

The Anticipated Effects of Alcohol Scale: Development and Psychometric Evaluation of a Novel Assessment Tool for Measuring Alcohol Expectancies

Meghan E. Morean
Yale University

William R. Corbin
Arizona State University

Teresa A. Treat
University of Iowa

Alcohol expectancy (AEs) research has enhanced our understanding of how anticipated alcohol effects confer risk for heavy drinking and alcohol-related problems. However, extant AE measures have limitations within 1 or more of the following areas: assessing a comprehensive range of effects, specifying the hypothetical number of drinks consumed, assessing AEs by limb of the blood alcohol curve (BAC), and/or not having undergone psychometric evaluation. Building upon the strengths of existing measures, we employed conceptual and statistical advances in measurement development to create the novel, psychometrically sound Anticipated Effects of Alcohol Scale (AEAS). Unique to this study, pilot data ensured that the AEAS comprised a comprehensive sampling of effects that varied in valence (positive/negative) and arousal (low/high) and were identified as plausible outcomes of drinking. The AEAS specified the number of drinks individuals imagined consuming (adjusted for sex) and the hypothetical drinking episode length (2 hr). AEs were also assessed separately by BAC limb. For validation purposes, the AEAS was included in several survey studies of young adults (ages 18–30). The validity argument for the proposed interpretation of AEAS test scores was based upon the following: (a) exploratory and confirmatory factor analyses ($N = 546$) identified a 22-item, 4-factor internal structure, categorizing alcohol effects into quadrants (high/low arousal crossed with positive/negative valence); (b) scalar measurement invariance was established for BAC limb, sex, and binge drinking status; (c) convergence/divergence was observed with alternative AEs measures and mood; and (d) test-criterion relationships were observed with several alcohol-related outcomes. The reliability argument was based on test–retest and internal consistency coefficients.

Keywords: alcohol expectancies, measurement development, measurement invariance, blood alcohol curve

Supplemental materials: <http://dx.doi.org/10.1037/a0028982.supp>

According to social learning theory (e.g., Bandura, 1977), alcohol use is contingent upon an individual's beliefs that the outcomes of alcohol use will be positive; namely, drinking will enhance a positive emotional state or improve a negative emotional state (e.g., Maisto, Carey, & Bradizza, 1999). These beliefs, commonly referred to as alcohol expectancies (AEs), are thought to be acquired both indirectly, through observing others' drinking behavior, and directly, through personal experience with alcohol (Bandura, 1977). The indirect acquisition of AEs begins very early in human development, well before individuals directly experience the effects of alcohol; AEs have even been observed in preschool-

ers (e.g., Dunn & Goldman, 1996). Children develop AEs through many sources (e.g., Maisto, Carey, & Bradizza, 1999), including exposure to the attitudes of their peers and family members (most often parents and siblings), as well as media depictions of drinking (e.g., advertisements, movies, television). Although young children's initial AEs tend to be negative, as children progress through the primary school years, AEs generally become increasingly positive, ultimately contributing to the initiation of alcohol consumption (e.g., Dunn & Goldman, 1998). Once alcohol use has begun, individuals' drinking experiences begin to influence and reinforce AEs directly.

Over the past 30 years, the alcohol expectancy (AE) field has sought to identify specific patterns of AEs that are most likely to motivate alcohol use and confer risk for heavy use and related problems. Broadly, positive AEs have been linked to the initiation of alcohol use, heavy drinking, the experience of negative alcohol-related consequences, alcohol use disorders, relapse, and other risk factors for negative alcohol outcomes including a family history of alcoholism (e.g., Brown, Christiansen, & Goldman, 1987; Brown, Goldman, & Christiansen, 1985; Connors, O'Farrell, & Pelcovits,

This article was published Online First June 18, 2012.

Meghan E. Morean, Department of Psychiatry, Yale University School of Medicine; William R. Corbin, Department of Psychology, Arizona State University; Teresa A. Treat, Department of Psychology, University of Iowa.

Correspondence concerning this article should be addressed to Meghan E. Morean, Department of Psychiatry, Yale University School of Medicine, Connecticut Mental Health Center, Office S-229, 34 Park Street, New Haven, CT 06519. E-mail: meghan.morean@yale.edu

1988; Cooper, Russell, & George, 1988; Kilbey, Downey, & Breslau, 1998; Lundahl, Davis, Adesso, & Lukas, 1997; Pastor & Evans, 2003; Rather & Goldman, 1994; Schuckit & Smith, 2001). The relationship between negative AEs and drinking behavior has been less consistent albeit less well studied. Several studies indicate that negative AEs are associated with reduced drinking (Leigh, 1989; Sharkansky & Finn, 1998; Wiers, Hoogveen, Sergeant, & Gunning 1997) and may protect against relapse in problem drinkers (e.g., Jones & McMahon, 1994), and others suggest that negative AEs (e.g., McMahon, Jones, & O'Donnell, 1994) or a mixed profile of negative and positive AEs (e.g., Lee, Oei, & Greeley, 1999; Leigh, 1989; Mann, Chassin, & Sher, 1987) may contribute to heavier drinking.

Some of the ambiguity across studies may be attributable to the numerous ways in which AEs have been assessed. Pooling the most desirable qualities of early AE measures, Leigh and Stacy (1993) suggested that an AE measure should be designed to capture the relative strength of AEs for both positive and negative alcohol effects. In addition to considering valence, Goldman and Darkes (2004) argued that AE measures should include items that vary in terms of arousal (i.e., stimulation vs. sedation). Although several measures contain item sets that appear to provide coverage of valence and arousal dimensions (e.g., AEMAX, Goldman & Darkes, 2004; CEOA, Fromme, Stroot, & Kaplan, 1993), coverage has been inferred rather than established through empirical evaluation.

Considering the affective characteristics of AEs has important implications beyond ensuring comprehensive coverage, as relationships between AEs and outcomes of interest may differ based on the affective characteristics of the effects. For example, imagining the arousal by valence space divided into quadrants, high arousal positive AEs (e.g., sociable, funny) may be associated with heavy drinking, whereas high arousal negative AEs (e.g., demanding, rude) may be associated with negative consequences like aggression. Unfortunately, aspects of extant AE measures make it difficult to discriminate AEs based on arousal and valence. For example, individual scale items often comprise several AEs spanning multiple affective quadrants (e.g., "People do stupid, strange, or silly things when they drink alcohol" [AEQ]). In this case, "silly" is likely to be perceived more positively than "stupid" or "strange," so it is unclear what type of emotional experience this item reflects. Items from different affective quadrants may also be grouped together in a single scale. For example, the Brief Comprehensive Effects of Alcohol (BCEOA; Ham, Stewart, Norton, & Hope, 2005) combines high arousal positive and negative AEs within its sexuality subscale (i.e., "I would enjoy sex more"; "I would be a better lover"; "I would act aggressively"). To help eliminate affective confounds, it would be helpful for individual scale items to reflect only one anticipated effect and for subscales to be based on distinct divisions of affective space (e.g., quadrants).

Irrespective of the affective characteristics of a given alcohol effect, it seems likely that AEs are dependent, at least in part, on the amount of alcohol an individual imagines consuming. Although there is little available research on this topic, there is some evidence to suggest that AEs vary based on the amount of alcohol an individual imagines consuming (e.g., George & Dermen, 1988; Southwick, Steele, Marlatt, & Lindell, 1981), the duration of the hypothetical drinking episode (Demmel, Klüsener, & Rist, 2004),

and the limb of the blood alcohol curve (BAC; e.g., Dunn & Earleywine, 2001; Earleywine & Martin, 1993). Increases in positive and negative AEs accompany increases in the imagined number of drinks consumed and in estimated blood alcohol levels (a function of estimated number of drinks and duration of the drinking episode; Demmel et al., 2004; Guarna & Rosenberg, 2000; Earleywine & Martin, 1993). In particular, positive, stimulant effects have been found to accompany moderate doses of alcohol and predominate on the ascending limb, and negative, sedative effects have been associated with larger doses of alcohol and predominate on the descending limb (e.g., Dunn & Earleywine, 2001; Earleywine & Martin, 1993; George & Dermen, 1988).

Only one existing AE measure, the Anticipated Biphasic Alcohol Effects Scale (A-BAES; Earleywine, 1994; Earleywine & Martin, 1993), accounts for alcohol dose, drinking episode length, and BAC limb. Although the A-BAES provides an important theoretical contribution to the literature, the effects assessed by the measure are largely restricted to high arousal positive (e.g., elated) and low arousal negative (e.g., sedated) effects. Thus, high arousal negative (e.g., aggressive) and low arousal positive (e.g., relaxed) effects are not well represented within this measure. As such, the A-BAES inextricably links BAC limb with the valence and arousal of effects; ascending limb effects are almost universally high arousal positive, and descending limb effects are almost universally low arousal negative. In addition, the A-BAES was derived from the BAES (Martin, Earleywine, Musty, & Perrine, 1993), a measure designed to assess subjective response to alcohol in lab-based studies, and has undergone limited psychometric evaluation as an AE measure. Thus, despite its important strengths, the A-BAES has not been widely adopted by researchers in the field.

The fact that important aspects of the imagined drinking episode have not been assessed regularly means that a vast AE literature has been built upon participants' reports of AEs related to an undisclosed number of drinks, consumed over an undisclosed period of time, without respect to BAC limb. As a result, it is difficult to determine which aspects of a hypothetical drinking experience contribute to study results. Observed AE differences attributed to factors such as group membership (e.g., heavy/light drinkers) may be accounted for, at least in part, by group-specific differences in the kinds of drinking parameters participants imagine. For example, consider the common finding that heavier drinkers have more positive AEs than lighter drinkers do (e.g., Jones, Corbin, & Fromme, 2001). Now imagine that a group of individuals is asked to report their AEs without any information about the amount of alcohol they are to imagine consuming or the duration of the drinking episode. Heavier drinkers may report their AEs imagining drinking 3 drinks over an hour, and lighter drinkers may report their AEs based on 1 drink. In this scenario, the dose imagined by lighter drinkers would lead to a low BAC and minimal reinforcing alcohol effects (e.g., stimulation), relative to the dose imagined by the heavier drinkers. In this hypothetical scenario, it would be no surprise if the heavier drinkers reported more positive AEs. This is a case in which differences in AEs attributed to drinking status may be better accounted for by group-specific differences in assumed drinking parameters. To ensure that study results reflect true differences in AEs, one would need to hold aspects of the imagined drinking episode constant.

