

A COMPARISON OF ALCOHOL SCREENING INSTRUMENTS AMONG UNDER-AGED DRINKERS TREATED IN EMERGENCY DEPARTMENTS

THOMAS M. KELLY^{1*}, JOHN E. DONOVAN¹, JANET M. KINNANE^{1,2} and DAVID M. C. D. TAYLOR³

¹University of Pittsburgh School of Medicine, Pittsburgh, ²Children's Hospital of Pittsburgh, Pittsburgh, PA, USA and

³Royal Melbourne Hospital, Victoria, Australia

(Received 19 September 2001; in revised form 1 February 2002; accepted 17 February 2002)

Abstract — **Aims:** Few studies have examined the adequacy of adult-validated alcohol screening measures when used with adolescents and young adults. A total of 103 subjects (55 males, 48 females) participated in a study of alcohol use among under-aged drinkers conducted in two emergency departments. **Methods:** Participants completed three brief screening instruments for problematic alcohol use: the Alcohol Use Disorders Identification Test (AUDIT); a modified version of the TWEAK; and the CAGE. **Results and Conclusions:** Missing data on the TWEAK, lower internal consistency for the TWEAK and CAGE, and the better ability of the AUDIT to differentiate problem drinkers from non-problem drinkers, suggest that the AUDIT performs best in screening for problematic alcohol use among under-aged drinkers treated in emergency departments.

INTRODUCTION

Adolescents who use alcohol are at increased risk for accidental and intentional injury (Clark *et al.*, 1997; Spain *et al.*, 1997; Kelly *et al.*, 2001) and illness (Hansell and White, 1991; Arria *et al.*, 1995). However, the passage of time reduces the impact that drinking-related injury has in motivating patients to seek or accept help (Nilssen *et al.*, 1994). Emergency departments (ED) are important settings for the early identification of adolescents who misuse alcohol, because injured adolescents with problematic drinking often present there first. There is evidence that referral for treatment for some alcohol-related problems at the time of an ED visit may improve the rate of referral acceptance and treatment follow-through (Spirito *et al.*, 1994).

Substance-use screening instruments specifically designed for adolescents are available (Leccese and Waldron, 1994; Blum, 1997), but are generally too long and time-consuming to be practical in the hectic environment of a hospital ED. Shorter scales have been developed for adolescents, but there is scant research on such instruments among adolescents who report a range of alcohol involvements. Knight *et al.* (1999) developed a 6-item screening instrument (the CRAFFT) for problematic drug and/or alcohol use by extracting items from several other, lengthier adolescent self-report instruments. These investigators found that the instrument possessed adequate internal consistency. Bastiaens *et al.* (2000) recently reported on a 5-item instrument, the RAFFT, and found it to possess good sensitivity and specificity for diagnosing substance use disorders among adolescents seen in an emergency room or ambulatory care centre. However, a selection bias was operating in both of these studies, as many of the subjects had apparent symptoms of substance misuse or had been referred for treatment of substance misuse pathology.

In contrast to these instruments developed for adolescents, there is a variety of measures designed to screen for problematic alcohol use in adults that are relatively short, can be used

in EDs, and have established reliability and validity. It seems reasonable to investigate the potential of these instruments for the early detection of problematic alcohol use among adolescents at high risk for injury and illness.

There are only a few reports on the utility of existing adult-validated alcohol screening instruments for use with adolescents in EDs. Shope *et al.* (1997) used the CAGE in an ED study of 263 injured adolescents (mean age 16.9 years), and found that CAGE scores correlated positively with age, alcohol misuse (negative consequences of drinking), and overall alcohol involvement. However, this study did not involve a comparison of the CAGE with other alcohol screening measures. Chung *et al.* (2000) described an ED study of adolescents 13–19 years old that compared the same three screening instruments for problematic alcohol use that are used in the current study: (1) the Alcohol Use Disorders Identification Test (AUDIT, Saunders *et al.*, 1993); (2) the TWEAK (Russell, 1994); (3) the CAGE (Ewing, 1984). Chung *et al.* (2000) used DSM-IV (American Psychiatric Association, 1994)-defined alcohol use disorders as the criterion measure, and found that the AUDIT performed best in predicting problematic alcohol use, the TWEAK performed satisfactorily at a cut-off score of 2, and the CAGE was relatively inefficient in its sensitivity and specificity for alcohol use disorders among the adolescents. Chung *et al.* (2000) excluded alcohol-positive adolescents and focused on the criterion validity of the instruments. Our method includes investigation of internal consistency and construct validity and includes alcohol-positive adolescents. As a result, we are able to compare alcohol-positive with alcohol-negative adolescents and to assess the criterion validity of the screening instruments using a standard that was not available in the Chung *et al.* (2000) investigation.

