Factor Structure of the Alcohol Urge Questionnaire Under Neutral Conditions and During a Cue-elicited Urge State

James MacKillop

**Background:** The Alcohol Urge Questionnaire (AUQ) is a promising multi-item measure of self-reported urges to drink in human laboratory studies; however, its factor structure has not been examined during an acute urge state. This study sought to validate the AUQ’s factor structure under neutral conditions and during a cue-elicited urge state in heavy drinkers.

**Methods:** Participants (248 heavy drinkers; 70% male) completed the AUQ, the Alcohol Dependence Scale (ADS), and the Positive and Negative Affect Scale (PANAS) under neutral conditions. A randomly selected subsample \( n = 61 \); 74% male) then underwent a multimodal alcohol cue exposure and completed the AUQ and PANAS a second time.

**Results:** Under neutral conditions, confirmatory factor analysis (CFA) replicated the previously reported single-factor structure, on which all items significantly loaded \( p < 0.001 \). Alcohol urges, as measured by the AUQ, exhibited significant positive associations with drinks per week and severity of dependence. Following the alcohol cue exposure, participants exhibited a significant increase in urge on the AUQ. Confirmatory factor analysis of the AUQ during the cue-elicited urge state also supported the single factor structure, on which all items significantly loaded \( p < 0.001 \). Positive and negative affect were positively associated with urges across the experimental protocol, but at greater magnitudes during an acutely elevated urge state.

**Conclusions:** These results further validate the use of the AUQ for real-time measurement of alcohol craving in human laboratory research.

**Key Words:** Alcohol, Craving, Urge, Cue Reactivity, Psychometric Validity.

Although alcohol craving, or an acute urge to drink alcohol (Kozlowski and Wilkinson, 1987; Kozlowski et al., 1989), is a prominent construct in contemporary alcoholism research (e.g., Kavanagh et al., 2005; Robinson and Berridge, 2001), it is also a highly contentious one. In empirical research, the majority of studies examining retrospective reports of precipitants of relapse report that craving played only a negligible role (e.g., Hodgins et al., 1995; Littman et al., 1983; Miller and Gold, 1994). Although recent applications of ecological momentary assessment that permit almost real-time data collection have affirmed an association between craving and relapse in smokers (Shiffman et al., 1996, 1997), parallel findings have yet to be reported following treatment for alcohol misuse. In addition, there is similar ambiguity in laboratory studies, where the relationship between self-reported craving and actual alcohol consumption has only been modest (Tiffany and Conklin, 2000). As a result, although a number of theories make predictions about craving (e.g., Ludwig et al., 1974; Siegel, 1998; Stewart et al., 1984; Tiffany, 1990), no single model has been clearly supported (Drummond, 2001; Tiffany, 1995).

The ambiguity in research on craving may be related to measurement error (Tiffany, 1997; Tiffany et al., 2000). The most common method of assessing craving is a single-item Likert scale, which has been criticized by Tiffany et al. (2000) on a number of counts. From a quantitative standpoint, a single-item approach precludes the calculation of internal reliability. More importantly, the phenomenon of craving is conceptually multidimensional (e.g., Stritzke et al., 2004) and single-item measures necessarily neglect facets of the construct. As a result, single-item measures not only fail to fully assess craving in general, but are likely to fall short of an individual’s personal semantic construal of craving. Finally, from a methodological standpoint, the administration of a single-item measure on multiple occasions during an experiment, a common approach in experimental research, may be contaminated by the memory of prior reports earlier in the procedure (i.e., recall bias). Although some of the preceding problems also apply to multi-item measures, on balance, these shortcomings suggest that a well-validated multi-item measure has the potential to be superior to a single-item measure and may...
clarify the relationship between craving and alcohol misuse (Tiffany, 1997; Tiffany et al., 2000).

