

Response burden, reliability, and validity of the CAGE, Short MAST, and AUDIT alcohol screening measures

RON D. HAYS and JON F. MERZ
RAND, Santa Monica, California

and

RONALD NICHOLAS
Center for Counseling, Tarzana, California

We administered the CAGE, the Short MAST, and the AUDIT to 832 clients at drinking driver treatment programs in Southern California. Correlations among the alcohol screening measures ranged from 0.62 (CAGE and AUDIT) to 0.70 (CAGE and Short MAST). As expected, response time for the CAGE was quicker than for the Short MAST and the AUDIT, but the internal consistency reliability of the CAGE was the lowest and the standard error of measurement largest. Moreover, greater support was found for the relative validity of the Short MAST and AUDIT than the CAGE. The CAGE provides a substantial amount of information quickly, but more reliable and valid information can be obtained with the short MAST or AUDIT measures, which require an additional 1–2 min to administer.

The most popular alcohol screening tools include the CAGE (Mayfield, McLeod, & Hall, 1974), the Michigan Alcoholism Screening Test (MAST; Selzer, 1971), and the Alcohol Use Disorders International Test (AUDIT; Babor & Grant, 1989). The CAGE acronym stands for the 4 yes/no items that constitute this screening test (Cut down, Annoyed, Guilty, Eye-opener). Item responses on the CAGE are scored 0 or 1, with a higher score an indication of alcohol problems. A total score of 2 or greater is considered clinically significant. The MAST items focus on symptoms of problematic drinking and negative consequences of alcohol use. The 13 items in the Short MAST are also presented with yes/no response choices and are often scored using unitary weighting of the items (Skinner, 1979). A score of 3 or greater is the clinical cutpoint for this measure. The AUDIT is a 10-item instrument with varying numbers of response options per item, ranging from 3 to 6 choices. Responses are differentially weighted such that between 0 and 4 points are possible per item. A score of 8 or greater is suggestive of alcohol problems.

Sensitivities and specificities of the MAST and the CAGE to alcohol problems have been found to be fairly comparable, with a slight edge overall to the MAST (Hays & Spickard, 1987; Magruder-Habib, Durand, & Frey,

1991). However, some reviewers of the literature have concluded that the brevity of the CAGE makes it a good choice for alcohol screening in primary care practices (Hays & Spickard, 1987; Magruder-Habib et al., 1991). "When different measures of the same dimension of drinking are about equally reliable and accurate, the more economical or efficient measure could be used" (Tucker, Vuchinich, Harris, Gavornik, & Rudd, 1991, p. 154).

In this study, we extend previous work (Hays et al., 1993) by using response time to quantify the relative burden of the CAGE, the Short MAST, and the AUDIT. In addition, we assess the reliability and relative validity of each measure.

METHOD

Subjects

The *admin per se* law in California requires the Department of Motor Vehicles to revoke an individual's driving privileges if a blood alcohol concentration (BAC) test reveals that this person has been driving with a BAC level exceeding the legal limit (i.e., 0.08 or above). When this study was conducted, all first and second offenders were required to undergo treatment as a condition of their offense.

Eight hundred and thirty-two clients (646 males, 186 females) from six West Coast treatment programs for drinking (impaired) drivers were sampled. Ninety-seven percent of the sample were in the program because of alcohol-related impaired driving, and the remainder were in the program because of other drug-related impaired driving. The average age of the clients was 33 years, with a range of 18–71. Sixty-nine percent of the sample were non-Hispanic white, 16% Hispanic, 6% non-Hispanic black, 5% Asian, 2% Native American, and 2% other. In addition, 84% were high school graduates, and 21% were married and living with their spouse. The median yearly household income was \$30,001–\$50,000 for the sample.

Sixty-six percent of the respondents were first offenders, and 34% were multiple offenders. The clients reported an average of

Preparation of this article was supported in part by Grant AA 07852 from the National Institute on Alcohol Abuse and Alcoholism. The opinions expressed are those of the authors and do not necessarily reflect the views of RAND or the sponsor. Thanks are due to Craig Barela for outstanding secretarial support. Correspondence concerning this article should be addressed to R. D. Hays, RAND, 1700 Main Street, Santa Monica, CA 90407-2138 (e-mail: ronald_hays@rand.org).

