Alcohol expectancies for social facilitation: A short form with decreased bias

Margaret-Anne Mackintosh a,*, Mitch Earleywine b, Michael E. Dunn c

a Seeley G. Madd Building, Room 518, Department of Psychology, University of Southern California, Los Angeles, CA 90089-1061, United States
b Department of Psychology, University of Albany, State University of New York, Albany, NY 12222, United States
c Department of Psychology, University of Central Florida, P.O. Box 161390, Orlando, FL 32816-1390, United States

Abstract

The social facilitation scale of the Alcohol Expectancy Questionnaire-Adolescent version predicts subsequent drinking behavior and covaries with other important constructs. An examination of the interval-level measurement properties and potential biases of items could help improve the scale. Responses in a large sample (N=518) of undergraduates confirmed that the items fit a Rasch model for a single-factor, interval scale. Two items were biased when comparing non-drinkers, social drinkers and extreme drinkers. Four additional items showed gender bias. Removal of these items shortened the scale without decreasing correlations with two measures associated with drinking, while the strength with one measure was slightly decreased. This short form of the questionnaire should prove useful in the study of alcohol expectancies for social facilitation without drinking- or gender-related bias. These results also support the utility of Rasch modeling.

© 2005 Elsevier Ltd. All rights reserved.

Keywords: Alcohol use; Alcohol expectancies; Adolescents; Rasch model; Psychometrics

* Corresponding author. Tel.: +1 213 821 2839; fax: +1 213 7465994.
E-mail addresses: mackinto@usc.edu (M.-A. Mackintosh), mearleywine@albany.edu (M. Earleywine), mdunn@pegasus.cc.ucf.edu (M.E. Dunn).

0306-4603/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.addbeh.2005.11.009
1. Introduction

Alcohol expectancies, beliefs about the drug’s impact on behavior and subjective experience, contribute to alcohol use and problems. Expectancies for alcohol-induced social changes play an important role in predicting alcohol consumption. They also help explain how risk for alcoholism relates to problem drinking (Darkes, Greenbaum, & Goldman, 2004; Fischer, Smith, Anderson, & Flory, 2000). Many aspects of risk for alcoholism, including a family history of the disorder or disinhibited personality traits, appear to lead to alcohol problems via expectancies (Sher, Walitzer, Wood, & Brent, 1991; Stacy, Newcomb, & Bentler, 1991). These expectancies are particularly important in young drinkers, leading researchers to develop the Alcohol Expectancy Questionnaire for Adolescents (AEQ-A). This scale predicts drinking concurrently and longitudinally in adolescents and young adults (e.g., Christiansen, Smith, Roehling, & Goldman, 1989; Smith, Goldman, Greenbaum, & Christiansen, 1995). The subscale of the AEQ-A, which assesses anticipated changes in social behavior, has shown good internal consistency as well as criterion and discriminant validity (Brown, Christiansen, & Goldman, 1987). Its replicable links to alcohol consumption suggest importance and utility. Expectancies for social enhancement contribute dramatically to current drinking and subsequent alcohol problems (Smith et al., 1995) as well as predicting longitudinal drinking patterns (Lundahl, Davis, Adesso, & Lukas, 1997; Shen, Lock-Wellman, & Hill, 2001). Because of these strong links between social expectancies and drinking patterns in adolescents, this study focuses on the Social Behavior scale of the AEQ-A.

Two psychometric questions remain about this important subscale: do the items fit an interval scale, and do any items show systematic biases? Answers to these questions can help refine the scale and improve its links to relevant constructs. Most analyses of this scale rely on parametric statistics, which assume interval measurement. If the items actually do not form an interval scale, they may alter links to other constructs. In addition, biased items may create misleading comparisons between groups. Groups might differ because of idiosyncratic content of items rather than genuine differences in expectancies about alcohol-induced social changes. Disentangling whether differences in groups are based on true score differences or due to differences in the measurement of the underlying construct is crucial. Establishing metric invariance (e.g. that a measure is assessing the same underlying construct across groups, ages, cultures or time) makes comparisons of differences between groups possible, since it establishes that groups respond to the measure in consistent ways.