The AUDIT was developed by the World Health Organization as a measure of alcohol consumption, alcohol dependence, and alcohol-related problems. The TWEAK (Russell, 1994) was developed to assess potentially hazardous drinking patterns among women and has demonstrated reliability and validity in these samples. The CAGE (Ewing, 1984) focuses on behaviours consistent with alcohol dependence. The objectives of the present study were to determine the reliability (internal consistency) and validity of these instruments in this

*Author to whom correspondence should be addressed at: Western Psychiatric Institute and Clinic, 3811 O'Hara St, Pittsburgh, PA 15213, USA.

sample, and to determine which of these instruments worked best in screening adolescents for problematic alcohol use in an ED.

SUBJECTS AND METHODS

Procedure

This protocol was approved by the Institutional Review Board of the University of Pittsburgh, Pittsburgh, Pennsylvania, USA. Dedicated research technicians staffed two regional Level-1 hospital EDs (one was a paediatric ED) at various times between 18:00 and 02:00 or between 21:00 and 06:00. Inclusion criteria for the study consisted of: (1) aged 12–20 years (inclusive); (2) Glasgow Coma Score = 15 (Jennet and Bond, 1975), i.e. no serious head injury; (3) patient accompanied by parent or legal guardian (if aged <18 years); (4) able to read and complete a self-report questionnaire. Patients whom the medical staff considered to be too seriously ill or too critically injured to approach were excluded.

Medical staff members informed the research technicians of patients meeting the above criteria. Parents or guardians of those aged <18 years were approached first, and the research project was explained. Parents were told that they would not be informed of the study findings for their adolescent child. If they gave written consent for their child to participate, the adolescent was then approached to obtain his/her written, informed assent. Following enrolment, an alcohol breathalyser test was administered using the Alco-Sensor III™ intoximeter (Intoximeters Inc., St Louis, MO, USA) to determine if the adolescent had been drinking within several hours of arrival at the ED. The medical staff were not informed of any results and there was no entry of this research data item into the medical record.

Following administration of the breathalyser test, a 29-item self-report questionnaire was administered in private to the adolescent. Although the questionnaire is in self-report format, the research technician remained with the subject to answer any questions or clarify issues related to completing the questionnaire. The alcohol screening items from the AUDIT, the TWEAK and the CAGE are embedded in the questionnaire. Participants were given a \$10 gift certificate for their participation.

Sample

A total of 554 patients between the ages of 12 and 20 years were screened for eligibility. The total sample included 298 males and 253 females (data on gender is missing for three patients who were screened for eligibility but not enrolled in the study). A total of 416 patients (75%) were screened at the paediatric ED, and 138 (25%) were screened at the general hospital ED; 359 patients could be approached according to protocol guidelines (195 patients were under the age of consent and had no parent with them or were too ill or injured to approach); 282 (78.5%) patients agreed to participate. No statistically significant age, gender or racial differences were found between those who participated and those who refused to participate. One case was missing from this analysis ($n = 358$).

Of the 282 participants, 170 (60.3%) indicated that they never drank alcohol and nine (3.2%) provided inconsistent responses regarding their alcohol use. In all, 103 participants (36.5%) indicated that they used alcohol, i.e. more than a sip

or taste on one occasion, and properly completed the alcohol screening questionnaire. These 103 alcohol-using participants (48 females, 55 males) are the focus of this report. There were 82 Caucasians (79.5%), 15 African-Americans (14.5%), five Asian-Americans (5%), and one Hispanic-American (1%) in the sample. The mean \pm SD age of the sample was 17.5 ± 2.1 years (range 12.2–20.9).

Data analytical procedures

The scale properties of the three alcohol screening measures were determined through the use of the Reliabilities procedure in SPSS 8.0. This procedure was used to calculate Cronbach α estimates of the inter-item reliabilities of the scales (Cronbach, 1951). The procedure also computed item means and SD, a scale mean and SD, a correlation matrix among the items, an average inter-item correlation, the correlation between each item and a corrected item-total correlation. The Correlation procedure of SPSS 8.0 was used to calculate Pearson correlations among AUDIT, TWEAK and CAGE summative scales scores. Student's *t*-test was used for group comparisons on the instruments. The assumption of homogeneity of mean differences was not met for some tests. For these tests, the *t*-statistic and degrees of freedom appropriate for an assumption of unequal variance are reported.

