Validation of the 30-Day Version of the Brief Young Adult Alcohol Consequences Questionnaire For Use in Longitudinal Studies

Christopher W. Kahler, Ph.D., **John Hustad, Ph.D.**, and **Nancy P. Barnett, Ph.D.**
Brown University, Center for Alcohol and Addiction Studies Box G-S121, Providence, RI 02912

David R. Strong, Ph.D. and
Butler Hospital and the Warren Alpert Medical School of Brown University 345 Blackstone Blvd., Providence, RI 02906

Brian Borsari, Ph.D.
Brown University, Center for Alcohol and Addiction Studies Providence Veterans Affairs Medical Center, 830 Chalkstone Ave., Providence RI 02908

Abstract

Objective—The Brief Young Adult Alcohol Consequences Questionnaire (B-YAACQ; Kahler et al., 2005) was developed using item response modeling to provide a brief and readily interpretable measure of negative alcohol consequences over the past year among college students. The purpose of the present study was to extend evaluation of the B-YAACQ by examining its psychometric properties when administered to college students cited for a university alcohol violation using a past 30-day timeframe of assessment.

Method—The B-YAACQ was administered at baseline and at a 6-week follow-up to 291 students cited for a university alcohol violation. Reliability and validity analyses, in addition to Rasch model (Rasch, 1960) analyses, were conducted using these data.

Results—Results demonstrated that the B-YAACQ was internally consistent, showed strong unidimensionality and additive properties, showed minimal item redundancy and minimal floor or ceiling effects, was reliable over a 6-week time period, and was sensitive to change in drinking following an alcohol intervention. In addition, the relative severity of items was preserved over time and generally consistent with results from Kahler et al. (2005).

Conclusions—The 30-day B-YAACQ appears valid for use with college students who have received an alcohol violation and for evaluating changes in alcohol consequences.

Introduction

Alcohol-related problems are common on college campuses (Wechsler et al., 2000), and instruments are needed to assess the full breadth of these problems and to evaluate the effects of interventions to reduce alcohol problems in college students. The two scales most commonly used for this purpose, the Young Adult Alcohol Problems Screening Test (Hurlbut and Sher, 1992) and the Rutgers Alcohol Problem Index (White and Labouvie, 1989), contain a large proportion of items involving alcohol problems that are rare and severe, providing limited information about common, less severe alcohol consequences in this population (Kahler et al., 2004; Neal et al., 2006). In response to these limitations, the Young Adult Alcohol Consequences Questionnaire (YAACQ; Read et al., 2006) was developed to assess a broader

range of alcohol consequences in college students. A brief version of this measure (the B-YAACQ; Kahler et al., 2005) was then created using Rasch model (Rasch, 1960) analyses to select those YAACQ items that (a) showed strong unidimensionality and covered a range of the problem severity continuum, (b) had minimal redundancy with other items, (c) showed high discrimination between levels of alcohol problem severity, and (d) showed no evidence of differential functioning by gender.

Although the psychometric properties of the B-YAACQ appeared strong, no test-retest reliability data for the measure were available, and the version tested used a past 12-month timeframe of assessment (Kahler et al., 2005). A measure with a 30-day assessment window may be more useful for intervention studies in which investigators want to track immediate changes in alcohol consequences. The initial validation of the B-YAACQ also was conducted with volunteer psychology students. Therefore, the utility of the B-YAACQ as a dependent measure of alcohol problems in college students receiving an alcohol intervention is unknown.

In the present study, we administered a 30-day version of the B-YAACQ to college students who had been cited for an alcohol violation and mandated by the university to attend an alcohol intervention. We examined the stability of the B-YAACQ over a 6-week period and its correlations with alcohol consumption and the Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993), a screening measure for general adult populations that is intended primarily for clinical purposes. We also examined whether changes in alcohol consumption over a 6-week period were associated with concurrent changes in B-YAACQ scores. Finally, we examined (a) whether items were unidimensional and fit the Rasch model, (b) the stability of the severity and rank ordering of items over 6 weeks, and (c) the degree to which severity estimates and ordering of items corresponded with results from Kahler et al. (2005). The relative severity and ordering of items is important for interpreting total scores, as it allows one to know which alcohol problems are likely to be endorsed with a given B-YAACQ score.

