Not All Alcohol Use Disorder Criteria Are Equally Severe:

Towards Severity Grading of Individual Criteria in College Drinkers

Cassandra L. Boness, Sean P. Lane, & Kenneth J. Sher

Author Note

Cassandra L. Boness, Department of Psychological Sciences, University of Missouri;

Sean P. Lane, Psychological Sciences, Purdue University; Kenneth J. Sher, Psychological

Sciences, University of Missouri.

Correspondence concerning this article should be addressed to Cassandra Boness, 146

Psychology Building, Department of Psychological Sciences, University of Missouri, Columbia,

MO 65202. E-mail: clmkdb@mail.missouri.edu; Phone: 480-324-6475

Abstract

Alcohol Use Disorder (AUD) diagnosis in the Fifth Edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5; APA, 2013) contains a severity gradient based on number of criteria endorsed, implicitly assuming criteria are interchangeable. However, criteria vary widely in endorsement rates, implying differences in the latent severity associated with a symptom (e.g., Lane et al., 2016) and demonstrating criteria are not interchangeable (Lane & Sher, 2015). We evaluated whether variation in the severity of criteria could be resolved by employing multiple indicators of each criterion varying in item-level severity. We assessed 909 undergraduate students aged 18 years or older with at least 12 drinking occasions in the past year. Participants self-administered questions on alcohol consumption and past year AUD symptoms via an online survey. For each of the 11 AUD criteria, we selected three indicators based on the difficulty values of the one-parameter logistic item response theory (IRT) model ranging from low to high. We first tested a higher-order AUD factor defined by 11 lower-order criterion factors ($\chi^2$ [551] = 2959.35, $p < .0001$; RMSEA = 0.09). The 33 items were used to create severity scores: a criterion count (0-11), symptom count (0-33), and factor scores derived from a bi-factor model. Though our new scores resulted in incremental validity over DSM-5 across a range of external validators, when the standardized regression estimates were compared, the new scores did not consistently outperform the DSM-5 suggesting this approach is viable for developing more sensitive diagnostic instruments but needs further refinement.

*Keywords:* Alcohol Use Disorder, Diagnosis, Assessment, Severity, Measurement, Psychometrics

Not All Alcohol Use Disorder Criteria Are Equally Severe:
Towards Severity Grading of Individual Criteria in College Drinkers

The major elements of alcoholism as a disease of addiction (i.e., loss of control over drinking and habitual/prolonged consumption) with varying degrees of severity were recognized as early as the 18th century by physicians such as Benjamin Rush and Thomas Trotter (for an overview see Li, Hewitt, & Grant, 2007). These elements influenced early editions of the *Diagnostic and Statistical Manual of Mental Disorders (DSM)*, which had a single category for diagnosis: alcoholism. Alcoholism in these first two versions of the DSM fell under the broader classification of personality disorders but offered no specific diagnostic criteria and no evaluation of severity. The DSM-III (APA, 1980) began to use the abuse/dependence classification, with the latter being presumed the more severe diagnosis, and made the distinction between these two types based on physiological dependence (i.e., tolerance and/or withdrawal), introducing the first attempt to grade the severity of AUD. The DSM-III-R (APA, 1987) made changes to the criteria and expanded the construct of dependence to include non-physiological symptoms (Edwards & Gross, 1976). The next revision of the DSM, the DSM-IV (APA, 1994), left most of the key changes of DSM-III-R intact, with minor revisions of the criteria sets and diagnostic algorithms.

In contrast, DSM-5's (APA, 2013) alcohol use disorder (AUD) and substance use disorders (SUDs), in general, are conceptualized as continuous, unitary constructs varying in severity based on the number of criteria endorsed. The threshold for diagnosis is endorsement of two criteria out of 11 occurring within a 12-month period leading to clinically significant impairment or distress. Diagnosis is graded as mild (2-3 criteria), moderate (4-5 criteria), and severe (6+ criteria). Although addressing AUD severity with criterion counts, DSM-5 fails to consistently consider inherent variation in criterion severity across disorders.

**Defining Severity**

It is important to distinguish between the three different types of severity that may be considered in AUD diagnosis. The first, *DSM-5 criterion count severity*, refers to the number of DSM-5 AUD criteria the individual endorses (APA, 2013). The second type, *latent severity*, refers to the underlying latent severity associated with AUD from a factor analytic perspective (Dawson & Grant, 2010). Therefore, in principle, someone may have a low DSM-5 criterion count, but a relatively high latent severity of AUD and vice versa (Cooper & Balsis, 2009; Lane, Steinley, & Sher, 2016; Schmulewitz, et al., 2010). The third type, *symptom severity,* refers to variation in the severity of specific symptoms used to assess a given AUD criterion. For example, a given criterion (e.g., craving) can be assessed as either mild (e.g., a strong urge) or severe (e.g., an overpowering urge).

**Potential Problems with DSM-5 Alcohol Use Disorder**

There are multiple, potential problems with using a simple criterion count as a measure of AUD severity. Research on differential symptom severity of DSM criteria, as operationalized via the Item Response Theory (IRT)-derived difficulty parameter, has found that criteria are arrayed along a continuum of severity (Beseler, Taylor, & Leeman, 2010; Hagman & Cohn, 2009; Lane et al., 2016; Saha, Stinson, & Grant, 2007). Using this measurement model, it is possible that someone with a mild AUD (i.e., a criterion count of 2-3) may experience more severe problems than someone with a moderate AUD (i.e., a criterion count of 4-5). Similarly, someone with a moderate AUD may have more latent severity than someone with a severe AUD (Lane & Sher, 2015). Despite this distinction, criterion counts and latent severity estimates can be highly correlated and use of one over the other is not likely to make much practical difference. Further, weighting symptoms for frequency of occurrence, as an alternate measure of severity, has been demonstrated to only modestly improve the association with AUD correlates such as quantity

and frequency of consumption, particularly among those diagnosed with AUD (Dawson & Grant, 2010). Nevertheless, it is still possible that in individual cases there may be meaningful divergence of severity assessed by criterion counts and latent severity from IRT analyses.

Perhaps more critically, the DSM-5's ability to accurately characterize the latent severity of AUD is further complicated by the operationalization of the criteria based on specific symptoms and the sub-algorithms used to combine them. Some operationalizations are narrow in that they require a higher threshold of latent severity before they are likely to be endorsed, while some are more liberal, being endorsed at a lower threshold. For example, more individuals would be likely to endorse the broader and more liberal indicator, "Strong desire or urge to drink," when compared to the higher threshold and narrower indicator, "Wanted to drink so badly you couldn't think of anything else" (e.g., Grant et al., 2011). Depending on the operationalization used, different rates of endorsement may be found, suggesting different latent severities. Consequently, IRT analyses of AUD criteria are highly conditional upon the specific operationalizations used to define each criterion and it is unclear how different operationalizations vary in severity. Although the research literature has been largely silent on this issue, IRT studies of DSM-IV and DSM-5 criteria suggest this is a nontrivial concern (Hasin et al., 2013). In a recent meta-analysis, Lane et al. (2016) showed that, depending on the instruments employed to assess AUD, estimated latent severities associated with the 11 AUD symptoms vary markedly, highlighting the importance of specific operationalizations.

The operationalization of psychological constructs is a significant issue in survey methodology more generally (e.g., Paulhus & Vazire, 2007; Schwarz, 1999). Minor changes to question wording (i.e., operationalization) has been demonstrated to result in major changes to study results (Schwarz, 1999). The lack of standardization of clinical assessment tools and the

absence of a "gold standard," has resulted in significant between-study heterogeneity in the measurement of constructs such as the 11 AUD criteria. Although methods such as meta-analysis and integrative data analysis (IDA) have been developed to help address this issue of varying operationalizations across studies, these processes have their own limitations (e.g., labor intensive, missing information) as well as the fact that some instruments are not harmonizationable with existing methodology. As such, there is still a need for further standardization across instruments, particularly in the assessment of AUD.