Toward improved measurement of alcohol craving, 3 validated multi-item measures have been independently developed: the Alcohol Urge Questionnaire (AUQ; Bohn et al., 1995), the Obsessive-Compulsive Drinking Scale (OCDS; Anton et al., 1995), and the Penn Alcohol Craving Scale (PACS; Flannery et al., 1999). All 3 are relatively short, unidimensional measures that assess various semantic and chronological dimensions of craving. The AUQ, OCDS, and PACS have each undergone substantial psychometric validation (Bohn et al., 1996; Drummond and Phillips, 2002; Flannery et al., 1999) and, in support of the hypothesis that inadequate measurement may have contributed to equivocal findings, all 3 measures have recently been demonstrated to predict ongoing alcohol use and posttreatment relapse (Flannery et al., 2003).

However, the AUQ, OCDS, and PACS are not redundant and may be differentially useful depending on the research question (Flannery et al., 2003). Based on their instructions and item content, Flannery et al. (2003) propose that the OCDS and PACS are better suited for retrospective assessment of urges in weekly administrations, whereas the AUQ, a state measure, may be better suited for real-time assessment in laboratory studies. Indeed, the AUQ has already been adopted in a number of human laboratory studies of craving (Hutchison et al., 2003; MacKillop and Lisman, 2005; O’Malley et al., 1997). In these studies, the AUQ has been shown to be sensitive to increases in craving via both alcohol administrations (O’Malley et al., 1997) and alcohol cue exposures (Hutchison et al., 2003; MacKillop and Lisman, 2005).

However, an important aspect of the psychometric validation of the AUQ for human laboratory research has not been studied. Although the internal consistency and single-factor structure of the AUQ have been validated under neutral conditions (Bohn et al., 1995; Drummond and Phillips, 2002), its factor structure has not been examined during an elevated urge state. This raises the possibility of a number of potential problems. It is possible that of the AUQ’s 8 items, some are not sensitive to transient changes in craving. For example, the item that directly inquires about craving may be less sensitive due to a perceived negative connotation. As scoring involves aggregating the item scores, insensitive items would diminish the AUQ’s overall sensitivity to changes.

Another potential problem is that of multiple factors emerging during an urge state. For example, studies on multi-item measures of craving for nicotine have revealed a 2-factor structure (Cox et al., 2001): the first relating to positive reinforcement (stimulation) and the second relating to negative reinforcement (alleviation of withdrawal). It is possible that the semantic representations of alcohol urges might be similarly multifactorial during a heightened urge state. If this were the case, previous studies using the AUQ that reported significant effects of laboratory manipulations on craving as a monolithic construct may have overlooked important nuances of the data.

To address these potential issues, the goal of the current study was to validate the AUQ’s factor structure under neutral conditions and during a cue-elicited urge state. Heavy drinkers were assessed in a neutral laboratory room and a subsample then underwent an alcohol cue exposure to elicit an acutely elevated urge state. Confirmatory factor analysis (CFA) was used to determine whether the previously reported single factor structure was robust in both contexts. The AUQ was hypothesized to be sensitive to changes in urge following the alcohol cue exposure and exhibit robust psychometric properties under both conditions. The associations between urges and both alcohol consumption and symptoms of alcohol dependence were also examined, with the hypothesis that the previously reported positive correlations (Drummond and Phillips, 2002) would be replicated. Finally, to contribute to the ongoing question of the relationship between craving and affect (Baker et al., 1986, 2004; Kavanagh et al., 2005; Rohsenow et al., 1992), associations between the AUQ and both positive and negative affect were examined under neutral conditions and in an urge state. Consistent with the notions that positive and negative affect are independent states that can cooccur (Cacioppo and Berntson, 1994) and that urges to drink are processed on distinct positive and negative affective networks (Baker et al., 1986), positive and negative affect were predicted to be significantly associated with urge across the experimental procedure but more prominently so during an acute elevation.