RESULTS

Scale properties of the alcohol screening measures

AUDIT. Table 1 presents the frequency distributions on the ten AUDIT items. Items with the highest level of response are those assessing frequency (no. 1), typical quantity per occasion (no. 2), and frequency of high-volume drinking (no. 3). Items assessing the experience of problems due to drinking are much less frequently endorsed in this young sample. Reliability for the AUDIT was high (standardized $\alpha = 0.88$). The corrected item-total correlations (Table 1) indicate that the alcohol intake items contribute most strongly to the total AUDIT scores. Items relating to alcohol dependence and consequences contribute more modestly. The mean score for the AUDIT is 7.74 ± 7.14 . This is high for this relatively young sample, as the generally accepted cut-off for optimal sensitivity/specificity of alcohol use disorders is a score of 8 (Conigrove *et al.*, 1995).

Table 2 presents the correlations among the ten AUDIT items. The average inter-item correlation among these items was 0.42. The strongest correlations were among the alcohol intake items (nos. 1, 2 and 3). Some noteworthy relationships between intake, dependence and problems included a moderately strong correlation between item 3, 'binge drinking', and item 5, 'failing to do what is expected' ($r = 0.61$). Item 4, 'being unable to stop drinking', and item 8, 'being unable to remember', are similarly correlated ($r = 0.58$). Item 10, 'others concerned', correlated 0.65 with item 6, 'needing an eye-opener'.

TWEAK. Table 3 presents the frequency distributions for the TWEAK scale. Consistent with the development of the instrument (Russell, 1994), the 'need to cut down' item and the 'eye-opener' item that are contained in the CAGE were included to construct the TWEAK scale. In order to avoid redundant concepts, the items measuring 'amnesia' and others 'worried' (concerned), as stated in the AUDIT, were used in

Table 1. Item distributions, means^a and corrected item-total correlations for the AUDIT scale

No.	Item	Score (<i>n</i>)					Mean ± SD	Item-total correlation ^b
		0	1	2	3	4		
1.	How often do you have a drink of alcohol?	—	52	30	16	5	1.75 ± 0.89	0.70
2.	How many drinks containing alcohol do you have on a typical day when you are drinking?	39	23	18	15	8	1.32 ± 1.32	0.59
3.	How often do you have six or more drinks on one occasion?	42	32	11	17	2	1.10 ± 1.16	0.79
4.	How often during the last year did you find that you were unable to stop drinking once you had started?	78	11	7	5	2	0.47 ± 0.96	0.61
5.	How often during the last year did you fail to do what you were expected to do because of drinking?	74	18	7	3	1	0.44 ± 0.83	0.63
6.	How often during the last year did you need a drink in the morning after a heavy drinking session to get yourself going?	96	3	0	2	2	0.17 ± 0.70	0.59
7.	How often during the last year did you feel guilty or remorseful after drinking?	59	35	5	3	1	0.57 ± 0.80	0.51
8.	How often during the last year were you unable to remember what happened the night before because of drinking?	64	28	7	1	3	0.56 ± 0.90	0.57
9.	Have you or someone else been injured as a result of your drinking?	78		7		18	0.84 ± 1.55	0.42
10.	Has a relative or friend, or a doctor or other health care worker been concerned about your drinking or suggested you cut down?	88		4		11	0.50 ± 1.27	0.55
Total scale score							7.74 ± 7.14	

n = 103.

^aMeans based on scoring continuum of 0–4 for items 1 and 3–8; 0 = never, 1 = monthly or less; 2 = monthly; 3 = weekly; 4 = daily or almost daily. Item 2: 0 = 1–2 drinks; 1 = 3–4 drinks; 2 = 5–6 drinks; 3 = 7–9 drinks; 4 = 10 or more drinks. Items 9 and 10 based on scoring 0 = never; 2 = yes, but not in the last year; 4 = yes, in the last year. Total scale scores range from 1 to 40.

^bCorrected for the presence of each item.

Table 2. Inter-item correlations^a for the AUDIT scale

	Items									
	1	2	3	4	5	6	7	8	9	10
AUDIT 1	—									
AUDIT 2	0.68	—								
AUDIT 3	0.75	0.79	—							
AUDIT 4	0.46	0.30	0.47	—						
AUDIT 5	0.50	0.43	0.61	0.46	—					
AUDIT 6	0.35	0.22	0.39	0.53	0.47	—				
AUDIT 7	0.33	0.32	0.45	0.33	0.48	0.39	—			
AUDIT 8	0.41	0.34	0.50	0.58	0.37	0.49	0.31	—		
AUDIT 9	0.39	0.30	0.41	0.32	0.22	0.27	0.22	0.27	—	
AUDIT 10	0.33	0.25	0.38	0.48	0.48	0.65	0.43	0.35	0.28	—

n = 103.

^aAll coefficients are statistically significant at $P \leq 0.05$ (two-tailed test).