We used Rasch modeling (Rasch, 1980) to examine the interval level scaling and potential biases in items. The Rasch model is a logistic item-response model that scales items and persons independently on a latent construct. The model estimates each person’s level of expectancies (like an ability or trait) and the strength of each item’s content (like item difficulty), to determine the probability that each person will endorse each item. Person parameters derived from Rasch models are similar to the total scale score used in classical test theory and represent one’s level of social expectancies about alcohol on the underlying trait or the ease of endorsement of positive expectations about alcohol. Item parameters are measured independently along the same theoretical continuum and represent increasingly more positive expectations about the effects of alcohol. These probabilities are used to form a log-odds scale with equal intervals. Items that fail to fit this model do not fit an interval scale.

Rasch models also can be used to assess unidimensionality. This is done via a principle components factor analysis using the standardized residuals from the Rasch model and certain Rasch fit statistics. In a
simulation study, factor analysis of Rasch residuals was found to be more effective in identifying multidimensionality than factor analysis of the original responses (Linacre & Wright, 1998).

Rasch modeling procedures can also be used to assess for differential item functioning (DIF). Items with dramatically different difficulties across groups suggest potential item bias. If separate analyses reveal comparable estimates of item difficulty in two different groups, the items likely function comparably. Nevertheless, significantly different item difficulties across groups suggest that the items may have different meanings and essentially do not tap the same construct (Holland & Wainer, 1993). An advantage of Rasch modeling procedures for differential item functioning compared to comparing mean differences in item responses, for example, is that Rasch models remove the effects of person ability and develop item difficulty ratings for each group separately. Thus, differences in total scale score, or person effects, are removed when comparing item difficulty estimates.

2. Method

2.1. Participants

Participants were 518 undergraduate students (34.4% males, 65.6% females) recruited through various undergraduate courses at a large state university in the Southeast United States. The age range was 17–59 years with a mean age of 21.5 years (SD=5.16). Of those who indicated their ethnicity (66.4% of the sample), 48.6% were Caucasian, 4.1% were African American, 7.7% were Hispanic, 2.1% were Asian, 3.9% reported “other” as their ethnicity, and 33.6% of the sample choose not to indicate their ethnicity. The large proportions of Caucasians and non-response to the ethnicity questions precluded examinations of bias across ethnic groups. All participants were offered extra credit in psychology courses for their participation.

To model expected differences in social expectancies about alcohol consumption, the sample was divided into three groups based on drinking behaviors. The extreme drinker group included males who reported consuming five or more drinks on any single day and females who reported consuming four or more drinks on any single day (Wechsler, Dowdall, Davenport, & Rimm, 1995). We chose to classify drinkers into groups based on their status as binge drinkers. Binge drinking has a generally accepted definition that would not vary with samples the way median or quartile splits would. Binge drinking is the most prevalent substance abuse problem among college students (Syre, Martino-McAllister, & Vanada, 1997). It covaries with many other serious problems, including unsafe sex, accidents, and psychological difficulties (Wechsler, Dowdall, Davenport, & Castillo, 1995; Wechsler et al., 1995). Thus, assessing the utility of the AEQ in binge drinkers has the potential to have dramatic impact.

The social drinkers reported some alcohol consumption on at least 1 day of the 2-week period, but no day with a binge episode. Finally, non-drinkers were individuals who reported no alcohol consumption on any of the recorded days. The current sample included 298 (57.5%) extreme drinkers, 101 (19.5%) social drinkers and 119 (23.0%) non-drinkers. Extreme drinkers significantly differed from social drinkers on several measures. Extreme drinkers drank on more days during the 2-week tracking period, \( M=4.9 \) (SD=2.8) days vs. \( M=2.9 \) (SD=2.6) days; \( t(449.2)=-14.02, p<0.0001 \). They also consumed more total drinks during the tracking period, \( M=25.8 \) (SD=19.2) vs. \( M=6.0 \) drinks (SD=6.0), \( t(397)=-6.38, p<0.0001 \) and had a higher maximum number of drinks on any single day, \( M=8.1 \)
vs. $M=2.5$ (SD=0.9), $t(368.7)=-23.49$, $p<0.0001$. Finally, more males than females were heavy drinkers, and more females were non-drinkers compared to males, $X^2(1)=9.49$, $p=0.009$.