Method

Participants attended a suburban four-year, private liberal arts university in the Northeast and were enrolled in a clinical trial examining stepped care (see Borsari et al., 2007). The school has an enrollment of 3,300 undergraduates (15% minority, 51% female, and 79% of students live on campus). Students were first-time alcohol offenders referred to the university's Alcohol Incident Referral Program and were invited to participate when they presented for their initial session. Out of 369 eligible students, 291 students (79%) enrolled (the rest received treatment as usual). Participants were 65% male, 96% Caucasian, and 66% freshman (mean age 19.0 years). Participants were paid for their baseline ($15) and 6-week ($45) assessments.

Participants completed a paper-and-pencil baseline assessment including a demographics questionnaire and measures of alcohol consumption in the past 30 days. They also provided information required to calculate peak blood alcohol concentration (pBAC) on their heaviest drinking day (Matthews and Miller, 1979). Participants completed the AUDIT and the 30-day B-YAACQ. The B-YAACQ assessed 24 consequences of alcohol consumption in the past 30 days using a dichotomous (no/yes) response format. In this text, we refer to items by the respective number shown in Table 1 of Kahler et al. (2005), in which higher numbered items were relatively more severe. The 6-week follow-up determined if the student was to receive the next step of care, a brief motivational intervention. It was identical to the baseline except that it was conducted via a web-based survey. Students were sent an email invitation to report on the previous 30 days of use. Of 291 students who completed the baseline assessment, 283 (97%) completed the 6-week assessment. Drop-outs were significantly older than completers.

Analysis Plan

We examined the internal consistency of the B-YAACQ at baseline and 6 weeks and the correlation between these assessments. We correlated B-YAACQ score with measures of alcohol use and the AUDIT and used regression analysis to examine change in B-YAACQ scores over the 6-week follow-up. We then conducted Rasch model analyses using BIGSTEPS (Linacre and Wright, 1998). We have described the Rasch model in depth in other papers (Kahler and Strong, 2006; Kahler et al., 2005; Kahler et al., 2004). Briefly, the Rasch model is a logistic item response model that independently scales both items and persons along a theorized underlying latent continuum. The severity of each item is defined as the degree of alcohol problem severity that is required before that item has a 50% probability of being endorsed. Rasch severity estimates mirror the frequency of endorsement of each item but are expressed in equal-interval log odds units and scaled so that the mean item severity is 0. The Rasch model assumes that less severe items have a lower probability of endorsement than more severe items across the full range of the continuum. With fit to the Rasch model, a scale can be considered truly additive, so that endorsing one item reflects a relatively equal increase in severity regardless of which other items are endorsed. Infit and outfit statistics for each item determine Rasch model fit (Wright and Masters, 1982). Because infit statistics are weighted locally, they are less susceptible to outlier influences and generally preferred (Bond and Fox, 2001). When data fit a Rasch model well, fit statistics will fall within an acceptable range of 0.6−1.4 (Linacre and Wright, 1994), and correlations among item residuals after fitting the model will be minimal.

We examined the extent to which the Rasch severity estimates for each item remained stable from baseline to 6 weeks. We also compared the severity estimates of items in this sample at baseline with results of the Kahler et al. (2005) study. We first examined the correlations between item severity estimates and examined the rank ordering of items, using Spearman rank-order correlations. As an a priori rule of thumb, we considered items that moved in rank more than 6 places (e.g., going from the 4th most severe item to the 11th most severe item) as being variable across assessments or timeframes; such items are moving in their placement at least one-quarter of the range of possible total scores (i.e., 0−24). We also tested whether there was significant drift in the severity estimates of the items across the two assessments and in comparison with the Kahler et al. (2005) study using a test for differential item functioning (DIF; Holland and Wainer, 1993). If items behave similarly over time or across studies, severity parameters estimated independently at different times or in different samples will fall within an acceptable range of agreement (95% confidence interval). These analyses do not, however assess the impact of DIF on total score interpretation. For that purpose, we first obtained model-based severity estimates for each person when items were freely estimated. We then used correlational analyses to compare these estimates to those obtained when the parameters for items that showed DIF were fixed at the values obtained at the other time point or in the other sample.

Results

Internal consistency of the B-YAACQ was high at baseline (alpha = .84) and 6 weeks (alpha = .89), with no items detracting from alpha. Mean B-YAACQ scores were 7.2 (SD = 4.7; range = 0−21) at baseline and 6.6 (SD = 5.4; range = 0−22) at 6 weeks. Only 15 participants had a score of 0 at baseline; 24 had a score of 0 at follow-up. Repeated measures ANOVA indicated that B-YAACQ scores significantly decreased from baseline to 6 weeks, $p < .05$. The correlation between baseline and 6 weeks was $r(283) = .70, p < .0001$, indicating high stability of scores. The most commonly endorsed item was item 2, (had a hangover; 77.3% at baseline, 68.6% at 6 weeks) followed by item 1 (did or said something embarrassing; 64.0% and 44.5%,

respectively). The least commonly endorsed item was item 24 (needed a drink after waking; 2.4% and 5.3%) followed by item 22 (overweight because of drinking; 8.7% and 11.3%).

