Readability of Self-Report Alcohol Misuse Measures

R. KATHRYN McHUGH, PH.D., a,b,* DAWN E. SUGARMAN, PH.D., a,b JULIA S. KAUFMAN, B.A., c SARA PARK, M.ED., a ROGER D. WEISS, M.D., a,b AND SHELLY F. GREENFIELD, M.D., M.P.H., a,b

aDivision of Alcohol and Drug Abuse, McLean Hospital, Belmont, Massachusetts
bDepartment of Psychiatry, Harvard Medical School, Boston, Massachusetts
cDepartment of Psychology, Miami University, Oxford, Ohio

ABSTRACT. Objective: Self-report measures of alcohol misuse and alcohol use disorders are valuable assessment tools for both research and clinical practice settings. However, readability is often overlooked when establishing the validity of these measures, which may result in measures written at a reading-grade level that is higher than the ability level of many potential respondents. The aim of the current study was to estimate the reading-grade level of validated measures of alcohol misuse and associated problems. Method: A total of 45 measures were identified, and reading-grade level was calculated using three validated readability formulas. Results: The majority of measures were written above the recommended reading-grade level for patient materials (5th–6th grade), with particularly poor readability for measure instructions. Conclusion: Given that many self-report alcohol misuse measures are written at a high reading-grade level, the consideration of readability is important when selecting measures for use in research and practice settings. Moreover, the development or modification of measures to target low-literacy populations may facilitate the broader applicability of these instruments. (J. Stud. Alcohol Drugs, 75, 328–334, 2014)

CLINICAL AND RESEARCH PROGRAMS for alcohol use disorders often rely on self-report assessments to screen for disorders, evaluate symptoms, and assess treatment progress. Although the development of these measures entails the assessment of their validity relative to other measures and outcomes, their validity is also dependent on respondents’ ability to read and comprehend the text. It is estimated that approximately 90 million adults in the United States demonstrate English literacy skills in the bottom two of five skill levels defined by the National Adult Literacy Survey (Kirsch et al., 2002). Individuals with a mental or physical health condition are more likely than the population as a whole to have lower literacy scores (Kirsch et al., 2002). Moreover, health literacy, which is defined as “the degree to which individuals have the capacity to obtain, process, and understand basic health information and services needed to make appropriate health decisions” (Institute of Medicine, 2004, p. 32), is often even lower than general literacy (Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, 1999). Approximately 36% of adults in the United States have basic to below-basic health literacy (Kutner et al., 2006).

Few studies have examined the reading ability of patients in treatment programs for alcohol and other drug use disorders. Several early studies found that treatment seekers and clinical research participants with substance use disorders (SUDs) had an average reading-grade level of between 8th and 10th grade (Davis et al., 1993; Johnson et al., 1995). Davis and colleagues (1993) found that, on average, patients in public SUD treatment settings had significantly lower reading-grade levels than those in private settings (8th-grade level vs. 10th-grade level). Of note, this study also examined patient education materials and found that written materials ranged from a 12th-grade to postgraduate (18th-grade) level.

Several other studies also have investigated the readability of SUD patient education materials. A study examining Alcoholics Anonymous literature found that the materials ranged from 8th-grade to college reading-grade levels (Rathер and Murphy, 1995). Greenfield and colleagues (2005) tested the readability of patient handouts in a random sample of SUD treatment programs in the United States and found an average readability level of 11.8 (i.e., nearly equivalent to 12th grade). Studies by Khazaal and colleagues found that Internet-based information on alcohol (Khazaal et al., 2010) and cocaine dependence (Khazaal et al., 2008) was written at an average reading-grade level of 11.5 and 10.5, respectively. Thus, there is a large gap between the reading-grade level of patients and the readability of written materials that are provided in these programs. Moreover, in all of these studies, the average reading-grade levels of SUD patient materials were well above the 5th- to 6th-grade level that is typically recommended for patient materials (Weiss and Coyne, 1997).