The issue is not merely academic. We have recently demonstrated (Hoffman et al., 2018) that using high versus low threshold symptoms to operationalize a criterion results in highly dissimilar symptom network structures when examining AUD. With increasing interest in the replicability of symptom networks (e.g., Forbes et al., 2017), and reproducibility in science more generally (Ioannidis, 2005), assessment of criterion-related symptoms of graded severity would greatly facilitate data harmonization and allow more nuanced evaluations of reproducibility.

These issues tend to be particularly pronounced among college student drinkers, and young adults more generally, (e.g., Harford et al., 2005; Lane et al., 2016), making them an ideal population for further study in this area. Research has demonstrated bias in AUD items (i.e., differential item functioning) and prevalence as a function of age (e.g., Harford et al., 2009; Mewton et al., 2011), several issues related to the misinterpretation of items in this population, resulting in inflated estimates and false positives (e.g., Chung et al., 2001), and lack of consistent psychometric properties within this age group as a result of different sample and study features (e.g., Lane et al., 2016; Rose et al., 2012). While some work supports the notion of convergent validity of DSM-5 criteria in college students (e.g., Hagman & Cohn, 2011), there remains room for refinement. Further, given this population has amongst the highest AUD rates (Harford et al.,

2005; Grant et al., 2004) with recent epidemiological estimates being approximately 27% (Grant et al., 2015), their progression to dependence may be more rapid than that of adults (Martin et al., 1995), and it has been suggested that this is a particularly critical time for AUD onset (e.g., Kessler et al., 2005), they represent an important population in which to apply this methodology. A more appropriate assessment of AUD severity in this population may allow providers to intervene more effectively and appropriately triage them into the most suitable intervention.

In sum, the DSM-5's formula for AUD severity scaling is potentially problematic in that (1) it assumes criteria are interchangeable in terms of latent severity, and (2) the way criteria are operationalized lends itself to considerable variation in the threshold for meeting each criterion.

**Alternative Methods for Evaluation of Severity**

Part of the problem with the current approach to evaluating severity involves reliance on one or two symptoms to assess a given criterion. Two symptoms might validly assess a criterion but differ in their severity. Choosing a symptom with a low relative severity may fail to distinguish a trivial from clinically significant pathology while choosing a symptom indicating marked pathology might result in false negatives. One approach to addressing this concern about where to draw the single threshold for determining the presence/absence of a criterion is to consider a scaled severity grading at the symptom level. This general approach is frequently employed in clinical medicine. The APGAR (Apgar, 1953) score, used to evaluate a newborn's health, is one example. This method assesses overall health (i.e., total score) and the severity (0, 1, or 2) of five health outcomes considered in the overall score (National Institute of Health, 2011). This is in comparison to summing across dichotomously endorsed criteria, as seen in the DSM-5 AUD diagnosis and alternative diagnostic frameworks. Similar approaches to evaluating severity are used for cancer (National Cancer Institute, 2013), chronic obstructive pulmonary disease (Funk, et al., 2009; Lange et al., 2012), and coma (Jalali & Rezaei, 2014). Using such

clinical assessment approaches, particularly their scalable summation method, as a model for AUD diagnosis could prove more useful than the dichotomous system currently in effect.

Notably, the DSM-5 attempts to consider individual symptom severity for select other diagnoses. For example, in the "Specifiers for Depressive Disorders" section of the DSM-5 (APA, 2013, p. 188), severity is determined by considering: (1) the number of symptoms endorsed, (2) *the severity of those individual symptoms* [emphasis added], and (3) the degree of functional disability in social/occupational functioning. However, there exists no equivalent consideration for AUD, or SUDs in general, whereby the severity of the individual symptoms is considered in addition to the number of symptoms endorsed and degree of functional disability, pointing to the importance of the current study.

By grading individual criteria and not treating each as a simple, present/absent decision, it is possible that several important issues in AUD diagnosis can be addressed. First, differences in operationalization might be minimized since a range of criterion severities could be resolved to facilitate harmonization. Second, summary indices (i.e., diagnosis severity) could reflect both the number of criteria endorsed as well as the severity of individual criteria. Such changes could help to both improve diagnostic consistency for clinical and research purposes and provide a more "fine-grained" assessment of AUD severity. These improvements would be especially beneficial in college students who have been demonstrated to be characterized by two classes of drinkers (i.e., less severe and more severe; Rinker & Neighbors, 2015) and a significant percentage of diagnostic orphans (e.g., Hagman et al., 2014).

**Current Study**

The purpose of the current study was to address the concerns with the current AUD DSM-5 criterion count severity algorithm by examining the degree to which ACCOUNTING FOR variation in the severity of individual criteria (i.e., symptom severity) improves

measurement of severity at the syndromal level (i.e. latent severity) in a college student sample.

The current study sought to accomplish this by assessing whether a series of new diagnostic

scores would better predict relevant clinical AUD correlates when compared to the current DSM-

5 criterion count technique. A more systematic approach to assessing AUD severity, such as

using an IRT-based scoring method, may be useful for better grading AUD severity and guiding

both treatment and research efforts, particularly in college students who are at high-risk for

things such as academic and legal problems, risky sexual behavior and sexual assault, physical

assault, and morbidity and mortality (Rinker & Neighbors, 2015). Further, a more precise

approach, such as the one described here, will allow the evaluation and determination of which

criteria are the critical indicators of alcohol use disorder in young adulthood, especially given

AUD criteria were largely developed and standardized in adult clinical populations (Deas et al.,

2000). To test the feasibility of a diagnostic score that tries to resolve severity at the criterion

level, we aimed to compare the new scoring methods with the current DSM-5 algorithm.

**Methods**

Participants were 909 undergraduates enrolled in Introduction to Psychology courses at a

large Midwestern university and recruited via an online research participation system during the

consecutive spring 2015 ($N = 143$) and fall 2016 ($N = 766$) semesters, and self-selected to

participate in the study. Inclusion criteria included being over the age of 18 and having

consumed alcohol on at least 12 separate occasions in the past 12 months. Compensation for

participation in the form of course credit was awarded upon completion. Forty percent ($N = 363$)

of the sample, chosen at random, were designated as a holdout sample for later replication of

exploratory models, resulting in a subsample of 546 for the primary analyses. The majority of

participants were female (55.5%) White (91.9%), and never married (98.1%). Further, the

participants reported heavy past year drinking with an average of 1.25 drinking days per week

and an average of 4.86 drinks per occasion. A full list of demographics can be found in Table 1.

All methods and procedures were reviewed by the university's Institutional Review Board.

**Measures**

           **Alcohol Use Disorder Symptom Indicators.** AUD symptom indicators for each of the

11 DSM-5 criteria were derived and/or adapted from pre-existing diagnostic interviews and self-

report scales, with the goal of identifying a range of symptom indicator operationalizations for

each criterion. Measures examined included the Alcohol Use Disorder and Associated

Disabilities Interview Schedule-IV and -5 (AUDADIS-IV, -5; Grant, Dawson & Hasin, 2001;

Grant et al., 2011), the Structured Clinical Interview for DSM-IV (SCID-I; Chung & Martin,

2005), the Semi-Structured Assessment for the Genetics of Alcoholism-IV (SSAGA-IV; Reich,

Herjanic, Welner, & Gandhy, 1982), the Alcohol Dependence Scale (ADS; Skinner & Allen,

1982), the Composite International Diagnostic Interview (CIDI; Robins et al., 1989; Cottler,

2000), the Alcohol Urge Questionnaire (AUQ; Bohn, Krahn, & Staehler, 1995), the Obsessive-

Compulsive Drinking Scale (OCDS; Kranzler, Mulgrew, Modesto-Lowe & Burleson, 1999), the

Short Alcohol Dependence Data Questionnaire (SADD; Raistrick, Dunbar, & Davidson, 1983),

and the Michigan Alcohol Screening Test (MAST; Selzer, 1971). See supplemental material for

a full description of the preliminary studies conducted to identify the final set of AUD indicators.

           **Alcohol Consumption.** Seven items, adapted from the IMPACTS survey (Sher &

Rutledge, 2007; NIAAA, 2003), were used to assess alcohol consumption in the past 12 months.