METHODS

Participants

Participants in this study were 248 heavy drinkers (mean age = 21.03; 70% male), who were recruited from the campus of the State University of New York at Binghamton and were participating in ongoing alcohol cue reactivity research (e.g., MacKillop and Lisman, 2005). Participants were required to be heavy drinkers, operationalized as consumption of a mean of 20 or more standard drinks per week for males and 14 or more for females. These criteria were used based on empirically demonstrated links to alcohol-related health problems (Dawson, 2000). In addition, participants were required to report beer as among their 3 most commonly consumed and favorite alcoholic beverages and to rate their enjoyment of beer as at least a 7 on a 10-point scale. These criteria were included to ensure that participants would be well suited for the standardized alcohol cue exposure (described below), which used beer cues. The screening also solicited participants’ most common reason for drinking from the following categories: “happy/celebrating with friends” (69%), “to relax” (16%), “for the taste” (5%), “out of habit” (5%), “when you’re bored” (3%), “when you’re nervous/anxious” (1%), or “when you’re depressed” (0%). Participants were matched to imaginal scenes during the alcohol cue exposure based on their most commonly reported reason for drinking.

Of the 248 total participants, all were randomly assigned to 1 of 4 experimental manipulations. Of these manipulations, 1 subsample of 61 (74% male) received a standardized alcohol cue exposure and was the focus of the CFA during an elevated urge state. Descriptive statistics are provided for the overall sample and subsample in Table 1.
FACTOR STRUCTURE OF THE ALCOHOL URGE QUESTIONNAIRE

All participants were informed that they were participating in a study of the environment and alcohol use and were compensated with $5 or required research credit (whichever they preferred) for their time.

Measures

The AUQ (Bohn et al., 1995) is an 8-question measure of drinking urges. Questions are in the form of a 7-point Likert scale and participants endorse the extent to which they agree or disagree with statements relating to desire to drink, expectation of a desired outcome from drinking, and inability to avoid drinking if alcohol was available. In 2 previous studies, the AUQ has been demonstrated to have a strong internal consistency (Bohn et al., 1995; Drummond and Phillips, 2002). In this study, high internal consistency was evident ($z = 0.86$) in the overall sample and was similar to the aforementioned previous assessments. Internal consistency was adequate but somewhat lower ($z = 0.73$) in the subsample.

In terms of drinking variables, the mean number of drinks per week was assessed using the Daily Drinking Questionnaire (DDQ; Collins et al., 1985). The DDQ is a short, face-valid measure of an individual’s average drinks consumed per day over the course of a week. The DDQ is a commonly used (e.g., Murphy et al., 2004) and reliable measure (Kivlahan et al., 1990).

The Alcohol Dependence Scale (ADS; Skinner and Horn, 1984) was administered at baseline to assess symptoms of alcohol dependence, although no entry criterion was used. The ADS is a 25-item multiple-choice questionnaire, yielding a score from 0 to 40, which is linearly related to severity of dependence; it has received extensive psychometric validation, indicating its reliability and validity (Skinner and Horn, 1984).

The Positive and Negative Affect Scale (PANAS; Watson et al., 1988) was used as a measure of transient mood. The PANAS requires the participant to rate their agreement with 20 descriptors (half relating to positive affect, half relating to negative affect) on a Likert scale from 1 to 5. This widely used instrument has been shown to have adequate reliability ($z = 0.85$), as well as convergent and discriminant validity (Schumkle et al., 2002).

Procedure

The following procedures were approved by the State University of New York at Binghamton Human Subjects Research Review Board (Institutional Review Board). Participants were recruited from the State University of New York at Binghamton via campus flyers and e-mail solicitations, followed by telephone screens conducted by research assistants. Participants who met criteria for the study were scheduled for an appointment between 6 PM and 10 PM in the evening. They were informed that they would have the opportunity to drink alcohol and were required to report that no preexisting circumstances (e.g., medication, athletic event, academic requirement) would preclude them from drinking if they so chose.

During the experimental session, participants first completed an approved informed consent form and a packet of questionnaires, including the AUQ, ADS, and PANAS, in a neutral laboratory room. Following completion of the questionnaire packet, the randomly selected subsample proceeded into the standardized cue exposure. Random assignment was based on order of enrollment into the study. Every fourth enrolled participant was assigned to receive the alcohol cue exposure; the remaining participants proceeded into alternative experimental conditions, not described here.