There were significant medium to large correlations of the baseline and 6-week B-YAACQ with concurrent drinking variables and large correlations with concurrent AUDIT scores (Table 1). To test whether change in B-YAACQ scores was sensitive to change in drinking, we regressed change in B-YAACQ (6-week value minus baseline) on baseline B-YAACQ value, baseline value of the drinking variable examined, and change in that drinking variable. The sr^2s for each respective analysis for changes in (a) drinks per week, (b) heavy drinking frequency, and (c) peak BAC were .11, .11, and .07, respectively, ps < .0001.

Rasch Model Analyses

Rasch models were fit with the baseline data and the 6-week data, respectively. Item responses fit a Rasch model well with infit values within the desired range of $0.6 - 1.4$. At baseline only, the more outlier-sensitive outfit statistic fell outside of this range for items 2 (had a hangover) and 3 (sick to stomach/vomited after drinking), outfits = 1.41 and 1.58, respectively. Correlations between item residuals obtained after fitting the data to a Rasch model were all less than .20, indicating minimal local dependence among items; there was one exception at 6 weeks where the residual correlation between items 22 (overweight because of drinking) and 23 (physical appearance harmed by drinking) was .27.

Item severity estimates from baseline were highly correlated with those obtained from 6-week data, $r(24) = .95$. Likewise, the rank order of each item across assessments was very well preserved, with a rank-order correlation coefficient of .96. To determine the degree to which the relative severity of each item was consistent over time, we plotted the severity estimates obtained at baseline against those obtained at 6 weeks. We then constructed 95% confidence interval bands around these estimates using their joint standard errors. Results are shown in the left half of Figure 1. Four items (1, 3, 12, 17) fell outside of the confidence intervals, suggesting significant DIF across assessments. To determine whether DIF in these 4 items affected the meaning of total scores, we re-ran the 6-week analyses while fixing the severity estimates of these 4 items at their baseline values. Person severity estimates derived from the freely estimated Rasch model were correlated .99 with those obtained from the model in which the severity estimates for the 4 items were fixed at their baseline values. The estimated severity of the sample when item parameters were freely estimated was -1.57 ($SD = 1.70$) vs. -1.60 ($SD = 1.71$) when the 4 items were fixed. Thus, drift in item severity estimates did not impact score interpretation.

Comparison with the 12-Month Version from Kahler et al. (2005)

We compared the item severity estimates at baseline to the estimates obtained by Kahler et al. (2005) with a 12-month assessment timeframe. Severity estimates and standard errors were obtained from Table 1 in Kahler et al. (2005). The correlation between item severity estimates in the different samples was very high, $r(24) = .91$. The rank-order of the symptoms was generally maintained as well, $r_s = .88$. Two items moved substantially in their relative order: item 17 (less energy due to drinking: ranked 8th most severe in Kahler et al. (2005) vs. 19th most severe in the present sample) and item 20 (neglected obligations to family: 5th most severe in Kahler et al. vs. 12th in the present sample). When these two items were not considered, the rank-order correlation between items in the two samples became very high, $r_s = .96$. Overall, all but one item (item 17) that was identified by Kahler et al. (2005) as being in the lower half of severity remained in lower half of severity in the present sample.

The right half of Figure 1 shows the comparison of severity estimates across the samples. Ten items (1, 3, 4, 5, 8, 10, 17, 20, 21, 23) fell outside of the 95% confidence interval, indicating

significant differences in severity estimates across these two studies. We re-ran the baseline analyses while fixing the severity estimates of these 10 items at their values obtained in Kahler at al. (2005). Person severity estimates derived from the freely estimated Rasch model were correlated .99 with those obtained from the model in which the severity estimates for the 10 items were anchored using Kahler et al. (2005). The estimated severity of the sample at baseline was −1.36 (SD = 1.48) when all items were freely estimated vs. −1.46 (SD = 1.65) when the 10 items were anchored. Thus, drift in item severity estimates across the samples did not affect the estimation of overall alcohol problem severity.

