Should the CRAFFT Substance Abuse Screening Test Be Shortened?

John R. Knight, MD, Lon Sherritt, MPH, Erin A. Gates, BA, and Sion Kim Harris, PhD

Abstract

• **Objective:** To determine if the 6-item CRAFFT substance abuse screening test can be shortened without sacrificing important psychometric properties.

• **Design:** Secondary analysis of data from a criterion standard validation study.

• **Setting and participants:** 538 patients aged 14 to 18 years arriving for routine medical care at a large, urban, hospital-based adolescent clinic.

• **Measurements:** The 6-item CRAFFT and a structured psychiatric diagnostic interview were administered. We computed the receiver-operating-characteristic (ROC) curve area for all possible combinations of the 6 CRAFFT items, selected the shorter version with the highest ROC curve area, and then computed sensitivity, specificity, predictive values, and accuracy for all shorter versions.

• **Results:** Accuracy of all models was significantly higher at a scoring cut-point of 2 compared to 1. Sensitivity of the 2-item (0.45 [95% confidence interval (CI), 0.35–0.56]) and 3-item (0.59 [95% CI, 0.48–0.69]) versions was significantly lower compared with the 6-item CRAFFT (0.80 [95% CI, 0.71–0.89]). Psychometric properties of the 4-, 5-, and 6-item versions did not differ significantly, although the 5-item version had more stable properties across demographic subgroups.

• **Conclusions:** The CRAFFT screening test could be shortened to 5 items (ie, CRAFT) without significant loss of psychometric properties, which may make it more practical and easier to remember.

Usage of alcohol and drugs by adolescents continues to be a serious problem. According to the 2001 Monitoring the Future study, 4 of 5 students have consumed alcohol by the end of high school, and more than half have done so by eighth grade [1]. By the end of high school, nearly two thirds of students have gotten drunk, and more than half have used an illicit drug at least once [1]. More than 30% of adolescents smoke marijuana, 14% drive after drinking alcohol, and 33% report riding with a driver who had been drinking [2]. Serious health risks are increasing as more youth are using illicit drugs with severe consequences, such as cocaine and 3,4-methylenedioxyamphetamine (MDMA or “Ecstasy”) [2].

More than 70% of all deaths among adolescents are due to accidents, homicides, and suicides, and many of these are linked to the use of alcohol and drugs [2]. More than 35% of fatal motor vehicle crashes involving drivers aged 15 to 20 years are related to drinking, and almost 25% involve a legally intoxicated driver [3]. Alcohol use is a factor in over one third of all suicides and almost two thirds of homicides [4]. Alcohol and drug use also are associated with serious but nonlethal health risks and problems, such as suicide attempts, violence, weapon carrying, aggressive behaviors, and high-risk sexual behaviors [5–11].

The clinician’s office provides an ideal setting for screening and early intervention for alcohol and drug use in conjunction with the evaluation of other mental and physical health problems. It places substance use in a health context and provides adolescents with a reasonable assurance of confidentiality. The American Medical Association guidelines for adolescent preventive services recommend that health care providers ask all adolescent patients about their use of alcohol and drugs yearly and further assess those who report any use [12]. In order for screening to be widely implemented, however, health care providers need a screening tool that is valid, reliable, developmentally appropriate for adolescents, and practical for use in busy medical offices. The ideal screen must be brief and easy to administer, score, and remember as well as be accurate across age-group, race, and sex [13]. The CRAFFT test meets these criteria. It consists of 6 orally administered yes/no questions that are easy to score (each “yes” answer = 1). Key words in the test’s 6 items form the mnemonic CRAFFT (Table 1). These items were designed to maximize sensitivity (ie, include the word “ever”) and to identify teens who engage in high-risk behaviors (eg, driving/riding while...

---

*From the Harvard Medical School, Boston, MA (Dr. Knight, Mr. Sherritt, and Dr. Harris) and the Children’s Hospital Boston, Boston, MA (Dr. Knight, Mr. Sherritt, Ms. Gates, and Dr. Harris).*

www.turner-white.com
Do you ever use alcohol or drugs while you are by yourself, or fit in?

