

**THE RELIABILITY OF ENVIRONMENTAL MEASURES
OF THE COLLEGE ALCOHOL ENVIRONMENT***

JOHN D. CLAPP, PH.D.

MIKE WHITNEY, MSW

AUDREY M. SHILLINGTON, PH.D.

*San Diego State University, California***ABSTRACT**

Much of what we know about students' drinking patterns and problems related to alcohol use is based on survey research. Although local and national survey data are important to alcohol-prevention projects, they do not sufficiently capture the complexity of the alcohol environment. Environmental prevention approaches to alcohol-related problems have been shown to be effective in community settings and researchers have begun to study and adapt such approaches for use on college campuses. Many environmental approaches require systematic scanning of the campus alcohol environment. This study assessed the inter-rater reliability of two environmental scanning tools (a newspaper content analysis form and a bulletin analysis form) designed to identify alcohol-related advertisements targeting college students. Inter-rater reliability for these forms varied depending across different rating categories and ranged from poor to excellent. Suggestions for future research are addressed.

INTRODUCTION

Systematic studies of college students' alcohol consumption and its attendant consequences began nearly 50 years ago with the seminal survey work of Straus

*This study was funded in part by grants from the National Institute of Alcoholism and Alcohol Abuse (RO1AA12540) and the U.S. Department of Education (S184H990014).

through surveys. Recent survey studies [9, 14-17], for instance, have examined environmental correlates of heavy episodic drinking and alcohol problems [14]. Wechsler et al. found drinking beer, living in a fraternity or sorority, reporting easy access to alcohol, and having access to alcoholic beverages priced below \$1.00 were all risk factors for heavy episodic drinking [2]. Clapp and Shillington found drinking with friends, drinking beer and hard liquor, and having many people intoxicated at an event were predictive of heavy episodic drinking events, while dating events were protective of heavy episodic drinking [16]. Finally, Clapp et al. reported that the presence of illicit drugs and playing drinking games were risk factors for alcohol-related problems [14]. In contrast, having food available, drinking with friends and drinking with roommates were protective of alcohol problems [16].

More direct measures of the alcohol environment include social indicators studies examining incident reports, campus arrest data and on campus advertising of alcoholic beverages [7]. To date, little research has been published examining such direct measures of the alcohol environment. Gebhardt, Kaphingist, and DeJong published a recent report on the use of a university and community coalition to reduce alcohol-related problems [18]. This case study collected social indicator data to monitor outcomes from the academic years of 1991/92-1998/99. Data were collected from three main areas: off-campus hotline calls; off-campus noise ordinance reports; and alcohol-related arrests. Results showed significant decreases in the off-campus reports of noise and hotline calls related to off-campus student drinking [18].

QA
Kaphingist
or
Kaphingist

Measuring Environmental Indicators

In 1994, Ryan and associates developed the College Alcohol Risk Assessment guide (CARA) [11]. The CARA was intended to help campuses identify environmental risk factors associated with alcohol-related problems. The CARA provides several data collection forms to facilitate this process. Two important areas of analysis include: 1) scanning of campus bulletin boards for party notices, posters, or flyers advertising or promoting alcohol related activities, advertisements for bars in the area, drink specials, and the like; and 2) content analyses of campus newspapers for messages focusing on alcohol consumption, alcohol related activities, and alcohol advertising [11]. The authors suggested these activities can help identify specific problems on campus, discover high-risk drinking environments, identify and enlist new allies to prevention efforts, and stimulate consideration for environmental factors that contribute to alcohol related problems on campus [11]. Although the CARA forms have been used widely, to date no one has tested their reliability.

This study assessed the inter-rater reliability of modified versions of the bulletin board scanning form and newspaper content analysis form found in the CARA. As more campuses embrace environmental approaches, studies such as these will

newspaper. These forms are slightly modified versions of the forms published in the CARA [11]. Sampling methods and data collection procedures are described below separately for each media source. Both data collection procedures are consistent with Holsti's definition of content analysis [20].