Discussion

Results demonstrate that the 30-day version of the B-YAACQ, whether administered by paper-and-pencil or by web-based survey, is internally consistent, shows strong unidimensionality and additive properties, shows minimal item redundancy, is reliable over a 6-week time period, is sensitive to change in drinking, and is valid for use in mandated students. The scale covers a wide range of problem severity; no participants reached the maximum score and a very small proportion endorsed no items. The relative severity and ordering of items also was quite stable over time. There was significant DIF for 10 out of 24 items when comparing the Rasch model item severity estimates obtained in the present sample to those obtained by Kahler et al. (2005). Two items became considerably less severe with a 30-day assessment timeframe: less energy due to drinking and neglected obligations to family, work, or school. Most important for pragmatic purposes, however, was that the rank order of the vast majority of items was similar across the studies, and estimated person severity did not change meaningfully due to item drift. Overall, scores obtained using the 30-day B-YAACQ and using the 12-month B-YAACQ appear quite comparable in terms of the items that contribute to specific total scores.

As discussed in depth by Kahler et al. (2005), the simple additive properties of the B-YAACQ and the ready translation of B-YAACQ scores to expected consequences endorsed makes the instrument particularly easy to score and interpret. We believe the ease of interpretation of the B-YAACQ, coupled with its extensive psychometric validation, make it an ideal instrument for evaluating changes in alcohol consequences over time among college students. It may prove particularly useful in tracking low to moderate severity alcohol problems in longitudinal investigations because it contains an ample number of items that are endorsed in the absence of relatively severe alcohol problems.

References

Bond, TG.; Fox, CM. Applying the Rasch Model: Fundamental measurement in the human sciences. Erlbaum; Mahway, NJ: 2001.

Borsari B, O'Leary Tevyaw T, Barnett NP, Kahler CW, Monti PM. Stepped care for mandated college students: a pilot study. Am J Addict 2007;16:131–7. [PubMed: 17453615]

Holland, PW.; Wainer, H. Differential Item Functioning. Lawrence Erlbaum; Hillsdale, NJ: 1993.

Hurlbut SC, Sher KJ. Assessing alcohol problems in college students. J Am Coll Health 1992;41:49–58. [PubMed: 1460173]

Kahler CW, Strong DR. A Rasch model analysis of DSM-IV Alcohol abuse and dependence items in the National Epidemiological Survey on Alcohol and Related Conditions. Alcohol Clin Exp Res 2006;30:1165–75. [PubMed: 16792564]

Kahler CW, Strong DR, Read JP. Toward efficient and comprehensive measurement of the alcohol problems continuum in college students: the brief young adult alcohol consequences questionnaire. Alcohol Clin Exp Res 2005;29:1180–9. [PubMed: 16046873]

Kahler CW, Strong DR, Read JP, Palfai TP, Wood MD. Mapping the continuum of alcohol problems in college students: a Rasch model analysis. Psychol Addict Behav 2004;18:322–33. [PubMed: 15631604]

Linacre JM, Wright BD. Reasonable mean-square fit values. Rasch Measurement Transactions 1994;8:370.

Linacre, JM.; Wright, BD. A User's Guide to BIGSTEPS: A Rasch-Model Computer Program. MESA Press; Chicago, ILL: 1998.

Matthews DB, Miller WR. Estimating blood alcohol concentration: two computer programs and their applications in therapy and research. Addict Behav 1979;4:55–60. [PubMed: 420046]

Neal DJ, Corbin WR, Fromme K. Measurement of alcohol-related consequences among high school and college students: application of item response models to the Rutgers Alcohol Problem Index. Psychol Assess 2006;18:402–14. [PubMed: 17154761]

Rasch, G. Probabilistic models for some intelligence and attainment test. Denmarks Paedagogiske Institut; Copenhagen: 1960.

Read JP, Kahler CW, Strong DR, Colder CR. Development and preliminary validation of the young adult alcohol consequences questionnaire. J Stud Alcohol 2006;67:169–77. [PubMed: 16536141]

Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption. II. Addiction 1993;88:791–804. [PubMed: 8329970]

Wechsler H, Lee JE, Kuo M, Lee H. College binge drinking in the 1990s: a continuing problem. Results of the Harvard School of Public Health 1999 College Alcohol Study. J Am Coll Health 2000;48:199–210. [PubMed: 10778020]

White HR, Labouvie EW. Towards the assessment of adolescent problem drinking. J Stud Alcohol 1989;50:30–7. [PubMed: 2927120]