Table 3. Item distributions, means^a and corrected item-total correlations for the TWEAK scale

Item	Score (<i>n</i>)				Item-total correlation
	0	1	2	Mean ± SD	
1. How many drinks can you hold without falling asleep or passing out? (Tolerance) ^b	23	—	47	1.37 ± 0.94	0.04
2. Has a relative or friend, or a doctor or other health worker been concerned about your drinking or suggested you cut down? (Worried) ^c	88	—	15	0.29 ± 0.71	0.29
3. Do you ever need a drink first thing in the morning to get going? (Eye-opener) ^d	101	2	—	0.002 ± 0.14	0.35
4. How often during the last year were you unable to remember what happened the night before because of drinking? (Amnesia) ^c	64	39	—	0.38 ± 0.49	0.17
5. Have you ever felt the need to cut down on your drinking? (Kut-down) ^d	71	32	—	0.32 ± 0.46	0.31
Total scale score				2.74 ± 1.7	

n = 103.

^aMeans based on scoring 0 for <6 drinks or 2 for ≥6 on item 1; 0 for 'no' or 2 for 'yes' to item 2; 0–1 for items 3 and 5. All zeros on item 4 were scored 0, all others were scored 1. Total scale scores range from 0 to 7.

^bThirty-three participants did not respond to this item.

^cAUDIT item.

^dCAGE item.

this version of the TWEAK. The scores on this scale are lower compared to the AUDIT due to the 0–1, 0–2 scaling and because most items measure problems related to very heavy drinking or alcohol dependence, e.g. tolerance, needing an eye-opener, and blacking out. Only two respondents endorsed needing an ‘eye-opener’ first thing in the morning (item 3). The mean score on the TWEAK scale was 2.74, which reflects a fairly high level of alcohol problems on this 7-point scale. The performance of this scale was compromised by the tolerance question, because 32% of these adolescent/young adult drinkers did not know how to respond to this item.

Correlations among the TWEAK items (Table 4) were much lower than those among the AUDIT items (Table 2), and the average inter-item correlation of 0.17 was much lower. The strongest correlations were between the ‘others worried’ item and the participant’s own concern for ‘needing to cut-down’ ($r = 0.43, P < 0.01$), and between ‘needing an eye-opener’ and the ‘others worried’ item ($r = 0.36, P < 0.01$).

Reliability based on internal consistency is low for the TWEAK scale (standardized item $\alpha = 0.50$), suggesting that the items are not measuring a single underlying construct. If the tolerance item is removed, internal consistency increases moderately (standardized item $\alpha = 0.62$). This is still low but approaches the range displayed by the CAGE (results below). According to the corrected item-total correlations (Table 3), item 1 (Tolerance) and item 4 (Amnesia) contribute very little to individual differences on the TWEAK scale. The other items contribute more, but still display only modest item-total correlations (range = 0.29–0.35).

CAGE. Table 5 displays the CAGE items, their frequency distributions, and item-total correlations.

The average correlation among the CAGE items was 0.35, much higher than that found for the TWEAK, but lower than found for the AUDIT. The item assessing ‘need to cut down’

correlated at 0.50 ($P < 0.01$) with ‘annoyance at criticism,’ 0.46 ($P < 0.01$) with ‘feeling guilty about your drinking,’ and 0.21 ($P < 0.05$) with the ‘eye-opener’ question. ‘Annoyance’ correlated at 0.24 ($P < 0.05$) with ‘guilt’ and 0.43 ($P < 0.01$) with ‘eye opener’; and ‘guilt’ and ‘eye-opener’ correlated at 0.23 ($P < 0.05$).

The estimate of internal consistency for the 4-item CAGE was $\alpha = 0.66$, which is satisfactory for such a short scale. The mean \pm SD score on the CAGE was just 0.71 ± 1.03 , reflecting the relative lack of alcohol problems as assessed by the CAGE in this young sample.

Concurrent validity of the screening measures

Concurrent validity of the instruments was assessed by correlations among the total scale scores. The AUDIT and TWEAK correlated at 0.83, the AUDIT and CAGE correlated at 0.43, and the TWEAK and CAGE correlated at 0.53. All three correlations were statistically significant ($P < 0.001$, two-tailed test).

Construct validity using ‘known group’ differences on alcohol use

We investigated the construct validity of the instruments by comparing younger and older participants, males and females, and ethnic groups on the screening instruments. Epidemiological studies showed that older adolescents are generally more involved in problem drinking than younger adolescents (Barnes and Welte, 1986; Centers for Disease Control and Prevention, 1996), that Caucasian adolescents are generally more involved in problem drinking than African-American adolescents (Rachal *et al.*, 1982; SAMHSA, 1997) and that male adolescents are more involved in problem drinking than female adolescents (Jessor, 1987; Windle, 1996). Construct validity of these instruments for assessing problematic alcohol use among adolescents would be supported to the extent that these same ‘known group’ differences are found on them.