Do you ever use alcohol or drugs to Relax, feel better about yourself, or fit in?

Do you ever Forget things you did while using alcohol or drugs?

Do your Family or friends ever tell you that you should cut down on your drinking or drug use?

Have you ever gotten into Trouble while you were using alcohol or drugs?

intoxicated) as well as those who manifest early signs of substance-related disorders (eg, social and interpersonal problems due to substance use) [14].

A CRAFFT score of 2 or higher has been shown to have good sensitivity and specificity for identifying substance-related problems and disorders in a general population of adolescent medical clinic patients [15]. Other studies found the CRAFFT test comparable to the Problem-Oriented Screening Instrument for Teenagers (POSIT) [16] and the Alcohol Use Disorders Identification Test (AUDIT) [17] and superior to the CAGE questionnaire [18] in identifying adolescents with alcohol problems [19]. However, CAGE may have one advantage over CRAFFT: it has 4 items instead of 6. The 2 items make little difference in the time required for administration or scoring, but it may make CRAFFT more difficult to remember for providers who wish to administer it orally without a written prompt. A shorter CRAFFT would have clear advantages in terms of clinical utility. The objective of this study was to determine if the CRAFFT test could be shortened without sacrificing any of its key psychometric characteristics. A secondary aim was to determine the validity of the CRAFFT without the C question for providers who serve populations of youth with limited access to cars.

Methods

Patients and Setting

This study was a secondary analysis of data collected during the CRAFFT validation study; a detailed description of the methods have been previously published [15]. The validation study was a criterion standard study conducted between March 1999 and September 2000 at a large, urban, hospital-based adolescent clinic. Health care providers invited patients aged 14 to 18 years to participate at the conclusion of the medical visit. A research assistant met with those who were interested, obtained informed assent, and administered an assessment battery. The Children’s Hospital Boston Committee on Clinical Investigation (institutional re-view board) waived the requirement for parental consent in accordance with current guidelines [20,21].

Assessment Battery

The study assessment battery included the 6 CRAFFT questions and the Adolescent Diagnostic Interview (ADI) [22]. The ADI is a structured interview with an administration time of 30 to 90 minutes that provides alcohol- and drug-related diagnoses according to the Diagnostic and Statistical Manual of Mental Disorders, 4th edition [23]. The ADI has been shown to have good reliability and validity among adolescents [24,25].

Analyses

Based on the ADI interview, participants were classified into 2 mutually exclusive categories: (1) no diagnosis or (2) any diagnosis (ie, abuse or dependence). For each of the 6 CRAFFT questions, we calculated frequencies of the participants’ positive responses and the following psychometric properties: the standardized alpha-if-item-deleted, sensitivity (ie, proportion of true positives identified by the screen), specificity (ie, proportion of true negatives identified by the screen), receiver-operating-characteristic (ROC) curve area, and the odds ratio (OR) with 95% confidence interval (CI) for the any diagnosis classification.

To produce shorter versions of the screen, we computed the ROC curve area of all possible combinations of the original 6 items and then selected the combination with the greatest ROC curve area as the optimal shorter version. We also performed principal components factor analysis of the 6-item CRAFFT to identify items that may be redundant and to ensure that no selected shorter version resulted in elimination of an important construct. Because dichotomous variables may yield distorted correlation coefficients (used in factor analyses), we also performed cluster analysis based on squared Euclidean distances. For each shorter test version, we calculated psychometric properties when total scores of 1 and 2 were used to indicate a positive screen; these score cut-points have previously been shown to be optimal [15,19]. We also calculated psychometric properties for a shorter version that did not include the Car question.