In addition to the issues outlined previously, most extant AE measures have not undergone psychometric evaluation at a level of rigor that is now permitted by advances in statistical methods. According to the set of standard guidelines for test development proposed by the American Education Research Association, the American Psychological Association, and the National Council on Measurement in Education (1999), test developers should provide theoretically and empirically supported arguments for the validity and reliability of the interpretation of test scores they have respectively proposed. Evidence for the reliability of test scores should be based upon the following: evaluations of the relationships among test scores (items and/or subscale scores) within one administration of the test (i.e., internal consistency coefficients), over time (i.e., test-retest coefficients), and, when applicable, across different proposed versions of a measure (i.e., alternate-form coefficients). Evidence for validity should be based on test content specification, confirmation and cross-validation of internal structure, and relationships to other variables of interest (i.e., convergence, divergence, and test-criterion relations). Moving beyond classical test theory approaches, a sound AE measure would also evidence measurement invariance (MI; for reviews, see Steenkamp & Baumgartner, 1998; Vandenberg & Lance, 2000). MI is critical for making valid comparisons of AEs across groups of interest (e.g., sex, binge drinking status). Evidence for MI demonstrates that a measure assesses the underlying construct similarly across groups, thereby ensuring that participants' subscale scores reflect their "true" position on corresponding latent factors. As such, individual differences in subscale scores and, consequently, group differences in mean subscale scores reflect true (latent) differences rather than discrepancies in factor structure or systematic measurement bias originating from sources like group differences in item interpretation. Although mean level comparisons of AEs across groups are common in the literature (e.g., heavier drinkers report stronger positive AEs than lighter drinkers), a level of MI that statistically ensures the ability to make these comparisons has yet to be established. As a result, "findings of differences between individuals and groups cannot be unambiguously interpreted" (Horn & McArdle, 1992, p. 117). Failure to attend to MI may result in artifactual group "differences" being interpreted as real group differences (e.g., Chen, 2008) or in real group differences being obscured by seemingly null findings.

In summary, existing AE measures arguably underrepresent the construct they are designed to assess. AE measures have not empirically ensured coverage of a broad range of alcohol effects based on valence and arousal dimensions, and variations in methodology and item content across measures have made it difficult to interpret results across studies. In addition, aspects of an imagined drinking episode (e.g., number of drinks) have not been assessed routinely, making it difficult to determine the extent to which these factors are important determinants of alcohol outcomes and whether systematic differences in these factors may account for AE differences that have (or have not) been identified across groups (e.g., drinking status). Finally, more stringent psychometric evaluation is necessary to ensure that reasonable validity and reliability arguments support the various interpretations of test scores proposed by extant AE measures.

The current study was designed to create an AE measure that addresses the aforementioned issues by (a) developing scale items that reflect comprehensive coverage of both valence and arousal

dimensions; (b) inquiring only about effects plausibly elicited by alcohol consumption rather than, for example, characteristics of alcohol or cultural alcohol use norms; (c) specifying the amount of alcohol an individual imagines consuming (adjusted for gender) and the length of the drinking episode; (d) assessing AEs separately for the ascending and descending BAC limbs; and (e) conducting stringent psychometric evaluations. Given that alcohol use, the experience of negative alcohol-related consequences (e.g., driving under the influence; accidental injuries; ignoring important academic, employment, or family obligations) and the development of alcohol use disorders peak during emerging adulthood (e.g., Grant et al., 2004), initial measurement development efforts were focused on developing and validating the Anticipated Effects of Alcohol Scale (AEAS) for use in a sample of young adults ranging in age from 18–30 (mean age = 20.42 years, $SD = 3.74$).

As described in detail in the pages that follow, the development and validation of the AEAS occurred via a three-phase process: (a) item development, (b) instrument development, and (c) psychometric evaluation. During the item development phase, 40 words were systematically selected from a larger sample of 215 words based on their coverage of affective space (i.e., positive, negative, stimulant, sedative), alcohol relatedness, and linguistic features (e.g., reading level). During the instrument development phase, the structural configuration and instructional set for the AEAS were determined. Finally, the validity and reliability of the proposed interpretation of test scores were assessed during the psychometric evaluation phase (i.e., evaluations of internal structure; measurement invariance; convergent, discriminant, and test-criterion relationships with external variables; internal consistency; and stability of test scores over a 2-week period).

Method

Item Set Development

Initial item pool selection. We identified 215 words that were potentially related to alcohol. Of these, 132 were from the Alcohol Expectancy Multiaxial Assessment (AEMAX; Goldman & Darkes, 2004) and 83 were from the Affective Norms for English Words, a large database of English words for which normative information on valence and arousal has been established (Bradley & Lang, 1999).

Pilot Study 1: Establishing affective norms of possible items. Normative data on the perceived valence and arousal of the initial 215 words were collected to ensure comprehensive coverage of affective space.

Participants. Fifty college students and individuals from the community (27 women; mean age = 21.2 years, $SD = 2.9$) served as study participants. Participants were eligible if they were at least 18 years old.

Procedure. All words were presented in counterbalanced blocks for valence and arousal ratings. Participants rated the affective connotations of each word using a 9-point graphic emotion rating scale (Self-Assessment Manikin; Bradley & Lang, 1994). For perceptions of valence, the word *happy* appeared on the screen accompanied by figures corresponding to the 1–9 rating scale that displayed emotions ranging from sadness (a frowning face corresponding to 1) to happiness (a smiling face corresponding to 9). Similarly, for perceptions of arousal, the word *stimulated* appeared

on the screen, with corresponding figures ranging from sedated (1) to stimulated (9).

Pilot Study 2: Establishing alcohol-relatedness of possible items. To ensure that the items used to develop the AEAS were perceived as plausible outcomes of drinking, we conducted a two-part study.

Participants. Fifty-five college students, who were at least 18 years of age and who had not previously participated in the first pilot study (29 female, mean age = 20.7 years, $SD = 1.2$; mean weekly drinks = 10.2, $SD = 6.8$), as well as 15 experts in the alcohol field (determined by Ph.D. status and alcohol-related research or clinical interests), participated in the study.

Procedure. Students rated all 215 words on a 10-point rating scale, ranging from *no relation to drinking alcohol* to *certain relation to drinking alcohol*. "Drinking alcohol" was explicitly defined as consuming any type of alcohol (e.g., beer, liquor). Student ratings were supplemented with ratings from experts in the alcohol field, who identified only those words they believed to be unrelated to alcohol.

Selection of the 40 AEAS items for measurement development. We dropped words if they had alcohol-relatedness ratings ≤ 3.9 (the midpoint of the range of relatedness scores), were marked for deletion by five or more experts, had multiple meanings (e.g., smooth), or were nonadjectives. This reduced the item pool by roughly 40%. Subsequent decisions about item retention occurred separately within each quadrant and were based on coverage of affective space (i.e., maximizing dispersion), relatedness ratings, linguistic features (i.e., reading level/frequency of usage), and breadth of coverage of behaviors and emotions. To facilitate this process, words were grouped based on similarity of meaning when possible (e.g., *attractive* and *sexy*). Next, alcohol relatedness and reading levels were considered for each word in the group. In cases in which words were synonymous (e.g., *sleepy*, *tired*), occupied a similar space in the affective grid, and evidenced a similar reading level, final decisions were based on alcohol relatedness. Ultimately, 10 items were retained in each affective quadrant. The mean alcohol relatedness score for all 40 items was 6.0, only four words were marked for deletion by more than one expert, and 78% of the items were within the 10,000 most commonly used words in the English language (Davies, 2011).

Instrument Development

The instructional set and physical formatting of the 40-item AEAS were modeled after the A-BAES (Earleywine & Martin, 1993), although several important alterations were made. Given the relative novelty of specifying drinking parameters regarding amount of alcohol consumed and the duration of the hypothetical drinking episode when assessing AEs, we felt that it was important to initially validate the AEAS based on an established drinking metric that is meaningful within the alcohol research community. Although we recognized that not every participant would have experienced drinking at a binge level, we made a decision to use the binge drinking criteria set forth by the National Institute on Alcohol Abuse and Alcoholism (4 drinks for women or 5 drinks for men within a 2-hr period).¹ Adopting these drinking parameters allowed us to account for gender, an important contributing factor to blood alcohol levels, when considering AEs. In addition, the binge drinking criteria conveniently approximate the number of

standard drinks an individual must consume over a 2-hr period to reach a blood alcohol level of .08g%, which is the legal limit of intoxication in the United States, and also represents a common target blood alcohol level used in alcohol administration studies. Use of this target blood alcohol level in AE research may facilitate future comparisons of AEs and the experience of subjective alcohol effects during an actual drinking episode. To address concerns about the differential experience of alcohol effects over the course of a drinking episode, we asked participants to imagine how they would expect to feel immediately after consuming their fourth/fifth drink (ascending limb) and 90 min after finishing their final drink (descending limb).

Psychometric Evaluation

Participants. Participants were drawn from four survey studies in which web-based versions of the AEAS were administered: (a) a follow-up to an alcohol administration study assessing the impact of drinking on gambling in a simulated bar ($n = 108$); (b) a survey examining young adults' substance use and eating behavior ($n = 107$); (c) a survey examining first-year college students' transition to campus life ($n = 331$); and (d) a survey evaluating the stability of AEAS test scores over a 2-week period ($n = 55$). Across studies, participants varied with respect to age and levels of engagement in a range of alcohol use behaviors. Measures of interest were common across the studies (although recruitment methods and study procedures differed). To maximize the variability of participant engagement in a range of alcohol-related behaviors that are important during the developmental period of emerging adulthood, we combined participants from the first three studies into a single sample for purposes of development and validation of the AEAS ($N = 546$; mean age = 20.42 years, $SD = 3.74$; 50.3% female; 58% Caucasian, 19% Asian, 6% African American, 17% other; average weekly drinks = 17.42, $SD = 14.41$; maximum drinks consumed on a single occasion = 7.58, $SD = 4.64$; binge drinking episodes = 4.71, $SD = 6.00$; alcohol-related problems experienced = 6.03, $SD = 8.87$). Test-retest reliability was evaluated with data from the fourth survey. (See Table 1 for a summary of participant characteristics and study methods; for a detailed description, see the online supplement, "Description of Studies Used to Evaluate the Reliability and Validity of the Proposed Interpretation of AEAS Test Scores.")

Measures.

