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}

^a*Division of Alcohol and Drug Abuse, McLean Hospital, Belmont, Massachusetts*

^b*Department of Psychiatry, Harvard Medical School, Boston, Massachusetts*

^c*Department 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. **Conclusions:** 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 (Rathner 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

Received: July 2, 2013. Revision: August 14, 2013.

This work was supported by the Brookline Mental Health Research Scholar Award (to R. Kathryn McHugh) and National Institute on Drug Abuse grants K24 DA019855 (to Shelly F. Greenfield), K24 DA022288 (to Roger D. Weiss), and U10 DA015831 (to Roger D. Weiss).

*Correspondence may be sent to R. Kathryn McHugh at the Division of Alcohol and Drug Abuse, McLean Hospital, 115 Mill Street, Proctor House 3, MS 222, Belmont, MA 02478, or via email at: kmchugh@mclean.harvard.edu.

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 PsycINFO 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 s range from .64 to .83, mean $r = .74$, all $ps < .01$) and items (r s range from .32 to .72, mean $r = .55$, all $ps < .05$).

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 1. Reading-grade level scores for alcohol misuse measures

Measure	Instructions				Items			
	FLE	SMOG	FCST	M	FLE	SMOG	FCST	M
Acute Hangover Scale (Rohsenow et al., 2007)	6.0	8.5	7.5	7.3	—	—	—	—
*Advanced Warning of Relapse (Miller and Harris, 2000)	8.5	12.5	9.5	10.2	6.0	6.4	9.1	7.2
Alcohol Dependence Data Questionnaire (Raistrick et al., 1983)	8.5	8.5	10.9	9.3	7.0	8.1	9.3	8.1
Alcohol Dependence Scale (Skinner and Allen, 1982)	7.0	8.5	9.9	8.5	7.0	8.0	9.5	8.2
*Alcohol Urge Questionnaire (Bohn et al., 1995)	8.5	11.1	10.6	10.1	5.0	4.9	6.4	5.4
Alcohol Use Disorders Identification Test (Babor et al., 1989)	8.5	10.8	10.3	9.9	7.0	9.0	10.1	8.7
Alcohol Use Inventory (Wanberg et al., 1977)	8.5	9.8	10.4	9.6	7.0	8.9	9.5	8.5
Alcohol Withdrawal Symptom Checklist-Revised (Pittman et al., 2007)	8.5	9.7	10.5	9.6	—	—	—	—
Alcoholic Involvement Scale (Aharan, 1966)	—	—	—	—	7.0	9.2	8.9	8.4
Athlete Drinking Scale (Martens et al., 2005)	—	—	—	—	6.0	9.0	8.6	7.9
CAGE (Ewing, 1984)	—	—	—	—	7.0	6.9	11.1	8.3
College Alcohol Problems Scale (O'Hare, 1997)	11.0	12.5	10.6	11.4	—	—	—	—
Drinker Inventory of Consequences (Miller et al., 1995)	7.0	7.9	10.0	8.3	7.0	8.1	9.9	8.3
Drinking Problems Index (Finney et al., 1991)	5.0	3.0	6.7	4.9	8.5	7.2	11.6	9.1
Drinking Restraint Scale (Curry et al., 1987)	—	—	—	—	7.0	9.3	9.9	8.7
Drinking Styles Questionnaire (Smith et al., 1995)	6.0	8.5	8.6	7.7	7.0	9.2	9.3	8.5
Ethanol Dependence Syndrome Scale (Babor, 1996)	11.0	10.8	11.2	11.0	7.0	8.1	9.0	8.0
Fast Alcohol Screening Test (Hodgson et al., 2002)	6.0	6.9	9.2	7.4	7.0	9.9	9.2	8.7
Five Shot Questionnaire (Seppä et al., 1998)	—	—	—	—	7.0	9.0	9.6	8.5
Form 90 QFV-90 Questionnaire (Miller, 1996)	7.0	7.5	9.0	7.8	7.0	10.8	8.7	8.8
*Impaired Control Scale (Heather et al., 1993)	11.0	11.7	9.9	10.9	6.0	6.2	8.0	6.7
Impaired Response Inhibition Scale for Alcoholism (Guardia et al., 2007)	—	—	—	—	7.0	8.8	9.0	8.3
Inventory of Drinking Situations (Annis, 1982)	8.5	10.8	11.0	10.1	7.0	10.1	9.0	8.7
Iowa Alcoholic Stages Index (Mulford, 1977)	—	—	—	—	7.0	8.1	9.5	8.2
Lubeck Alcohol Dependence and Abuse Screening Test (Rumpf et al., 1997)	—	—	—	—	7.0	7.1	11.0	8.4
Michigan Alcohol Screening Test (Selzer, 1971)	—	—	—	—	8.5	8.4	10.5	9.1
Missouri Alcoholism Severity Scale (Evenson et al., 1973)	—	—	—	—	—	—	—	—
Mortimer-Filkins Test (Mortimer et al., 1971)	6.0	5.6	8.7	6.8	6.0	7.5	9.4	7.6
Munich Alcoholism Test (Feuerlein et al., 1979)	11.0	11.2	10.8	11.0	7.0	8.8	9.6	8.5
Obsessive-Compulsive Drinking Scale (Morgan et al., 2004)	8.5	8.5	11.2	9.4	11.0	10.6	10.7	10.8
Penn Alcohol Craving Scale (Flannery et al., 1999)	8.5	8.5	10.8	9.3	7.0	8.5	8.8	8.1
Rapid Alcohol Problems Screen (Cherpitel, 1995)	—	—	—	—	7.0	9.1	8.8	8.3
Research Institute on Addictions Self-Inventory (Nochajski et al., 1995)	8.5	10.8	10.0	9.8	6.0	8.2	9.0	7.7