1.4 arrests in the last 7 years for impaired driving, and 22% reported one or more automobile accidents during the last 12 months. The average blood alcohol level reported for the time of the arrest was 0.15, which is similar to peak BACs for adult drivers involved in fatal crashes (Zobeck, Williams, Grant, & Bertolucci, 1990) and the average BAC of convicted drinking drivers in previous studies (Wilson, 1993). In the sample, 17% categorized themselves as alcoholic, 7% as problem drinkers, 9% as heavy drinkers, 40% as social drinkers, 12% as light drinkers, 12% reported drinking only on special occasions, and 3% reported that they never drank alcohol. (The clients who reported never drinking alcohol were reporting this status since their arrest for impaired driving.) Seventy percent of the clients exceeded the clinical cutpoint for problematic alcohol use on one or more of the three alcohol screening measures (see Measurement section below) included in the study.

Measurement

The Microcomputer Assessment System, a microcomputer program for computer administration of questionnaires (Hays et al., 1992), was used to field a 112-item interview. The beginning of the interview instructed respondents that their participation was part of the usual intake procedure at the treatment center and, per human subjects requirements, the information they provided would be kept confidential with the exception of the information that the courts, DMV, probation, and state and county offices need in order to assure completion of program requirements and compliance with the law. A single-page summary of the interview was produced for the clinical staff for all clients completing the interview. After completing the interview, clients were asked by microcomputer to provide consent for the data to be used for this research project. Eighty-nine percent of the 935 clients ($n = 832$) who completed the computer interview provided consent to allow their data to be used for the study.

The interview contained sociodemographic items, a socially desirable response set scale (Hays, Hayashi, & Stewart, 1989), questions about frequency of alcohol and drug use, the Drug Abuse Screening Test (Skinner, 1982), and alcohol use screening measures including the CAGE (Mayfield et al., 1974), a revised version of the AUDIT (Babor & Grant, 1989), and a short form of the MAST (Selzer, 1971; Selzer, Vinokur, & Van Rooijan, 1975). A reading skills assessment (two-paragraph story followed by four factual questions) was also included to ensure that clients could read at a minimum of seventh grade level. For frequency of alcohol use questions, we randomized the location in the interview (earlier vs. later), the nature of the response options presented (qualitative vs. quantitative), and the order of presentation of response options (light use to heavy use vs. heavy use to light use). All other measures were presented in a constant manner to each respondent (Hays et al., 1995). Average time to complete the microcomputer interview was 26 min ($SD = 9$, range = 9–64).

Although this study did not compare the computer interview directly with other methods of assessment, clients were asked to contrast the interview with answering the same questions by a counselor. Thirty-nine percent of the clients reported that they found the

computer interview to be easier than a counselor interview, and only 14% felt it was harder.

Analysis Plan

Pearson product-moment correlations were computed between the CAGE, Short MAST, and the AUDIT. We also examined the extent to which each measure was susceptible to socially desirable responding by correlating these scales with a five-item socially desirable response set scale (Hays et al., 1989). Internal consistency reliability estimates were calculated for each scale (Cronbach, 1951), as well as the standard error of measurement: "The expected standard deviation of scores for any person taking a large number of randomly parallel tests" (Nunnally, 1978, p. 218). The Microcomputer Assessment System used to administer the interview in this study logs the time to complete each question using the computer's internal clock (BIOS timer). We estimated the number of seconds to complete each alcohol screening scale using this timer, which is accurate to $1/18$ of a second.

We examined the association of the alcohol screening scales with six criterion variables: drinker self-label (*alcoholic or problem drinker; heavy drinker; social drinker; light drinker or special occasions drinker*), frequency of consuming five or more drinks during the last 30 days (recoded into five categories), number of arrests for drinking and driving in the last 7 years (observed range of one to five), frequency of drinking and driving in the last 12 months (recoded into five categories), likelihood of drinking and driving in the future (five response choices ranging from *definitely will not to definitely will*), and number of automobile accidents in the last 12 months (*none, one, more than one*). Relative validity was estimated by calculating F statistics for each measure to determine in proportional terms the strength of associations with known group differences (Liang, Larson, Cullen, & Schwartz, 1985).

RESULTS

The CAGE and the Short MAST were the most strongly intercorrelated ($r = .70, p < .01$) of the alcohol screener scales, followed by the Short MAST and the AUDIT ($r = .66, p < .01$), and then the CAGE and the AUDIT ($r = .62, p < .01$). These three correlations differed significantly ($p < .05$) from one another (Steiger, 1980). All three screening scales correlated significantly with the socially desirable response set scale (product-moment correlations ranging from $-.21$ to $-.25, ps < .001$), indicating that those scoring higher on the socially desirable response scale tended to score lower on the alcohol screening measures.