We computed sensitivity, specificity, positive predictive value, negative predictive value, and accuracy (ie, proportion of all cases identified correctly by the screen) for each shorter version. We calculated the area under the ROC curve, sensitivity, specificity, and accuracy of the 4-, 5-, and 6-item screens for sex, age, and race/ethnicity subgroups. In order to preserve adequate cell sizes for these analyses, we transformed participant age into a dichotomous variable (ie, younger youth and older youth) based on the sample median (16.7 years). We used the bootstrap technique to estimate 95% CIs [26–28]. Data were analyzed in SPSS v10.0 (SPSS Inc., Chicago, IL).
Results

538 of 670 eligible patients (80.3%) agreed to participate; 2 were excluded from analysis due to missing data. Participants were nearly equally distributed across years of age, and 68.4% were female. The racial/ethnicity proportions were as follows: black non-Hispanic, 50.6%; white non-Hispanic, 24.2%; Hispanic, 18.8%; and Asian/Other, 6.5%. Approximately one half of the study participants had used alcohol or drugs during the past year, and more than one quarter had experienced alcohol- or drug-related problems. Refusers did not differ significantly from study participants in age, sex, race/ethnicity, or clinician impressions of substance use. The study sample was also representative of the primary care clinic population at large in age and race/ethnicity, but it included a significantly greater proportion of females (68.4% versus 59.4%, P < 0.001).

The Car question had a higher affirmative endorsement rate (42.6%) than any other single item (Table 2). It also had significantly higher sensitivity and lower specificity compared to all other items. The Car question was the first to form a cluster in factor analyses, which indicated that it may be redundant with the Alone item. In cluster analyses, the Car and Alone items loaded on a single factor separate from the other items. Accuracy of all optimal models was significantly higher at a cut-point of 2 compared with a cut-point of 1. We therefore dropped the 1 cut-point from further consideration. Accuracy, specificity, and predictive values did not differ significantly among optimal versions at a cut-point of 2 (Table 3). However, sensitivity was significantly lower for the 2-item and 3-item versions compared with the 6-item CRAFFT. We therefore dropped these 2 shorter versions from further consideration.

Psychometric characteristics of the 4-, 5- and 6-item versions for sex, age, and race/ethnicity subgroups are presented in Table 4. This study is relatively underpowered for examining these differences with statistical precision (ie, 95% CIs are wide and overlapping), but certain differences (eg, 0.10 or greater) are strongly suggestive. ROC curve areas appeared virtually identical across sex, age, and race/ethnicity for all 3 versions. The 6- and 4-item versions appeared to have higher sensitivity in boys compared with girls, and all versions appeared to have higher sensitivity in older youth compared.

### Table 2. Positive Responses to and Psychometric Characteristics of Individual CRAFFT Items

<table>
<thead>
<tr>
<th>Item</th>
<th>Prevalence, %</th>
<th>Alpha-if-Item-Deleted</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>ROC Area (95% CI)</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Car</td>
<td>42.6</td>
<td>0.66</td>
<td>0.86 (0.79–0.94)</td>
<td>0.66 (0.62–0.70)</td>
<td>0.76 (0.71–0.81)</td>
<td>5.46 (2.62–11.41)</td>
</tr>
<tr>
<td>Relax</td>
<td>15.6</td>
<td>0.62</td>
<td>0.51 (0.41–0.62)</td>
<td>0.91 (0.88–0.93)</td>
<td>0.71 (0.64–0.78)</td>
<td>3.98 (2.03–7.81)</td>
</tr>
<tr>
<td>Alone</td>
<td>10.8</td>
<td>0.62</td>
<td>0.40 (0.30–0.50)</td>
<td>0.95 (0.93–0.97)</td>
<td>0.68 (0.61–0.78)</td>
<td>3.32 (1.48–7.42)</td>
</tr>
<tr>
<td>Forget (F1)</td>
<td>12.3</td>
<td>0.62</td>
<td>0.44 (0.33–0.54)</td>
<td>0.94 (0.91–0.96)</td>
<td>0.69 (0.62–0.76)</td>
<td>3.07 (1.48–6.35)</td>
</tr>
<tr>
<td>Family (F2)</td>
<td>8.4</td>
<td>0.65</td>
<td>0.31 (0.22–0.41)</td>
<td>0.96 (0.94–0.98)</td>
<td>0.64 (0.56–0.71)</td>
<td>3.47 (1.45–8.29)</td>
</tr>
<tr>
<td>Trouble</td>
<td>10.6</td>
<td>0.64</td>
<td>0.38 (0.28–0.48)</td>
<td>0.95 (0.92–0.97)</td>
<td>0.66 (0.59–0.73)</td>
<td>3.37 (1.55–7.31)</td>
</tr>
</tbody>
</table>

CI = confidence interval; ROC = receiver-operating-characteristic.