Items measured frequency (e.g. "how often did you drink alcohol?"), binge frequency (e.g. "how

many times have you had 5+ [male] or 4+ [female] drinks in a single sitting?"), frequency of 12

or more drinks in a single sitting (e.g. "how many times have you had 12 or more drinks in one

sitting?"), frequency high (e.g., "how often were you 'buzzed' or light-headed due to alcohol

consumption?"), and frequency of drunkenness (e.g. "how often were you drunk?"). These items

were evaluated on a nine-point scale as follows: 0 = "did not occur/did not drink in past 12

months," 1 = "1-5 times a year", 2 = "6-11 times a year," 3 = "about once a month," 4 = "2-3

times a month," 5 = "once or twice a week," 6 = "3-4 times a week," 7 = "5-6 times a week," 8 =

"every day." Quantity (e.g. "…how many drinks did you usually have on any one occasion?")

was assessed on an 11-point scale from "did not drink" to "12 or more drinks." Maximum drinks

(e.g., "what is the maximum amount of drinks you have had in one sitting?") was assessed via

text entry with a possible range of 0 to 98. The variable was truncated to a set value of 36 for the

sake of plausibility ($n$ = 10 reports above 36 were winsorized to a value of 36).

To have a meaningful standard for assessing validity, we created a composite measure of

past year consumption by summing the standardized values for quantity, frequency, binge

frequency, frequency of 12 or more drinks in a sitting, frequency high, frequency of

drunkenness, and maximum drinks ($\alpha$ = 0.89). This decision was made because none of our

AUD correlates alone are thought of as a "gold standard," each holds considerable unique

variance, and each indexes a range of important drinking behaviors. This "heaviness" composite

is considered appropriate and suitable for the purpose of the current study because: (1) heavy use

over time is the most parsimonious construct for explaining the neurobiological changes that

occur with substance use disorders and for contextualizing the varied social and physical

consequences that occur in substance users (Grant et al., 2009; Rehm & Roerecke, 2013a; Rehm

et al., 2013b) and (2) heaviness of consumption shows a strong monotonic relationship with the

DSM-5 criterion count and this relation is more robust than other possible alternative correlates

(e.g., general functioning, psychiatric comorbidity; see Dawson, Saha, & Grant, 2010; Lane &

Sher, 2015; Saha et al., 2007). Other work has demonstrated that factor scores derived from a

comparable past 12-month consumption composite are heritable, influenced by genetic factors that influence heavy drinking, and stable across time (Agrawal et al., 2011).

Several other constructs including hangover, blackout, self- and other-identified problem use, treatment utilization, legal problems, sexual consequences, and general functioning (via the SF-12) were assessed for the purposes of the external validation analyses. The details of these measures can be found in supplemental material.

**Data Analysis**

**Item Response Theory Analyses.** We applied 11 separate one-parameter (1PL) logistic IRT models, one for each AUD criterion, to estimate item difficulty (severity) using the MLR estimator in Mplus 8 (Brown, 2006). The severity parameter represents the point along the latent-trait continuum at which 50% of the sample is likely to endorse an item. It provides a rank ordering of the severity of each item along the continuum (Embretson & Reise, 2000). While a two-parameter logistic model would allow estimation of item discrimination as well, this significantly complicates the goal of selecting high, medium, and low severity items because it introduces the additional complexity of discrimination. As in previous work (e.g., Kahler & Strong, 2006), the 1PL model is preferred given the assumption of equal discrimination across items which, given adequate fit, provides a more parsimonious description of the data and is consistent with how AUD items are used in practice (i.e., determined to be present/absent without weighting based on discrimination).

The primary assumptions of IRT are local independence (i.e., responses are independent of one another) and unidimensionality (i.e., the 11 criteria approximate unidimensional structure). When unidimensionality is met, the assumption of local independence is also likely to be met (Embretson & Reise, 2000). To confirm that these assumptions were met, we conducted a Confirmatory Factor Analysis (CFA) on the item indicators for each of the 11 AUD criteria (i.e.,

89 items total) and specified a single AUD factor using the WLSMV estimator. This model

indicated a unidimensional AUD construct with standardized factor loadings ranging from 0.40

to 0.97. Model fit indices were adequate: $\chi2$ [3827] = 5514.24, p < .0001, and RMSEA = 0.03[1].

Next, to select symptom indicators for each of the 11 AUD criteria to be included in the

new diagnostic scores, a series of 11 IRT analyses, one per criterion, were conducted using the

procedure described above. Individual item endorsement rates and severity parameter estimates

can be found in Supplemental Table 1. Model fit and individual item fit was acceptable in the

majority of cases (see Supplemental Table 1). In those cases where model fit was marginal, or on

the single occasion nominal, this may be explained by the low number of items for some criteria

and the high correlation among items assessing the same criterion. Misfit or marginal fit should

not be considered problematic, however, given that the goal of the current IRT analyses was item

selection in the service of developing an instrument that would ultimately demonstrate desirable

model fit. That is, we were interested in items that showed empirically low, moderate, and high

severity, and not in illustrating that the intentionally diverse initial item sets were

psychometrically sound. Each of the initial IRT analyses supported unidimensionality and

generally did fit well, which allowed us to focus more on selecting items that provided more

complete coverage of the entire distribution of the latent trait (i.e., generalizability).

Based on the 11 1PL IRT models, three symptom indicators, representing a range of

severities (relatively low to high), were chosen for each of the 11 criteria (Table 2). For each

criterion, we selected three items representing relatively low, medium, and high latent severity

(represented by the "Rank" column in Table 2; 1 = low and 3 = high. For example, an IRT

analysis was run on the six symptom indicators for the larger/longer criterion. Each of the

symptoms fell along a latent severity continuum, indicated by the difficulty parameter. Three

indicators across this range of difficulty were chosen for use in the next stage of analyses[2];

"Have you gone out drinking and ended up drinking more than you intended?" (low), "Have you

gotten drunk when you didn't mean to?" (medium), and "After taking one or two drinks, have

you been UNABLE to stop?" (high).

## Results

The results of the 11 separate IRT analyses demonstrate that each set of symptom

indicators across the criteria produced a wide range of base rates and item severities (Table 2).

For some criteria, it was intrinsically more difficult to achieve a broad range of item severities.

For example, withdrawal is such an inherently severe criterion that the range in difficulties (i.e.,

the IRT severity parameter) was restricted (1.40-3.00), while for other criteria, such as

social/interpersonal, this was less of a problem (0.76-3.41; See Supplemental Table 1). We were

unable to have comparable ranges of thresholds across all criteria for this reason. However,

regardless of the total range of observed thresholds, the three chosen indicators were assigned a

value one (low), two (medium), or three (high). Trait coverage was poor for some criteria, so

items were selected such that they best covered the range.

To evaluate the plausibility of a unidimensional structure of these 33 items, a higher-

order CFA was conducted (see Figure 1)[3] using the WLSMV estimator in Mplus 8 (Brown,

2006). Eleven primary latent variables representing each of the DSM-5 AUD criteria were

estimated along with a higher order AUD factor indicted by the 11 primary factors. Outliers ($n =$

10) were detected using the OUTINFL option and were excluded for observations > 4, resulting

in a total sample size of 536. The model demonstrated adequate fit ($\chi^2$ [551] = 2959.35, $p <$

.0001; RMSEA = 0.09), suggesting the 33 items indicated by the IRT analyses can be described

by a unidimensional AUD factor.

**Diagnostic Scores**

The first stage of the analyses involved creating several diagnostic severity scores. These scores included a traditional DSM-5 AUD criterion count, a criterion count, a symptom count, and three bi-factor scores. Each score represented a different approach to scaling AUD severity.

**DSM-5 Alcohol Use Disorder.** Using the AUDADIS-5 algorithm following the DSM-5 description of criteria, a DSM-5 AUD criterion count was produced (Grant et al., 2003). Symptoms of a given criterion must have occurred 3+ times in the past year to be considered "present." This standard is based on the scoring algorithm used in SSAGA-IV (Bucholz et al., 1994). This score ranged from 0 to 11 (all items endorsed) ($M = 2.41$, $SD = 2.27$, $\alpha = 0.75$).