Age differences. To test for age differences on the alcohol screening measures, adolescents in the sample were divided into two groups at the mean age of 17.5 years. There were 51 participants in the younger group and 52 in the older group. The groups differed in the expected direction on the TWEAK (younger group mean = 2.23 ± 1.5 vs older group mean = 3.10 ± 1.8 ; $t = -2.1$, $df = 68, P < 0.04$) and on the AUDIT (younger group mean = 5.02 ± 4.9 vs older group mean = 10.39 ± 8.0 ; $t = -4.0$, $df = 82.7, P < 0.001$). The groups did not differ significantly on the CAGE (younger group mean = 0.60 ± 0.85 vs older group mean = 0.84 ± 1.17 ; $t = -1.2$, $df = 90, P = 0.22$).

Table 4. Inter-item correlations for TWEAK

	1	2	3	4	5
TWEAK 1	—				
TWEAK 2	0.01	—			
TWEAK 3	0.12	0.36*	—		
TWEAK 4	0.08	0.17	0.17	—	
TWEAK 5	-0.01	0.43*	0.21	0.15	—

$n = 70$, but 33 cases were excluded due to missing data on Item 1 (Tolerance).

* $P < 0.05$ (two-tailed test).

Table 5. Item distributions, means^a and corrected item-total correlations for the CAGE scale

Item	Score (n)			Item-total correlation
	0	1	Mean \pm SD	
1. Have you ever felt the need to cut down on your drinking? (Cut down)	71	32	0.31 \pm 0.47	0.60
2. Have you felt annoyed by criticism of your drinking? (Annoyed)	92	11	0.11 \pm 0.31	0.51
3. Do you ever feel guilty about your drinking? (Guilty)	75	28	0.27 \pm 0.45	0.47
4. Do you ever need a drink first thing in the morning to get going? (Eye-opener)	101	2	0.002 \pm 0.14	0.34
Total scale score			0.72 \pm 1.03	

$n = 103$.

^aMeans based on scoring 0 for ‘no’ responses; 1 for ‘yes’ responses to all items. Total scale scores range from 0 to 4.

Gender differences. Males and females did not differ on the AUDIT or the CAGE. They did differ significantly only on the TWEAK (male mean = 3.9 ± 1.5 vs female mean = 2.1 ± 1.8 ; $t = -2.8$, $df = 67$, $P < 0.009$).

Ethnic/racial differences. Caucasian adolescents ($n = 82$) were compared to African American adolescents ($n = 15$) on each screening measure (other adolescents were not included). The two racial groups differed significantly only on the AUDIT, with Caucasian adolescents scoring higher than African-American adolescents (Caucasian mean = 8.56 ± 7.6 versus African-American mean = 5.07 ± 3.9 ; $t = 2.7$, $df = 37$, $P < 0.02$).

Criterion validity

Hazardous drinking. In order to assess potentially hazardous drinking, the participants were assigned to groups based on whether they drank six or more drinks on one occasion 'less than once a month' versus more frequently. Bohn *et al.* (1995) used a similar 'binge drinking' criterion with adults, and defined hazardous drinking as drinking more than six drinks on one occasion weekly or more often. Given the younger age and greater inexperience with alcohol among adolescents, it seemed reasonable to set the criterion here at 'monthly or more often'.

Based on the specified criterion, there were 73 'non-hazardous' drinkers and 30 'hazardous' drinkers in the AUDIT and CAGE analyses, and 42 'non-hazardous' drinkers and 28 'hazardous' drinkers in the TWEAK analysis. All three screening instruments significantly discriminated between potentially hazardous drinkers and non-hazardous drinkers (AUDIT, mean for hazardous drinkers = 15.9 ± 7.3 vs 4.3 ± 3.2 for non-hazardous drinkers, $t = -8.4$, $df = 33.8$, $P < 0.001$; TWEAK, mean for hazardous drinkers = 3.93 ± 1.5 vs 1.93 ± 1.3 for non-hazardous drinkers, $t = -5.8$, $df = 68$, $P < 0.001$; CAGE, mean for hazardous drinkers = 1.1 ± 1.3 vs 0.56 ± 0.85 for non-hazardous drinkers, $t = -2.1$, $df = 39.6$, $P < 0.05$). The criterion for hazardous drinking is also the third AUDIT item and may have inflated differences on this scale. Therefore, we removed this item and computed a 9-item AUDIT score to compare the groups. Hazardous drinkers were still higher on the AUDIT: mean = 14.3 ± 6.4 vs 3.8 ± 2.8 for non-hazardous drinkers ($t = -8.7$, $df = 33.8$, $P < 0.001$).