Despite findings suggesting low average reading ability in SUD populations and high reading-grade levels of SUD treatment information, there have not been any published studies to date that have examined the readability of self-report assessments. Examinations of the readability
of self-report measures of other mental health symptoms have suggested that these measures are often written at a higher reading-grade level than recommended (Andrasik et al., 1981; Richards et al., 2013). For example, a study of five self-report clinical outcome measures showed that measures ranged from 6th-grade to college level (Beckman and Lueger, 1997). McHugh and Behar (2009) assessed the readability of self-report measures of depression and anxiety. On average, measures were written at a grade level higher than 7th grade for depression measures and higher than 8th grade for anxiety measures, with even higher grade levels for measure instructions. Although there was wide variability across measures (ranging from 6th- to 14th-grade level), few met recommendations for readability. Similar to the literature on SUD patient education materials, the literature on the readability of self-report measures in other areas of mental health suggests that many of these measures are written at reading-grade levels at or above the literacy level of the average adult in the United States, thus limiting their applicability to a large portion of the population.

The readability of alcohol self-report measures has important implications for treatment and research. Given the limitations of biological testing for alcohol consumption and associated symptoms (e.g., cost of tests that can detect alcohol use beyond very recent use), research and clinical practice rely heavily on self-report measures. Research findings are used to inform treatment on a broader scale, and self-report clinical measures are used to inform alcohol use disorder treatment and to monitor patient progress. Thus, data on the nature and treatment of alcohol use disorders and clinical decisions about treatment planning and progress are predicated on the validity of these assessments.

The overarching aim of the current study was to examine the readability of self-report measures of alcohol misuse that may be used in both clinical and research settings. We identified validated self-report measures and calculated reading-grade levels using standardized readability formulas for both measure instructions and items. Consistent with previous studies, we hypothesized that these measures would, on average, be written at a grade level higher than that recommended for patient materials (5th–6th grade).

Method

Measures were initially identified using several sources. First, large, publicly available Internet-based repositories of substance use measures were searched, such as the Center on Alcoholism, Substance Abuse, and Addictions website (http://casaa.unm.edu/inst.html). Second, a search of PubMed and PsychINFO databases was conducted for clinical trials of treatment for alcohol use disorders published during the previous 12 months to identify measures not captured in the first search. Finally, a panel of two expert SUD clinical researchers (SFG and RDW) reviewed the list of identified measures to determine any widely used measures that may have been overlooked.

The following criteria were used for selection of the initial pool of measures: (a) self-report format, (b) assessed symptoms of alcohol misuse, and (c) specific to alcohol. We chose to focus specifically on measures of symptoms of alcohol misuse that may be used to guide screening, assessment, and treatment planning and evaluation in clinical practice and research settings. Self-report measures are used to assess a wide variety of constructs related to alcohol misuse (e.g., treatment utilization, drinking expectancies, motives for drinking). For the purpose of this study and to maximize comparisons among measures, we selected measures that shared a focus on alcohol misuse symptoms specifically and not these related constructs. The examination of the readability of such measures is an important future research direction given their important role in research settings.

We restricted our analysis to measures that had support for validity and acceptability in the field, defined as the presence of both (a) at least one peer-reviewed validation paper with data on psychometric properties (broadly defined as any reliability and/or validity data) and (b) use by at least two independent investigator groups.

Based on these criteria, 45 measures of alcohol misuse were identified for analysis. We analyzed measure instructions and measure items separately, consistent with previous studies of readability (e.g., Andrasik et al., 1981; McHugh and Behar, 2009). All analyses were conducted using a computerized readability software package (Readability Calculations, Micro Power and Light Co.).

Readability assessment

A number of formulas have been used to examine readability of written text. Most formulas estimate reading-grade level based on the number of syllables per sentence or proportion of words with a certain number of syllables. We chose to use three different formulas and to calculate a readability score representing the average of these formulas, to minimize the limitations of any individual formula. The outcome reported is a reading-grade level.