Wright, BD.; Masters, GN. Rating scale analysis. MESA; Chicago: 1982.
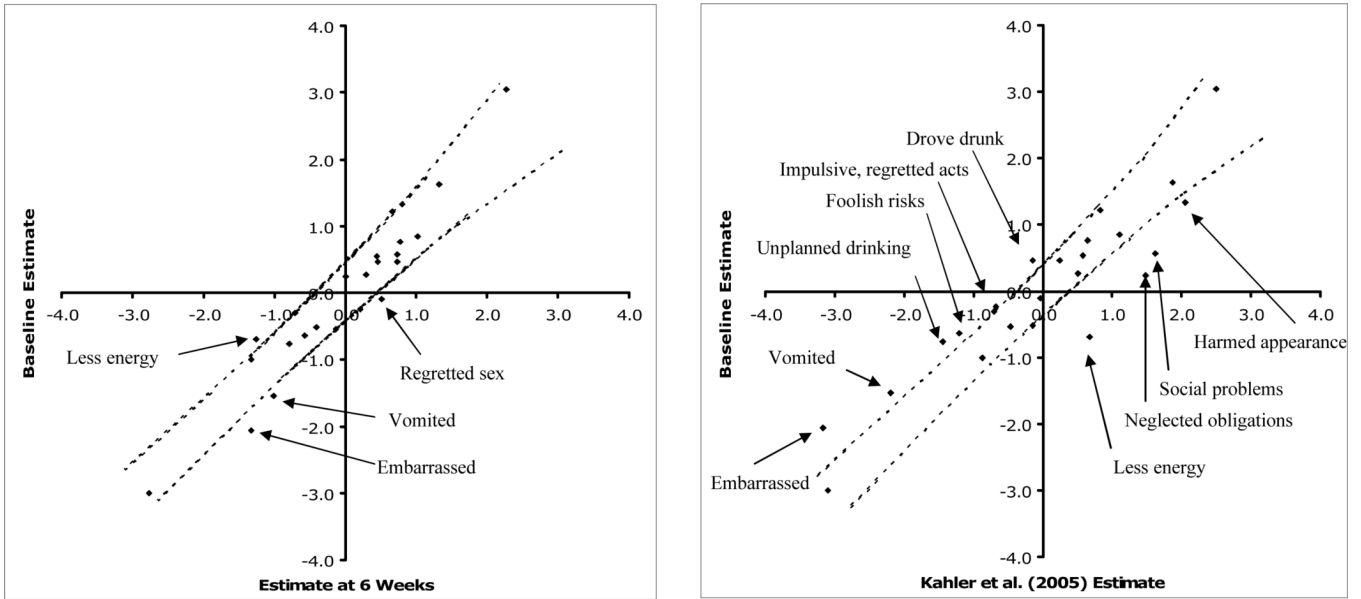
Figure 1.
Plots of item severity estimates and 95% confidence intervals (CI) from differential item functioning (DIF) analyses comparing estimates obtained at baseline to those obtained at 6-weeks and those obtained by Kahler et al. (2005). Items that fall below the lower 95% CI reflect that these items were relatively less severe at baseline relative to their respective comparison estimates; items that fall above the upper 95% CI reflect that the items were relatively more severe at baseline relative to their respective comparison estimates. Items with significant DIF are labeled according to the item content.

Table 1

Correlations Among B-YAACQ Scores, Measures of Past-Month Alcohol Consumption, and AUDIT Scores At Baseline and 6-Week Follow-up

Measure	1	2	3	4	5	6	7	8	9	10
1. Baseline B-YAACQ	---									
2. Baseline drinks per typical week	**.41**	---								
3. Baseline heavy drinking frequency	**.52**	**.65**	---							
4. Baseline peak BAC	**.38**	**.36**	**.44**	---						
5. Baseline AUDIT	**.64**	**.53**	**.59**	**.43**	---					
6. 6-week B-YAACQ	.70	.33	.44	.27	.55	---				
7. 6-week drinks per typical week	.40	.52	.51	.32	.42	**.51**	---			
8. 6-week heavy drinking frequency*	.36	.44	.53	.18	.39	**.49**	**.60**	---		
9. 6-week peak BAC	.26	.28	.39	.52	.32	**.35**	**.53**	**.49**	---	
10. 6-week AUDIT	.67	.45	.55	.35	.71	**.73**	**.56**	**.57**	**.42**	---

Note. All correlations significant at $p < .01$. Correlations in bold represent concurrent correlations. *ns* range from 279 to 291 due to missing data.

*
defined as 5+ drinks per occasion for men; 4+ drinks per occasion for women.