**Criterion Count.** Using results from the IRT analyses described above (see Table 2), 33 items (3 for each AUD criterion), were used to derive a criterion count ranging from 0 to 11 ($M = 2.68$, $SD = 2.56$, $\alpha = 0.79$). If any of the three items for each criterion were endorsed, the participant received a '1' for that criterion. Consistent with the decision rule employed by the Collaborative Studies on Genetics of Alcoholism (COGA; Bierut et al., 1998) with the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), symptoms of a given criterion must have occurred 3+ times in the past year to be considered as "present." The total number of AUD criterion endorsed were summed to create a criterion count.

**Symptom Count.** Using results from the IRT analyses (see Table 2), the 33 items identified were used to derive a symptom count ranging from 0 to 33 ($M = 3.81$, $SD = 4.48$, range = 0-29, $\alpha = 0.89$). Symptoms of a given criterion must have occurred 3+ times in the past year to be considered as "present." The total number of items endorsed were summed to create a symptom count[4].

**Bi-Factor Scores.** Bi-factor scores were derived through a series of steps, beginning with an exploratory bi-factor analysis of the 33 items[3], utilizing the bi-quartimax estimator, in Mplus

8 (Muthen & Muthen, 2017). The exploratory bi-factor model is a factor analysis approach that

uses the covariance among a set of items to derive (1) a single general factor reflecting the

common variance among all items with similar content and (2) a set of group factors or "residual

factors" reflecting the common variance among subsets of items with similar content, above and

beyond the general factor. The general factor represents the construct the instrument intends to

measure and the residual factors represent conceptually specific subdomain constructs (Reise,

2012; Rodriguez, Reise, & Haviland, 2016). This approach is well-suited for representing the

multidimensionality and heterogeneity of broad constructs such as AUD. The best fitting model

(see Table 3) was a three-factor solution ($\chi^2$ [432] = 540.69, $p < .001$; CFI = 0.96; TLI = 0.95;

RMSEA = 0.02) with one general factor and two residual factors. To derive factor scores from

the exploratory bi-factor analysis, an exploratory structural equation modeling (eSEM) approach

was used, resulting in three factor scores per participant (i.e., general [$M = 0.13$, $SD = 0.82$],

residual 1 [$M = 0.02$, $SD = 0.68$], and residual 2 [$M = -0.06$, $SD = 0.63$]) [5]. The general factor can

be labeled as a general AUD factor while residual 1 and residual 2 are consistent with the notions

of compulsive use and physiological dependence (i.e., tolerance and withdrawal), respectively.

As an additional validation check of our exploratory model, a confirmatory factor

analysis (CFA) was conducted, utilizing the holdout sample described above ($N = 363$), to

replicate the three-factor EFA solution. This resulted in acceptable model fit ($\chi^2$ [475] = 801.56,

$p < .0001$; CFI = 0.88; TLI = 0.87; RMSEA = 0.04).

**Implications for Prevalence Rates**

Prior to looking at correlates of the scores, we examined the prevalences implied by each

of the scoring approaches. For comparative purposes, the DSM-5 algorithm gave a prevalence of

57.9% in this collegiate sample, of which the majority (30.0%) were indicated as mild (i.e., 2-3

criteria). Similarly, when our criterion score approach was used, a roughly comparable

prevalence of 59.3% was found, with slightly less than half (27.8%) being in the mild category.

The prevalences of AUD monotonically decreased as a function of systematically higher

symptom difficulties based on the IRT severity rankings. For example, if we only counted those

items indicative of the lowest severity (i.e., rank = 1; "Low AUD" in Table 5) for each of the 11

criteria, 52.4% of the sample would diagnose with slightly more than half (27.1%) in the mild

category. In contrast, only 11.4% would diagnose if we only counted those items indicative of

the highest severity (i.e., rank = 3; "Severe AUD" in Table 5). Unlike prevalences based on less

severe symptom grading, the proportion of mild cases (i.e., criterion counts of 2 or 3) represented

only 7.9% of the sample. These findings are consistent with expectations and the larger literature

examining IRT differences across interviews (Lane et al., 2016), and demonstrate the significant

differences in prevalence rates based on the severity of the items used. For comparison purposes,

these additional scores were included in the external validation analyses (Table 5).

**Association among Diagnostic Scores**

Table 4 shows the intercorrelations among the DSM-5 AUD criterion count and the new

diagnostic scores described above. All four general scores (i.e., DSM-5 criterion count, criterion

count, symptom count, and general factor of the bi-factor model) were highly correlated with

each other ($r$ range = 0.85-0.93). By design, the residual scores were not strongly related to the

bi-factor general score ($r = 0.08$) and only modestly related to the other overall scores.

**External Validation**

To examine the external validity of our newly derived scoring methods, we conducted a

series of hierarchical regression analyses across the AUD correlates described above (see Table

5). The outcomes of interest included alcohol consumption, hangover, blackout, self- and other-

identified problem use, treatment usage, legal problems, sexual consequences, and the SF-12

(Ware et al., 2002). Sex was included as a covariate in each model. Change in $R^2$ ($\Delta R^2$) was

examined to determine whether any of the new scores resulted in incremental validity over the

DSM-5. Further, the standardized regression estimates for the new score and DSM-5 were

compared to determine if there were significant differences between the two.

Results from the external validation (Table 5) demonstrate that the new diagnostic scores

are generally similar to the DSM-5 but do possess incremental validity over DSM-5 in some

cases. For example, each of the new scores resulted in significant incremental validity in

predicting consumption, hangover, and blackout. However, when the standardized regression

estimates were compared, the new scores and DSM-5 were either (1) not significantly different,

or (2) the DSM-5 estimate was significantly larger than that of the new score (e.g., criterion

count predicting treatment usage). Note that, when the Moderate, and Severe AUD diagnoses

(i.e., those that were derived by using only items ranked 2 or 3, respectively) were subjected to

the external validation, results differed slightly. Similar to the other scores, the Moderate and

Severe AUD diagnoses resulted in significant incremental validity over DSM-5 across the

majority of the external validators. In most cases where there was a significant difference

between the DSM-5 and new score standardized estimates, the new score estimates were larger.

**Discussion**

Although the DSM-5 uses a criterion count to index AUD severity, there has been very

little consideration of heterogeneity related to the AUD criteria themselves and their

operationalizations. Although alternatives to criterion counts (e.g., latent severity scores derived

from IRT; Dawson & Grant, 2010) have been explored, symptom-severity heterogeneity within

each criterion has been neglected. As noted earlier, recent evidence highlights considerable

variation in criterion severities across instruments (Lane et al., 2016) and variations in criterion

operationalization have dramatic effects in estimating diagnostic structure (Hoffman et al., 2018)

which, in turn, impacts the reliability and validity of actual diagnosis.

Not surprisingly, the estimated prevalences of AUD in this sample systematically varied as a function of the severity of the items used to assess each criterion and the more severe diagnoses (i.e., Moderate and Severe AUD in Table 5) were more related to outcomes such as treatment usage and legal problems. However, accounting for variation in the latent severity of individual criteria (i.e. symptom severity) did not improve the measurement of severity at the syndromal level (i.e. latent severity) when judged from the standpoint of external validity coefficients. To some extent, this speaks to the general robustness of a linear model for characterizing multi-indicator constructs (Wiggins, 1973). It is worth noting that, in some cases, these scores did demonstrate incremental validity over the DSM-5. The lack of consistent improvement in predicting relevant outcomes using the IRT-based scores may be explained by the conceptual issues with a dimensional approach to AUD diagnosis. For example, since disorder is a two-dimensional concept consisting of both harmful symptoms of significant severity and underlying dysfunction (i.e., failure of a psychological mechanism to perform its biological function), Wakefield (2016) argues severity dimensions are insufficient indicators of disorder because they fail to consider the inherent dysfunction that may be causing the symptoms. Wakefield (1997) also suggests that when harmful symptoms are mislabeled disorders without also requiring dysfunction, false positives result.