The Flesch Reading Ease formula (Flesch, 1948) is a widely used readability formula that is based on average sentence and word length. The formula reports scores that range from 0 (unreadable) to 100 (very easy to read). These scores can be converted to an approximate reading-grade level. The Flesch formula has been validated against indices of reading comprehension (McCall and Crabbs, 1961) and correlates highly with other readability formulas (Meade and Smith, 1991).

The Simple Measure of Gobbledygook (SMOG) Readability Formula (McLaughlin, 1969) is based on the proportion of polysyllabic words (words with three or more syllables) in a text. The SMOG formula has been validated...
relative to measures of reading comprehension (Ley and Florio, 1996) and other readability formulas (Meade and Smith, 1991). Higher grade levels on the SMOG are also associated with poorer reading efficiency, defined as comprehension divided by the time needed to read a passage (McLaughlin, 1969).

The FORCAST Readability Formula (Kern et al., 1976) was developed for nonnarrative texts, such as questionnaires, forms, and lists (Zraick et al., 2012). It is calculated based on the number of monosyllabic words. This formula also has been validated with measures of reading comprehension (Hooke et al., 1979) and correlates highly with other readability formulas (Caylor et al., 1973).

All three readability formulas require full sentences for analysis. Thus, we included measures that consisted of full sentences and excluded those that did not. However, if a measure had a clearly implied full sentence (e.g., the use of a stem, such as “When I drink . . .”) followed by a list of phrases, full sentences were generated and the measure was analyzed ($n = 2$). For any measure that included full sentences as response options (instead of numerical or Likert-type scales), these responses were also analyzed as part of the measure items.

**Results**

Based on the criteria listed above, 40 measures had analyzable item text, of which 25 also had analyzable standardized instructions. An additional five measures had only analyzable instructions. See Table 1 for the reading-grade-level results for each measure. The three readability formulas were significantly correlated for both instructions ($r$ range from .64 to .83, mean $r = .74$, all $p < .01$) and items ($r$ range from .32 to .72, mean $r = .55$, all $p < .05$).

**Instructions**

The 30 measures for which instructions were available had a mean word count of 79.5 words ($SD = 87.3$, range: 9–451). The mean reading-grade level (averaged across the three formulas) was 9.2 ($SD = 1.5$, range: 4.9–11.6). Only one measure was written below a 7th-grade reading level; 60% of measures were written at a 9th-grade level or higher. There was no association between word count and the average reading-grade level ($r = -.06$, n.s.); one measure was excluded because of an outlier word count value ($>3 SD$ from the mean).

**Items**

The 40 measures for which items were available and calculated had a mean word count of 356.6 words ($SD = 435.6$, range: 37–2,657). Word count analyses included the increased word count for two measures for which incomplete sentences were completed by the authors to allow for the use of readability formulas. The mean reading-grade level for measure items was 8.3 ($SD = 1.0$, range: 5.4–11.5), with only one measure written below a 7th-grade level and 32% of measures written higher than an 8th-grade level. After excluding one measure for an outlier word count value, the association between word count and reading-grade level was significant ($r = .37$, $p < .05$), characterized by higher reading-grade level for longer measures.

**Discussion**

Consistent with previous studies of other psychiatric disorder symptom measures (e.g., Andrasik et al., 1981; McHugh and Behar, 2009), this study found that the majority of alcohol misuse measures were written at reading-grade levels above recommended levels for patient materials. Only one measure from this study had instructions written at the 5th- to 6th-grade level, and only one measure had items written at this level. Measure items, on average, required an 8th-grade reading level. These findings are consistent with other evidence suggesting that attention to literacy and health literacy is often lacking in the development of patient materials (e.g., Greenfield et al., 2005) and further highlight the importance of considering literacy in mental health settings.

The results from this study have several implications for the use of self-report measures in SUD treatment and research settings. Self-report measures can be valid only if they can be understood by the individual completing them. Thus, consideration of the reading ability of treatment-seeking patients and research participants is crucial when selecting self-report measures. It is important to note that educational attainment level (i.e., highest grade completed) is often poorly correlated with reading ability (Davis et al., 1993; Johnson et al., 1995). Thus, the use of measures written at the lowest reading-grade level available may be preferable to maximize validity across the broadest possible sample. Evidence also suggests that people at all levels of reading ability prefer materials that are written at lower reading-grade levels (Sudore et al., 2007).