2.2. Measures

2.2.1. Expectancies

The social facilitation subscale of the Alcohol Expectancy Questionnaire–Adolescent Version (AEQ-A) consists of 17 items measuring anticipated alcohol-induced enhancement of social interactions. Items are rated on a dichotomous scale (i.e. true or false) based on what respondents believe generally happens in response to drinking alcohol. The scale showed good internal consistency, Cronbach’s $\alpha=0.75$. Internal consistency could not be improved by dropping any items. Thus, Cronbach’s $\alpha$ offered little help for shortening the scale.

2.2.2. Drinking habits

Students completed a timeline follow-back measure for the previous 2 weeks, documenting the total number of drinks and maximum number of drinks on a single day (Sobell & Sobell, 1992). This approach led to a reasonable estimate of recent drinking without taxing memory.

2.3. Procedure

Participants were administered informed consent forms prior to survey administration. Informed consent forms and questionnaire packets were kept separate to ensure that participants would remain anonymous. Participants were informed that their participation was voluntary and they could discontinue their participation at any time and still receive full credit. After providing consent, participants completed the AEQ-A measure first, followed by the drinking measure. After completing all measures, participants were provided with debriefing information.

2.4. Data analysis

Rasch models were used to assess the fit and reliability of the AEQ-A scale to identify differential item functioning (DIF), and to assess unidimensionality. Unidimensionality was assessed via a principle components factor analysis using the standardized residuals from the Rasch model and certain Rasch fit statistics. Eigenvalues for the residual factors were used to assess unidimensionality. Simulation studies indicate that eigenvalues less than 1.4 are at the random level (Smith & Miao, 1994) and first residual factors that explain less than 3.0 units of residual variance indicate unidimensionality (Linacre, 2004).

Next, we examined differences in item functioning in different groups. Specifically, we were interested in whether excessive drinkers, social drinkers, and non-drinkers responded to items in a different ways as well as if there were differences in response patterns based on gender. We used DIF procedures to identify items that showed bias among groups. Using the set of person estimates derived from a run using the entire dataset as anchor points, separate models were developed for item estimates for each level of drinker type and subsequently for each gender. $T$-tests comparing mean level of item difficulty between groups were used to identify significant levels of item bias (Draba, 1977; Smith, 2004). Additional criteria for assessing significant differences between groups were (1) a difference in
item difficulty estimates of at least 0.5 logits; (2) a standardized effect size (d; Cohen, 1977) of at least 0.2; and (3) theoretical considerations of pattern of expected results. Wright and Douglas (1975) showed that differences in item parameters of less than 0.5 logits had little effect on the accuracy of tests with tests with 20 or more non-DIF items and Smith (2004) has shown that differences smaller than 0.5 are hard to detect even with large sample sizes.

Next, because the large sample size can affect probability statements, a measure of effect size was also selected and the criterion set to be at least a small effect size (0.2; Rosenthal & Rosnow, 1991). Finally, because the separate group calibration process which centers the distribution on zero often leads to an equal numbers of items to be identified with \( t \)-statistics above +2.0 and below −2.0, it is suggested that results can be interpreted in light of theoretical meaning (Smith, 2004).

In the case of the type of drinker analyses, research and theory both suggest that higher levels of drinking behavior (e.g. moving from non-drinkers to social drinkers to extreme drinkers) is predicted by higher levels of positive expectations about alcohol consumption in both concurrent (Christiansen et al., 1989) and longitudinal studies (Shen et al., 2001; Werner, Walker, & Greene, 1995), especially beliefs about increases in social assertiveness (Ohannessian & Hesselbrock, 2004). Thus, it would be expected that item difficulty rating (or the likelihood that an item will be endorsed) will be lower for those in more extreme drinking categories. While items exhibiting these statistical characteristics would be identified as showing DIF, these items were retained as they represent fundamental nature of the construct. That is, we would expect mean differences between drinking groups on these items because of true differences in positive expectancies. Items that do not follow theoretically meaningful patterns were dropped as biased items in the traditional sense. For DIF analyses based on gender, while gender differences do exist in alcohol expectancies (Lundahl et al., 1997), there is no theoretical grounding to keep these items and, therefore, all items showing any form of DIF related to gender were dropped from the final scale. WINSTEPS (Linacre, 2004) was used for all Rasch analyses. SPSS version 9 (SPSS, 1999) was used for all parametric analyses.