Table Continued

TABLE 1. Reading-grade level scores for alcohol misuse measures

Measure	Instructions				Items			
	FLE	SMOG	FCST	M	FLE	SMOG	FCST	M
Restrained Drinking Scale (Ruderman and McKirnan, 1984)	—	—	—	—	6.0	8.0	9.0	7.7
Rutgers Alcohol Problems Index (White and Labouvie, 1989)	8.5	9.7	10.7	9.6	—	—	—	—
Self-Administered Alcoholism Screening Test (Swenson and Morse, 1975)	8.5	11.4	9.7	9.9	8.5	8.5	10.6	9.2
Severity of Alcohol Dependence Questionnaire (Stockwell et al., 1983)	11.0	12.2	11.0	11.4	7.0	8.2	10.3	8.5
Severity Scales of Alcohol Dependence (John et al., 2003)	8.5	8.5	10.9	9.3	6.0	8.2	8.6	7.6
Short Inventory of Problems (Feinn et al., 2003)	7.0	7.9	9.9	8.3	7.0	8.0	10.1	8.4
Student Alcohol Questionnaire (Engs, 1975)	6.0	8.2	8.7	7.6	11.0	12.1	11.3	11.5
*T-ACE (Sokol et al., 1989)	—	—	—	—	6.0	6.9	7.9	6.9
Temptation and Restraint Inventory (Collins and Lapp, 1992)	7.0	8.9	10.1	8.7	7.0	8.3	10.3	8.5
TWEAK (Chan et al., 1993)	—	—	—	—	6.0	8.5	8.4	7.6
Veterans Alcoholism Screening Test (Magruder-Habib et al., 1982)	—	—	—	—	7.0	7.2	9.5	7.9
Young Adult Alcohol Problems Screening Test (Hurlbut and Sher, 1992)	11.0	12.5	11.3	11.6	8.5	9.4	10.2	9.4

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 (*).

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

- Ad Hoc Committee on Health Literacy for the Council on Scientific Affairs, American Medical Association. (1999). Health literacy: Report of the council on scientific affairs. *Journal of the American Medical Association*, 281, 552–557.
- Aharan, C. (1966). *The interdependence of personal and cultural variables in alcoholism* (Unpublished). University of Western Ontario, London, Ontario, Canada.
- Andrasik, F., Heimberg, R. G., Edlund, S. R., & Blankenberg, R. (1981). Assessing the reliability levels of self-report assertion inventories. *Journal of Consulting and Clinical Psychology*, 49, 142–144.
- Annis, H. M. (1982). *Inventory of drinking situations*. Toronto, Canada: Addiction Research Foundation of Ontario.
- Babor, T. F. (1996). Reliability of the Ethanol Dependence Syndrome scale. *Psychology of Addictive Behaviors*, 10, 97–103.
- Babor, T. F., de la Fuente, J. R., Saunders, J., & Grant, M. (1989). *AUDIT: The Alcohol Use Disorders Identification Test: Guidelines for use in primary health care*. Geneva, Switzerland: World Health Organization.
- Beckman, H. T., & Lueger, R. J. (1997). Readability of self-report clinical outcome measures. *Journal of Clinical Psychology*, 53, 785–789.
- Bohn, M. J., Krahn, D. D., & Stachler, B. A. (1995). Development and initial validation of a measure of drinking urges in abstinent alcoholics. *Alcoholism: Clinical and Experimental Research*, 19, 600–606.
- Caylor, J. S., Sticht, T. G., Fox, L. C., & Ford, J. P. (1973). *Methodologies for determining reading requirements of military occupational specialties*. Alexandria, VA: Human Resources Research Organization.
- Chan, A. W., Pristach, E. A., Welte, J. W., & Russell, M. (1993). Use of