Demographics. This measure assessed factors including age, sex, and race.

¹ For the first 212 participants, sex and weight were entered into a computerized blood alcohol level calculator for each individual to determine the personalized number of drinks he or she would need to consume in 2 hours to reach a blood alcohol level of at least .08g%. However, an examination of measures of central tendency challenged the utility of determining a personalized number of drinks for each participant. For men and women (men $n = 129$; women $n = 83$), the mean (4.97, $SD = .88$; 4.20, $SD = .94$), median (5.00; 4.00), and modal (5.00; 4.00) number of drinks necessary to reach a BAC of .08 were extremely close to the 4/5 drink binge criteria, and less than 6% of all participants needing to drink more to reach .08 g%. To facilitate data collection in a survey study format, we decided to adopt the more parsimonious 4/5 criteria.

Table 1
Study Design, Inclusion Criteria, AEAS Version, and Sample Characteristics Across the Survey Studies Used to Validate the AEAS

Study characteristic	Alcohol and gambling persistence	Health behaviors survey	2012 survey	Test-retest reliability survey
Study design	AEAS included as part of a 2-week follow-up survey to an alcohol administration study conducted in a simulated bar setting	AEAS included as part of a large, Web-based survey assessing engagement in various health-related behaviors (e.g., drinking, smoking)	AEAS included as part of a large, Web-based survey of first-year college students' transition to life on campus	AEAS included in a two-part Web-based survey to assess test-retest reliability
Inclusion criteria	Consume ≥ 3 alcoholic drinks at least 1 time per week No contraindications to alcohol (e.g., flushing) Played poker in the past year (note: Participants completed a computerized gambling task not relevant to the current study aims) No current/past enrollment in abstinence-oriented alcohol or gambling program Not pregnant	≥ 18 years old	First-year college students	≥ 18 years old
AEAS version	Personalized no. drinks specified based on gender, weight	Personalized no. drinks specified based on gender, weight	Standard specification of 5 drinks for men or 4 drinks for women, consumed within a 2-hr period	Standard specification of 5 drinks for men or 4 drinks for women, consumed within a 2-hr period
Sample characteristics	108 college students and community participants 73.6% male	107 college students and community participants 49.5% male	331 first-year college students 42.3% male	55 college students and community participants 50% male
Gender				
Age	22.88 (2.36)	23.50 (5.64)	18.41 (0.62)	23.5 (3.10)
Ethnicity	75% non-Hispanic Caucasian	54% non-Hispanic Caucasian	56% non-Hispanic Caucasian	51.7% non-Hispanic Caucasian
Weekly drinks	18.16 (10.84)	19.54 (13.81)	15.37 (10.99)	
Maximum drinks	6.08 (2.95)	8.58 (5.25)	7.82 (4.82)	
Binge drinking	7.75 (6.19)	3.97 (5.05)	4.94 (6.31)	
Alcohol problems	11.56 (10.70)	6.91 (10.40)	4.09 (7.19)	

Note. Parenthetical data are standard deviations. AEAS = Anticipated Effects of Alcohol Scale.

Assessing alcohol expectancies. We carefully considered which existing measures of expectancies to include in our efforts to examine the validity of the AEAS. The Alcohol Expectancy Questionnaire (AEQ; Brown, Christiansen, & Goldman, 1987) and the Comprehensive Effects of Alcohol (CEOA; Fromme, Stroot, & Kaplan, 1993) are the two most widely used measures of alcohol expectancies; both have been cited roughly 20 times each year since publication, with total citation counts of greater than 350 for each. We selected the CEOA over the AEQ primarily because it includes both positive and negative anticipated alcohol effects, a critical feature given our interest in examining the full valence by arousal space. In addition, the CEOA has established psychometric properties in short form (B-CEOA; Ham et al., 2005). The Alcohol Multi-Axial Assessment–Short Form (AEMAX; Goldman & Darkes, 2004) was included as a second expectancy measure, given the strong measurement development procedures used in its creation and the fact that it is the most similar of the existing measures of expectancies to the AEAS (e.g., efforts to capture both valence and arousal of anticipated alcohol effects). Thus, we felt that it was important to demonstrate incremental validity relative to this well-developed scale.

The Anticipated Effects of Alcohol Scale (AEAS). The AEAS is described in detail later in the text. A copy of the measure is included in Figure 1.

The B-CEOA (Ham et al., 2005). Using a 5-point rating scale (*disagree to agree*), participants rated 15 AEs using the stem “After a few drinks of alcohol, I would be more likely to . . .” Subscales included Sociability/Liquid Courage (Cronbach’s $\alpha = .89$), Change in Self-Perception/Cognitive-Behavioral Impairment ($\alpha = .70$), Sexual Experiences ($\alpha = .59$), and Tension Reduction ($\alpha = .80$).

The AEMAX–Short Form (Goldman & Darkes, 2004). Using a 7-point rating scale (*never to always*), participants rated 24 AEs using the stem “Drinking alcohol makes one . . .” Subscales included attractive ($\alpha = .89$), social ($\alpha = .91$), horny ($\alpha = .78$), sleepy ($\alpha = .85$), egotistical ($\alpha = .84$), woozy ($\alpha = .78$), dangerous ($\alpha = .83$), and sick ($\alpha = .91$).

Assessing alcohol use and related problems.

Weekly alcohol use and maximum daily alcohol use. Participants reported the number of drinks they typically consume in a week, the maximum number of drinks consumed in a single drinking episode during the past 3 months, and the frequency of engaging in binge drinking.

The Rutgers Alcohol Problem Index (RAPI; White & Labouvie, 1989). Participants rated how frequently they experienced 23 alcohol-related social/health problems over the past 3 months

AEAS (female version*)												
Instructions: The following is a list of feelings that people may experience after drinking alcohol. People often report different experiences during a drinking episode and after stopping drinking for the night. Therefore, you will be asked to indicate 1) how you expect to feel immediately after drinking 4 drinks within a 2 hour period and 2) how you expect to feel an hour and a half (90 minutes) after finishing your final (4 th) drink.												
If you have never consumed 4 drinks in a 2 hour period, please indicate how you think you would feel in this situation.												
Remember, we want to know what feelings you expect to experience as a result of drinking not whether you are currently experiencing them.												
On a scale of 0 – 10 please rate the extent to which you believe you would experience each of these feelings immediately after drinking 4 drinks.												
		Not At All			Moderately				Extremely			
1.	Sociable	0	1	2	3	4	5	6	7	8	9	10
2.	Moody	0	1	2	3	4	5	6	7	8	9	10
3.	Demanding	0	1	2	3	4	5	6	7	8	9	10
4.	Mellow	0	1	2	3	4	5	6	7	8	9	10
5.	Carefree	0	1	2	3	4	5	6	7	8	9	10
6.	Rude	0	1	2	3	4	5	6	7	8	9	10
7.	Relaxed	0	1	2	3	4	5	6	7	8	9	10
8.	Woozy	0	1	2	3	4	5	6	7	8	9	10
9.	Fun	0	1	2	3	4	5	6	7	8	9	10
10.	Lively	0	1	2	3	4	5	6	7	8	9	10
11.	Calm	0	1	2	3	4	5	6	7	8	9	10
12.	Aggressive	0	1	2	3	4	5	6	7	8	9	10
13.	Dizzy	0	1	2	3	4	5	6	7	8	9	10
14.	Anxious	0	1	2	3	4	5	6	7	8	9	10
15.	Attractive	0	1	2	3	4	5	6	7	8	9	10
16.	Ill	0	1	2	3	4	5	6	7	8	9	10
17.	Wobbly	0	1	2	3	4	5	6	7	8	9	10
18.	Funny	0	1	2	3	4	5	6	7	8	9	10
19.	Talkative	0	1	2	3	4	5	6	7	8	9	10
20.	Confident	0	1	2	3	4	5	6	7	8	9	10
21.	Happy	0	1	2	3	4	5	6	7	8	9	10
22.	Drunk	0	1	2	3	4	5	6	7	8	9	10
Using the same 0 – 10 scale, please rate the extent to which you believe you would experience each of these feelings an hour and a half (90 minutes) after finishing your final (4th) drink. (list of 22 items repeats)												

Figure 1. The Anticipated Effects of Alcohol Scale (AEAS).

using the following prompts: never, 1–2 times, 3–5 times, 6–10 times, >10 times ($\alpha = .89$). Sample items include “neglected responsibilities” and “tried to cut down or stop drinking.”

Assessing mood.

The Short Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). Using a 5-point rating scale (*not at all/very slightly to extremely*), participants rated the extent to which they were currently experiencing 10 positive (e.g., enthusiastic) and 10 negative (e.g., ashamed) affective states. Subscales included positive ($\alpha = .90$) and negative affect ($\alpha = .86$).

Data Analytic Plan

Data from all participants from the first three studies who had completed the measures included in the current study were merged into a single data file. All alcohol-related variables were log-transformed to approximate normal distributions (weekly drinks: skewness of raw data = 4.34, skewness of transformed data = .46; maximum drinks: raw = 1.02, transformed = .40; binge drinking: raw = 2.51, transformed = .94; alcohol-related problems: raw = 2.80, transformed = .23).

A combination of exploratory factor analysis (EFA), theoretically and data-driven confirmatory factor analysis (CFA), and internal consistency analyses (Cronbach’s α) was used to develop the preliminary version of the AEAS. An EFA of the 40 items for the ascending limb (i.e., AEAS-A) was conducted for a randomly selected sample of 273 (50% of the total sample). A combination of eigenvalues, model fit indices, solution interpretability, number of items per factor, and consistency with theoretical expectations was used to determine the preliminary factor structure. Next, a CFA model (i.e., the 22-item EFA model) that was developed within the first random sample ($N = 273$) was cross-validated within the remaining sample ($n = 273$). The model was then fit to the ascending and descending limb data from the full sample ($N = 546$).² To ensure that meaningful comparisons of AE could be made across the ascending and descending limbs of the BAC, we used a multiple-group CFA approach to assess measurement invariance (for a review, see Vandenberg & Lance, 2000). After evaluating MI for limb of the BAC, we computed factor scores and evaluated the reliability and validity of the proposed interpretation of the AEAS test scores. Subsequently, MI was evaluated for gender and binge drinking status with the same multigroup CFA approach.

