

Validity of the CAGE in Screening for Problem Drinking in College Students

Edward J. Heck

Department of Counseling Psychology, University of Kansas

James W. Lichtenberg

Department of Counseling Psychology, University of Kansas

This study examined the ability of an established screening test for alcoholism to identify accurately problem drinkers among college students.

Public concern over alcohol use among college students has existed for some time and will likely increase since the prevalence of heavy drinking has recently been reported to be higher in this population than in their college-age cohorts (Johnston, O'Malley, & Bachman, 1986). There is evidence that colleges and universities are responding increasingly to campus alcohol abuse problems through a variety of means, including policy changes, the development of training and intervention programs, and increasing numbers and variety of studies on students' drinking activities (Anderson & Gadaletto, 1984; Klein, 1989). With this increased interest in the issue of problem drinking, a need has grown to develop screening methods that allow for the early identification of high-risk students.

Although several screening instruments have been developed and validated for the detection of alcoholism in adults, such as the CAGE (described later in text; Ewing & Rouse, 1970), Michigan Alcohol Screening Test (MAST) (Selzer, 1971), Brief MAST (Pokorny, Miller, & Kaplan, 1972), and Trauma Scale (Skinner, Holt, Schuller, Roy, & Israel, 1984), there is evidence that the use of these instruments for detecting problem drinking in a college population may not be justified (Smith, Collins,

Kreisberg, Volpicelli, & Alterman, 1987). These data, along with the opinions of others (Burns & Sloane, 1987), suggest that the differential screening adequacy of these instruments could be the result of differences in the problem-drinking behaviors found in most college students as compared with the more severe behaviors characterizing the syndromes of alcohol abuse and alcoholism. Although the evidence questioning the screening adequacy for problem drinking of these three instruments is limited to the study by Smith et al. (1987), proposals have been made suggesting that these instruments can provide an easy means for the rapid screening of alcohol problems in all individuals (Kinney & Meilman, 1987). Although these instruments have established validity in detecting more severe alcohol-related problems, we question whether these instruments are sufficiently sensitive to identify accurately the less severe pattern of problem drinking that occurs in a college population.

One of the most efficient and effective instruments used for the routine and rapid screening of alcohol problems is a brief, 4-item questionnaire known as the CAGE. The term "CAGE" is an acronym with each letter representing one of the four items that composes the instrument (see Table 1). Developed by Ewing and Rouse (1970), the CAGE has been demonstrated to have a high degree of validity in identifying alcoholism and excessive drinking in an adult, psychiatric population (Bernadt, Taylor, Mumford, Smith, & Murray, 1982; Mayfield, McLeod, & Hall, 1974). Use of the CAGE as a definitive diagnostic test with a college population is generally not recommended (Kinney & Meilman, 1987). It is recommended, however, to be used as a screening test that is useful in identifying individuals whose alcohol use warrants further evaluation (Clark, 1985). The

Edward J. Heck and James W. Lichtenberg are professors of counseling psychology in the Department of Counseling Psychology and can be contacted at the University of Kansas, Lawrence, KS 66045. The authors gratefully acknowledge the financial support of the Division of Student Affairs for the alcohol survey project of which this study was a part.

TABLE 1
Quantity-Frequency Categories, Negative Effects, and CAGE Items

| |
|--|
| <p>Quantity-Frequency Categories</p> <p>QF1. Occasional drinking (1–2 drinks less than once/week)</p> <p>QF2. Light drinking (3–4 drinks less than once/week or 1–2 drinks 1–2 times/week)</p> <p>QF3. Light–Moderate (5–6 drinks less than once/week, 3–4 drinks 1–2 times/week, or 1–2 drinks more than twice/week)</p> <p>QF4. Moderate (7+ drinks less than once/week, 5–6 drinks 1–2 times/week, or 3–4 drinks more than twice/week)</p> <p>QF5. Moderately heavy (7+ drinks 1–2 times/week or 5–6 drinks more than twice/week)</p> <p>QF6. Heavy drinking (7+ drinks more than twice/week)</p> <p>Negative Effect Items</p> <p>Missed school</p> <p>Blacking out</p> <p>Arguments with close friend</p> <p>Arguments with boyfriend/girlfriend</p> <p>Fights while intoxicated</p> <p>Acts of stolen or damaged property</p> <p>Indiscriminate sexual activity</p> <p>Physical injuries</p> <p>CAGE Items</p> <p>Feeling the need to Cutdown on your drinking</p> <p>Becoming Annoyed at criticism of your drinking</p> <p>Feeling Guilty about your drinking</p> <p>Needing a drink first thing in the morning to get going (<i>Eye-opener</i>)</p> |
|--|

CAGE has been shown to be superior to either the MAST or Trauma Scale for screening purposes in a college freshman population based on its superior sensitivity (true positive rate) and specificity (true negative rate) values (Smith et al., 1987). Yet these same researchers still do not consider the sensitivity and specificity values high enough to serve as an accurate screening device for identifying problem drinkers.