- the TWEAK test in screening for alcoholism/heavy drinking in three populations. *Alcoholism: Clinical and Experimental Research*, 17, 1188–1192.
- Cherpitel, C. J. (1995). Screening for alcohol problems in the emergency room: A rapid alcohol problems screen. *Drug and Alcohol Dependence*, 40, 133–137.
- Collins, R. L., & Lapp, W. M. (1992). The Temptation and Restraint Inventory for measuring drinking restraint. *British Journal of Addiction*, 87, 625–633.
- Curry, S., Southwick, L., & Steele, C. (1987). Restrained drinking: Risk factor for problems with alcohol? *Addictive Behaviors*, 12, 73–77.
- Davis, T. C., Jackson, R. H., George, R. B., Long, S. W., Talley, D., Murphy, P. W., . . . Truong, T. (1993). Reading ability in patients in substance misuse treatment centers. *International Journal of the Addictions*, 28, 571–582.
- Engs, R. C. (1975). *The Student Alcohol Questionnaire*. Bloomington, IN: Department of Health and Safety Education, Indiana University.
- Evenson, R. C., Altman, H., Cho, D. W., & Montgomery, J. (1973). Development of an alcoholism severity scale via an iterative computer program for item analysis. *Quarterly Journal of Studies on Alcohol*, 34, 1336–1341.
- Ewing, J. A. (1984). Detecting alcoholism: The CAGE questionnaire. *Journal of the American Medical Association*, 252, 1905–1907.
- Feinn, R., Tennen, H., & Kranzler, H. R. (2003). Psychometric properties of the short index of problems as a measure of recent alcohol-related problems. *Alcoholism: Clinical and Experimental Research*, 27, 1436–1441.
- Feuerlein, W., Ringer, C., Küfner, H., & Antons, K. (1979). Diagnosis of alcoholism: The Munich Alcoholism Test (MALT). *Currents in Alcoholism*, 7, 137–147.
- Finney, J. W., Moos, R. H., & Brennan, P. L. (1991). The Drinking Problems Index: A measure to assess alcohol-related problems among older adults. *Journal of Substance Abuse*, 3, 395–404.
- Flannery, B. A., Volpicelli, J. R., & Pettinati, H. M. (1999). Psychometric properties of the Penn Alcohol Craving Scale. *Alcoholism: Clinical and Experimental Research*, 23, 1289–1295.
- Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology*, 32, 221–233.
- Greenfield, S. F., Sugarman, D. E., Nargiso, J. E., & Weiss, R. D. (2005). Readability of patient handout materials in a nationwide sample of alcohol and drug abuse treatment programs. *The American Journal on Addictions*, 14, 339–345.
- Guardia, J., Trujols, J., Burguete, T., Luquero, E., & Cardús, M. (2007). Impaired response inhibition scale for alcoholism (IRISA): Development and psychometric properties of a new scale for abstinence-oriented treatment of alcoholism. *Alcoholism: Clinical and Experimental Research*, 31, 269–275.
- Heather, N., Tebbutt, J. S., Mattick, R. P., & Zamir, R. (1993). Development of a scale for measuring impaired control over alcohol consumption: A preliminary report. *Journal of Studies on Alcohol*, 54, 700–709.
- Hodgson, R., Alwyn, T., John, B., Thom, B., & Smith, A. (2002). The FAST Alcohol Screening Test. *Alcohol and Alcoholism*, 37, 61–66.
- Hooke, L. R., DeLeo, P. J., & Slaughter, S. L. (1979). *Readability of Air Force publications: A criterion referenced evaluation* [Final Report AFHRL-TR-79-21]. Washington, DC: U.S. Air Force. Retrieved from <http://eric.ed.gov/?q=ED177512>
- Hurlbut, S. C., & Sher, K. J. (1992). Assessing alcohol problems in college students. *Journal of American College Health*, 41, 49–58.
- Institute of Medicine. (2004). *Health literacy: A prescription to end confusion*. Washington, DC: The National Academies Press.
- John, U., Hapke, U., & Rumpf, H. J. (2003). A new measure of the alcohol dependence syndrome: The severity scale of alcohol dependence. *European Addiction Research*, 9, 87–93.