These findings also have implications for measure development standards. Although several indices of validity (e.g., predictive) and reliability (e.g., internal consistency) are expected as part of measure development, readability is often overlooked. Attention to writing items and instructions at a reading-grade level approximating 5th–6th grade (e.g., by minimizing use of complex and multisyllabic words and long sentences, minimizing measure length, and avoiding complex sentence structures) would improve measure validity for many populations.

There are several limitations to the current study. First, although we used several methods to identify validated alcohol measures, it is possible that some measures were not identified by these methods. Second, readability formulas
<table>
<thead>
<tr>
<th>Measure</th>
<th>Instructions</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Hangover Scale</td>
<td>6.0 8.5 7.5 7.3</td>
<td>7.0 8.5 9.9 8.5</td>
</tr>
<tr>
<td>*Advanced Warning of Relapse</td>
<td>8.5 12.5 9.5 10.2</td>
<td>6.0 6.4 9.1 7.2</td>
</tr>
<tr>
<td>Alcohol Dependence Data Questionnaire</td>
<td>8.5 8.5 10.9 9.3</td>
<td>6.0 8.0 9.5 8.2</td>
</tr>
<tr>
<td>Alcohol Dependence Scale</td>
<td>7.0 8.5 9.9 8.5</td>
<td>7.0 8.0 9.5 8.2</td>
</tr>
<tr>
<td>Alcohol Use Disorders Identification Test</td>
<td>8.5 10.8 10.3 9.9</td>
<td>7.0 9.0 10.1 8.7</td>
</tr>
<tr>
<td>Alcohol Use Inventory</td>
<td>8.5 9.8 10.4 9.6</td>
<td>7.0 8.9 9.5 8.5</td>
</tr>
<tr>
<td>Alcohol Withdrawal Symptom</td>
<td>8.5 9.7 10.5 9.6</td>
<td>– – – –</td>
</tr>
<tr>
<td>Alcohol Involvement Scale</td>
<td>7.0 9.2 8.9 8.4</td>
<td>7.0 9.2 8.9 8.4</td>
</tr>
<tr>
<td>Athlete Drinking Scale</td>
<td>11.0 12.5 10.6 11.4</td>
<td>– – – –</td>
</tr>
<tr>
<td>Drinker Inventory of Consequences</td>
<td>7.0 7.9 10.0 8.3</td>
<td>7.0 8.1 9.9 8.3</td>
</tr>
<tr>
<td>Drinking Problems Index</td>
<td>5.0 3.0 6.7 4.9</td>
<td>8.5 7.2 11.6 9.1</td>
</tr>
<tr>
<td>Drinking Restraint Scale</td>
<td>– – – –</td>
<td>7.0 6.9 11.1 8.3</td>
</tr>
<tr>
<td>Drinking Styles Questionnaire</td>
<td>6.0 8.5 8.6 7.7</td>
<td>7.0 9.2 9.3 8.5</td>
</tr>
<tr>
<td>Ethanol Dependence Syndrome Scale</td>
<td>11.0 10.8 11.2 11.0</td>
<td>7.0 8.1 9.0 8.0</td>
</tr>
<tr>
<td>Fast Alcohol Screening Test</td>
<td>6.0 6.9 9.2 7.4</td>
<td>7.0 9.9 9.2 8.7</td>
</tr>
<tr>
<td>Five Shot Questionnaire</td>
<td>– – – –</td>
<td>7.0 9.0 9.6 8.5</td>
</tr>
<tr>
<td>Form 90 QFV-90 Questionnaire</td>
<td>7.0 7.5 9.0 7.8</td>
<td>7.0 10.8 8.7 8.8</td>
</tr>
<tr>
<td>*Impaired Control Scale</td>
<td>11.0 11.7 9.9 10.9</td>
<td>6.0 6.2 8.0 6.7</td>
</tr>
<tr>
<td>Impaired Response Inhibition Scale</td>
<td>– – – –</td>
<td>7.0 8.8 9.0 8.3</td>
</tr>
<tr>
<td>Inventory of Drinking Situations</td>
<td>8.5 10.8 11.0 10.1</td>
<td>7.0 10.1 9.0 8.7</td>
</tr>
<tr>
<td>Iowa Alcoholic Stages Index</td>
<td>8.5 10.8 11.0 10.1</td>
<td>7.0 8.1 9.5 8.2</td>
</tr>
<tr>
<td>Lubeck Alcohol Dependence and Abuse Screening Test</td>
<td>– – – –</td>
<td>7.0 7.1 11.0 8.4</td>
</tr>
<tr>
<td>Michigan Alcohol Screening Test</td>
<td>– – – –</td>