With the increased focus in higher education on the issue of problem drinking and the necessity of developing valid instruments for future research, the purpose of this study was to evaluate further the use of the CAGE as an initial screening instrument in identifying problem drinking in a college population. The study extends and follows up the Smith et al. (1987) study with freshmen by evaluating the identification capability of the CAGE with a random sample of the entire student body at a large public university.

METHOD

Sample and Questionnaire Design

During the spring 1988 semester, a random sample of 1,000 degree-seeking students attending

a large public midwestern university were mailed a confidential questionnaire soliciting responses concerning their alcohol use. Excluded from the sample were medical and nondegree-seeking students. The 17-item survey contained several demographic items plus weekly quantity and frequency of alcohol consumption items from Hickenbottom, Bissonette, and O'Shea (1987), which were combined to form several quantity-frequency (QF) drinking categories. Also included were negative effects items from Smith et al. (1987) and the four items of the CAGE. The several QF categories, negative effects items, and CAGE items are presented in Table 1.

A total of 582 questionnaires were returned (58.2%), with the sample not significantly differing from the population proportions in either class standing, $\chi^2(4, N=582) = 5.74, p = \text{NS}$ or sex, $\chi^2(1, N=582) = .001, p = \text{NS}$. Because the sample proportions did not differ significantly from the population figures, no efforts to increase the response rate were made. Of the 582 students, 60 students (10.3%) reported being nondrinkers, leaving 522 students who reported the regular use of alcohol to varying degrees. Of the 522 surveys from students reporting reg-

ular use of alcohol, 508 surveys had complete data.

Identification of Problem Drinkers

Because no standardized definition or validated reference standard for problem drinking exists in the literature, a definition of problem drinking highly similar to the one used by Smith et al. (1987) was used. The definition and criteria used were based on combining certain quantity-frequency (QF) categories of drinking with ranges of negative effects. *Problem drinkers* were defined as individuals in QF categories 5 and 6 with a frequency range of 3–8 negative effects (see Table 1). *Normal drinkers* were defined as those in QF categories 1 and 2 with a negative effects range of 0–2 (see Table 1). To avoid ambiguity in the criteria and data analysis, students between these extremes were excluded from analysis.

Because this study was a partial replication of the Smith et al. (1987) study, it is important to note the slight differences in the problem-drinking criterion groups. In both studies the normal-drinking group represents about the lower 40% in quantity-frequency indexes; but in contrast with the current study, the Smith study did not include any students reporting negative effects in this group. For the problem-drinking group, the Smith study investigated students in the upper quartile in quantity-frequency who reported at least one negative effect. Although both studies surveyed identical negative effects, the Smith et al. problem-drinking criterion is more restrictive than that used in the current study, in the sense of allowing the possibility of fewer negative effects to define the group. Both studies excluded from analysis moderate users or students falling between these extremes, to avoid ambiguity in the criteria and data analysis.

Data Analysis

As was done by Smith et al. (1987), sensitivity, specificity, and positive predictive values of the CAGE were calculated at various cutoff points using procedures developed by Griner, Mayewski, Mushlin, and Greenland (1981). This was done to examine the capability of the CAGE to discriminate between problem-drinking and normal-drinking groups. Scores for the CAGE range from 0–4, and the literature suggests using a score of 2 or more as a threshold for

screening for potential alcoholism. Using this threshold score, those individuals with a score of 2 or more had a positive test, whereas those with a score of 0–1 had a negative test. When cross-classified with the previously discussed problem-drinking criteria, those who were identified as problem drinkers and who had a positive test were considered *true positive*. Conversely, those who were classified as normal drinkers and who had a negative test were considered *true negatives*.