Results

Evidence for the Validity of the Proposed Interpretation of AEAS Subscale Scores Based on Internal Structure

EFA. We hypothesized that a four-factor internal structure would emerge corresponding to quadrants of affective space (i.e., high arousal positive [HIGH+]; high arousal negative [HIGH-]; low arousal positive [LOW+]; low arousal negative [LOW-]). Nonetheless, formal evaluation of the AEAS began with an EFA to examine the dimensionality of the 40 AEAS–Ascending (i.e., AEAS-A) items in the absence of a priori restrictions and to identify items that violated the simple structure requirements of CFA by loading strongly onto multiple factors. Using Mplus 5.1

(Muthén & Muthén, 1998–2008), we fit an EFA model to a random subsample of 273 participants (50% of the total sample) with the intention to cross-validate the resulting factor structure with the second half of the sample. Several of the AEAS items were nonnormally distributed. Therefore, robust maximum likelihood estimation was used, as this estimation method is robust to nonnormality and produces a range of fit indices that are helpful in determining latent factor structure. Given the hypothesized four-factor solution, associations between the latent factors based on shared valence and/or arousal were anticipated. For example, HIGH+ (e.g., sociable) and LOW+ (e.g., relaxed) factors were expected to be correlated, given their shared valence. Therefore, an oblique rotation (i.e., CF-Varimax [oblique]) was specified. Missing data were handled with full information maximum likelihood estimation; 4.1% of the overall data was missing. Based on study hypotheses and the factor structures of existing AE measures, the EFA was permitted to extract up to eight factors. Modification indices were requested to identify theoretically justifiable alterations that could improve model fit.

A variety of indices influenced model selection: eigenvalue magnitudes, examination of the scree plot (e.g., Tabachnick & Fidell, 2001), model fit indices, prior theoretical expectations, factor interpretability, and whether factors contained enough items to be considered a scale (e.g., Jöreskog & Sörbom, 1989). Although chi-square tests have historically been used to evaluate model fit, they are overly sensitive to sample size, leading to an overrejection of potentially viable models when sample sizes are large (e.g., Chen, 2007; Cheung & Rensvold, 2002). As such, chi-square was reported as a matter of convention but was not used as a diagnostic fit index. Instead, the following fit indices were used: Bentler’s comparative fit index (CFI), the Tucker–Lewis index (TLI), the root-mean-square error of approximation (RMSEA), and the standardized root-mean-square residual (SRMR). These fit indices were chosen because (a) they represent both absolute and relative fit indices (Newsom, 2005); (b) CFI, RMSEA, and SRMR are among the most stable fit indices even when sample size, loadings, and sample distribution vary (Chen, 2007); and (c) TLI is relatively insensitive to sample size and encourages selection of parsimonious models (Newsom, 2005). Goodness of model fit was evaluated with the following criteria: CFI and TLI indices $> .90$ (Bentler, 1990), a 90% RMSEA confidence interval with an upper limit $< .08$ (e.g., Brown, 2006; Newsom, 2005), and SRMR $< .08$ (Hu & Bentler, 1999; Newsom, 2005).

Consideration of the eigenvalues, the scree plot, and the model fit indices suggested that the four-, five-, and six- factor solutions merited further evaluation. Within each model, items were retained if the latent factor accounted for at least 30% of the variance in the item (factor loadings $\geq .55$) and no other latent factor accounted

² All analyses were also completed starting with the AEAS-D to ensure that the latent factor structure and factor indicators identified through EFA were similar irrespective of the starting point. Three items differed across the analyses: “lonely” and “depressed” were identified only by the EFA starting with the AEAS-A, and “silly” was identified only by the EFA starting with the AEAS-D. Ultimately, “lonely” and “depressed” were removed from the measure, so “silly” was the only item that differed when the analytic process began with the AEAS-D.

for more than 9% of the variance (cross-loadings $< .30$). Next, the viability of each model was evaluated based on the number of items comprised by each factor. The five- and six-factor models were rejected because each contained a factor with fewer than three items, the minimum number considered sufficient for forming a scale (e.g., Jöreskog & Sörbom, 1989) and estimating latent variables (e.g., Levitt, Sher, & Bartholow, 2009). Only a 24-item, four-factor model remained.

The four-factor model comprised two positive and two negative factors. The two positive factors (LOW+ and HIGH+) and the LOW- factor generally corresponded to the hypothesized quadrants of affective space; all of the LOW+ items fell within the anticipated quadrant, the HIGH+ factor contained only one “borderline” item (carefree), which had been rated as falling just within the LOW+ quadrant during instrument development, and the LOW- quadrant included two borderline items (dizzy, drunk), which fell just within the high arousal quadrant during instrument development. However, the factor that best captured HIGH- effects did not correspond as clearly to the anticipated affective quadrant. This factor contained two items (lonely, depressed) that had been rated previously as low arousal. In addition to a mismatch in terms of arousal, “lonely” and “depressed” were identified as problematic in several important ways. In the EFA evaluating the potential five- and six-factor models, “lonely” and “depressed” formed a two-item factor suggestive of a potential negative psychological factor. However, these models were inconsistent with the scree plot, and the additional factor(s) accounted for little variance above the four-factor model. Further, within the four-factor AEAS-A model, “lonely” and “depressed” evidenced considerable overlap; the modification index indicated that their re-

sidual variances were highly correlated ($MI = 90.0$). Given the concerns noted above, these items were dropped from the scale. At this point, each of the four subscales was distinguishable from the others based on arousal and valence (e.g., the arousal level of all items in the LOW- quadrant was lower relative to those falling into the HIGH- quadrant; see Figure 2). The four factors of the 22-item AEAS-A were intercorrelated (HIGH+ with LOW+ [.650]; HIGH- with LOW- [.658]; HIGH+ with HIGH- [.198]; LOW+ with LOW- [.283]). However, the magnitudes of these correlations were below the established criteria for multicollinearity ($r > .80$, Meyers, Gamst, & Guarino, 2006; $r > .90$, Tabachnick & Fidell, 2001), reducing concerns in later validity analyses. The mean alcohol-relatedness score (6.10, $SD = .85$) indicated that the average item fell within the quartile of words most strongly related to alcohol.

CFA. Based on study hypotheses and the EFA, the 22-item four-factor model identified within Random Sample 1 ($N = 273$) was fit with CFA within Random Sample 2 ($N=273$). For each factor, the highest loading item defined the factor metric; the factor loading was set to 1.0. The HIGH+ and LOW+ factors were allowed to correlate, as were the HIGH- and LOW- factors. Robust maximum likelihood estimation was specified. The goodness-of-fit indices indicated acceptable model fit within Random Sample 2, $\chi^2(231) = 3,165.749$, SRMR = .08, RMSEA = .069, 90% CI [0.061, 0.077], TLI = .900, CFI = .912. The CFA model was then fit to the ascending limb data from the full sample, $\chi^2(231) = 5,452.391$, SRMR = .074, RMSEA = .063, 90% CI [0.058, 0.069], TLI = .905, CFI = .916. Within the full sample, the AEAS-A subscales demonstrated adequate to excellent internal consistency (mean $\alpha = .860$; range = .739–.944). In preparation

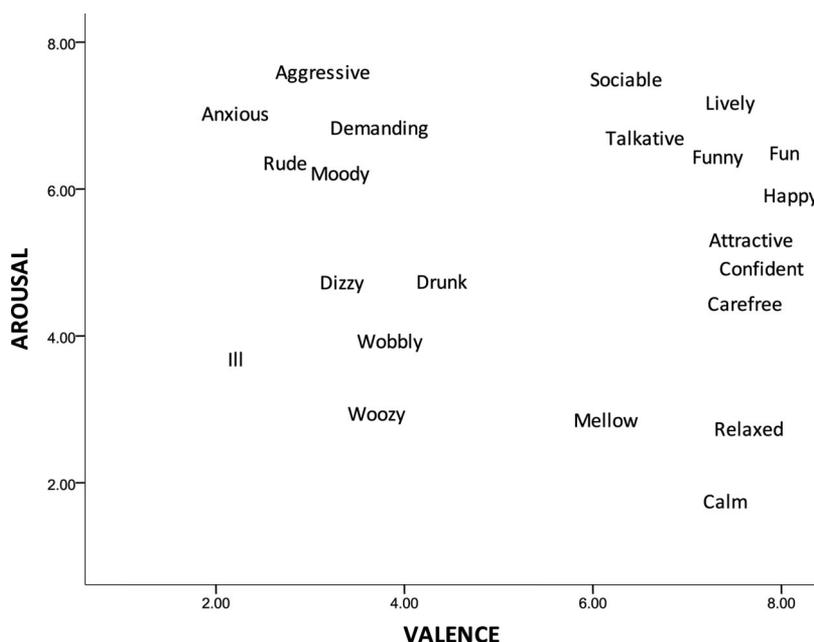


Figure 2. The 22 items of the Anticipated Effects of Alcohol Scale (AEAS) plotted by arousal and valence ratings. The 22 AEAS items are plotted by the valence and arousal ratings provided by the 50 participants who completed the affective norms study, which occurred during the item development phase. All words were presented in counterbalanced blocks for valence and arousal, and participants rated the affective connotations of each word using a 9-point graphic emotion rating scale (Self-Assessment Manikin; Bradley & Lang, 1994).

for testing MI, the AEAS-A model was fit to the descending limb data (AEAS-D) for the full sample. Model fit was acceptable, $\chi^2(231) = 6,333.873$, SRMR = .085, RMSEA = .072, 90% CI [0.067, 0.078], TLI = .896, CFI = .909, and the AEAS-D subscales demonstrated excellent internal consistency (mean $\alpha = .890$; range = .801–.956). In sum, all prerequisites for testing measurement invariance were met: (a) model fits were acceptable for assessments of AEs on the ascending and descending limbs when assessed independently, (b) all freely estimated factor loadings were significant (see Table 2 for factor loadings), and (c) there were no problems identified via the modification indices.

Measurement invariance (MI) across limbs of the blood alcohol curve. The CFA approach to assessing MI across the ascending and descending BAC limbs occurred in three steps: evaluating configural, metric, and scalar invariance. First, it is essential to demonstrate that a similar global latent factor structure is shared across groups when no equality constraints are in place. Therefore, configural invariance was tested to assess whether the same four-factor model fit the data across the ascending and descending limbs, the loadings of all items comprising the latent factors were significant across limbs, and the correlations among latent factors were not so strong as to indicate collinearity (e.g., Steenkamp & Baumgartner, 1998). It is also important that latent variables are measured on the same scale (i.e., are based on the same *metric*). Therefore, metric invariance was tested to establish whether the strength of the relationships of the latent factors to their constituent items (i.e., factor loadings) was comparable across the limbs. Finally, scalar invariance was tested to establish whether items' origins were invariant across limbs. Establishing scalar invariance ensures that limb-specific differences in factor scores reflect differences in latent factor means rather than measurement bias (Chen, 2008; Steenkamp & Baumgartner, 1998; Widaman & Reise, 1997).