Newspaper

Sampling

The newspaper used for this study publishes four days a week (Monday through Thursday) during the academic year. A random sample ($n = 40$ issues) of the newspaper was selected for analysis. The issues were selected from the period of 2/7/00 to 12/7/00, corresponding with the Spring 2000 and Fall 2000 semesters. During the Spring 2000 semester, 57 issues of the newspaper were published. Sampling for the Spring 2000 semester involved randomly selecting one issue for each week. For the Fall 2000 semester there were a total of 57 issues published. In order to increase the sample size, two issues were randomly selected from each of the 14 weeks when four issues were published. The additional 28 issues selected during the Fall 2000 semester made a total of 43 issues. Three issues were omitted from the sample because they could not be found in the campus library, making a final sample size of 40 issues.

Procedures for the Newspaper Analysis

For the purposes of this study, three categories were analyzed: total number of ads; number of alcohol and other drug (AOD) ads; and number of AOD articles. As noted above, two coders participated in this analysis. Each was trained based on a set of rules which were employed in scanning the newspaper for AOD content. First, all articles in the issue were read and articles with AOD content were counted. Next, the issue was scanned for AOD specific ads. These included any alcohol promotional ads for bars, restaurants, stores, social activities, as well as any promotion of tobacco or any other drug. Also included were any AOD prevention ads. These were all tallied. The final step involved counting all the ads in the issue, including non-AOD ads. The observations were recorded on the standard form.

Campus Bulletin Boards

Sampling

The BBs assessed in this study can be used by anyone to advertise or post any type of message (see Figure 1 for an example of the boards). BBs are located in various areas around campus and six locations in high pedestrian traffic areas

in the upper right-hand quadrant of the BB were counted. This total was multiplied by 4 to gain an estimate of the overall number of messages on a board. This procedure was used given the high volume of messages on each board.

After counting, each coder scanned the entire BB for AOD promotion messages and counted these. The names of the bars that advertised were also recorded. Last, the coders scanned the entire board for AOD prevention messages and counted these. Scanning always proceeded from left to right. Coders scanned the BBs an average of 40 minutes apart. Observations were recorded on a standard form.

Data Analysis

The data sets produced from each type of media were subjected to two statistical tests of inter-rater reliability: the Kappa (κ) statistic and Pearson's Product Moment Correlation Coefficient (r). Berk addressed the use of the Kappa statistic for reliability studies and indicated it is more accurate than simply comparing agreement percentages, as Kappa adjusts for chance agreement [23]. Kappa is limited to assessment of inter-rater reliability for categorical data.

To compute Kappa scores, each rater's counts for all the categories of interest were compared by date and data source. Cases where raters had identical ratings were coded as 1 (agreement), while cases where raters had different ratings were coded as 0 (disagreement). For instance, if both raters counted 120 advertisements on BB one on 12/10/00, a score of 1 was recorded for that BB, on that category on that day. In addition, we computed adjusted Kappa scores. For the adjusted scores, we coded raters as being in agreement if their counts were within ± 2 . Using the same example as above, in our adjusted coding scheme, rater 1 might have counted 120 ads on BB 1 on 12/10/00, while rater 2 counted 118. This situation was also coded as 1 (adjusted agreement). We calculated the adjusted scores to account for the complexity of counting ads on several BBs (see Figure 1). Thus, κ is reported for overall agreement between raters on counts for each category of interest. Kirk and Kutchins noted that Kappa is difficult to interpret [24]. For the purposes of this study, we set acceptable levels for Kappa at .50 to .70. Good levels were considered to be .71 or higher.

Berk suggested the need for a methodology capable of treating quantitative as well as categorical data with equal precision [23]. This can be achieved through the use of the Pearson r correlation between sets of data.

The Pearson r statistics reported here reflect the overall correlations between raters' 1 and 2 observations (estimates of the number overall ads, the number of AOD ads, and the number of prevention messages) for each study. Pedhazur and Schmelkin suggested good levels of r to be .70 or higher in reliability studies [25]. The Pearson r was considered good at .70 or above and excellent at .90 or above. All statistics were computed using SPSS version 10.

Table 2. Inter-Rater Agreement: Reported Agreement on the Total Number of Messages, by Bulletin Board Location ($N = 23$)

Location	Agreement (raw)		Agreement (adjusted)		Kappa (raw)	Kappa (adjusted) ^a
	(n)	(%)	(n)	(%)		
BB #1	1	04.3	1	04.3	-.913	-.913
BB #2	3	13	4	17.4	-.739	-.652
BB #3	3	13	3	13	-.739	-.739
BB #4	2	08.7	2	08.7	-.826	-.826
BB #5	1	04.3	2	08.7	-.913	-.826
BB #6	13	56.5	18	78.3	.130	.565

^aAdjusted Kappa based on perfect agreement (± 2) for each location.