The standardized alcohol cue exposure was oriented around beer and used visual, tactile, olfactory, proprioceptive, and imaginal stimuli. Per the recommendation of Staiger and White (1991), participants were matched to their favorite category of beer (imported, domestic, light) to increase the likelihood of responsivity to the cues. Participants were also matched to imaginal scenes related to their most common reason for drinking. These standardized scenes evocatively described common environmental, interpersonal, and affective contexts of drinking as well as the orosensory properties of beer. The scenes were developed based on previous research reports demonstrating evocation of cue-reactive responses to imaginal scenes (Maude-Griffin and Tiffany, 1996; Payne et al., 1992). Non-alcoholic beer was rebottled as commercial brands based on its orosensory and perceptual similarity to alcoholic beer (Keane et al., 1980) and Institutional Review Board concerns regarding intoxicated participants. The nonalcoholic beer used was based on pilot testing to determine the nonalcoholic brand that most closely corresponded to the 3 categories of beer used. Based on Wertz and Sayette’s (2001) recommendation, all participants were explicitly informed that the beer used in the cue exposure would be available at the conclusion of the procedure.

The alcohol cue exposure involved introducing participants to an experimental room in which the lights were dimmed and the decor was of alcohol-related images. A 90-second acclimation period was provided for the participants to become familiar with the stimuli. An experimenter then brought 3 bottles of beer of the participant’s favorite category and poured 1 bottle into a beer glass. Participants listened to the matched imaginal scene while holding the beer up to their nose and continuously smelling it. The exposure lasted approximately 5 minutes, which was also the duration of the olfactory exposure.

Following the cue exposure, participants then completed the AUQ and PANAS a second time in the cue exposure room. Before the consumption period, participants were also provided with an equivalent amount of bottled water as an alternative beverage. They were informed they could drink as much or as little beer, water, or both beverages as they wanted. Although not informed of the duration, all participants were then provided a consumption period for 5 minutes to drink ad libitum. To assess beer and water consumption, glassware, bottles, and liquid for each beverage were weighed in grams before use with each individual participant and reweighed following the procedure to assess consumption in grams. All participants were informed before entering the cue exposure that the beer would be available for drinking following the exposure.

Following the procedure, participants completed a debriefing interview and were compensated with either $5 in cash or research credits toward undergraduate research requirements. During the debriefing, participants were asked to rate the similarity of the beer used to their usual beer on a Likert scale from 1 to 10. Participants rated the beer a mean of 7.35 ($SE = 0.34$), suggesting that they were unaware that the beer was nonalcoholic. After the completion of the study, follow-up debriefing sessions were held to inform participants that nonalcoholic beer was used in the study. These debriefing sessions were conducted after the study was complete to reduce the

<table>
<thead>
<tr>
<th>Table 1. Subject Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>Body mass index</td>
</tr>
<tr>
<td>Drinks/week(^a)</td>
</tr>
<tr>
<td>Severity of dependence(^b)</td>
</tr>
</tbody>
</table>

\(^a\)Drinking Days Questionnaire. \(^b\)Alcohol Dependence Scale.
likelihood that information about the use of nonalcoholic beer would influence prospective participants’ expectations about the study.

Statistical Analysis

Data screening for distribution normality and outliers was undertaken following the recommendations of Tabachnick and Fidell (2001). All variables exhibited adequately normal distributions and no transformations were necessary.

Confirmatory factor analysis was undertaken using AMOS 4.0 (Arbuckle, 1999). Goodness-of-fit was assessed using 4 commonly used indices. First, the comparative fit index (CFI; Bentler, 1990), a direct measure of the tested model relative to the data, was calculated. A CFI value of >0.90 was used as the criterion for sufficiently good fit (Ullman, 2001). Second, the standardized root mean square residual (RMR) was calculated; an RMR value of <0.10 was used as an indicator of small discrepancies between the observed and proposed covariance matrices (Marsh and Hocevar, 1985). Third, a chi-square test was used to test whether the observed covariance matrix significantly diverged from the proposed covariance matrix. Importantly, the chi-square test examines the relative difference between the 2 matrices; therefore, a nonsignificant chi-square test value supports a proposed model. Because the significance of the chi-square test is highly dependent on the number of degrees of freedom, the ratio of chi-square test to degrees of freedom was also calculated. A ratio of χ2/df > 3 is considered reflective of a good model fit (Byrne, 1989). Finally, based on the recommendation of MacCallum and Austin (2000), the root mean square error of approximation (RMSEA) was calculated with 90% confidence intervals (CI); RMSEA values of 0.08 and below reflect a good model fit (Brown and Cudeck, 1993).

