, Krull JL, Agocha VB, Flanagan ME, Orcutt HK, Grabe S, Maudette J. Motivational pathways to alcohol use and abuse among Black and White adolescents. *Journal of Abnormal Psychology*. 2008; 117:485–501. [PubMed: 18729604]
- Cooper ML, Russell M, Skinner JB, Windle M. Development and validation of a three-dimensional measure of drinking motives. *Psychological Assessment*. 1992; 4:123–132.
- Cox WM, Klinger E. A motivational model of alcohol use. *Journal of Abnormal Psychology*. 1988; 97:168–180. [PubMed: 3290306]
- Grant VV, Stewart SH, O'Connor RM, Blackwell E, Conrod PJ. Psychometric evaluation of the Five-Factor Modified Drinking Motives Questionnaire- Revised in undergraduates. *Addictive Behaviors*. 2007; 32:2611–2632. [PubMed: 17716823]
- Helms JE, Henze KT, Sass TL, Mifsud VA. Treating Cronbach's alpha reliability coefficients as data in counseling research. *The Counseling Psychologist*. 2006; 34:630–660.
- Hingson RW, Zha W, Weitzman ER. Magnitude of and trends in alcohol-related mortality and morbidity among U.S. college students ages 18–24, 1998–2005. *Journal of Studies on Alcohol and Drugs, Supplement*. 2009; 16:12–20. [PubMed: 19538908]
- Hoyt WT, Melby JN. Dependability of measurement in counseling psychology: An introduction to generalizability theory. *The Counseling Psychologist*. 1999; 27:325–352.
- Johnston, LD.; O'Malley, PM.; Bachman, JG.; Schulenberg, JE. Monitoring the Future National Survey results on drug use, 1975–2007: Vol. I, Secondary school students. Bethesda, MD: National Institute on Drug Abuse; 2008.
- Kuntsche E, Knibbe R, Gmel G, Engels R. Why do young people drink? A review of drinking motives. *Clinical Psychology Review*. 2005; 25:841–861. [PubMed: 16095785]
- Kuntsche E, Knibbe R, Gmel G, Engels R. Who drinks and why? A review of socio-demographic, personality, and contextual issues behind the drinking motives in young people. *Addictive Behaviors*. 2006; 31:1844–1857. [PubMed: 16460883]
- MacLean MG, Lecci L. A comparison of models of drinking motives in a university sample. *Psychology of Addictive Behaviors*. 2000; 14:83–87. [PubMed: 10822750]

- Martens MP, Cox RH, Beck NC, Heppner P. Measuring motivations for intercollegiate athlete alcohol use: A confirmatory factor analysis of the Drinking Motives Measure. *Psychological Assessment*. 2003; 15:235–239. [PubMed: 12847784]
- Martens MP, Rocha TL, Martin JL, Serrao HF. Drinking motives and college students: Further examination of a four-factor model. *Journal of Counseling Psychology*. 2008; 55:289–295.
- Mushquash C, O'Connor BP. SPSS and SAS programs for generalizability theory analyses. *Behavior Research Methods*. 2006; 38:542–547. [PubMed: 17186766]
- Newcomb MD, Chou C, Bentler PM, Huba GJ. Cognitive motivations for drug use among adolescents: Longitudinal tests of gender differences and predictors of change in drug use. *Journal of Counseling Psychology*. 1988; 35:426–438.
- Shrout PE, Fleiss JL. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*. 1979; 86:420–428. [PubMed: 18839484]
- Simons JS, Correia CJ, Carey KB, Borsari BE. Validating a five-factor marijuana motives measure: Relations with use, problems, and alcohol motives. *Journal of Counseling Psychology*. 1998; 45:265–273.
- Stewart SH, Zeitlin SB, Samoluk SB. Examination of a three-dimensional drinking motives questionnaire in a young adult university student sample. *Behaviour Research and Therapy*. 1996; 34:61–71. [PubMed: 8561766]
- Thompson, R. *Score reliability: Contemporary thinking on reliability issues*. Thousand Oaks, CA: Sage; 2003.
- Volpe RJ, Briesch AM, Gadow KD. The efficiency of behavior rating scales to assess inattentive-overactive and oppositional-defiant behaviors: Applying generalizability theory to streamline assessment. *Journal of School Psychology*. 2011; 49:131–155. [PubMed: 21215839]
- Webb, NM.; Shavelson, RJ.; Haertel, EH. Reliability coefficients and generalizability theory. In: Rao, CR.; Sinharay, S., editors. *Handbook of statistics: Vol. 26. Psychometrics*. Amsterdam, Netherlands: Elsevier; 2007. p. 81-124.
- Wechsler H, Lee JE, Kuo M, Seibring M, Nelson TF, Lee H. Trends in college binge drinking during a period of increased prevention efforts. *Journal of American College Health*. 2002; 50:203–217. [PubMed: 11990979]
- Winterstein BP, Willse JT, Kwapil TR, Silvia PJ. Assessment of score dependability of the Wisconsin Schizotypy Scales using generalizability analysis. *Journal of Psychopathology and Behavioral Assessment*. 2010; 32:575–585.