Table 3. Inter-Rater Agreement: Reported Agreement on the Total Number of AOD Prevention Messages, by Bulletin Board Location ($N = 23$)

Location	Agreement (raw)		Agreement (adjusted)		Kappa (raw)	Kappa (adjusted) ^a
	(n)	(%)	(n)	(%)		
BB #1	19	82.6	23	100	.652	1.00
BB #2	21	91.3	23	100	.826	1.00
BB #3	18	78.3	23	100	.565	1.00
BB #4	23	100	23	100	1.00	1.00
BB #5	18	78.3	23	100	.565	1.00
BB #6	22	95.7	22	95.7	.913	.913

^aAdjusted Kappa based on perfect agreement (± 2) for each location.

Table 5. Inter-Rater Agreement: Reported Number of Days Out of Total ($n = 23$) When Both Raters Identified the Same Bar Advertisements, by Bulletin Board Location ($N = 23$)

Location	Agreement (raw)		Agreement (adjusted)		Kappa	Kappa
	(n)	(%)	(n)	(%)	(raw)	(adjusted) ^a
BB #1	15	65.2	22	95.7	.304	.913
BB #2	22	95.7	23	100	.913	1.00
BB #3	21	91.3	23	100	.826	1.00
BB #4	17	73.9	23	100	.478	1.00
BB #5	19	82.6	21	91.3	.652	.826
BB #6	21	91.3	22	95.7	.826	.913
Total	115	83.3	134	97.1	.667	.942

^aAdjusted Kappa based on perfect agreement (± 1) for each location.

Kappa score (.478), but a perfect adjusted Kappa score (1.00). Location 5 had an acceptable raw Kappa score (.652) and a good adjusted Kappa score (.826)

Daily Newspaper

There were 40 issues of the campus newspaper rated and for each issue the total ads, AOD ads, and AOD articles were counted. Based on an average of both raters' assessments, the mean average count for total ads was 35.9, ($sd = 18.6$). The mean count for AOD ads was 4.1, ($sd = 3.8$). Finally, the mean count for AOD articles was 1.30 ($sd = 1.8$).

Results are reported with raw and adjusted Kappa scores. Raw agreement is the number and percent of issues where the raters had exactly the same count, for each category. Similar to the BB data, we adjusted agreement scores by ± 2 in cases where raters were off in their count by two or fewer ads or articles.

Table 6 reveals low raw Kappa score's for all categories with a drastic increase in Kappa score for adjusted agreement. Specifically, the total number of ads had the greatest increase, going from a raw Kappa score of $-.35$, to an adjusted Kappa score of $.85$. The AOD ads went from a raw Kappa score of $.40$ to an adjusted Kappa score of $.90$. The AOD articles increased from a raw Kappa score of $.20$ to an adjusted Kappa score of $.95$. The correlation between ratings for total ads and

chance, while Pearson's Product Moment Correlation examines the covariation between two variables (in the case of inter-rater reliability between raters' counts). Thus, raters tended not to agree perfectly, but tended to have similar ratings that across observations.

The results for the newspaper scanning tool suggest that there might have been a systematic coding error. Although minor, it consistently kept the raters from agreeing exactly. The raw Kappa scores were very low. However, when a minor adjustment was made, the Kappa scores dramatically increased. To further examine this issue, the raters conducted a post-hoc content analysis of the newspaper and found that they made two consistent coding errors. First, there was an ad placed regularly by the campus paper that had drinking as one of the things to do at college in a list of things to do. Second, there was an ad that regularly appeared near the classified section of the paper that advertised a defense attorney specializing in drunk driving. One rater consistently missed both these ads when counting total and AOD ads.

It was also noted that raters had some difficulty deciding if some articles were AOD related. Many articles mentioned AOD issues, but were not primarily focused on AOD issues. It was difficult to decide which of these articles to include as AOD articles in the count. These systematic errors help explain why exact agreement between raters was poor, while adjusted agreement was very high.