Raw and 0–100 linear transformed means, standard deviations, observed ranges, internal consistency reliability estimates, and response times for the alcohol screening scales are provided in Table 1. The linear transformation

Table 1
Descriptive Statistics, Reliabilities, and Response Times for Alcohol Scales

Measure	Raw		Observed Range	Internal Consistency Reliability	0–100 Transformed			Response Time (in seconds)
	<i>M</i>	<i>SD</i>			<i>M</i>	<i>SD</i>	<i>SEM</i>	
CAGE	1.7	1.3	0–4	.69	42.8	32.4	18.0	31
Short MAST	3.7	3.1	0–13	.84	28.7	23.8	9.5	97
AUDIT	8.6	6.9	0–35	.83	21.6	17.3	7.1	131

Note— M = mean, SD = standard deviation, SEM = standard error of measurement. Internal consistency reliability is Cronbach's (1951) alpha. Possible range and clinical cutpoints (in parentheses) for screener scales are as follows: CAGE (4 items) is 0–4 (>1), Short MAST (13 items) is 0–13 (>2), and AUDIT (10 items) is 0–40 (>7).

Table 2
Relative Validity (*F* ratios) Estimates for Alcohol Scales

Criterion Variable	Alcohol Screening Scale		
	CAGE	Short MAST	AUDIT
Drinker self-label	215.5	426.3	166.8
Binge drinking in last 30 days	14.6	13.6	54.5
Number of arrests for drinking and driving in last 7 years	16.2	21.7	3.0
Frequency of drinking and driving in last 12 months	7.4	6.0	16.4
Likelihood of drinking and driving in future	6.9	6.7	14.9
Automobile accidents in last 12 months	6.7	6.2	26.4

Note—Table entries are *F* ratios from one-way ANOVAs with alcohol screening scales as the dependent variables and row entries as the classification (known groups) variables. All *F* ratios were statistically significant ($p < .05$).

of raw scores to 0–100 scales facilitates comparisons across scales and does not influence interval-level properties of the measures.

The estimated reliability of the CAGE was the lowest (internal consistency reliability of .69) and its standard error of measurement was the highest of the three scales (see Table 1). Reliabilities (.84 and .83, respectively) and standard errors of measurement were similar for the Short MAST and the AUDIT. The average response times for the CAGE, the Short MAST, and the AUDIT were 31, 97, and 131 sec, respectively.

As shown in Table 2, the Short MAST was more strongly related to drinker self-label than were the CAGE and the AUDIT. The relative validity of the Short MAST and the CAGE were 2.6 and 1.3 times, respectively, that of the AUDIT. The Short MAST was also 1.3 times more sensitive than the CAGE and 7.2 times more sensitive than the AUDIT to number of arrests for drinking and driving during the last 7 years. In contrast, the AUDIT was approximately 4 times more sensitive than the CAGE and the Short MAST to binge drinking in the last 30 days and number of automobile accidents in the last 12 months and more than 2 times as sensitive to frequency of drinking and driving during the last 12 months and perceived likelihood of drinking and driving in the future.

DISCUSSION

The results of this study indicate that different alcohol screening measures are correlated significantly with one another, but they clearly yield unique information. The higher reliability and lower standard error of measurement for the Short MAST and the AUDIT relative to the CAGE is consistent with the fact that both of the former scales have more items than the latter. These standard errors can be used to set confidence intervals around an individual's scores. For example, for someone scoring at the mean of each scale, the 95% confidence interval around their observed raw score would be 0.3–3.1 (1.7 ± 1.4) for the CAGE, 1.2–6.2 (3.7 ± 2.5) for the Short MAST, and 2.9–14.3 (8.6 ± 5.7) for the AUDIT. These confidence intervals encompass 70% of the observed

range for the CAGE, 38% of the observed range for the Short MAST, and 33% of the observed range for the AUDIT.

The associations of alcohol screening scales with criterion variables suggest that the Short MAST is most sensitive to long-term drinking patterns (drinker self-label and number of arrests for drinking and driving during the last 7 years), whereas the AUDIT is more sensitive to recent drinking experience (binge drinking in last 30 days, frequency of drinking and driving in the last 12 months). The Short MAST and the AUDIT intercorrelated at a moderate level ($r = .66$), sharing 44% of variance in common.

When selecting an alcohol screening measure for clinical applications, the user needs to decide whether the administration time saved by using the CAGE versus the Short MAST and the AUDIT offsets the cost of lower reliability, greater error of measurement, and diminished sensitivity to alcohol-related problems. In fact, the results of this study indicate that administering both the Short MAST and the AUDIT could be accomplished in less than 4 min, and these alcohol screening scales provide distinct information. Future evaluations of these and other alcohol screening measures (e.g., Hays & Revetto, 1992; Rost, Burnam, & Smith, 1993) are needed to help quantify the tradeoffs between response burden and psychometric performance.