Alcohol use prior to presentation in the ED. Eleven participants (11%) scored positively on the alcohol breathalyser test. The alcohol breathalyser score (BrAC) represents a criterion for problematic alcohol use that is external to the scores the participants received on the screening instruments. The breathalyser-positive participants were compared to the breathalyser-negative participants on the screening instruments, and were found to differ significantly only on the AUDIT (BrAC positive mean \pm SD: 15.6 ± 9.6 versus BrAC negative: 6.8 ± 6.2 ; $t = -2.96$, $df = 11.04$, $P < 0.02$; two cases were missing from this analysis, $n = 101$).

DISCUSSION

The findings in this study indicate that alcohol screening instruments previously validated with adults vary in their performance with an older adolescent sample. The primary objective of this study was to determine the utility of the instruments as assessments for problematic alcohol use in

adolescents treated in EDs. In this study, the AUDIT displayed high internal consistency and demonstrated validity for differentiating alcohol users along 'known group' lines: Caucasians scored higher on the AUDIT than African-Americans, and older adolescents scored higher than younger adolescents. The AUDIT also discriminated between hazardous and non-hazardous drinking and was the only screening instrument in this study that differentiated alcohol-positive versus alcohol-negative adolescents based on breathalyser readings in the ED.

Utilizing the same screening instruments, Chung *et al.* (2000) also found that the AUDIT performed best in screening for problematic alcohol use among adolescents treated in an emergency department. These investigators used DSM-defined alcohol use disorders as the criterion measure for assessing the performance of these instruments. However, Chung *et al.* (2000) studied only alcohol-negative adolescents. Our method included screening for alcohol use at the time of the participant's presentation at the emergency department as an additional criterion for problematic alcohol use. Our findings on this criterion contribute additional evidence that the AUDIT performs best as a screening instrument among adolescents seen in emergency departments.

The TWEAK was too low on internal consistency to be considered an adequate measure of problematic alcohol use in the current sample. However, if the tolerance item is removed, internal consistency for the TWEAK increases to an $\alpha = 0.62$. Chung *et al.* (2000) pointed out that tolerance is quite variable and it is, therefore, difficult to develop a definition of tolerance that is appropriate for both adults and adolescents. These investigators concluded that an item that assesses high-volume drinking, e.g. frequency of drinking five or more drinks, may be preferable in this population to one that assesses tolerance.

The TWEAK was, however, the only instrument that differentiated male and female drinkers, with males scoring higher. Interestingly, the highest degree of variance between the genders on the TWEAK was on item 1, the tolerance item. Sixty-five per cent of the males who responded to this item indicated that they can hold six or more drinks without passing out, whereas only 39% of the females gave this response. However, many adolescents responded that they did not have enough experience with alcohol to know how to answer the tolerance item. Thus, the tolerance item may work well in discriminating between the genders, but only among those participants who know how to answer it.

The CAGE did not perform well as a screening measure for problematic alcohol use in our adolescent/young adult sample. This finding is consistent with that by Chung *et al.* (2000). Heck and Williams (1995) similarly found the CAGE to be a poor predictor of problem drinking in college students. They suggested that its emphasis on items indicative of alcohol dependence may have been responsible for its poor performance in their relatively young college sample. Fleming *et al.* (1991), on the other hand, found that the AUDIT had good internal consistency ($\alpha = 0.80$) among college students, and that it correctly classified 78% of those diagnosed with alcohol misuse by DSM-III criteria. These investigators suggested that the AUDIT has advantages over the CAGE, because it assesses frequency of use and quantity and frequency of binge drinking along a continuum, which is an advantage for the early detection of drinking problems.

Our assessment of concurrent validity may have been affected by the use of some of the same items for constructing the various scales. Notwithstanding this, the analysis indicated that all the instruments were measuring alcohol misuse, but the varying sizes of the correlations suggest that the instruments are not measuring the same underlying construct(s). The high correlation between the AUDIT and the TWEAK may be related to increased variance in the scores, due to their larger scales (AUDIT range 1–40; TWEAK range 0–7). Furthermore, the AUDIT measures alcohol intake and the tolerance item on the TWEAK may be a proxy for intake, since it is based on number of drinks. In contrast, the CAGE focuses on items related to alcohol dependence and this may explain its lower correlation with the other instruments.

Cherpitel and Clark (1995) studied these instruments in an adult ED sample and found that a significantly higher percentage of African-American males scored positively on all three measures compared to African-American females. The small number of participants who reported any alcohol use in our study precluded testing for gender differences within the ethnic/racial groups, but we did find that African-Americans scored lower on the AUDIT. Consistent with our findings, Clements (1998) found that only the AUDIT differentiated between the races in a college sample, with Caucasians scoring higher than African-Americans. Our findings are also in line with a recent epidemiological study (Grant, 1997) that reported lower rates of alcohol misuse and dependence among African-Americans.