Configural invariance. A two-group CFA model was specified in Mplus to fit the four-factor model to the AEAS-A and the AEAS-D data simultaneously.³ All observed dependent variables were continuous, so maximum likelihood estimation with robust standard errors and chi-squares was used (ESTIMATOR = MLR). The factor loadings of the four factor metrics (i.e., the highest loading items for each factor) were set to 1.0, and the factor means were set to zero. All remaining model parameters (e.g., factor loadings, intercepts, variances, covariances) were freely estimated. The resulting model evidenced good fit, $\chi^2(406) = 1,399.289$, SRMR = .079, RMSEA = .068, 90% CI [.064, .072], TLI = .900, CFI = .912. All items significantly loaded onto their respective factors on the ascending and descending limbs. As expected, the HIGH- and LOW- subscales as well as the HIGH+ and LOW+ subscales were significantly correlated, but the magnitudes of these correlations were below the established criteria for multicollinearity ($r > .80$, Meyers et al., 2006; $r > .90$, Tabachnick & Fidell, 2001). The fit of the configurally invariant model served as a benchmark against which the fit of the metric invariant model was compared.

Metric invariance. The factor loadings of matching items on the AEAS-A and the AEAS-D were constrained to equality (e.g., factor loadings of "sociable" on the AEAS-A and AEAS-D were set to be equal), and the latent factor means were set to zero. Based on a series of statistical cutoffs established by Chen (2007) indicating that noninvariance exists only in cases where the decrement

in model fits exceeds SRMR $\geq .030$, RMSEA $\geq .015$, or CFI $\geq -.01$, the resulting model did not evidence significant decrement in fit, $\chi^2(424) = 1,447.717$, SRMR = .080, RMSEA = .067, 90% CI [.064, .071], TLI = .901, CFI = .910, when compared to the configurally invariant model (Δ SRMR = .001, Δ RMSEA = $-.001$, Δ CFI = $-.002$, Δ TLI = .001). Thus, individual items related to their respective latent factors equivalently across BAC limbs.

Scalar invariance. Factor loadings and intercepts (item means) of matching items on the AEAS-A and AEAS-D were constrained to equality while the latent factor means were allowed to be freely estimated. Chen (2007) suggested that unique change in fit indices be considered when evaluating scalar invariance, with changes in CFI $\geq -.010$ accompanied by a change in SRMR $\geq .010$ or RMSEA $\geq .015$ indicating noninvariance. Based on these indices, the resulting model evidenced no significant decrement in fit, $\chi^2(442) = 1,627.287$, SRMR = .082, RMSEA = .071, 90% CI [.067, .075], TLI = .891, CFI = .895, when compared to the model testing metric invariance (Δ CFI = $-.015$, Δ TLI = $-.010$, Δ SRMR = .002, Δ RMSEA = .004).

Computation of factor scores. A summary scale approach to scoring the AEAS, in which subscale scores are computed from raw item responses, inherently gives all items on a given factor equal weight or "importance" (i.e., all factor loadings are essentially set to 1). However, as demonstrated by the CFA, individual items actually vary in the strength of their relation to the latent factor (i.e., have different factor loadings) such that certain items are functionally more "important" than others. Unlike summary scores, factor scores derived from the CFA provide information about participants' positions on each factor based on their responses to individual items comprised by the subscale. Therefore, factor scores for the scalar invariant model derived within Mplus (using the modal posterior estimator) were used in all subsequent analyses.⁴

Measurement invariance by gender and binge drinking status. Using a CFA multigroup approach similar to that described above, we evaluated configural, metric, and scalar MI for gender and for binge drinking status, respectively. Models were run separately for the ascending and descending limb data with gender or binge drinking status identified as the grouping variable. Scalar MI was achieved for gender and binge drinking

³ A series of specialized syntax commands was specified within Mplus to account for the dependent sampling nature of the data (i.e., each individual provided responses to the AEAS for both the ascending and descending limbs). First, within the "variable" section of the Mplus syntax, commands were entered to indicate that the data were clustered by participant (i.e., CLUSTER = ID), with each participant providing data for the ascending and descending limbs of the AEAS. Second, in order to ensure that the clustered nature of the data was taken into account when computing standard errors and tests of model fit, the analysis type was specified as complex.

⁴ Bivariate correlations between corresponding AEAS subscale scores derived with either Mplus factor scores or a summary approach ranged between .97 and .99. The very strong correlations suggest that a summary score approach to obtaining subscale scores would be practical for future use. However, given the current study's commitment to evaluating the psychometric properties of the AEAS as precisely as possible, Mplus factor scores were used for all statistical analyses.

Table 2

Confirmatory Factor Analysis: AEAS Subscale Internal Consistencies and Item Factor Loadings When the Four-Factor AEAS Model Was Fit to the AEAS-Ascending and AEAS-Descending Limb Data From the Full Sample (N = 546)

Subscale	Internal consistency			
	Ascending limb		Descending limb	
HIGH+	$\alpha = .944$		$\alpha = .956$	
HIGH-	$\alpha = .853$		$\alpha = .883$	
LOW+	$\alpha = .739$		$\alpha = .801$	
LOW-	$\alpha = .904$		$\alpha = .924$	

Items by subscale	Factor loadings							
	Ascending limb				Descending limb			
	Estimate	SE	Est./SE	StdYX	Estimate	SE	Est./SE	StdYX
HIGH+								
Lively	1.000				1.000			
Sociable	0.954	.038	25.055	.811	0.925	.038	24.592	.826
Carefree	0.968	.043	22.368	.745	0.899	.041	22.146	.784
Fun	1.084	.043	25.382	.862	1.025	.031	33.465	.874
Attractive	0.952	.048	19.639	.722	0.909	.041	22.347	.782
Funny	0.984	.039	25.475	.802	0.996	.036	27.722	.856
Talkative	1.021	.040	25.590	.823	1.005	.040	25.261	.836
Confident	1.006	.046	21.852	.827	0.971	.037	26.035	.861
Happy	1.101	.045	24.368	.878	1.062	.036	29.382	.892
HIGH-								
Demanding	1.000				1.000			
Moody	0.919	.062	14.817	.681	1.074	.054	19.792	.773
Rude	1.184	.065	18.315	.818	1.076	.048	22.533	.842
Aggressive	1.018	.066	15.530	.728	0.899	.055	16.430	.751
Anxious	0.834	.095	8.780	.669	0.901	.059	15.234	.695
LOW+								
Mellow	1.000				1.000			
Calm	1.164	.107	10.880	.659	0.898	.058	15.552	.687
Relaxed	1.331	.156	8.503	.762	0.986	.080	12.393	.765
LOW-								
Wobbly	1.000				1.000			
Woozy	0.815	.038	21.493	.769	0.920	.032	29.011	.852
Dizzy	0.930	.029	31.771	.872	0.934	.026	35.421	.912
Ill	0.861	.037	23.110	.818	0.902	.031	28.777	.840
Drunk	0.731	.037	19.593	.687	0.727	.039	18.616	.679

Note. Bolded AEAS items are factor metric items. AEAS = Anticipated Effects of Alcohol Scale; SE = standard error; Est./SE = a ratio of the factor estimate and the standard error; StdXY = standardized factor loading.

status, respectively, assuring our ability to make meaningful comparisons of AEs within these groups using the AEAS (see online supplemental Table 1 for MI by gender and binge drinking status). For both gender and binge drinking status, there were latent mean differences across a number of AEAS subscales. With respect to gender, men expected stronger LOW+ AEs and weaker LOW- AEs on the ascending limb than women did and stronger HIGH+, stronger LOW+, and weaker LOW- AEs on the descending limb than women did. With regard to drinking status, binge drinkers expected stronger HIGH+ and weaker LOW- effects on the ascending limb and stronger HIGH+, stronger LOW+, and weaker LOW- effects on the descending limb than non-binge drinkers did. Although establishing MI for gender and binge drinking status permits a number of interesting research questions to be addressed in the future, a thorough explication of group-level differences in AEs is beyond the scope of the current paper.

Evidence for the Validity of the Proposed Interpretation of AEAS Subscale Scores Based on Relations With Other Variables

Convergent and discriminant evidence. The convergence and divergence of associations of the AEAS scores with extant measures of AEs (BCEOA; Ham et al., 2005; AEMAX-S; Goldman & Darkes, 2004) and affective state (PANAS; Watson et al., 1988) were examined with bivariate correlations. Correlations were designated as small (.10-.29), moderate (.30-.49), or large ($\geq .50$; Cohen, 1992). Predicted relations among the test scores of related AE measures were based on evaluations of the individual items comprised by the subscales of the existing measures rather than on the subscale names, because subscale names did not always adequately reflect the valence or arousal of the items. For example, the BCEOA's sexual subscale combines positive and negative experiences: "I would be a better lover," "I would enjoy

sex more,” and “I would be aggressive.” In general, the observed relationships provided convergent and discriminant evidence for the validity of the AEAS test scores (see online supplemental Table 2 for bivariate correlations). In support of convergence, the HIGH+ subscales correlated moderately with the B-CEOA sociability and liquid courage subscale and the AEMAX sociable subscale; the AEAS HIGH- subscales correlated moderately with the B-CEOA change in self-perception/cognitive behavioral impairment subscale; the AEAS LOW+ subscales correlated moderately with the B-CEOA tension reduction subscale; and the AEAS LOW- subscales correlated moderately with the B-CEOA change in self-perception/cognitive behavioral impairment subscale and the AEMAX sick and sleepy subscales. As anticipated, the AEAS subscales correlated weakly with the PANAS subscales, demonstrating the independence of the AEs test scores and general affective states.

Test-criterion evidence.