![Fig. 1. Mean item difficulty estimates for non-, social, and extreme drinkers for each AEQ-A item.](image-url)
3. Results

3.1. Assessing model fit and unidimensionality

Values for each person’s level of alcohol expectancies (θ’s) were estimated from a single run of the Rasch model for all 518 participants. Infit and outfit mean square values for individual items all fell within the expected range of 1 ± 0.4 (Wright, Linacre, Gustafson, & Martin-Löf, 1994), indicating that the items were functioning reliably according to Rasch model assumptions. The Rasch model (first factor) explained 47.1% of the variance in AEQ-A responses. The first residual factor explained 2.3 units (i.e. eigenvalue) of the 17 units of residual variance. The absolute values of all outfit statistics were less than 1.3, below the critical value of 3 (Wright, 1996). Thus, the scale appears to assess a single, interval scale that fits the assumptions for Rasch models.

3.2. Differential Item Functioning (DIF)

Results of the DIF analyses showed that responses to six of the AEQ-A items differed based on type of drinker. Fig. 1 displays the item parameter estimates for each of the drinking groups for each item and Table 1 shows descriptive and statistical results for DIF analyses based on level of drinking. Four of the items showing DIF followed the pattern expected based on previous research and theory with increasing levels of social expectations about drinking as level of drinking increased. The item difficulty estimates for Item 4 (“teens drink in order to get attention”) were significantly different among the three drinking groups. Heavy drinkers found it harder to endorse item 4 compared to both non-drinkers and social drinkers. Non-drinkers were more likely to endorse item 12 (“most alcohol tastes terrible”) compared to extreme drinkers. However, the estimate for social drinkers fell in between the two other groups and did not significantly differ from either. There were two items where non-drinkers had significantly lower item difficulty estimates compared to extreme drinkers; item 14 (“it’s fun to watch others act silly”) and item 16 (“alcohol makes parties more fun”). In addition, two items exhibited DIF and followed patterns not expected based on theory. Specifically, for item 10 light drinkers were significantly more likely to endorse the item (“alcohol tastes good”) compared to both non-drinkers and extreme drinkers. A similar pattern was identified for the item 15 (“teens feel forced to drink”). Thus, these two questions were dropped from the model and then items were tested for DIF based on gender.

Table 1
Results from t-tests comparing non-drinkers (Group 1), social drinkers (Group 2), and heavy drinkers (Group 3) with differences in parameter estimates (Delta) and standardized estimates of effect size (d)

<table>
<thead>
<tr>
<th>Item</th>
<th>1 vs. 2 delta</th>
<th>d</th>
<th>1 vs. 3 delta</th>
<th>d</th>
<th>2 vs. 3 delta</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>3.97***</td>
<td>0.55</td>
<td>-0.98</td>
<td>0.44</td>
<td>0.94</td>
<td>0.24</td>
</tr>
<tr>
<td>10</td>
<td>2.27*</td>
<td>0.31</td>
<td>0.25</td>
<td>0.19</td>
<td>2.05*</td>
<td>-0.52</td>
</tr>
<tr>
<td>12</td>
<td>0.92</td>
<td>0.13</td>
<td>0.58</td>
<td>0.25</td>
<td>1.10</td>
<td>0.29</td>
</tr>
<tr>
<td>14</td>
<td>0.87</td>
<td>0.12</td>
<td>2.35*</td>
<td>0.67</td>
<td>1.16</td>
<td>0.37</td>
</tr>
<tr>
<td>15</td>
<td>2.91**</td>
<td>0.17</td>
<td>0.38</td>
<td>0.50</td>
<td>3.11**</td>
<td>0.80</td>
</tr>
<tr>
<td>16</td>
<td>1.26</td>
<td>0.40</td>
<td>0.80</td>
<td>0.33</td>
<td>1.36</td>
<td>0.40</td>
</tr>
</tbody>
</table>

Degrees of freedom for 1 vs. 2 tests = 210, for 1 vs. 3 tests = 401, and for 2 vs. 3 = 389.