The present findings should not be generalized beyond adolescents treated in EDs. Regional variations in alcohol use which affect samples and results have been reported from one ED to another in adult samples (Cherpitel and Clark, 1995). However, it is not clear that geographic region impacts our findings, since Chung *et al.* (2000) reported similar findings in an ED-treated sample of adolescents from a different region of the country. The fact that the current sample was primarily Caucasian requires us to be cautious in applying our findings to African-American or other race participants.

It is expected that level of illness and injury will affect the sample in studies such as ours, and we were unable to approach as many patients at the adult ED due to serious illness or injury. It is possible that more seriously ill or injured adolescent patients are treated at adult facilities. As a result, the AUDIT, the TWEAK, and the CAGE may perform differently from what was found in the current study, if used with critically injured adolescents.

Limitations notwithstanding, we were able to examine the performance of three widely used alcohol screening instruments in a sample of ED-treated adolescents. The success of prevention and treatment efforts relies on the early identification of substance use problems (Winters *et al.*, 1993). This is a population that is at high risk for problematic alcohol use and a setting in which adolescents are likely to present first for treatment of alcohol-related illness or injury. Therefore, the potential for early detection of problematic alcohol use among ED-treated adolescents is great and this can have a positive impact on prevention and treatment efforts.

Experienced investigators have also noted that the population under investigation is an important consideration in choosing the appropriate instrument (Maisto *et al.*, 1995; Cherpitel, 1998). The higher internal consistency for the AUDIT, as compared

to the TWEAK and CAGE, as well as its more consistent performance in differentiating alcohol-using adolescents along 'known group' lines and in discriminating between problematic and non-problematic alcohol users, lead us to conclude that the AUDIT performs best as an alcohol screening instrument among adolescents treated in EDs. The consistency of findings between our study and others, especially those from other EDs and studies of samples close in age to our subjects (e.g. college students) adds to the validity of our findings and supports this conclusion.

Acknowledgements — The authors wish to express their appreciation to Drs Duncan Clark, Oscar Bukstein and Kevin Lynch for their assistance in the preparation of this manuscript. This work was supported by the National Institute on Alcohol Abuse and Alcoholism (NIAAA) Grant No. P50 AA08746 and a NIAAA Mentored Clinical Scientist Development Award to the first author (Grant No. K08-AA00280).

REFERENCES

- American Psychiatric Association (1994) *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn. American Psychiatric Association, Washington, DC.
- Arria, A. M., Dohey, M. A., Mezzich, A. C., Bukstein, O. G. and Van Theil, D. H. (1995) Self-reported health problems and physical symptomatology in adolescent alcohol abusers. *Journal of Adolescent Health* **16**, 226–231.
- Barnes, G. M. and Welte, J. W. (1986) Adolescent alcohol abuse: subgroup differences and relationships to other problem behaviors. *Journal of Adolescent Research* **1**, 79–94.
- Bastiaens L., Francis, G. and Lewis, K. (2000) The RAFFT as a screening tool for adolescent substance use disorders. *American Journal of Addictions* **9**, 10–16.
- Blum, R. (1997) Adolescent substance use and abuse. *Archives of Pediatric and Adolescent Medicine* **151**, 805–808.
- Bohn, M. J., Babor, T. F. and Kranzler, H. R. (1995) The alcohol use disorders identification test (AUDIT): validation of a screening instrument for use in medical settings. *Journal of Studies on Alcohol* **56**, 423–432.
- Centers for Disease Control and Prevention (1996) *CDC Surveillance Summaries. Youth Risk Behavior Surveillance — United States, 1995*. September 27, 1996. *MMWR Morbidity and Mortality Weekly Report* 1996, **45** (No. SS-4).
- Cherpitel, C. J. (1998) Performance of screening instruments for identifying alcohol dependence in the general population, compared with clinical populations. *Alcoholism: Clinical and Experimental Research* **22**, 1399–1404.
- Cherpitel, C. J. and Clark, W. B. (1995) Ethnic differences in performance of screening instruments for identifying harmful drinking and alcohol dependence in the emergency room. *Alcoholism: Clinical and Experimental Research* **19**, 628–634.
- Chung, T., Colby, S., Barnett, N., Rohsenow, D., Spirito, A., and Monti, P. (2000) Screening adolescents for problem drinking: performance of brief screens against DSM-IV alcohol diagnoses. *Journal of Studies on Alcohol* **61**, 579–587.
- Clark, D. B., Lesnick, L. A. and Hegedus, A. M. (1997) Traumas and other adverse life events in adolescents with alcohol abuse and dependence. *Journal of the American Academy of Child and Adolescent Psychiatry* **36**, 1744–1751.
- Clements, R. (1998) A critical evaluation of several alcohol screening instruments using the CIDI-SAM as a criterion measure. *Alcoholism: Clinical and Experimental Research* **22**, 985–993.
- Conigrave, K. M., Hall, W. D. and Saunders, J. B. (1995) The AUDIT questionnaire: choosing a cut-off score. *Addiction* **90**, 1349–1356.
- Cronbach, L. J. (1951) Coefficient alpha and the internal structure of tests. *Psychometrika* **16**, 297–334.
- Ewing, J. A. (1984) Detecting alcoholism — the CAGE questionnaire. *Journal of the American Medical Association* **252**, 1905–1907.