Concurrence. Multiple regression was employed to investigate whether there was concurrent evidence for the validity of the proposed interpretation of the AEAS scores. Concurrence was evaluated separately for the AEAS-A and AEAS-D scores. Sex was included as a covariate, given significant correlations with AEs observed in the current data and in the AE literature. Separate regression analyses (by limb) were used to evaluate the cross-sectional test-criterion relationships between the AEAS subscale scores and the following: average number of drinks per week, the maximum number of drinks consumed on a single occasion, frequency of binge drinking, and the experience of alcohol-related

problems. All variables referred to behavior over the past three months. All models were assessed for multicollinearity using the following criteria: a variance inflation factor for a given subscale above 10 (e.g., Chatterjee, Hadi, & Price, 2000) and a condition index ≥ 30 with the subscale accounting for considerable variance in ≥ 2 outcome variables (i.e., variance proportions $> .50$; Tabachnick & Fidell, 2001). Based on these criteria, there were no significant concerns with multicollinearity. The AEAS accounted for 8–22% of the variance in alcohol use variables (see Table 3), with the majority of variance accounted for by scores on the HIGH+, HIGH-, and LOW- subscales. Of note, an interesting pattern of results emerged based on the arousal level of negative AEs. Where significant relationships emerged, scores on the HIGH- subscale were positively associated with alcohol outcomes. However, high scores on the LOW- subscale seemed to serve a protective function and were inversely related to all outcomes.

Incremental evidence. Hierarchical multiple regression analyses were performed to examine the whether the proposed test score interpretations for the AEAS-A and AEAS-D, respectively, were able to account for incremental variance in predicting alcohol-related outcomes above and beyond the B-CEOA and the AEMAX, respectively. Predictor variables were entered simultaneously into the regression model: sex, the subscales of the alternative AE measure, and the four AEAS-A or AEAS-D subscale scores. The AEAS accounted for 2–9% of the variance in alcohol-related outcomes above and beyond that accounted for by the other predictors, with similar results across the CEOA and the AEMAX. Participant scores on the AEAS-A

Table 3
Hierarchical Regression Analyses Demonstrating the Predictive Validity of the AEAS

Predictor	AEAS, ascending limb subscales							
	Weekly drinks		Max. no. drinks		Binge (2 hr)		Problems	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.10***		.16***		.08***		.05***	
Sex		-.32***		-.41***				-.24***
Step 2	.18***		.20***		.13***		.14***	
HIGH+		.25***		.20*		.21***		.17***
HIGH-		.05		.24**		.17**		.25***
LOW+		-.04		-.08		-.12*		.04
LOW-		-.44***		-.52***		-.42***		-.48***
Total R^2	.28***		.36***		.22***		.20***	
n	473		447		488		508	
	AEAS, descending limb subscales							
Step 1	.09***		.09***		.08***		.05***	
Sex		-.31***		-.29***		-.30***		-.24***
Step 2	.13***		.08***		.08***		.10***	
HIGH+		.28***		.23***		.17**		.17**
HIGH-		.10		.08		.21**		.25***
LOW+		-.03		-.05		-.04		.04
LOW-		-.35***		-.26***		-.37***		-.48***
Total R^2	.23***		.17***		.16***		.20***	
n	450		427		463		497	

Note. Dependent variables are weekly drinks (average no. drinks consumed per week); max. no. drinks (the maximum number of drinks consumed on a single occasion); binge (2 hr) (drinking 4 drinks for women or 5 drinks for men within a 2-hr period); problems (the experience of alcohol-related problems). Step 1 variable was sex, and Step 2 variables were the AEAS subscales HIGH+ (high arousal positive), HIGH- (high arousal negative), LOW+ (low arousal positive), and LOW- (low arousal negative). AEAS = Anticipated Effects of Alcohol Scale.

* $p < .05$. ** $p < .01$. *** $p < .001$.

and AEAS-D LOW- subscales were the strongest predictors, accounting for significant incremental variance in all outcomes above and beyond both the BCEOA and the AEMAX, respectively. HIGH+ and HIGH- subscale scores each accounted for significant incremental variance in some, but not all, of the relationships assessed. Given space constraints and the greater familiarity of the CEOA, data are presented only for the B-CEOA (see Table 4; data for the AEMAX are presented in online supplemental Table 3).

Evidence for the Reliability of the Proposed Interpretation of AEAS Subscale Scores

Stability of test scores over 2 weeks. Bivariate correlations between corresponding subscales (e.g., ascending limb HIGH+ subscale scores at baseline and follow-up) ranged from .80 to .93 ($M = .85$), indicating good to excellent stability of the AEAS subscales scores over a 2-week period.

Discussion

The current study built upon the strengths of previous AE measures by utilizing advances in approaches to measurement development, including assessment of measurement invariance, to develop a novel theoretically and psychometrically sound AE measure. Development of the AEAS involved a three-phase process: item development, instrument development, and psychometric evaluation. For each phase, the current study's methodological and theoretical contributions to the AE field are discussed. A discussion of study limitations and important areas for future research follows.

With respect to the item set, the current study was the first to ensure that the items selected for measurement development provided adequate coverage of affective space and were, in fact, perceived to be plausible outcomes of drinking; data were collected on individuals' perceptions of the arousal and valence of each word as well as the alcohol relatedness of each word. The initial 40 items were carefully chosen such that they provided broad

Table 4
Hierarchical Regression Analyses Demonstrating the Incremental Validity of the AEAS Above the BCEOA

Predictor	AEAS, ascending limb subscales							
	Weekly drinks		Max. no. drinks		Binge (2 hr)		Problems	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Step 1	.23***		.17***		.15***		.18***	
Sex		-.23***		-.23***		-.22***		-.10*
Sociable/LQ		.11		.00		.10		.19**
Δ Percep.		-.08		-.14**		-.04		.10*
Sexuality		.14**		.10		.08		.11*
TR		.03		-.08		-.11		-.03
Step 2	.08***		.07***		.08***		.07***	
HIGH+		.19**		.23***		.15**		.11*
HIGH-		-.02		.02		.12†		.14*
LOW+		-.02		-.05		-.05		.05
LOW-		-.33***		-.31***		-.39***		-.40***
Total R^2	.32***		.24***		.23***		.25***	
<i>n</i>	418		415		355		468	
	AEAS, descending limb subscales							
Step 1	.23***		.16***		.14***		.18***	
Sex		-.24***		-.25***		-.24***		-.13**
Sociable/LQ		.13*		.06		.14*		.25***
Δ Percep.		-.13*		-.20***		-.10		.06
Sexuality		.14*		.10		.08		.10
TR		.05		-.08		-.12*		.01
Step 2	.05***		.04***		.04***		.04***	
HIGH+		.22***		.18**		.10		.09
HIGH-		.05		.08		.18*		.19**
LOW+		-.05		-.03		-.01		-.05
LOW-		-.25***		-.21***		-.33***		-.32***
Total R^2	.28***		.20***		.19***		.23***	
<i>n</i>	418		410		346		462	

Note. Dependent variables are weekly drinks (average no. drinks consumed per week); max. no. drinks (the maximum number of drinks consumed on a single occasion); binge (2 hr) (drinking 4 drinks for women or 5 drinks for men within a 2-hr period); problems (the experience of alcohol-related problems). Regression steps: Step 1, participant sex and the CEOA subscales Sociable/LQ (Sociability and Liquid Courage); Δ Percep. (Change in Self-Perception); Sexuality (Sexuality); TR (Tension Reduction); Step 2, the AEAS subscales HIGH+ (high arousal positive), HIGH- (high arousal negative), LOW+ (low arousal positive), and LOW- (low arousal negative). AEAS = Anticipated Effects of Alcohol Scale; BCEOA = Brief Comprehensive Effects of Alcohol.
 † $p < .01$. * $p < .05$. ** $p < .01$. *** $p < .001$.

coverage of the affective quadrants, were among the words identified as most strongly related to alcohol, covered a wide range of possible behavioral and affective outcomes of alcohol use, and were accessible to a broad range of participants based on reading level.

Most notable among the measure's structural features, the AEAS explicitly specifies the number of drinks an individual imagines consuming (adjusted for gender) and the duration of the hypothetical drinking episode and assesses AEs separately by BAC limb. The benefits of these structural elements are evident in the results discussed below.

Once the initial item set was chosen and the structure of the measure was established, the proposed AEAS test score interpretation underwent extensive psychometric evaluation. With respect to the internal structure, the four-factor EFA model was generally consistent with study hypotheses and the CFA model adequately fit the data. EFA models that are subsequently fit in CFA often require additional alterations (as indicated by model modification indices) before adequate fit is demonstrated, effectively reverting a confirmatory modeling approach to one that is exploratory in nature (Browne, 2001). The fact that no alternations to the CFA model were necessary in the current study was consistent with the confirmatory nature of the approach and provided further justification for the model categorizing AEs into affective quadrants.

Capitalizing on advancements in test theory and analytic approaches, the current study was the first to ensure that comparisons of mean AE subscale scores (in this case across limbs of the BAC, by gender, and by binge drinking status) could be interpreted meaningfully. To our knowledge, scalar MI has not been established for patterns of test scores associated with any extant measure, calling into question the interpretability of the extant body of research on differences in AEs by BAC, gender, or drinking status. For example, group differences in AEs (e.g., men vs. women) that have been identified, as well as relationships between AEs and other constructs (e.g., family history), could be due, at least in part, to violations of configural, metric, or scalar invariance. There is no way of knowing, for example, whether discrepancies in internal structure or systematic measurement bias originating from sources such as group differences in item interpretation may have produced or obscured gender differences in latent scores.

Evidence for the validity and reliability of the proposed interpretations of AEAS scores was drawn from a number of sources. With respect to evidence based on test-criterion relationships, the AEAS is one of few AE measures whose subscale scores account for significant variance in outcomes beyond alcohol use (e.g., alcohol-related problems). In addition, study results indicated that discriminating AEs based on arousal and valence is important in relation to alcohol outcomes. For example, strong HIGH- AEs (e.g., aggressive, rude) were associated with heavier alcohol use and alcohol-related problems. Conversely, strong LOW- AEs (e.g., drunk, woozy) may serve a protective function and were associated with decreased alcohol-related behavior across all outcomes assessed. These results suggest that the specificity of the proposed AEAS test scores may be helpful in clarifying some of the conflicting findings in the AE literature, especially those regarding associations between negative AEs and drinking behavior.