<td>7.0 8.1 9.5 8.2</td>
</tr>
<tr>
<td>Missouri Alcoholism Severity Scale</td>
<td>8.5 8.4 10.5 9.1</td>
<td>7.0 8.1 9.5 8.2</td>
</tr>
<tr>
<td>Mortimer–Filkins Test</td>
<td>6.0 5.6 8.7 6.8</td>
<td>6.0 7.5 9.4 7.6</td>
</tr>
<tr>
<td>Munich Alcoholism Test</td>
<td>11.0 11.2 10.8 11.0</td>
<td>7.0 8.8 9.6 8.5</td>
</tr>
<tr>
<td>Obsessive–Compulsive Drinking Scale</td>
<td>8.5 8.5 11.2 9.4</td>
<td>11.0 10.6 10.7 10.8</td>
</tr>
<tr>
<td>Penn Alcohol Craving Scale</td>
<td>8.5 8.5 10.8 9.3</td>
<td>7.0 8.5 8.8 8.1</td>
</tr>
<tr>
<td>Rapid Alcohol Problems Screen</td>
<td>– – – –</td>
<td>7.0 9.1 8.8 8.3</td>
</tr>
<tr>
<td>Research Institute on Addictions Self-Inventory</td>
<td>8.5 10.8 10.0 9.8</td>
<td>6.0 8.2 9.0 7.7</td>
</tr>
</tbody>
</table>
have inherent limitations, including using a simple metric that may not sufficiently account for other linguistic complexities. We attempted to minimize these limitations by using the average of three widely used and validated formulas, and the fact that the three formulas produced similar results strengthens our findings. Third, the ability to comprehend a measure is associated with a number of other factors (e.g., formatting, grammatical complexity) that also should be considered when developing and evaluating measures because of their contribution to readability (see McHugh et al., 2011; Schinka, 2012). Finally, some self-report measures use word checklists, and readability formulas do not allow for evaluation of these types of formats. Although alternative strategies, such as expert reviews or comparison to vocabulary lists, may provide a rough estimate of reading-grade level for such measures, we elected to exclude these measures because of the absence of a validated strategy for determining their readability. The validation of more broadly applicable assessment strategies for reading-grade level is needed to improve the ability to estimate readability across varied measure formats.

The validity of self-report measures is crucial to their accurate use and interpretation. The ability of respondents to read measures provides a ceiling on their validity; the inclusion of reading-grade level among the basic psychometric properties evaluated during measure development and validation is needed. Many alcohol misuse measures may be most appropriate for populations with higher-than-average reading abilities. Thus, the revision of existing measures to decrease reading-grade levels or the development of new, low reading-grade-level measures is needed to increase measure validity across diverse populations and broaden the applicability of these instruments.

References


Notes: FLE = Flesch Reading Ease translated grade-level scores; SMOG = Simple Measure of Gobbledygook Readability Formula; FCST = FORCAST Readability Formula; M = mean grade level for the three formulas. The four measures with the lowest reading-grade level for items are marked with an asterisk (*).
the TWEAK test in screening for alcoholism/heavy drinking in three populations. Alcoholism: Clinical and Experimental Research, 17, 1188–1192.