Sensitivity refers to the true positive rate or the probability of a positive CAGE when problem drinking is present. It reflects the instrument's ability to identify true problem drinkers. The sensitivity of the CAGE was calculated as the number of true positives divided by the combined number of true positives and false negatives. *Specificity* refers to the false positive rate or the probability of a negative CAGE when problem drinking is not present. It reflects the ability of the instrument not to misclassify normal drinkers as problem drinkers. The specificity of the CAGE was calculated as the number of true negatives divided by the combined number of true negatives and false positives. *Positive predictive value* refers to the probability that problem drinking is present when the instrument is positive. It reflects the prevalence of problem drinking and was calculated as the number of true positives divided by the combined number of true positives and false positives.

The three indexes were calculated for each of three cutoff points of the CAGE. Unlike the Smith et al. (1987) study that calculated values for each of four cutoff points, there were no CAGE scores of 4 in this sample, thus precluding the use of this cutoff point. Calculations of these indexes were made for both the total sample and by sex.

RESULTS

Using the stipulated criteria of problem and normal drinking, there were 69 problem drinkers (13.6%) and 204 normal drinkers (40.2%) among the 508 students. There were 235 students (46.3%) who fell between these criterion groups. Within the problem-drinking group, there were 50 men and 19 women, whereas the normal-drinking group had 85 men and 119 women. Comparative profiles of sensitivity,

specificity, and positive predictive values of the CAGE, for both this study and the Smith et al. (1987) study, are presented in Table 2.

The results of this study suggest that the CAGE has an optimal sensitivity and specificity for problem drinking among college students, 57% and 76%, at a cutoff score of ≥ 1 for a positive test. These are similar percentages to the 57% and 85% found by Smith et al. (1987). At the recommended cutoff score of ≥ 2 for suspecting alcoholism (Mayfield et al., 1974), the CAGE has a sensitivity of only 26%, but an increased specificity of 95%. Again, these percentages are close to the 18% and 99% figures found previously. Similarly, the percentage figures noting the probability that problem drinking is present when the CAGE was positive (i.e., positive predictive value), between both studies, were similar.

Analysis of the sex data suggests similar sensitivity and specificity values for both the total sample of this study and the Smith et al. sample. The positive predictive value, however, was considerably lower at the ≥ 1 cutoff level for women in both studies, and this was especially the case in this study. It seems that a positive CAGE is a poorer predictor of problem drinking in women at the lower cutoff levels.

DISCUSSION

This study was, in part, a partial replication of the Smith et al. (1987) study in the sense in that it used a negative effects criterion for defining problem drinking, and it evaluated the CAGE as an alcohol-screening questionnaire among college students. This study, however, expanded the sample beyond freshmen by sampling students at all educational levels within the institution. Using identical measures of negative effects and highly similar criteria of problem drinking, the same conclusion was reached, namely, that the CAGE does not perform well enough to serve as a screening tool for problem drinking within a college population.

The CAGE was shown to have an optimal sensitivity of 57% and specificity of 76% using a cutoff score of 1, and 26% and 95% using a cutoff of 2. This means that at a cutoff score of 1, it will fail to identify 43% of the problem drinkers while inaccurately categorizing 24% of all normal drinkers as problem drinkers. If the cutoff score of 2 is used, then the misclassification of normal drinkers is fairly well resolved (i.e., 5% will be mislabeled as problem drinkers), but the sensitivity drops so that it now fails to identify 74% of the problem drinkers. This

TABLE 2
Comparative Test Characteristics of the CAGE by Varying Scores and Groups

| Score | Sensitivity (%) | Specificity (%) | Positive Predictive Value (%) |
|-----------------------|----------------------|-----------------|-------------------------------|
| Total Group | | | |
| ≥ 1 | 57 (57) ^a | 76 (85) | 44 (40) |
| ≥ 2 ^b | 26 (18) | 95 (99) | 64 (73) |
| ≥ 3 | 15 (7) | 99 (99.6) | 83 (78) |
| >4 | (1) | (99.6) | (40) |
| Men | | | |
| ≥ 1 ^c | 56 (60) | 75 (87) | 57 (45) |
| ≥ 2 | 26 | 93 | 68 |
| ≥ 3 | 12 | 98 | 75 |
| Women | | | |
| ≥ 1 | 58 (50) | 76 (84) | 28 (36) |
| ≥ 2 | 26 | 97 | 56 |
| ≥ 3 | 21 | 100 | 100 |

^aFigures in parentheses are reported by Smith et al. (1987).