The finding that the Moderate and Severe AUD diagnoses resulted in significant incremental validity over DSM-5 suggests that more clear operationalization of AUD with a range of severities for each criterion allows us to better discriminate individuals with more significant AUDs, particularly in our sample of college students. This is important both theoretically and practically. In the diagnostic imposter/orphan literature, there seems to be a tendency to assume that the nuance between diagnostic impostors and orphans (i.e.,

sensitivity/specificity) occurs at the transition to mild AUD. The assumption is that if someone has a severe AUD, they endorse all problems, and any heterogeneity in presentation is assumed to be random and not clinically relevant. In contrast, these results suggest that if researchers and clinicians took the time to precisely measure the severity continuum of AUD, one could predict meaningful differences (i.e., certain types of "really severe" are worse than others). From a clinical standpoint, this could be potentially important for identifying those with severe AUDs that might benefit from treatment or some sort of intervention and those that that might require more serious treatment options or alternatives, possibly resulting in more effective triage practices. Given the median age of onset for AUD has been estimated to be 20 years of age (Kessler et al., 2005), this is a particularly critical time for identifying at-risk college students.

It is important to remember that there are additional conceptual problems with the DSM-5 criteria set that the current study fails to address, which may, in part, explain the lack of significant findings. One problem is that symptom indicators are loosely operationalized, with no standardization across diagnostic instruments. The result of this is that some symptom indicators may imply more than one diagnostic criterion. For example, the item, "Was there ever a period when your drinking or being intoxicated interfered with your responsibilities at work, school, or at home?", might be conceptualized as a failure to fulfill role obligations, a significant amount of time spent drinking, being sick, or even getting over the effects of alcohol, and giving up or reducing important social, occupational, or recreational activities. Martin et al. (2014) have argued that items such as these create a problem regarding which diagnostic criteria are being met by a given symptom, and the potential problem of "double-dipping." This state of affairs jeopardizes the validity of diagnosis and the assessment of severity via a symptom count.

Despite the general ambiguities inherent in the current diagnostic enterprise, we maintain there is considerable value in continuing to attempt to resolve diagnostic severity by attending to the severities of symptoms endorsed or observed, particularly in college students. As noted in the case of depression, there seems to be clear recognition that symptoms vary in their severity and have implications for syndromal severity. Practically, failure to consider specific operationalizations of criteria (an issue closely allied with the problem of symptom severity) appears to lead to reduced comparability across studies (Lane et al., 2016) and inconsistent empirical findings (Hoffman et al., 2018). Clinically, this suggests that the current diagnostic algorithm for AUD may not be capturing the entire range of severities, which may be resulting in at-risk cases going undetected. For college students, this means that they might "fall under the radar" during a time when it might be particularly beneficial to intervene with treatment services or interventions. This is problematic given research demonstrating the effectiveness of interventions aimed at reducing college student drinking and related problems, particularly over long intervals (e.g., Carey, Scott-Sheldon, Carey, & DeMartini, 2007).

These conclusions should be considered within the context of the study's limitations. The most notable limitation includes the collegiate nature of the sample, which significantly limits generalization to other types of populations. Slutske (2005) demonstrated that (1) college attendance is associated with experiencing consequences of drinking to a greater extent than those not enrolled in college and (2) college attendance is associated with higher rates of alcohol abuse when compared to non-college attending peers. This is potentially because certain symptoms are less relevant or noticeable in a college environment, or because college provides some sort of protective effect again certain alcohol-related problems (Slutske, 2005). Regardless of the explanations for these differences, this suggests that collegiate samples may not be the

most representative of the general population given college students are only a fraction of the population. This sample is further limited by geographical location and lack of diversity, as most the participants were White. Therefore, these results should be generalized with caution.

An additional concern with the collegiate nature of the sample is related to concerns about false positive diagnosis of AUD, specifically alcohol dependence (DSM-IV). For example, Caetano and Babor (2006) suggested that young adults may be reporting higher rates of tolerance and withdrawal symptoms, when compared to older adults. They hypothesized that this may be because of the wording of structured interview schedules, which leads to a misunderstanding of binge drinking (i.e., acute intoxication), and its corollaries with physical symptoms of alcohol dependence (i.e., withdrawal). The authors instruct clinicians and researchers to remain cautious of the tendency of structured interviews to classify young adults as alcohol-dependent and to consider the possibility of measurement error in their assessments. In fact, Boness, Lane, and Sher (2016), demonstrated that use of the AUDADIS-IV overestimates the prevalence of withdrawal in college students by failing to adequately distinguish withdrawal from hangover. Given the high mean of our criterion count variables and, by extension our high prevalence rates, we suspect that many, if not most, of our undergraduate drinkers do not have problems of clinical severity. However, we note this was still a fairly heavy drinking sample. Although only requiring that participants drank on 12 separate occasions in the past year, participants reported drinking a mean of 5 drinks per occasion, on average bingeing on a weekly basis, and only a small minority reporting having never been drunk (~3%). Thus, while our prevalence rates are high they are not necessarily inconsistent with the heavy-drinking of the sample and the assessment strategy employed. Another research group that recruited a heavy-drinking college student sample found nearly 63% met criteria for DSM-5 AUD (Rinker & Neighbors, 2015) and, in general, research

has demonstrated that college students have higher rates of heavy drinking and alcohol use disorders when compared to their non-college attending peers (e.g., Dawson et al., 2004; Slutske, 2005). We also note that the mode of administration (i.e., a web-based survey) precluded use by a clinician who would be well-trained to appropriately probe individual responses to ensure participants were interpreting and responding to the questions correctly.

Future research should include a broader range and larger number of participants, both population- and clinically-based samples, to ensure representation from the entire range of alcohol use behaviors and consequences. Broader representation could identify items better able to cover the latent trait associated with each criterion and provide greater separation between low, moderate, and severe difficulties at the symptom-level. It is possible that this would have resulted in the latent severity scores performing better and accounting for more variability in consumption. Further, given there is no "gold standard" of external criteria for AUD severity, future work should include additional clinically relevant indicators of overall well-being, distress, and functioning to examine the external validity of these alternative scoring methods.

Nevertheless, results of the current paper point to the potential importance of considering individual symptom indicator severities (i.e., symptom severities) in estimating the prevalence of AUD. The wide range of symptom indicator severities observed in this population suggests that some AUD criteria may be more inherently severe than others. These are important considerations for prevalence because current DSM-5 algorithm, which is often used to estimate population prevalence, fails to consider the differential performance of symptom indicators across criteria. Therefore, prevalence estimates may change based on the symptom indicators used and the population assessed. Consequently, the analyses presented in the current paper are intended to not only expose problems with the current approach to DSM-5 AUD diagnosis but

also as a potential means for conceptualizing diagnostic performance and enhancing symptom

indicators, criteria sets, and algorithms in college students. This approach may also have

important implications for data harmonization across studies, which has become a scientific

priority in recent years.

From a clinical perspective, these results suggest careful consideration of the instruments

and tools used to assess AUD, particularly how estimated prevalence rates may differ as a result

of the items used to operationalize the criteria. While DSM-5 is currently the gold standard, this

work demonstrates that clinicians might consider the assessment of other indicators of AUD

severity (e.g., treatment utilization, injury and sexual assault) in order to bolster the diagnostic

severity indicated by a criterion count. This is particularly true if the clinician is unaware of the

relative severity of items used to assess each criterion. If high-severity items are used to assess a

given criterion, individuals at risk may fall through the cracks despite clinically relevant

symptomatology. Assessment of severity is important for determining next steps in treatment, for

example, and is therefore paramount to clinical diagnosis. The clinical utility of any given AUD

assessment depends, in part, on how accurately the assessment identifies those in need of

treatment or those at risk for future problems. A further implication of this work may be in the

development of screening tools that utilize lower severity items to detect individuals, particularly

college students, who may be overlooked by instruments relying on more severe items. Such

screening tools could be used in college health centers, for example. Overall, this work

demonstrates an important starting place for examining alternative approaches to scaling AUD

severity and offers evidence for the usefulness of further exploring this issue.

<div align="center">References</div>

Agrawal, A., Grant, J. D., Littlefield, A., Waldron, M., Pergadia, M. L., ... & Todd, R. D. (2009).

    *Developing a quantitative measure of alcohol consumption for genomic studies on*

    *prospective cohorts. Journal of Studies on Alcohol and Drugs, 70*(2), 157-168.