REFERENCES

- BABOR, T. F., & GRANT, M. G. (1989). From clinical research to secondary prevention: International collaboration in the development of the Alcohol Use Disorders Identification Test (AUDIT). *Alcohol Health & Research World*, 13, 371-374.
- CRONBACH, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- HAYS, J. T., & SPICKARD, W. A. (1987). Alcoholism: Early diagnosis and intervention. *Journal of General Internal Medicine*, 2, 420-427.
- HAYS, R. D., BELL, R., GILLOGLY, J. J., HILL, L., GIROUX, D., DAVIS, C., LEWIS, M. W., DAMUSH, T. M., & NICHOLAS, R. (1995). *Impact of response options and feedback about response inconsistencies on alcohol use self-reports by microcomputer*. Submitted for publication.
- HAYS, R. D., GILLOGLY, J. J., HILL, L., LEWIS, M. W., BELL, R. M., & NICHOLAS, R. (1992). A Microcomputer Assessment System (MAS) for administering computer-based surveys: Preliminary results from administration to clients at an impaired-driver treatment program.

- Behavior Research Methods, Instruments, & Computers*, **24**, 358-365.
- HAYS, R. D., HAYASHI, T., & STEWART, A. L. (1989). A five-item measure of socially desirable response set. *Educational & Psychological Measurement*, **49**, 629-636.
- HAYS, R. D., HILL, L., GILLOGLY, J. J., LEWIS, M. W., BELL, R. M., & NICHOLAS, R. (1993). Response times for the CAGE, Short MAST, AUDIT, and JELLINEK Alcohol Scales. *Behavior Research Methods, Instruments, & Computers*, **25**, 304-307.
- HAYS, R. D., & REVETTO, J. P. (1992). Old and new MMPI-derived scales and the Short-MAST as screening tools for alcohol disorder. *Alcohol & Alcoholism*, **27**, 685-695.
- LIANG, M. H., LARSON, M. G., CULLEN, K. E., & SCHWARTZ, J. A. (1985). Comparative measurement efficiency and sensitivity of five health status instruments for arthritis research. *Arthritis & Rheumatism*, **28**, 542-547.
- MAGRUDER-HABIB, K., DURAND, M., & FREY, K. A. (1991). Alcohol abuse and alcoholism in primary health care settings. *Journal of Family Practice*, **32**, 406-413.
- MAYFIELD, D., MCLEOD, G., & HALL, P. (1974). The CAGE questionnaire: Validation of a new alcoholism screening instrument. *American Journal of Psychiatry*, **131**, 1121-1123.
- NUNNALLY, J. C. (1978). *Psychometric theory*. New York: McGraw-Hill.
- ROST, K., BURNAM, A., & SMITH, G. R. (1993). Development of screeners for depressive disorders and substance disorder history. *Medical Care*, **31**, 189-200.
- SELZER, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic instrument. *American Journal of Psychiatry*, **127**, 1653-1658.
- SELZER, M. L., VINOKUR, A., & VAN ROOIJAN, L. (1975). A self-administered Short Michigan Alcoholism Screening Test (SMAST). *Journal of Studies on Alcohol*, **36**, 117-126.
- SKINNER, H. A. (1979). A multivariate evaluation of the MAST. *Journal of Studies on Alcohol*, **40**, 831-844.
- SKINNER, H. A. (1982). The Drug Abuse Screening Test. *Addictive Behaviors*, **7**, 363-371.
- STEIGER, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin*, **87**, 245-251.
- TUCKER, J. A., VUCHINICH, R. E., HARRIS, C. V., GAVORNIK, M. G., & RUDD, E. J. (1991). Agreement between subject and collateral verbal reports of alcohol consumption in older adults. *Journal of Studies on Alcohol*, **52**, 148-155.
- WILSON, R. J. (1993). Drinking and driving: In search of solutions to an international problem. *Alcohol Health & Research World*, **17**, 212-220.
- ZOBECK, T. S., WILLIAMS, G. D., GRANT, B. F., & BERTOLUCCI, D. (1990). Years of potential life lost and other trends in alcohol-related fatal traffic crashes: 1977-1987. *Alcohol Health & Research World*, **14**, 63-68.

(Manuscript received November 14, 1994;
revision accepted for publication January 27, 1995.)