^bCutoff scores recommended for detecting alcoholism.

^cSmith et al. (1987) have reported values for only cutoff score of ≥ 1 .

is a problem when it is used as a screening device, because the sensitivity index is more important for a screen function than is the specificity or mislabeling index (Griner et al., 1981). The primary requirement of a screening test should be its ability to identify accurately a high percentage of true positives, even at the expense of including some false positives.

The evidence from this study, as well from others that examined the screening validity of the CAGE, has demonstrated that the CAGE is differentially sensitive depending on the nature of the problem (i.e., problem drinking versus alcoholism). This fact supports the distinction made between the problem-drinking behaviors of college students and the behavioral patterns that define alcoholism. A likely reason for this differential sensitivity is that the CAGE contains items that reflect behavioral events that occur further along in the developmental chain of events leading to alcoholism. This notion seems supported by results in the current study when the frequencies of the four items of the CAGE were examined.

Specifically, the following was noted: (a) no student among the 508 checked more than 3 items and (b) there was considerable variation in frequencies of response between the CAGE items, ranging from a frequency of 3 for "needing a drink first thing in the morning" to 145 for "feeling a need to cut down." This latter point may be particularly important in that the low frequency item is one that is suggestive of physical dependency, a discriminating feature between problem drinking/alcohol abuse and alcoholism.

If this interpretation is correct, "needing a drink first thing in the morning" reflects an event or behavior occurring further along in the development chain of abuse and is not useful as a screening item for problem drinking. The identification of item content that might prove useful in constructing a sensitive screening test is suggested by other research. For example, there are several studies, including this one, implicating certain social influence variables such as peer pressure (Harford & Spiegler, 1983; Sherry & Stolberg, 1987; Stumphauzer & Perez, 1982) or friend's approval of problem-drinking behavior (Donovan, Jessor, & Jessor, 1983). Furthermore, research on alcohol-related expectancies of adolescents (Christiansen, Smith, Roehling, & Goldman, 1989) suggests that certain social expectancy scales or items (i.e., alcohol-en-

hancing social functioning) are highly correlated with a variety of drinking variables including the presence of problem drinking.

Another interpretation of these results involves the nature of the samples used in studies with the CAGE. Previous research has demonstrated that the CAGE can sensitively detect alcoholism; but this research has been typically conducted with adult, identified, or self-acknowledged alcoholic patients in a hospital setting. This study, as well as that of Smith et al. (1987), surveyed college students who were neither identified nor self-acknowledged as alcoholics.

If the "problem-drinking" rate in these two studies is about 13% to 15%, then it is likely that a certain percentage of the problem drinkers are alcoholics. Because the CAGE did not identify these individuals, it is plausible that college students who have a dependency problem are using some kind of a response set (e.g., social desirability set) to certain CAGE items. There is some indirect evidence on this point provided by Ewing (1984) who found that among a general hospital group of nonalcoholic men, 15% and 9% responded positively to the CAGE items of "feeling guilty about drinking" and "feeling a need to cut down," with negative responses to the other two items. In the current study, a similar pattern emerged among the normal-drinking group in which positive response percentages of 15% (feeling guilty), 17% (need to cut down), and 0% and 3% to the other items were found. All of this suggests that the CAGE's lack of sufficient validity for detecting problem drinking within a nonalcoholic sample is likely a result of the insensitivity of certain CAGE items.

In conclusion, although the availability of an efficient and sensitive screening instrument for problem drinking certainly would assist student development personnel in the early identification of students who may be at risk for developing more severe problems of alcohol abuse or alcoholism, the CAGE does not seem to be such an instrument. Specifically, issues pertaining to the instrument's sensitivity, specificity, and positive predictive value raise questions about the adequacy of the instrument as a problem-drinking screening device for a college population. Difficulties with the CAGE seem to involve, at least in part, issues related to its item content—items that may not reflect the behavioral criteria of problem drinking as a phenom-

enon distinct from the categories of alcohol abuse, alcoholism, or both.

Additionally, unlike alcohol abuse and alcoholism that have been defined by the *Diagnostic and Statistical Manual of Mental Disorders, Third Edition, Revised* (American Psychiatric Association, 1980), the numerous studies that have appeared on problem drinking—including studies that have used the CAGE—each have used different definitions and criteria (e.g., Donovan et al., 1983; Engs & Hanson, 1985; Hay, 1988; Hickenbottom et al., 1987; Hughes & Dodder, 1983; Klein, 1989; Sherry & Stolberg, 1987). The shortcomings of the CAGE notwithstanding, what is needed in this area are a standardized definition and set of criteria for problem drinking to test effectively the adequacy of any screening device.