Agrawal, A., Lynskey, M. T., Heath, A. C., & Chassin, L. (2011). Developing a genetically

    informative measure of alcohol consumption using past-12-month indices. *Journal of*

    *Studies on Alcohol and Drugs*, *72*(3), 444-452.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*

    (5th Ed.). Arlington: VA: American Psychiatric Publishing.

American Psychiatric Association. (1994). *Diagnostic and statistical manual of mental disorders*

    (4th Ed.). Washington, DC: Author.

American Psychiatric Association. (1987). *Diagnostic and statistical manual of mental disorders*

    (3rd Ed., text rev.). Arlington, VA: Author.

American Psychiatric Association. (1980). *Diagnostic and statistical manual of mental disorders*

    (3rd Ed.). Arlington, VA: Author.

Apgar, V. (1953).  A proposal for the new method of evaluation of the newborn infant. *Current*

    *Researches in Anesthesia and Analgesia, 32*(4), 260-267.

Beseler, C. L., Taylor, L. A., & Leeman, R. F. (2010). An item-response theory analysis of

    DSM-IV alcohol-use disorder criteria and "binge" drinking in undergraduates. *Journal of*

    *Studies on Alcohol and Drugs, 71*(3), 418-423.

Bierut, L. J., Dinwiddie, S. H., Begleiter, H., Crowe, R. R., Hesselbrock, V., Nurnberger, J. I., ...

    & Reich, T. (1998). Familial transmission of substance dependence: Alcohol, marijuana,

    cocaine, and habitual smoking: A report from the Collaborative Study on the Genetics of

    Alcoholism. *Archives of General Psychiatry, 55*(11), 982-988.

Bohn, M. J., Krahn, D. D., & Staehler, B. A. (1995). Development and initial validation of a measure of drinking urges in abstinent alcoholics. *Alcoholism: Clinical and Experimental Research, 19*(3), 600–60.

Boness, C. L., Lane, S. P., & Sher, K. J. (2016). Assessment of withdrawal and hangover is confounded in the Alcohol Use Disorder and Associated Disabilities Interview Schedule: Withdrawal prevalence is likely inflated. *Alcoholism: Clinical and Experimental Research, 5*, 1–9.

Brown, T. (2006). *Confirmatory factor analysis for applied research.* New York: Guildford.

Bucholz, K. K., Cadaret, R., Cloniger, C. R., Dinwiddie, …, & Schuckit, M. A. (1994). A new semi-structured psychiatric interview for use in genetic linkage studies: A report on the reliability of the SSAGA. *Journal of Studies on Alcohol*, *55,* 149-158.

Bucholz, K. K., Heath, A.C., & Madden, P.A.F. (2000). Transitions in drinking in adolescent females: Evidence from the Missouri Adolescent Female Twin Study. *Alcohol Clinical Experimental Research, 24,* 914-923.

Caetano, R., & Babor, T. F. (2006). Diagnosis of alcohol dependence in epidemiological surveys: an epidemic of youthful alcohol dependence or a case of measurement error? *Addiction*, *101*(s1), 111-114.

Carey, K. B., Scott-Sheldon, L. A., Carey, M. P., & DeMartini, K. S. (2007). Individual-level interventions to reduce college student drinking: A meta-analytic review. *Addictive Behaviors, 32*(11), 2469-2494.

Chung, T. & Martin, C.S. (2005). What were they thinking? Adolescents' interpretations of DSM-IV alcohol dependence symptom queries and implications for diagnostic validity. *Drug and Alcohol Dependence, 80,* 191-200.

Chung, T., Martin, C. S., Winters, K. C., & Langenbucher, J. W. (2001). Assessment of alcohol tolerance in adolescents. *Journal of Studies on Alcohol, 62*(5), 687-695.

Colton Meier, S. L., Fitzgerald, K. M., Pardo, S. T., & Babcock, J. (2011). The effects of hormonal gender affirmation treatment on mental health in female-to-male transsexuals. *Journal of Gay & Lesbian Mental Health*, *15*(3), 281-299.

Cooper, L.D., & Balsis, S. (2009). When less is more: How fewer diagnostic criteria can indicate greater severity. *Psychological Assessment, 21*(3), 285-293.

Cottler, L. B. (2000). Composite International Diagnostic Interview (CIDI) - Substance Abuse Module (SAM). St. Louis, MO: Department of Psychiatry, Washington University.

Dawson, D. A., & Grant, B. F. (2010). Should symptom frequency be factored into scalar measures of alcohol use disorder severity? *Addiction, 105,* 1568-1579.

Dawson, D.A., Grant, B.F., Stinson, F.S., & Chou, P.S. (2004). Another look at heavy episodic drinking and alcohol use disorders among college and noncollege youth. *Journal of Studies on Alcohol, 65*(4), 477-488.

Dawson, D. A., Saha, T. D., & Grant, B. F. (2010). A multidimensional assessment of the validity and utility of alcohol use disorder severity as determined by item response theory models. *Drug and Alcohol Dependence, 107*, 31-38.

Deas, D., Riggs, P., Langenbucher, J., Goldman, M., & Brown, S. (2000). Adolescents are not adults: Developmental considerations in alcohol users. *ACER, 24*, 232–237

Edwards, G., & Gross, M. M. (1976). Alcohol dependence: provisional description of a clinical syndrome. *British Medical Journal, 1,* 1058-1061.

Embretson, S.E., & Reise, S. (2000). *Item Response Theory of Psychologists.* Mahwah, NJ: Erlbaum Publishers.

Forbes, M. K., Wright, A. G., Markon, K. E., & Krueger, R. F. (2017). Evidence that

        psychopathology symptom networks have limited replicability. *Journal of Abnormal*

        *Psychology, 126*(7), 969-988.

Funk, G., Kirchheiner, K., Burghuber, O. C., & Hartl, S. (2009). BODE index versus GOLD

        classification for explaining anxious and depressive symptoms in patients with COPD- a

        cross-sectional study. *Respiratory Research, 10*(1), 1-8.

Grant, J. D., Agrawal, A., Bucholz, K. K., Madden, P. A., Pergadia, M. L., Nelson, E. C., ... &

        Whitfield, J. B. (2009). Alcohol consumption indices of genetic risk for alcohol

        dependence. *Biological Psychiatry*, *66*(8), 795-800.

Grant, B.F., Dawson, D., & Hasin, D. (2001). The Alcohol Use Disorder and Associated

        Disabilities Interview Schedule-IV (AUDADIS-IV). *NIAAA, Rockville, MD.*

Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, S. P., … & Pickering, R. P. (2004). The 12-

        month prevalence and trends in DSM-IV alcohol abuse and dependence: United States,

        1991-1992 and 2001-2002. *Drug and Alcohol Dependence, 74*(3), 223–234.

Grant, B. F., Dawson, D. A., Stinson, F. S., Chou, P. S., Kay, W., & Pickering R. (2003). The

        Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (AUDADIS-

        IV): Reliability of alcohol consumption, tobacco use, family history of depression and

        psychiatric diagnostic modules in a general population sample. *Drug and Alcohol*

        *Dependence, 71*(1), 7-16.

Grant, B. F., Goldstein, R. B., Chou, S. P., Saha, T. D., Ruan, W. J., Huang, B., ... & Aivadyan,

        C. (2011). The Alcohol Use Disorder and Associated Disabilities Interview Schedule-

        Diagnostic and Statistical Manual of Mental Disorders, Version (AUDADIS-5).

        *Rockville, MD: National Institute on Alcohol Abuse and Alcoholism.*

Grant, B. F., Goldstein, R. B., Saha, T. D., Chou, S. P., Jung, J., Zhang, H., ... & Hasin, D. S.

    (2015). Epidemiology of DSM-5 alcohol use disorder: Results from the National

    Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry, 72*(8),

    757-766.

Hagman, B. T., Cohn, A. M., Schonfeld, L., Moore, K., & Barrett, B. (2014). College students

    who endorse a sub-threshold number of DSM-5 alcohol use disorder criteria: Alcohol,

    tobacco, and illicit drug use in DSM-5 diagnostic orphans. *American Journal on

    Addictions, 23*(4), 378–385.