In terms of statistical power, because the sample sizes for the overall sample and subsample were smaller than those typically reported (MacCallum and Austin, 2000), special consideration was given to whether the study was adequately powered for the CFA analyses. Recent reviews have criticized structured equation modeling sample size “rules of thumb” as generally invalid and recommend focusing on the anticipated quantitative properties of the analysis (e.g., MacCallum and Austin, 2000), which was the approach used in this case. Specifically, given the conceptual relatedness of the items on the AUQ, parameter effect sizes were anticipated to be large and communalities were anticipated to be high. Based on these assumptions and the relatively small number of parameters to be estimated, both the overall sample and the subsample were anticipated to be adequately powered for the CFAs. Finally, Pearson’s product–moment correlations and analysis of variance (ANOVA)-based statistics were used as necessary. In the case of the latter, partial η2 was used as a measure of effect size.

Table 2. Alcohol Urge Questionnaire Standardized (SFL) and Unstandardized Factor Loadings (UFL) with Standard Errors Under Neutral Conditions and During a Cue-elicited Urge State

<table>
<thead>
<tr>
<th>Items</th>
<th>Neutral context</th>
<th>Post–cue exposure</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;All I want to do now is have a drink&quot;</td>
<td>0.77**</td>
<td>0.77**</td>
</tr>
<tr>
<td>&quot;I do not need to have a drink right now&quot;</td>
<td>0.51**</td>
<td>0.46**</td>
</tr>
<tr>
<td>&quot;It would be difficult to turn down a drink this minute&quot;</td>
<td>0.70**</td>
<td>0.69**</td>
</tr>
<tr>
<td>&quot;Having a drink right now would make things seem perfect&quot;</td>
<td>0.82**</td>
<td>0.84**</td>
</tr>
<tr>
<td>&quot;I want a drink so bad I can almost taste it&quot;</td>
<td>0.84**</td>
<td>0.90**</td>
</tr>
<tr>
<td>&quot;Nothing would be better than a drink right now&quot;</td>
<td>0.85**</td>
<td>0.84**</td>
</tr>
<tr>
<td>&quot;If I had the chance to have a drink, I don’t think I would drink it&quot;</td>
<td>0.31**</td>
<td>0.43**</td>
</tr>
<tr>
<td>&quot;I crave a drink right now&quot;</td>
<td>0.81**</td>
<td>0.83**</td>
</tr>
</tbody>
</table>

*Reverse scored. **p < 0.001.

RESULTS

Neutral Context

In the neutral context, participants yielded a mean AUQ score of 25.55 (SE = 0.66; 46% of scale maximum). Performance on the AUQ was significantly correlated with the number of drinks consumed per week (r = 0.26; p < 0.001) and degree of alcohol dependence (r = 0.34; p < 0.001).

Confirmatory factor analysis identified the single-factor structure reported by both Bohn et al. (1995) and Drummond and Phillips (2002) on 3 of the goodness-of-fit indices: CFI = 0.99; RMR = 0.08; RMSEA = 0.053 (CI = 0.02–0.08). Of note, the chi-square test value was marginally statistically significant, χ2 (20, N = 248) = 34.12; p = 0.03; however, the ratio of the chi-square test value to degrees of freedom was 1.71, reflecting a good fit. All items significantly loaded on the single factor (p < 0.001); unstandardized and standardized factor loadings with standard errors are presented in Table 2.

Post–alcohol Cue Exposure

Performance before and following the cue exposure for the AUQ is presented in Fig. 1. Following the alcohol cue exposure, participants yielded mean AUQ scores of 33.46 (SE = 1.70; 60% of scale maximum; 34% increase). A one-way within-subjects ANOVA demonstrated that this increase was statistically significant with a large effect size, F(1, 60) = 36.90, p < 0.001. η2 = 0.38.