REFERENCES

- American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders*, third edition, revised. Washington, DC: Author.
- Anderson, D., & Gadaletto, A. (1984). Progress or illusion: The 1979 and 1982 college alcohol surveys. *Journal of College Student Personnel*, 25, 332–337.
- Bernadt, M., Taylor, C., Mumford, J., Smith, B., & Murray, R. (1982). Comparison of questionnaire and laboratory tests in the detection of excessive drinking and alcoholism. *Lancet*, 1, 325–328.
- Burns, W., & Sloane, D. (1987). Commentary. *Journal of American College Health*, 36, 57–59.
- Christiansen, B., Smith, G., Roehling, P., & Goldman, M. (1989). Using alcohol expectancies to predict adolescent drinking behavior after one year. *Journal of Consulting and Clinical Psychology*, 57, 98–99.
- Clark, W. (1985). The medical interview: Focus on alcohol problems. *Hospital Practice*, 20, 59–68.
- Donovan, J., Jessor, R., & Jessor, L. (1983). Problem drinking in adolescence and young adulthood. *Journal of Studies on Alcohol*, 44, 109–137.
- Engs, R., & Hanson, D. (1985). The drinking patterns and problems of college students: 1981. *Journal of Alcohol and Drug Education*, 31, 65–83.
- Ewing, J. (1984). Detecting alcoholism, the CAGE questionnaire. *Journal of the American Medical Association*, 252, 1905–1907.
- Ewing, J., & Rouse, B. (1970, February). *Identifying the hidden alcoholic*. Paper presented at the 29th International Congress on Alcoholism and Drug Dependence, Sydney, Australia.
- Griner, P., Mayewski, R., Mushlin, A., & Greenland, P. (1981). Principles of test selection and use. *Annals of Internal Medicine*, 94, 559–592.
- Harford, T., & Spiegler, D. (1983). Developmental trends of adolescent drinking. *Journal of Studies on Alcohol*, 44, 181–188.
- Hay, R. (1988). Screening counseling center clients for drinking problems. *Journal of College Student Development*, 29, 79–80.
- Hickenbottom, J., Bissonette, R., & O'Shea, R. (1987). Preventive medicine and college alcohol abuse. *Journal of American College Health*, 36, 67–72.
- Hughes, S., & Dodder, R. (1983). Alcohol consumption patterns among college students. *Journal of College Student Personnel*, 24, 257–264.
- Johnston, L. D., O'Malley, P. M., & Bachman, J. G. (1986). *Drug use among American high school students, college students, and other young adults: National trends through 1985* (Report No. DHS, ADM-86-1450). Washington, DC: National Institute on Drug Abuse.
- Kinney, J., & Meilman, P. (1987). Alcohol use and alcohol problems: Clinical approaches for the college health service. *Journal of American College Health*, 36, 73–82.
- Klein, H. (1989). Helping the college student problem drinker. *Journal of College Student Development*, 30, 323–331.
- Mayfield, D., McLeod, G., & Hall, P. (1974). The CAGE questionnaire: Validation of a new alcoholism screening instrument. *American Journal of Psychiatry*, 131, 1121–1123.
- Pokorny, A., Miller, B., & Kaplan, H. (1972). The brief MAST: A shortened version of the Michigan Alcoholism Screening Test. *American Journal of Psychiatry*, 129, 342–345.
- Selzer, M. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry*, 27, 1643–1648.
- Sherry, P., & Stolberg, V. (1987). Factors affecting alcohol use by college students. *Journal of College Student Personnel*, 28, 350–355.
- Skinner, H., Holt, S., Schuller, R., Roy, J., & Israel, Y. (1984). Identification of alcohol abuse using laboratory tests and a history of trauma. *Annals of Internal Medicine*, 101, 847–851.
- Smith, D., Collins, M., Kreisberg, J., Volpicelli, J., & Alterman, A. (1987). Screening for problem drinking in college freshmen. *Journal of American College Health*, 36, 89–94.
- Stumphauzer, J., & Perez, P. (1982). Learning to drink: Peer survey of normal adolescents. *International Journal of the Addictions*, 17, 1363–1372.