The results regarding scores on the LOW+ subscale also merit consideration. The tension reduction model of alcohol use, a prominent theory for 60 years, suggests that individuals consume alcohol for its anxiolytic properties (for a review, see Conger,

1951, 1956; Greeley & Oei, 1999). However, the results of the current study suggested that anticipating the types of LOW+ effects consistent with the tension reduction model (e.g., relaxed, calm, mellow) was not positively associated with any outcome assessed. In contrast to the tension reduction model, subscale scores reflecting stronger stronger LOW+ AEs on the ascending limb were associated only with decreased binge drinking. Although the tension reduction model was not supported in the current study, it is important to consider that assessing LOW+ AEs may have greater utility within certain populations (e.g., clinical samples, individuals high in anxiety sensitivity). Future work evaluating LOW+ AEs as they apply to the tension reduction model is needed.

Finally, providing direct evidence of its incremental utility, the AEAS subscale scores accounted for significant variance in each of the outcomes assessed above and beyond two prominent AE measures, the CEOA and the AEMAX. The incremental utility of the AEAS was evident in two important but distinct ways. First, scores on the novel AEAS subscales accounted for incremental variance in alcohol-related outcomes (e.g., HIGH- scores predicted weekly drinks and alcohol-related problems on the ascending limb and binge drinking and problems on the descending limb). Second, AEAS subscale scores accounted for incremental variance in alcohol-related outcomes even when the AEAS subscale scores evidenced moderate to strong correlations with subscales from alternative AE measures. For example, AEAS LOW- subscale scores correlated quite strongly with scores on the CEOA woozy subscales (.50 ascending; .48 descending). However, LOW- scores accounted for variance in all outcomes assessed but scores on the woozy subscales did not predict any of the outcomes. It seems likely that the incremental utility of the AEAS relates to the greater specificity of the proposed test score interpretations with respect to the number of drinks imagined, the duration of the hypothetical drinking episode, the assessment of AEs across the BAC limbs, and its better delineation of affective space.

The current study has a number of important strengths, but several limitations merit note. Although there is strong evidence supporting the utility of self-report measures (Del Boca & Noll, 2000), it is important to acknowledge that the current study relied upon self-report data to develop and validate the proposed AEAS score interpretations and was, therefore, limited by participants' willingness to respond honestly. It is also of note that the current data were cross-sectional. Future longitudinal studies should evaluate whether the proposed interpretation of test scores can predict outcomes of interest over time (independently and above and beyond extant AE measures).

With respect to the study sample, the AEAS was explicitly developed and validated for use in a sample of young adults ranging in age from 18 to 30 (mean age = 20.42, *SD* = 3.74). Understanding the AEs and drinking behavior of individuals within this age range is of particular import because alcohol use, the experience of alcohol-related problems, and the development of alcohol use disorders peak during this developmental period (Grant et al., 2004). However, it is unclear to what extent the results would generalize to other age groups, pending future research in this area. Given that AEs develop prior to alcohol use and previously have been identified as important contributing factors to the onset of drinking, we recently began a study assessing AEs

using the AEAS in an adolescent sample, which has a larger percentage of individuals who have yet to initiate alcohol use.

Future evaluations of the psychometric properties of the AEAS in this sample will help to answer the question of whether the proposed interpretation of AEAS scores in the current study generalized to younger drinkers and/or abstainers.

The current study was also limited by the fact that 95% of study participants reported consuming at least one drink in the past month, making it infeasible to evaluate the psychometric properties of the AEAS independently for drinkers and abstainers. Rather than eliminate abstainers, a decision was made to develop and validate the proposed interpretation of the AEAS test scores in a sample that combined drinkers and nondrinkers based, in part, on a promising study that found support for configural and metric invariance of AEs across drinkers and abstainers using a modified version of the Adolescent Alcohol Expectancy Questionnaire (Martino, Collins, Ellickson, Schell, & McCaffrey, 2006). However, future research evaluating MI of the four-factor AEAS in abstainers (e.g., those who have never consumed alcohol, individuals in recovery who made a decision to abstain from alcohol) will be important.

A final study limitation is that all participants were asked to report ascending limb AEs prior to descending limb AEs, making it impossible to examine the potential impact of order effects. However, the decision not to counterbalance the order of presentation was a conscious one we made prior to data collection. We decided to universally ask about ascending limb effects first because this order best reflects the way that alcohol effects are actually experienced; people never experience descending limb effects before ascending limb effects. Given our instructional set, we were worried that participants would be confused if they were asked to report about their descending limb AEs first. Also contributing to our decision, there would be an inherent contrast by limb irrespective of the order in which questionnaires were presented. We were concerned that if participants were asked about the ascending limb first, they might feel naturally compelled to report weaker expectancies on the descending limb. We were similarly concerned that, if participants were asked to answer the descending limb questions first, they might be primed to report stronger effects when asked about more immediate outcomes. Thus, we felt that, if order effects were present, they would likely produce the same outcome—stronger ascending limb effects. Nonetheless, we acknowledge that it is impossible to know if and how mean AE levels might differ if the order of their presentation were counterbalanced.

Despite study limitations, the development of the AEAS advances the literature in several important ways. The current study was the first to ensure empirically that the AEs assessed were plausible outcomes of alcohol use and that they provided adequate coverage of both arousal and valence dimensions. Further, to ensure that findings were not accounted for by variability in the drinking parameters imagined by participants, the current study specified gender-specific drinking parameters corresponding to binge drinking criteria (4 drinks for women, 5 drinks for men in 2 hr) and assessed AEs separately by BAC limb. To ensure our ability to meaningfully interpret group differences in AEs across limbs of the BAC, by gender, and by binge drinking status, the current study was the first to establish scalar measurement invariance across these domains. Ultimately, study results indicated that

discriminating effects based on affective quadrant and limb of the BAC is important in relation to outcomes of interest. For example, high HIGH+ scores (e.g., sociable, confident) were positively associated with all outcomes assessed, and high LOW- scores (e.g., drunk, woozy) were inversely associated with all outcomes.

Moving forward, the strengths of the AEAS will permit a number of exciting research endeavors. The proposed interpretation of AEAS scores could be used to establish a clear dose-response profile for AEs by systematically manipulating the number of imagined drinks (increasing or decreasing from the levels in the current study), the duration of the hypothetical drinking episode, and/or the BAC limb. Establishing a dose-response profile for AEs may help clarify past research findings obtained when nonspecific prompts were used to characterize alcohol quantity (e.g., “too much”) or no information about alcohol quantity was provided. It will also be interesting to see at what point AEs normatively begin to shift from positive to negative as the number of drinks increases. Variability in this “tipping point” may be related to important risk factors for alcohol-related problems including tolerance (e.g., Morean & Corbin, 2008).

Future research could also evaluate the proposed interpretation of AE test scores based on participants’ personal drinking behavior. For example, participants could be asked to identify the typical number of drinks they consume over a 2-hr period and subsequently report their AEs based on this self-generated number. In this case, the number of drinks may vary considerably across participants, but the fixed duration of the drinking episode allows standardized estimates of blood alcohol level to be calculated for each individual. These estimates could be particularly helpful in evaluating AEs when differences in pharmacological response to alcohol would be expected (e.g., by gender, by weight). For example, based on their physiological composition, women reach higher blood alcohol levels than do men at comparable doses of alcohol (e.g., Mumenthaler, Taylor, O’Hara, & Yesavage, 1999). Although a number of studies have reported gender differences in AEs, a study that controlled for differences in estimated blood alcohol level found that men and women’s AEs were not different (Demmel et al., 2004). These findings highlight the concern that the many group differences in AEs commonly noted in the literature may be attributable to differences in participants’ self-generated referents. Future work examining how differences in estimated blood alcohol level impact a range of alcohol-related outcomes is essential.

In summary, there have been numerous important advances in the AE field over the past 30 years, but many important questions have remained. Perhaps best conceptualized as a victim of its own success, the AE field has produced a large number of methodologically and theoretically diverse measures that have likely contributed to the equally diverse and often conflicting research findings. As a result, it has been difficult to reach consensus about fundamental aspects of the construct of AEs ranging from latent internal structure to identifying constellations of anticipated alcohol effects that confer risk for heavy drinking, the experience of alcohol-related problems, and the development of alcohol use disorders. Moving forward, collaborative efforts within the AE field will be essential if the shared goals of understanding the construct, its relation to alcohol-related behavior, and, ultimately, its role in informing the treatment of alcohol problems are to be realized. We

hope that the newly developed AEAS will serve as an impetus for future work in this direction.