- Fleming, M. F., Barry, K. L. and MacDonald, R. (1991) The Alcohol Use Disorders Identification Test (AUDIT) in a college sample. *International Journal of the Addictions* **26**, 1173–1185.
- Grant, B. (1997) Prevalence and correlates of alcohol use and DSM-IV alcohol dependence in the United States: results of the National Longitudinal Alcohol Epidemiologic Survey. *Journal of Studies on Alcohol* **58**, 464–473.
- Hansell, S. and White, H. W. (1991) Adolescent drug use, psychological distress and physical symptoms. *Journal of Health and Social Behavior* **32**, 288–301.
- Heck, E. J. and Williams, M. D. (1995) Using the CAGE to screen for drinking-related problems in college students. *Journal of Studies on Alcohol* **56**, 282–286.
- Jennet, B. and Bond, M. (1975) Assessment of outcome after severe brain damage. *Lancet* **i**, 480–484.
- Jessor, R. (1987) Problem-behaviour theory, psychosocial development, and adolescent problem drinking. *British Journal of Addiction* **82**, 331–342.
- Kelly, T. M., Donovan, J. E., Lynch, K. G. and Clark, D. B. (2001) Alcohol use disorders and other risk factors for adolescent suicidal ideation and attempts. *Suicide and Life-Threatening Behavior* **31**, 181–193.
- Knight, J. R., Shrier, L. A., Bravender, T. D., Farrell, M., Vander Bilt, J. and Shaffer, H. J. (1999) A new brief screen for adolescent substance abuse. *Archives of Pediatric and Adolescent Medicine* **153**, 591–596.
- Leccese, M. and Waldron, H. B. (1994) Assessing adolescent substance use: a critique of current measurement instruments. *Journal of Substance Abuse Treatment* **11**, 553–563.
- Maisto, S. A., Connors, G. J. and Allen, J. P. (1995) Contrasting self-report screens for alcohol problems: a review. *Alcoholism: Clinical and Experimental Research* **19**, 1510–1516.
- Nilssen, O., Ries, R. K., Rivara, F. P., Gurney, J. G. and Jurkovich, G. J. (1994) The CAGE questionnaire and The Short Michigan Alcohol Screening Test in trauma patients: comparison of their correlations with biological alcohol markers. *Journal of Trauma* **36**, 784–788.
- Rachal, J. V., Guess, L. L., Hubbard, R. L. and Maisto, S. A. (1982) Facts for planning no. 4: alcohol misuse by adolescents. *Alcohol Health and Research World* **6**, 61–68.
- Russell, M. (1994) New assessment tools for risk drinking during pregnancy: T-ACE, TWEAK, and others. *Alcohol Health and Research World* **18**, 55–61.
- SAMHSA (1997) *National Household Survey on Drug Abuse Main Findings 1995*. US Department of Health and Human Services, Rockville, MD.
- Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R. and Grant, M. (1993) Development of the alcohol use disorders identification test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption — II. *Addiction* **88**, 791–804.
- Shope, J. T., Copeland, L. A., Male, R. F., Blow, F. C. and Brockmann, L. M. (1997) Correlates of alcohol involvement among injured adolescents (abstract). *Alcoholism: Clinical and Experimental Research* **21** (Suppl. to no. 3), 29A.
- Spain, D. A., Boaz, P. W., Davidson, D. J., Miller, F. B., Carrillo, E. H. and Richardson, J. D. (1997) Risk-taking behaviors among adolescent trauma patients. *Journal of Trauma: Injury, Infection and Critical Care* **43**, 423–426.
- Spirito, A., Lewander, W. J., Levy, S., Kurkjian, J. and Fritz, G. (1994) Emergency department assessment of adolescent suicide attempters: factors related to short-term follow-up outcome. *Pediatric Emergency Care*, **10**, 6–12.
- Windle, M. (1996) An alcohol involvement typology for adolescents: convergent validity and longitudinal stability. *Journal of Studies on Alcohol* **57**, 627–637.
- Winters, K. C., Stinchfield, R. D. and Henly, G. A. (1993) Further validation of new scales measuring adolescent alcohol and other drug abuse. *Journal of Studies on Alcohol* **54**, 534–541.