- Johnson, M. E., Fisher, D. G., Davis, D. C., & Cagle, H. H. (1995). Reading abilities of drug users in Anchorage, Alaska. *Journal of Drug Education*, 25, 73–80.
- Kern, R., Sticht, T., Welty, D., & Hauke, R. (1976). *Guidebook for the development of Army training literature* (Human Resources Research Organization Tech. Rep. No. 75-3). Alexandria, VA: Human Resources Research Organization.
- Khazaal, Y., Chatton, A., Cochand, S., Coquard, O., Fernandez, S., Khan, R., & Zullino, D. (2010). Quality of web-based information on alcohol dependence. *Drugs: Education, Prevention, and Policy*, 17, 248–260.
- Khazaal, Y., Chatton, A., Cochand, S., & Zullino, D. (2008). Quality of Web-based information on cocaine addiction. *Patient Education and Counseling*, 72, 336–341.
- Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (2002). *Adult literacy in America: A first look at the findings of the National Adult Literacy Survey* (NCES 1993-275). Washington, DC: National Center for Education Statistics.
- Kutner, M., Greenberg, E., Jin, Y., & Paulsen, C. (2006). *The health literacy of America's adults: Results from the 2003 National Assessment of Adult Literacy*. Washington, DC: U.S. Department of Education.
- Ley, P., & Florio, T. (1996). The use of readability formulas in health care. *Psychology, Health and Medicine*, 1, 7–28.
- Magruder-Habib, K., Harris, K. E., & Fraker, G. G. (1982). Validation of the Veterans Alcoholism Screening Test. *Journal of Studies on Alcohol*, 43, 910–926.
- Martens, M. P., Watson, J. C., II, Royland, E. M., & Beck, N. C. (2005). Development of the athlete drinking scale. *Psychology of Addictive Behaviors*, 19, 158–164.
- McCall, W. A., & Crabbs, L. M. (1961). *Standard test lessons in reading*. New York, NY: Teachers College.
- McHugh, R. K., & Behar, E. (2009). Readability of self-report measures of depression and anxiety. *Journal of Consulting and Clinical Psychology*, 77, 1100–1112.
- McHugh, R. K., Rasmussen, J. L., & Otto, M. W. (2011). Comprehension of self-report evidence-based measures of anxiety. *Depression and Anxiety*, 28, 607–614.
- McLaughlin, G. H. (1969). SMOG grading: A new readability formula. *Journal of Reading*, 12, 639–646.
- Meade, C. D., & Smith, C. F. (1991). Readability formulas: Cautions and criteria. *Patient Education and Counseling*, 17, 153–158.
- Miller, W. R. (1996). *Form 90: A structured assessment interview for drinking and related behaviors: Test manual* (Project MATCH Monograph Series, Vol. 5, NIH Publication No. 96-4004). Bethesda, MD: Department of Health and Human Services.
- Miller, W. R., & Harris, R. J. (2000). A simple scale of Gorski's warning signs for relapse. *Journal of Studies on Alcohol*, 61, 759–765.
- Miller, W. R., Tonigan, J. S., & Longabaugh, R. (1995). *The Drinker Inventory of Consequences (DrInC): An instrument for assessing adverse consequences of alcohol abuse* (Volume 4, NIH Publication No. 95-3911). Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.
- Morgan, T. J., Morgenstern, J., Blanchard, K. A., Labouvie, E., & Bux, D. A. (2004). Development of the OCDS—revised: A measure of alcohol and drug urges with outpatient substance abuse clients. *Psychology of Addictive Behaviors*, 18, 316–321.
- Mortimer, R., Filkins, L., & Lower, J. S. (1971). *Development of court procedures for identifying problem drinkers: Final report*. Washington, DC: Department of Transportation, National Highway Traffic Safety Administration.
- Mulford, H. A. (1977). Stages in the alcoholic process. Toward a cumulative, nonsequential index. *Journal of Studies on Alcohol*, 38, 563–583.
- Nochajski, T., Miller, B., & Wiczorek, W. (1995). *Training manual for the Research Institute on Addictions Self-Inventory*. Buffalo, NY: Research Institute on Addictions.