The Pearson r correlation for the campus paper was good for all three categories, suggesting overall agreement was extremely good. These findings indicate the newspaper scanning tool has strong inter-rater reliability.

IMPLICATIONS FOR FUTURE RESEARCH

Of primary concern is the need for the tools to be subjected to tests of validity. Of particular concern is how data collected by the tools evaluated here are to be used. If such data were to be used as an indicator of alcohol promotion on campus, for instance, the reliability of scanning tools might preclude valid estimates of this construct. At best, if used for the purposes of estimating alcohol promotion on campus, the tools assessed here would yield gross estimates. That is, these tools do not appear to have great sensitivity. Conceptually, however, such data do reflect domains presented in the model by Clapp and associates [7].

In contrast, if environmental data such as those collected here are used to identify specific bars that advertise high-risk promotions, the tools evaluated here might have sufficient specificity and reliability to be useful. These tools have been used in applied prevention settings to identify high-risk price promotions and establishments [26]. Future studies might examine the construct validity of such measures and further examine how the data are to be used within a conceptual framework. In addition, the reliability of these instruments should be tested further in other settings.

11. B. Ryan, T. Colthurst, and L. Segars, *College Alcohol Risk Assessment Guide*, University of California San Diego Press, San Diego, 1994.
12. H. Holder and J. Blose, Reduction of Community Alcohol Problems: Computer Simulation Experiments in Three Countries, *Journal of Studies on Alcohol*, 48, pp. 124-135, 1986.
13. R. Saltz, Research in Environmental and Community Strategies for the Prevention of Alcohol Problems, *Contemporary Drug Problems*, 15:1, pp. 67-81, 1988.
14. J. D. Clapp, A. M. Shillington, and L. Segars, Deconstructing Contexts of Heavy Episodic Drinking among College Students, *The American Journal of Drug and Alcohol Abuse*, 26, pp. 139-154, 2000.
15. J. D. Clapp and A. L. McDonnell, The Relationship of Alcohol Promotion and Peer Drinking Norms to Alcohol Problems reported by College Students, *Journal of College Student Development*, 41:1, pp. 19-26, 2000.
16. J. D. Clapp, A. M. Shillington, and L. Heidt, Student Perceptions of Alcohol Promotion on Campus, *Journal of Health Education*, 31:5, pp. 252-256, 2000.
17. H. Wechsler, J. Kuo, J. Lee, and J. Dowdall, _____, 2000.
18. T. Gebhardt, K. Kaphingst, and W. DeJong, A Campus-Community Coalition to Control Alcohol-Related Problems Off Campus: An Environmental Management Case Study, *College Health*, 48, pp. 211-215, 2000.
19. J. C. Kronick and S. Silver, Using the Computer for Content Analysis, *Quantitative Methods in Social Work*, 10, pp. 41-62, 1992.
20. O. R. Holsti, *Content Analysis for the Social Sciences and Humanities*, (Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1969.)
21. J. W. Anastas, *Research Design for Social Work and the Human Services*, Columbia University Press, New York, 1999.
22. W. Reid and C. Bailey-Dempsey, Content Analysis in Design and Development, *Research on Social Work Practice*, 4, pp. 101-114, 1994.
23. R. Berk, Generalizability of Behavioral Observations: A Clarification of Interobserver Agreement and Interobserver Reliability, *American Journal of Mental Deficiency*, 18, pp. 460-472, 1979.
24. J. Kirk and J. Kutchins, _____, 1992.
25. E. Pedhazur and L. Schmelkin, *Measurement, Design, and Analysis: An Integrated Approach*, Lawrence/Erlbaum Associates, Hillsdale, New Jersey, 1991.
26. J. D. Clapp and M. Novak, *Community Specific Case Studies*, a panel presentation at Alcohol Beverage Control: Partnerships for Prevention, an invitational symposium sponsored by the U.S. Department of Education, San Diego, California, pp. 7-21, 2000.

QA: Please complete →

← QA: Kaphingst or Kaphingist

Roman

QA: 1979 or 1969

← QA: please complete

Direct reprint requests to:

Dr. John D. Clapp
 San Diego State University
 5500 Campanile Drive
 San Diego, CA 92182-4119
 e-mail: jdclapp@mail.sdsu.edu