Hagman, B. T., & Cohn, A. M. (2009). Toward DSM-V: Mapping the alcohol use disorder

    continuum in college students, *Drug and Alcohol Dependence, 6*(813), 247–253.

Hagman, B. T., & Cohn, A. M. (2011). Toward DSM-V: Mapping the alcohol use disorder

    continuum in college students. *Drug and Alcohol Dependence, 118*(2–3), 202–208.

Harford, T. C., Grant, B. F., Yi, H., & Chen, C. M. (2005). Patterns of DSM-IV alcohol abuse

    and dependence criteria among adolescents and adults: Results From the 2001 National

    Household Survey on Drug Abuse. *ACER, 29*(5), 810–828.

Harford, T. C., Yi, H. Y., Faden, V. B., & Chen, C. M. (2009). The dimensionality of DSM-IV

    alcohol use disorders among adolescent and adult drinkers and symptom patterns by age,

    gender, and race/ethnicity. *ACER, 33*(5), 868–878.

Hasin, D. (2003). Classification of alcohol use disorders. *Alcohol Research and Health, 27*, 5-17.

Hasin, D. S., O'Brien, C. P., Auriacombe, M., Borges, G., Bucholz, K., Budney, A., Compton,

    W.M., Crowley, T., Ling, W., Petry, N. M., Schuckit, M., & Grant, B. F. (2013). DSM-5

    criteria for substance use disorders: Recommendations and rationale. *American Journal

    of Psychiatry*, *170*(8), 834-851.

Hoffman, M., Steinley, D., Trull, T. J., & Sher, K. J. (2018). Criteria definitions and network

relations: The importance of criterion thresholds. *Clinical Psych Science*, *6*, 506-516.

Hurlbut, S. C., & Sher, K. J. (1992). Assessing alcohol problems in college students. *Journal of American College Health, 41*, 49-58.

Hussong, A. M., Curran, P. J., & Bauer, D. J. (2013). Integrative data analysis in clinical psychology research. *Annual Review of Clinical Psychology, 9*, 61-89.

Ioannidis, J.P. (2005). Why most published research findings are false. *PLoS Medicine*, *2*, e124.

Jalali, R., & Rezaei, M. (2014). A comparison of the Glasgow Coma Scale score with Full Outline of Unresponsiveness Scale to predict patients' traumatic brain injury outcomes in intensive care units. *Critical Care Research and Practice,* 1-4.

Kahler, C. W., & Strong, D. R. (2006). A Rasch model analysis of DSM-IV alcohol abuse and dependence items in the National Epidemiological Survey on Alcohol and Related Conditions. *Alcoholism: Clinical and Experimental Research, 30*(7), 1165-1175.

Kessler, R. C., Berglund, P., Demler, O., Jin, R., Merikangas, K. R., & Walters, E. E. (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry, 62*(6), 593-602.

Kranzler, H.R., Mulgrew, C.L., Modesto-Lowe, V., & Burleson, J.A. (1999). Validity of the Obsessive Compulsive Drinking Scale (OCDS): Does Craving Predict Drinking Behavior? *Alcoholism: Clinical and Experimental Research, 23,* 108-114.

Lane, S. P., & Sher, K. J. (2015). Limits of current approaches to diagnosing severity based on criterion counts: An example with DSM-5 Alcohol Use Disorder. *Clinical Psych Science, 3*(6), 819-835.

Lane, S. P., Steinley, D., & Sher, K. J. (2016). Meta-analysis of DSM alcohol use disorder

criteria severities: Structural consistency is only 'skin deep'. *Psych Medicine*, *46*(8),

1769-1784.

Lange, P., Marott, J. L., Vestbo, J., Olsen, K. R., Ingebrigtsen, T. S., Dahl, M., & Nordestgaard,

B. G. (2012). Prediction of the clinical course of Chronic Obtrusive Pulmonary Disease,

using the new GOLD classification. *Am. Journal of Respiratory & Critical Care

Medicine, 186,* 975-981.

Li, T., Hewitt, B. G., & Grant, B. F. (2007). The Alcohol Dependence Syndrome, 30 years later:

A commentary. *Addiction, 102,* 1522-1530.

Martin, C. S., Kaczynski, N. A., Maisto, S. A., Bukstein, O. M., & Moss, H. B. (1995). Patterns

of DSM-IV alcohol abuse and dependence symptoms in adolescent drinkers. *Journal of

Studies on Alcohol, 56*(6), 672–680.

Martin, C. S., Langenbucher, J. W., Chung, T., & Sher, K. J. (2014). Truth or consequences in

the diagnosis of substance use disorders. *Addiction*, *109*(11), 1773-1778.

Mewton, L., Teesson, M., Slade, T., & Cottler, L. (2011). Psychometric performance of DSM-IV

alcohol use disorders in young adulthood: Evidence from an Australian general

population sample. *Journal of Studies on Alcohol and Drugs, 72*, 811–822.

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus User's Guide (8ᵗʰ Ed.)*. Los Angeles, CA.

National Cancer Institute (2013). *Cancer staging*. Retrieved from

http://www.cancer.gov/cancertopics/factsheet/detection/staging

National Institute of Health: MedlinePlus (2011). *APGAR*. Retrieved from

http://www.nlm.nih.gov/medlineplus/ency/article/003402.htm

NIAAA Council Task Force on Recommended Alcohol Questions (2003). Recommended
    alcohol questions. Retrieved from http://www.niaaa.nih.gov/research/guidelines-and-
    resources/recommended-alcohol-questions

Paulhus, D.L., Vazire, S. (2005). The self-report method. In R.W. Robins, R.C. Fraley, & R.F.
    Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 224–
    239). New York: Guilford

Raistrick, D., Dunbar, G., & Davidson, R. (1983). Development of a questionnaire to measure
    alcohol dependence. *British Journal of Addiction, 78*, 89-95.

Rehm, J. & Roerecke, M. (2013a). Reduction of drinking in problem drinkers and all-cause
    mortality. *Alcohol and Alcoholism, 48*, 509-513.

Rehm, J., Marmet, S., Anderson, P., Gual, A., Kraus, L., ... & Gmel, G. (2013b). Defining
    substance use disorders: do we really need more than heavy use? *Alcohol and
    Alcoholism, 48*, 633-640.

Reich, W., Herjanic, B., Welner, Z., & Gandhy, P. R. (1982). Development of a structured
    psychiatric interview for children: Agreement on diagnosis comparing child and parent
    interviews. *Journal of Abnormal Child Psychology, 10,* 325–326.

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral
    Research*, *47*(5), 667–696.

Rinker, D. V., & Neighbors, C. (2015). Latent class analysis of DSM-5 Alcohol Use Disorder
    criteria among heavy-drinking college students. *Journal of Substance Abuse Treatment,
    57*, 81–88.

Robins, L.N., Cottler, L., Bucholz, K., and Compton, W. (1995). Diagnostic Interview Schedule
    for DSM-IV (DIS-IV). St Louis, MO: Department of Psychiatry, Washington University.

Robertson, B.M., Piasecki, T.M., Slutske, W. S., Wood, P. K., Sher, K. J., Shiffman, S., &
    Heath, A. C. (2012). Validity of the Hangover Symptoms Scale: Evidence from an
    electronic diary study. *Alcoholism: Clinical and Experimental Research*, *36*(1), 171-177.

Robins, L.N., Wing, J., Wittchen, H.U., Helzer, J.E.. Babor, T. F., Burke, J., Farmer, A., ... &
    Towle, L. H. (1989). The Composite International Diagnostic Interview: An
    epidemiologic instrument suitable for use in conjunction with different diagnostic
    systems and in different cultures. *Arch Gen Psychiatry, 45*, 1069-1077.

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating
    and interpreting statistical indices. *Psychological Methods, 21*(2), 137-150.

Rose, J. S., Lee, C. T., Selya, A. S., & Dierker, L. C. (2012). DSM-IV alcohol abuse and
    dependence criteria characteristics for recent onset adolescent drinkers. *Drug and Alcohol
    Dependence, 124*(1–2), 88–94.