- O'Hare, T. (1997). Measuring problem drinking in first time offenders. Development and validation of the College Alcohol Problem Scale (CAPS). *Journal of Substance Abuse Treatment, 14*, 383–387.
- Pittman, B., Gueorgieva, R., Krupitsky, E., Rudenko, A. A., Flannery, B. A., & Krystal, J. H. (2007). Multidimensionality of the Alcohol Withdrawal Symptom Checklist: A factor analysis of the Alcohol Withdrawal Symptom Checklist and CIWA-Ar. *Alcoholism: Clinical and Experimental Research, 31*, 612–618.
- Raistrick, D., Dunbar, G., & Davidson, R. (1983). Development of a questionnaire to measure alcohol dependence. *British Journal of Addiction, 78*, 89–95.
- Rather, B. C., & Murphy, J. D. (1995). Alcoholics anonymous and rational recovery: Readability of the "big book" vs the small book. *Perceptual and Motor Skills, 81*, 1313–1314.
- Richards, L. K., McHugh, R. K., Pratt, E. M., & Thompson-Brenner, H. (2013). Readability and comprehension of self-report binge eating measures. *Eating Behaviors, 14*, 167–170.
- Rohsenow, D. J., Howland, J., Minsky, S. J., Greece, J., Almeida, A., & Roehrs, T. A. (2007). The Acute Hangover Scale: A new measure of immediate hangover symptoms. *Addictive Behaviors, 32*, 1314–1320.
- Ruderman, A. J., & McKirnan, D. J. (1984). The development of a restrained drinking scale: A test of the abstinence violation effect among alcohol users. *Addictive Behaviors, 9*, 365–371.
- Rumpf, H. J., Hapke, U., Hill, A., & John, U. (1997). Development of a screening questionnaire for the general hospital and general practices. *Alcoholism: Clinical and Experimental Research, 21*, 894–898.
- Schinka, J. A. (2012). Further issues in determining the readability of self-report items: Comment on McHugh and Behar (2009). *Journal of Consulting and Clinical Psychology, 80*, 952–955.
- Selzer, M. L. (1971). The Michigan alcoholism screening test: The quest for a new diagnostic instrument. *American Journal of Psychiatry, 127*, 1653–1658.
- Seppä, K., Lepistö, J., & Sillanaukee, P. (1998). Five-shot questionnaire on heavy drinking. *Alcoholism: Clinical and Experimental Research, 22*, 1788–1791.
- Skinner, H. A., & Allen, B. A. (1982). Alcohol dependence syndrome: Measurement and validation. *Journal of Abnormal Psychology, 91*, 199–209.
- Smith, G. T., McCarthy, D. M., & Goldman, M. S. (1995). Self-reported drinking and alcohol-related problems among early adolescents: Dimensionality and validity over 24 months. *Journal of Studies on Alcohol, 56*, 383–394.
- Sokol, R. J., Martier, S. S., & Ager, J. W. (1989). The T-ACE questions: Practical prenatal detection of risk-drinking. *American Journal of Obstetrics and Gynecology, 160*, 863–868, discussion 868–870.
- Stockwell, T., Murphy, D., & Hodgson, R. (1983). The severity of alcohol dependence questionnaire: Its use, reliability and validity. *British Journal of Addiction, 78*, 145–155.
- Sudore, R. L., Landefeld, C. S., Barnes, D. E., Lindquist, K., Williams, B. A., Brody, R., & Schillinger, D. (2007). An advance directive redesigned to meet the literacy level of most adults: A randomized trial. *Patient Education and Counseling, 69*, 165–195.
- Swenson, W. M., & Morse, R. M. (1975). The use of a self-administered alcoholism screening test (SAAST) in a medical center. *Mayo Clinic Proceedings, 50*, 204–208.
- Wanberg, K. W., Horn, J. L., & Foster, F. M. (1977). A differential assessment model for alcoholism. The scales of the Alcohol Use Inventory. *Journal of Studies on Alcohol, 38*, 512–543 http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=859335&dopt=Abstract.
- Weiss, B. D., & Coyne, C. (1997). Communicating with patients who cannot read. *The New England Journal of Medicine, 337*, 272–274.
- White, H. R., & Labouvie, E. W. (1989). Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol, 50*, 30–37.
- Zraick, R. I., Atcherson, S. R., & Brown, A. M. (2012). Readability of patient-reported outcome questionnaires for use with persons who stutter. *Journal of Fluency Disorders, 37*, 20–24.