### Table 3. Psychometric Characteristics of CRAFFT and Shortened Versions (Cut-point = 2)

<table>
<thead>
<tr>
<th>Test</th>
<th>Sensitivity (95% CI)</th>
<th>Specificity (95% CI)</th>
<th>Positive Predictive Value (95% CI)</th>
<th>Negative Predictive Value (95% CI)</th>
<th>Accuracy (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CRA,F,F,T</td>
<td>0.80 (0.71–0.89)</td>
<td>0.87 (0.83–0.90)</td>
<td>0.54 (0.45–0.63)</td>
<td>0.96 (0.94–0.98)</td>
<td>0.86 (0.82–0.89)</td>
</tr>
<tr>
<td>CRA,F,T</td>
<td>0.78 (0.69–0.86)</td>
<td>0.88 (0.85–0.91)</td>
<td>0.55 (0.46–0.64)</td>
<td>0.95 (0.93–0.97)</td>
<td>0.86 (0.83–0.89)</td>
</tr>
<tr>
<td>CRA,T</td>
<td>0.70 (0.61–0.80)</td>
<td>0.90 (0.87–0.93)</td>
<td>0.58 (0.48–0.67)</td>
<td>0.94 (0.92–0.96)</td>
<td>0.87 (0.84–0.90)</td>
</tr>
<tr>
<td>CRF,</td>
<td>0.59 (0.48–0.69)</td>
<td>0.93 (0.90–0.95)</td>
<td>0.61 (0.50–0.71)</td>
<td>0.92 (0.90–0.95)</td>
<td>0.87 (0.84–0.90)</td>
</tr>
<tr>
<td>CR</td>
<td>0.45 (0.35–0.56)</td>
<td>0.96 (0.93–0.97)</td>
<td>0.66 (0.53–0.78)</td>
<td>0.90 (0.87–0.93)</td>
<td>0.87 (0.84–0.90)</td>
</tr>
</tbody>
</table>

CI = confidence interval.
with younger youth. Differences were small among race/ethnicity subgroups, although sensitivity appeared to diminish substantially for white adolescents for the 4-item version only.

Although the 5-item RAFFT was not identified as an optimal shorter version, we separately computed its psychometric properties (Table 5). For this version intended for youth without access to motor vehicles, a cut-point of 1 yielded significantly higher sensitivity compared with a cut-point of 2 (0.85 versus 0.60, P < 0.05). The psychometric properties of this version did not differ from the 6-item CRAFFT (with a cut-point of 2), although it appeared to be somewhat less specific among boys (0.78 versus 0.90, P = 0.05).

**Discussion**

This study shows that the CRAFFT test could be shortened to 5 items by deleting the family question without adversely affecting its psychometric characteristics. The 4-item version did not differ significantly from the 6-item CRAFFT, but changes in properties among demographic subgroups suggest that further testing is needed before it is widely recommended.

Clinical relevance of individual items and issues of practical utility should be considered along with psychometric data in making decisions about shortening or changing scales. For example, we found that the Car question, which screens for the very serious risk of driving under the influence of alcohol...
or drugs, had both the highest frequency of positive endorsement and the greatest sensitivity for identifying alcohol and drug disorders. We were somewhat surprised by the latter finding, as the content of this question relates to risk of alcohol-related car crashes rather than to pathology. Some adolescents will endorse this item even when the impaired driver in question is a parent, older sibling, or other family member.

We found that accuracy was significantly improved by using a score of 2 or more to indicate a positive screening as opposed to using a cut-point of 1. This may be due at least in part to the high sensitivity (ie, 86%) and moderately low specificity (ie, 66%) of the Car question alone. Most adolescents screening positive at a cut-point of 1 do so based on a positive response to the Car question only, so that accuracy changes little with addition of other items. A cut-point of 2, in contrast, raises accuracy significantly by substantially improving the proportion of correctly identified negative cases (ie, specificity), which likely are those youth who report driving/riding after drinking but not at least 1 additional risk factor.