Biographies

Brooke J. Arterberry is a graduate student in the Counseling Psychology program at the University of Missouri. Her primary research interests include addictive behaviors, interventions, and psychometrics.

Matthew P. Martens is an Associate Professor in the Counseling Psychology program at the University of Missouri. His primary research interests include addictive behaviors, health psychology, and assessment issues.

Jennifer M. Cadigan is a graduate student in the Counseling Psychology program at the University of Missouri. Her primary research interests include addictive behaviors and interventions.

Ashley E. Smith is a graduate student in the Counseling Psychology program at the University of Missouri. Her primary research interests include health psychology and intervention-related issues.

Table 1

Internal Consistency and Test–Retest Estimates for the DMQ-R Subscales

Subscale	α Baseline	α 6 Months	α 12 Months	ICC
Conformity	.83	.88	.87	.70
Coping	.83	.85	.85	.75
Enhancement	.85	.85	.85	.81
Social	.87	.86	.89	.78

Note: DMQ-R = Drinking Motives Questionnaire–Revised; ICC = intraclass correlation coefficient.

Table 2G-study G-coefficients and Phi-coefficients for $p \times i \times o$ Design

Subscale	G-Coefficient	Phi-Coefficient
Conformity	.66	.66
Coping	.69	.66
Enhancement	.74	.67
Social	.73	.72

Table 3
Estimated Variance Components for the G-Study $p \times i \times o$ Design for Each Subscale

Effect	Conformity		Coping		Enhancement		Social	
	Variance	Percentage	Variance	Percentage	Variance	Percentage	Variance	Percentage
a_p	.17	26.2	.30	26.2	.55	26.6	.52	33.0
b_i	.01	1.7	.07	6.3	.42	20.1	.03	1.6
c_o	.00	0.4	.00	0.0	.00	0.1	.01	0.7
$d_{p \times i}$	.05	7.7	.14	12.3	.24	11.6	.16	10.5
$e_{p \times o}$	.19	28.2	.25	22.2	.32	15.7	.38	24.0
$f_{i \times o}$	.00	0.0	.00	0.0	.00	0.2	.00	0.0
$g_{p \times i \times o, e}$	.24	35.8	.38	33.0	.53	25.7	.47	30.2

Note: ^a p persons effect;

^b i items effect;

^c o occasions effect;

^d $p \times o$ persons by occasions effect;

^e $p \times i$ persons by items effect;

^f $i \times o$ items by occasions effect;

^g $p \times i \times o, e$ persons by items by occasions effect, plus error.

Table 4

Conformity and Coping Subscales: D-Study G-Coefficients for $p \times I \times O$ design

Items	Conformity				Coping			
	Measurement Occasions				Measurement Occasions			
	1	3	5	5	1	3	5	5
1	.27	.47	.56	.28	.46	.46	.53	
3	.38	.62	.71	.41	.63	.71		
5	.42	.66	.75	.46	.69	.76		
7	.43	.68	.77	.48	.71	.79		
10	.45	.70	.79	.50	.73	.81		
12	.45	.70	.79	.50	.74	.81		
15	.46	.71	.80	.51	.75	.82		

Table 5

Enhancement and Social Subscales: D-Study G-Coefficients for $p \times I \times O$ Design

Items	Enhancement					Social				
	Measurement Occasions					Measurement Occasions				
	1	3	5	1	3	5	1	3	5	1
1	.33	.51	.57	.34	.54	.61				
3	.49	.69	.75	.47	.69	.76				
5	.55	.74	.80	.51	.73	.80				
7	.56	.77	.83	.53	.75	.82				
10	.58	.79	.85	.54	.77	.84				
12	.59	.79	.85	.55	.77	.84				
15	.59	.80	.86	.55	.78	.85				