*p-value < 0.05, **p-value < 0.01, ***p-value < 0.001.
Item difficulty estimates for the remaining 15 items were derived for each gender separately. Four items were found to show significant bias between genders. Males found it easier to endorse two of these questions. Fig. 2 depicts mean item difficulty estimates for each gender for each item and Table 2 shows descriptive and statistical results from the DIF analyses based on gender. Results indicated that females had significantly lower item difficulty estimates compared to males on three questions; item 4 (“teens drink to get attention”), item 6 (“people are more caring”) and item 9 (“sweet alcohol tastes good”). Item difficulty estimates were significantly lower for males on one question; item 2 (“alcohol creates problems”). All four items were dropped from the scale.

3.3. Assessing reliability and convergent validity of 11-item Rasch measure

After dropping the six items that showed significant bias based on drinker type and gender, a shortened 11-item AEQ-A (Rasch AEQ-A) measure was assessed through a final Rasch model. Overall scale reliability was 0.97, and infit and outfit mean squares for all items were within the acceptable range. See Table 3 for final scale items.

Using Rasch person ability estimates for the 11-item scale, several analyses were run to compare the shortened scale to the full 17-item scale. Because the homogeneity of variances assumption was
violated, a Welch’s one-way analysis of variance (ANOVA) was conducted to compare scores on the AEQ-A measures among drinking groups. Results indicated significant differences among the groups based on drinking level, $F(2, 208.23)=53.23$, $p<0.001$. Post-hoc comparisons using Dunnett T3 comparisons indicated that non-drinkers had significantly lower scores on the Rasch AEQ-A compared to social drinkers with a mean difference of -1.49 logits, $p<0.001$. Also, non-drinkers scored significantly lower on the Rasch AEQ-A compared to excessive drinkers, with a mean difference in Rasch AEQ-A scores of -1.78, $p<0.001$. Excessive and social drinkers did not significantly differ, mean difference = -0.29, $p=0.12$. Excessive drinkers had a mean of 1.17 (SD=1.25) on the Rasch measure; the mean for social drinkers was 1.37 (SD=1.24), while the non-drinkers averaged -0.12 (SD=1.71). Similar results were found using the total scale score from the 17-item measure for the overall omnibus test, Welch’s ANOVA, $F(2, 196.89)=63.68$, $p<0.001$. Post hoc comparisons using Dunnett T3 tests found significant differences among all three groups. The mean difference on the 17-item AEQ-A measure between non-drinkers ($M=7.9$, SD=3.88) and extreme drinkers ($M=12.3$, SD=3.43) was -3.36, $p<0.001$. The mean difference between non-drinkers and social drinkers ($M=11.23$, SD=2.66) was -4.4, $p<0.001$. Interestingly, the mean difference (1.04) between social drinkers and extreme drinkers was significant for the original 17-item version of the AEQ-A, $p=0.003$.

Next, analyses of the effects of gender using both the 11-item Rasch AEQ-A measure and the original 17-item measure were compared. Because the homogeneity of variance assumption was violated for the
original 17-item measure, but not for the Rasch AEQ-A measure, a Welch’s ANOVA was used on both measures. Both analyses showed significant differences between genders on social expectations about drinking; \( F(1, 389.35)=14.00, p<0.001 \) for Rasch-measure and \( F(1, 415.62)=11.53, p<0.001 \) for the 17-item version of the AEQ-A. Men scored significantly higher than women on both measures. The average effect size based on gender was 0.34 for the Rasch AEQ-A measure and 0.30 for the original 17-item measure.