There are no set rules on how high sensitivity or specificity must be for a screening test to be considered valid. However, for alcohol and drug use, where the risk posed by a false-positive screen is low (eg, further assessment), we believe that a screening test should correctly identify no less than 70% of those truly at risk. Both the 4- and 5-item versions meet this requirement, and no other psychometric characteristics were significantly lower for these 2 shorter versions compared with the 6-item CRAFFT. However, the lower limit of 95% confidence for sensitivity falls below the 70% level for both shorter versions. Further testing in a larger sample, which has greater power to detect significant differences, may therefore be warranted.

Demographic subgroup analyses suggest that the 5-item model may possess the most stable criterion validity across subgroups because its sensitivity was somewhat less affected by sex and race/ethnicity. The 4-item screen had poor sensitivity (ie, 59%) among white adolescents in this study, although our design was relatively underpowered to fully examine subgroup differences. Future studies should address this limitation and also test additional questions aimed at improving sensitivity among girls, who account for a majority of medical clinic visits.

Some providers may wish to administer a screen that does not include a question about driving/riding after drinking risk. This might be appropriate when screening adolescents who live in very large cities where most families do not own motor vehicles (eg, New York City), incarcerated youth, or those who live in underdeveloped countries. The 5-item RAFFT can be recommended for this situation, although the test’s cut-point should be lowered to 1. We would urge caution, however, in using this version among adolescents who do not drive (eg, younger adolescents) but still have access to motor vehicles driven by others. One does not have to be a driver to be killed in an alcohol-related motor vehicle crash.

This study was conducted in a primary care outpatient clinic among a group of general adolescent patients, and it used a valid and reliable diagnostic interview as the criterion standard. However, our approach also had limitations, the greatest of which involved power. The size of the study sample was chosen to assure that we would have adequate ability to ascertain that the sensitivity of the CRAFFT was sufficiently high among all adolescents. Power was not adequate to fully examine smaller subgroup differences. In addition, this study, like many others, relied on adolescents self-report. The extent to which some adolescents may have under-reported or over-reported their use of substances is unknown. However, prior studies have shown that self-report compares favorably with other means of substance abuse detection [29,30]. Finally, the study was conducted in a single, urban, adolescent clinic. The generalizability of our findings to other adolescent samples (eg, rural youth, small private office practices, psychiatric patients) is therefore unknown. More studies of the CRAFFT are needed to both attempt to replicate these findings and determine their durability.

We conclude that the CRAFFT test could be shortened to 5 items without sacrificing internal consistency or criterion validity. Serendipitously, the new acronym retains its mnemonic sound (ie, CRAFT). This may be an important consideration, since the test is widely known by this name, and it is so listed in a number of policy statements and guidelines [31–33]. Retaining the mnemonic also may be of benefit to those clinicians who wish to administer the screen orally, without a written prompt. We do not recommend that the test be shortened further. We believe that the 5-item test may present a small but important advantage, compared with the
The CRAFFT SCREENING TEST

6-item version, for busy primary care providers who now need only remember one “F” in the spelling of the test and in its mnemonic acronym.

The authors thank Betsy Gates, BA, Sarah Rosenberg, BA, Katherine O’Connor, BA, and Allison Arnell, MA, for their assistance in study implementation; Lydia Shrier, MD, MPH, and the clinicians and staff of the Adolescent/Young Adult Medical Practice at Children’s Hospital Boston for assistance in recruitment; Ken C. Winters, PhD, for consultation on the study measurement battery; and S. Jean Emans, MD, for review of the manuscript.

Corresponding author: John R. Knight, MD, Director, Center for Adolescent Substance Abuse Research, Children’s Hospital Boston, 300 Longwood Ave., Boston, MA 02115, john.knight@tch.harvard.edu.

Financial disclosures: None.

Author contributions: conception and design, JRK, LS, EAG, SKH; analysis and interpretation of data, JRK, LS, EAG, SKH; drafting of the article, JRK, LS, EAG, SKH; provision of study materials or patients, JRK; statistical expertise, LS, SKH; obtaining of funding, JRK; administrative, technical, or logistic support, JRK, LS; collection and assembly of data, EAG.

References