References

- American Educational Research Association, American Psychological Association, & the National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Bandura, A. (1977). *Social learning theory*. Oxford, England: Prentice-Hall.
- Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238–246.
- Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The Self-Assessment Manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*, 49–59. doi:10.1016/0005-7916(94)90063-9
- Bradley, M. M., & Lang, P. J. (1999). Fearfulness and affective evaluations of pictures. *Motivation and Emotion*, *23*, 1–13. doi:10.1023/A:1021375216854
- Brown, S. A., Christiansen, B. A., & Goldman, M. S. (1987). The Alcohol Expectancy Questionnaire: An instrument for the assessment of adolescent and adult alcohol expectancies. *Journal of Studies on Alcohol*, *48*, 483–491.
- Brown, S. A., Goldman, M. S., & Christiansen, B. A. (1985). Do alcohol expectancies mediate drinking patterns of adults? *Journal of Consulting and Clinical Psychology*, *53*, 512–519. doi:10.1037/0022-006X.53.4.512
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford Press.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, *36*, 111–150. doi:10.1207/S15327906MBR3601_05
- Chatterjee, S., Hadi, A. S., & Price, B. (2000). *Regression analysis by example*. New York, NY: Wiley.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. doi:10.1080/10705510701301834
- Chen, F. F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of Personality and Social Psychology*, *95*, 1005–1018. doi:10.1037/a0013193
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi:10.1207/S15328007SEM0902_5
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155–159.
- Conger, J. J. (1951). The effects of alcohol on conflict behavior in the albino rat. *Quarterly Journal of Studies on Alcohol*, *12*, 1–29.
- Conger, J. J. (1956). Reinforcement theory and the dynamics of alcoholism. *Quarterly Journal of Studies on Alcohol*, *17*, 296–305.
- Connors, G. J., O'Farrell, T. J., & Pelcovits, M. A. (1988). Drinking outcome expectancies among male alcoholics during relapse situations. *British Journal of Addiction*, *83*, 561–566. doi:10.1111/j.1360-0443.1988.tb02575.x
- Cooper, M. L., Russell, M., & George, W. H. (1988). Coping, expectancies, and alcohol abuse: A test of social learning formulations. *Journal of Abnormal Psychology*, *97*, 218–230. doi:10.1037/0021-843X.97.2.218
- Davies, M. (2011). *Word frequency data from the Corpus of Contemporary American English (COCA)* [Data file]. Retrieved from <http://wordfrequency.info>
- Del Boca, F. K., & Noll, J. A. (2000). Truth or consequences: The validity of self-report data in health services research on addictions. *Addiction*, *95*(Suppl. 3), S347–S360.
- Demmel, R., Klüsener, J., & Rist, F. (2004). Anticipated levels of alcohol-induced sedation and stimulation in relation to estimated blood alcohol concentration. *Journal of Studies on Alcohol*, *65*, 22–26.
- Dunn, M. E., & Earleywine, M. (2001). Activation of alcohol expectancies in memory in relation to limb of the blood alcohol curve. *Psychology of Addictive Behaviors*, *15*, 18–24. doi:10.1037/0893-164X.15.1.18
- Dunn, M. E., & Goldman, M. S. (1996). Empirical modeling of an alcohol expectancy memory network in elementary school children as a function of grade. *Experimental and Clinical Psychopharmacology*, *4*, 209–217.
- Dunn, M. E., & Goldman, M. S. (1998). Age and drinking-related differences in the memory organization of alcohol expectancies in 3rd-, 6th-, 9th-, and 12th-grade children. *Journal of Consulting and Clinical Psychology*, *66*, 579–585. doi:10.1037/0022-006X.66.3.579
- Earleywine, M. (1994). Confirming the factor structure of the Anticipated Biphasic Alcohol Effects Scale. *Alcoholism: Clinical and Experimental Research*, *18*, 861–866. doi:10.1111/j.1530-0277.1994.tb00051.x
- Earleywine, M., & Martin, C. S. (1993). Anticipated stimulant and sedative effects of alcohol vary with dosage and limb of the blood alcohol curve. *Alcoholism: Clinical and Experimental Research*, *17*, 135–139. doi:10.1111/j.1530-0277.1993.tb00738.x
- Fromme, K., Stroot, E. A., & Kaplan, D. (1993). Comprehensive effects of alcohol: Development and psychometric assessment of a new expectancy questionnaire. *Psychological Assessment*, *5*, 19–26. doi:10.1037/1040-3590.5.1.19
- George, W. H., & Dermen, K. H. (1988). Self-reported alcohol expectancies for self and other as a function of behavior type and dosage set. *Journal of Substance Abuse*, *1*, 71–78. doi:10.1016/S0899-3289(88)80010-5
- Goldman, M. S., & Darkes, J. (2004). Alcohol expectancy multiaxial assessment: A memory network-based approach. *Psychological Assessment*, *16*, 4–15. doi:10.1037/1040-3590.16.1.4
- Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., Dufour, M. C., & Pickering, R. P. (2004). The 12-month prevalence and trends in DSM-IV alcohol abuse and dependence: United States, 1991–1992 and 2001–2002. *Drug and Alcohol Dependence*, *74*, 223–234. doi:10.1016/j.drugalcdep.2004.02.004
- Greeley, J., & Oei, T. (1999). Alcohol and tension reduction. In H. T. Blane (Ed.), *Psychological theories of drinking and alcoholism* (2nd ed., pp. 14–53). New York, NY: Guilford Press.
- Guarna, J., & Rosenberg, H. (2000). Influence of dose and beverage type instructions on alcohol outcome expectancies of DUI offenders. *Journal of Studies on Alcohol*, *61*, 341–344.
- Ham, L. S., Stewart, S. H., Norton, P. J., & Hope, D. A. (2005). Psychometric assessment of the Comprehensive Effects of Alcohol Questionnaire: Comparing a brief version to the original full scale. *Journal of Psychopathology and Behavioral Assessment*, *27*, 141–158. doi:10.1007/s10862-005-0631-9
- Horn, J. L., & McArdle, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, *18*, 117–144. doi:10.1080/03610739208253916
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55. doi:10.1080/10705519909540118
- Jones, B. T., Corbin, W., & Fromme, K. (2001). A review of expectancy theory and alcohol consumption. *Addiction*, *96*, 57–72. doi:10.1046/j.1360-0443.2001.961575.x
- Jones, B. T., & McMahon, J. (1994). Negative alcohol expectancy predicts post-treatment abstinence survivorship: The whether, when and why of relapse to a first drink. *Addiction*, *89*, 1653–1665. doi:10.1111/j.1360-0443.1994.tb03766.x
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications*. Chicago, IL: SPSS.
- Kilbey, M. M., Downey, K., & Breslau, N. (1998). Predicting the emergence and persistence of alcohol dependence in young adults: The role

- of expectancy and other risk factors. *Experimental and Clinical Psychopharmacology*, 6, 149–156. doi:10.1037/1064-1297.6.2.149
- Lee, N. K., Oei, T. P. S., & Greeley, J. D. (1999). The interaction of alcohol expectancies and drinking refusal self-efficacy in high and low risk drinkers. *Addiction Research*, 7, 91–102. doi:10.3109/16066359909004377
- Leigh, B. C. (1989). Attitudes and expectancies as predictors of drinking habits: A comparison of three scales. *Journal of Studies on Alcohol*, 50, 432–440.
- Leigh, B. C., & Stacy, A. W. (1993). Alcohol outcome expectancies: Scale construction and predictive utility in higher order confirmatory models. *Psychological Assessment*, 5, 216–229. doi:10.1037/1040-3590.5.2.216
- Levitt, A., Sher, K. J., & Bartholow, B. D. (2009). The language of intoxication: Preliminary investigations. *Alcoholism: Clinical and Experimental Research*, 33, 448–454. doi:10.1111/j.1530-0277.2008.00855.x
- Lundahl, L. H., Davis, T. M., Adesso, V. J., & Lukas, S. E. (1997). Alcohol expectancies: Effects of gender, age, and family history of alcoholism. *Addictive Behaviors*, 22, 115–125. doi:10.1016/S0306-4603(96)00022-6
- Maisto, S. A., Carey, K. B., & Bradizza, C. M. (1999). Social learning theory. In H. T. Blane (Ed.), *Psychological theories of drinking and alcoholism* (Vol. 2., pp. 106–163). New York, NY: Guilford Press.
- Mann, L. M., Chassin, L., & Sher, K. J. (1987). Alcohol expectancies and the risk for alcoholism. *Journal of Consulting and Clinical Psychology*, 55, 411–417. doi:10.1037/0022-006X.55.3.411
- Martin, C. S., Earleywine, M., Musty, R. E., & Perrine, M. W. (1993). Development and validation of the Biphasic Alcohol Effects Scale. *Alcoholism: Clinical and Experimental Research*, 17, 140–146. doi:10.1111/j.1530-0277.1993.tb00739.x
- Martino, S. C., Collins, R. L., Ellickson, P. L., Schell, T. L., & McCaffrey, D. (2006). Socio-environmental influences on adolescents' alcohol outcome expectancies: A prospective analysis. *Addiction*, 101, 971–983. doi:10.1111/j.1360-0443.2006.01445.x
- McMahon, J., Jones, B. T., & O'Donnell, P. (1994). Comparing positive and negative alcohol expectancies in male and female social drinkers. *Addiction Research*, 1, 349–365. doi:10.3109/16066359409005202
- Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Thousand Oaks, CA: Sage.
- Morean, M. E., & Corbin, W. R. (2008). Subjective alcohol effects and drinking behavior: The relative influence of early response and acquired tolerance. *Addictive Behaviors*, 33, 1306–1313. doi:10.1016/j.addbeh.2008.06.007
- Mumenthaler, M. S., Taylor, J. L., O'Hara, R., & Yesavage, J. A. (1999). Gender differences in moderate drinking effects. *Alcohol Research & Health*, 23, 55–64.
- Muthén, L. K., & Muthén, B. O. (1998–2008). *Mplus user's guide* (5th ed.). Los Angeles, CA: Authors.
- Newsom. (2005). *Some clarifications and recommendations on fit indices*. Retrieved from http://www.upa.pdx.edu/IOA/newsom/semclass/ho_fit.pdf
- Pastor, A. D., & Evans, S. M. (2003). Alcohol outcome expectancies and risk for alcohol use problems in women with and without a family history of alcoholism. *Drug and Alcohol Dependence*, 70, 201–214. doi:10.1016/S0376-8716(03)00007-3
- Rather, B. C., & Goldman, M. S. (1994). Drinking-related differences in the memory organization of alcohol expectancies. *Experimental and Clinical Psychopharmacology*, 2, 167–183. doi:10.1037/1064-1297.2.2.167
- Schuckit, M. A., & Smith, T. L. (2001). A comparison of the correlates of DSM-IV alcohol abuse or dependence among more than 400 sons of alcoholics and controls. *Alcoholism: Clinical and Experimental Research*, 25, 1–8. doi:10.1111/j.1530-0277.2001.tb02120.x
- Sharkansky, E. J., & Finn, P. R. (1998). Effects of outcome expectancies and disinhibition on ad lib alcohol consumption. *Journal of Studies on Alcohol*, 59, 198–206.
- Southwick, L., Steele, C., Marlatt, A., & Lindell, M. (1981). Alcohol-related expectancies: Defined by phase of intoxication and drinking experience. *Journal of Consulting and Clinical Psychology*, 49, 713–721. doi:10.1037/0022-006X.49.5.713
- Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research*, 25, 78–90. doi:10.1086/209528
- Tabachnick, B. G., & Fidell, L. S. (Eds.). (2001). *Using multivariate statistics* (4th ed.). Needham Heights, MA: Allyn & Bacon.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, 3, 4–70. doi:10.1177/109442810031002
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54, 1063–1070. doi:10.1037/0022-3514.54.6.1063
- White, H. R., & Labouvie, E. W. (1989). Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol*, 50, 30–37.
- Widaman, K. F., & Reise, S. P. (1997). Exploring the measurement invariance of psychological instruments: Applications in the substance use domain. In S. G. West (Ed.), *The science of prevention: Methodological advances from alcohol and substance abuse research* (pp. 281–324). Washington, DC: American Psychological Association.
- Wiers, R. W., Hoogveen, K., Sergeant, J. A., & Gunning, W. B. (1997). High- and low-dose alcohol-related expectancies and the differential associations with drinking in male and female adolescents and young adults. *Addiction*, 92, 871–888. doi:10.1111/j.1360-0443.1997.tb02956.x

Received September 14, 2011

Revision received April 24, 2012

Accepted May 7, 2012 ■