Confirmatory factor analysis successfully identified the previous reported single-factor structure on all of the goodness-of-fit indices: CFI = 1.0; RMR = 0.09; RMSEA = 0.00 (CI = 0.00–0.00); χ2 (20, n = 61) = 10.83, p > 0.90, and χ2 ratio = 0.54. All items significantly loaded on the single factor (p < 0.001). Unstandardized and standardized factor loadings with standard errors are presented in Table 2.

Following the acute urge state assessment, 95% of participants consumed the beer provided during the
ad libitum drinking period. Participants consumed a mean of 393.08 g (SE = 29.86 g) of beer or approximately 1 full 16-ounce beer. The AUQ following the alcohol cue exposure was moderately correlated with the amount of beer consumed, $r = 0.31, p < 0.01$.

A notable finding in this study was that although urges exhibited a significant association with beer consumption, the relationship only accounted for about 10% of the variance. Thus, even with improved measurement, considerable variation between self-reported urge and alcohol consumption remained unexplained. One potential explanation for this relatively small magnitude association, and a limitation of the study, is the use of nonalcoholic beer.

Alternatively, another possible explanation is that the relationship between urges and beer consumption may
have actually been obscured by the use of ad libitum access. Behavioral economic research has demonstrated that substance use is highly price sensitive (Jacobs and Bickel, 1999; Murphy and MacKillop, 2006), exhibiting steep decreases in consumption as price increases. As a result, it is likely that when heavy drinkers are provided with free alcohol in circumstances with no negative contingencies, they will drink it, even if they are not in a particularly high urge state. This generalized probability of consumption may have eclipsed more subtle effects of urge state, reducing observed associations. As such, the effect of urge states on actual alcohol consumption may be more readily evident in operant or behavioral economic preparations (e.g., Griffiths et al., 1996; Higgins et al., 1994) where consumption involves some form of response cost. This possibility also applies to previous studies that have reported equivocal findings using an ad libitum approach (e.g., Brandon et al., 1996) and is an empirical question that is worthy of being pursued.

Two additional aspects of the present study should be considered. Given the use of CFA, the subsample that received the cue exposure was smaller than is typically used (MacCallum and Austin, 2000; Tabachnick and Fidell, 2001), which may be considered a limitation. To an extent, however, this potential limitation is moot because the AUQ's factor structure was validated during an urge state in spite of the relatively small sample. This is consistent with the initial assumptions that loading magnitudes and item communalities would be high and, as a result, power would be adequate. Indeed, confirming the AUQ's factor structure in a relatively small sample further underscores the robustness of the measure.

A second consideration is that both the overall sample and the subsample were predominantly male, reducing the degree to which the samples were representative. Although previous studies have not found significant gender differences in terms of the expression of craving (e.g., Bohn et al., 1995) or alcohol cue reactivity (e.g., Rubonis et al., 1994), the disproportionately high number of males is nonetheless a limitation of the current study and the results should be interpreted cautiously for generalizability. As with all CFA studies, it should be recognized that these findings are limited to this sample and are best substantiated by replication in multiple independent samples (MacCallum and Austin, 2000).

Nonetheless, the AUQ's factor structure appears to be stable and valid across experimental contexts, and it appears to provide an impressive balance of semantic variation and item parsimony in the assessment of alcohol urges. The present study suggests that it is an effective real-time measure of alcohol urges and is well suited for use in human laboratory research. Furthermore, this study lends credence to the notion that, given the limitations of single-item measures and viable multi-item alternatives, the practice of exclusively assessing craving via a single-item measure is inadvisable.

ACKNOWLEDGMENTS

The author would like to thank Emily Anderson, Bryan Castelda, Mathew Johnson, Stephen Lisman, and James Murphy for helpful comments on earlier drafts of this manuscript.

REFERENCES


Arbuckle JL (1999) AMOS 4.0. [Computer Software]. Smallwaters Corporation, Chicago, IL.