Correlations between the three drinking behaviors (maximum number of drinks per day, total drinks during the 2-week assessment period and the number of days drinking during the assessment period) and both the Rasch version of the social facilitation subscale and the full version were calculated. See Table 4 for correlation coefficients. Using methods described by Meng, Rosenthal, and Rubin (1992) to compare correlated correlation coefficient, no differences were found in the strength of the correlation coefficients between the shortened Rasch scale and the total number of drinks, \( Z=0.88, p>0.05 \), and for number of days drinking, \( Z=-0.06, p>0.05 \). However, a significant effect was found for the maximum number of drinks consumed during the 2-week assessment period, \( Z=2.13, p=0.02 \), indicating that the original 17-item measure showed a significantly stronger correlation with the dependent variable compared to the 11-item Rasch measure. However, the difference in correlation coefficients was only 0.03, indicating that as little as 0.09% of variance in total drinks was lost when changing versions of the AEQ-A measure.

4. Discussion

This study used Rasch modeling to investigate the interval scaling and potential item bias for the AEQ-A’s social behavior subscale, an important index of alcohol expectancies that predicts drinking habits and accounts for links between risk and alcohol consumption. These results can help further investigations of alcohol expectancies and support the utility of Rasch modeling. Although the current form of the scale is certainly brief, deleting items can prove helpful in large studies involving the assessment of multiple constructs. Decreasing participant burden undoubtedly leads to better data. In addition, the removal of biased items improves estimates of the true relations among underlying constructs and lends support to establishing invariance of the measure across groups. Confirming the interval level of measurement also can assure researchers that the scale can meet assumptions for parametric statistics.

The items formed an interval scale, but some showed biases related to drinker status or gender. The content of the biased items deserves comment. Heavier drinkers found it harder to endorse an item that suggested that teens drink in an effort to get attention. The dissonance such an item might induce in heavy drinkers may function independently of genuine beliefs about alcohol’s impact on social behavior. Few adolescents long to admit that their behavior is designed to get attention. As previous critiques of the AEQ have emphasized, the best approach to assessment of expectancies probably requires a specific focus on personal experiences of acute effects (Leigh & Stacy, 1993). In contrast, heavy drinkers found it easier to endorse an item that implied that drinking was okay because it allowed a drinker to join with others having fun. Note that this item may confound the social facilitation of alcohol with the judgment of whether or not drinking for this reason is acceptable.

The content of the items that showed gender bias also proved intriguing. Adolescent males found it easier to endorse the idea that alcohol creates problems between people. Perhaps males access
stereotypes about intoxication and aggression when answering this item, whereas females may not. Men also found it easier to endorse the idea that alcoholic drinks taste good, which may be a sexual stereotype. In contrast, women found it easier to endorse the idea that alcohol makes people more caring. “Caring” may not be a stereotypically masculine feeling, particularly among adolescents. This item may reflect more about sex biases in ‘caring’ than in alcohol’s effects. Females also found it easier to endorse ‘sweet alcoholic drinks taste good’, which may say more about sexual stereotypes of appropriate beverages than alcohol-related social behaviors. This item’s bias seems comparable to the bias found in ‘alcoholic drinks taste good’. The unbiased version of the scale may have considerable utility in studies of alcohol expectancies. Despite the potential for sample-independent estimates of difficulty with the Rasch model, generalizing from this sample to treatment samples remains an empirical question. Although the AEQ has a comparable factor structure in college students and alcohol-dependent clients, rates of endorsement of items appear to differ across these groups, potentially leading to different estimates of item difficulty (Demmel & Hagen, 2004). Further work with an alcohol-dependent sample can help confirm the utility of this short form.

While omitting the biased items shortened the scale, meaningful links to measures of alcohol consumption were not sacrificed. Six of 17 items were removed without meaningfully decreasing correlations with measures of drinking habits. The failure to improve the correlations may stem from the deletion of items that were biased in both directions. Deleting an item biased against one group has the potential to make mean comparisons across groups more meaningful. With less bias in the measurement, any mean differences should stem from genuine differences in the construct of interest rather than from potentially biased aspects of the question.

These results suggest that Rasch modeling can improve other measures of alcohol expectancies. This Rasch analysis of the AEQ-A showed that a 35% decrease in the length of the scale appears could be made without sacrificing the strong correlations with drinking variables found in the long form. One can imagine that if Rasch modeling could help decrease the length of other scales measuring other constructs, researchers could shorten their questionnaires by a third and potentially improve participation, save time, and decrease participant burden.

Acknowledgements

Data analysis for this project was funded by NIMH F31-MH069034.

References