Saha, T. D., Stinson, F. S., & Grant, B. F. (2007). The role of alcohol consumption in future
    classifications of alcohol use disorders. *Drug and Alcohol Dependence*, *89,* 82-92.

Schmulewitz, D., Keyes, K., Beseler, C., Aharonovich, E., Aivadyan, C., Spivak, B., & Hasin, D.
    (2010). The dimensionality of alcohol use disorders: Results from Israel. *Drug and
    Alcohol Dependence, 111*(1-2), 146-154.

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American Psychologist*,
    *54*(2), 93-105.

Selzer, M. L. (1971). The Michigan Alcoholism Screening Test: The quest for a new diagnostic
    instrument. *American Journal of Psychiatry*, *127*(12), 1653-1658.

Sher, K. J., & Rutledge, P. (2007). Heavy drinking across the transition to college: Predicting
    first-semester heavy drinking from precollege variables. *Addictive Behavior, 32,* 819-835.

Sher, K. J., Walitzer, K. S., Wood, P. K., & Brent, E. E. (1991). Characteristics of children of alcoholics: putative risk factors, substance use and abuse, and psychopathology. *Journal of Abnormal Psychology*, *100*(4), 427.

Shevlin, M., Miles, J. N. V., Davies, M. N. O., & Walker, S. (2000). Coefficient alpha: A useful indicator of reliability? *Personality and Individual Differences*, *28*(2), 229-237.

Skinner, H.A., & Allen, B.A. (1982). Alcohol dependence syndrome: Measurement and validation. *Journal of Abnormal Psychology, 91*, 199-209.

Slutske, W. S. (2005). Alcohol use disorders among US college students and their non-college-attending peers. *Archives of General Psychiatry, 62*, 321-327.

Wakefield, J. C. (1997). Diagnosing DSM-IV—Part I: DSM-IV and the concept of disorder. *Behaviour research and therapy*, *35*(7), 633-649.

Wakefield, J. C. (2016). Diagnostic issues and controversies in DSM-5: Return of the false positives problem. *Annual Review of Clinical Psychology*, *12*(1), 105–132.

Ware, J. E., Kosinski, M., Turner-Bowker, D. M., & Gandek, B. (2002). User's manual for the SF-12v2 health survey with a supplement documenting SF-12 health survey. *Lincoln, RI: QualityMetric Incorporated*, 1-250.

White, H. R., & Labouvie, E. W. (1989). Toward the assessment of adolescent problem drinking. *Journal of Studies on Alcohol, 50*(50), 30–37.

Wiggins, J. W. (1973). *Personality and prediction: Principles of personality assessment.* Reading, Mass.: Addison-Wesley.

Footnotes

[1] In addition to a CFA with all 89 items, we also conducted 11 separate CFAs specifying a single factor to test for unidimensionality within each AUD criterion. Model fit was adequate in most cases. In those cases where model fit was poorer, this was likely because the items within a single criterion were so highly correlated (see Supplementary Material for details).

[2] It is useful to point out that there are two issues to consider when choosing items: (1) (criterion) trait coverage and (2) discrimination. It is important to ensure that the continuum of severity for each set of items (i.e., items covering the criterion) is maximized. Further, items selected should adequately discriminate between persons in different regions of the latent continuum. Both can be assessed with a 2PL model; however, we used 1PL models, which assume equivalent discrimination across items. This assumption is unlikely to hold but introduces a level of objectivity that is lost when ranking items based on both difficulty and discrimination. Of note, the average correlation between the 1PL and 2PL thresholds across items within a criterion was 0.95, and we confirmed that item selection would not have changed if a 2PL model was adopted.

[3] In some cases (Table 3), the item with the next highest difficulty was chosen because the original IRT-indicated item used in the other diagnostic scores had too low of a base rate. Therefore, the item pool used to estimate the higher-order CFA and derive bi-factor scores was slightly different from that used to derive the other scores described.

[4] In addition to a symptom count, we also created a weighted symptom count score that was ultimately excluded from the presented results (see Supplementary Material for details).

[5] Due to low base rates, the eSEM model excluded the following items to achieve convergence: "Find yourself shaking (when the effects of alcohol were wearing off)?"; "Were you unable to sleep?"; and "Have you ever given up or greatly reduced important activities because of drinking, like participating in sports, work or associating with friends or relatives?".

Table 1

*Demographic Characteristics of Sample (N = 909)*

| Characteristic | Main Sample (*N* = 546) | | | | Holdout Sample (*N* = 363) | | | |
|---|---|---|---|---|---|---|---|---|
| | % | *M* | *SD* | Range | % | *M* | *SD* | Range |
| Age | | 18.66 | 1.00 | 18.00-22.00 | | 18.69 | 1.05 | 18.00-26.00 |
| Sex | | | | | | | | |
|   Male | 44.49 | | | | 43.25 | | | |
| Race | | | | | | | | |
|   White | 91.94 | | | | 91.18 | | | |
|   African-American | 5.49 | | | | 5.51 | | | |
|   Other | 2.56 | | | | 3.31 | | | |
| Relationship Status | | | | | | | | |
|   Married | 0.38 | | | | 0.30 | | | |
|   Divorced | 0.19 | | | | 0.00 | | | |
|   Separated | 0.19 | | | | 0.00 | | | |
|   Never married | 98.09 | | | | 97.92 | | | |
|   Living with partner | 1.15 | | | | 1.78 | | | |
| DSM-5 Alcohol Use Disorder | 57.88 | 2.41 | 2.27 | 0.00-11.00 | 55.92 | 2.37 | 2.17 | 0.00-8.00 |
| Age of first full drink | | 15.51 | 1.54 | 10.00-21.00 | | 15.49 | 1.57 | 8.00-19.00 |
| Age first drunk | | 16.02 | 1.42 | 10.00-21.00 | | 15.98 | 1.38 | 12.00-19.00 |
| Past year consumption | | | | | | | | |
|   Maximum drinks in single sitting past year | | 11.90 | 6.88 | 1.00-36.00 | | 12.05 | 6.84 | 1.00-36.00 |
|   Drinking days/week | | 1.25 | 1.04 | 0.06-5.00 | | 1.30 | 1.12 | 0.06-5.00 |
|   Average drinks per occasion | | 4.86 | 2.29 | 1.00-12.00 | | 5.10 | 2.45 | 1.00-12.00 |
|   Days "drunk"/week | | 0.79 | 0.88 | 0.00-3.50 | | 0.81 | 0.96 | 0.00-5.00 |
|   Days "buzzed" or "high"/week | | 1.09 | 1.13 | 0.00-7.00 | | 1.04 | 1.06 | 0.00-5.00 |
|   Days with 5+/4+ drinks in single sitting/week | | 0.90 | 1.01 | 0.00-5.00 | | 0.96 | 1.07 | 0.00-5.00 |
|   Days with 12+ drinks in single sitting/week | | 0.31 | 0.76 | 0.00-7.00 | | 0.33 | 0.70 | 0.00-5.00 |

*Note.* Individuals who have never been drunk were excluded from "Age first drunk" for the main sample (*n* = 17; 3.11%) and the holdout sample (*n* = 9; 2.48%).

Table 2

*Alcohol Use Disorder Diagnostic Criteria Symptom Indicators*

| Alcohol Use Disorder Criterion (*N* = 546) | Measure | Base Rate | 1PL Unstandardized Difficulty Estimate | Rank | Standardized Alpha (Unstandardized Alpha) |
|---|---|---|---|---|---|
| **Tolerance** | | | | | |
| Did you ever find that you could drink a lot more before you got drunk? | SSAGA-IV | 21.52% | 0.91 | 1 | |
| Did you ever find that you could drink a lot more before you passed out? | Authors | 15.64% | 1.15 | 2 | |
| Have you ever drank as much as a fifth of liquor in one day, that would be about 20 drinks, or 3 bottles of wine, or as much as 3 six packs of beer in a single day? | AUDADIS-IV; AUDADIS-5 | 8.89% | 1.47 | 3 | |
| *Mean* | | 15.35% | 1.18 | | 0.73(0.73) |
| **Hazardous Use** | | | | | |
| Have you ever been under the influence of alcohol in a situation where it increased your chances of getting hurt - e.g., when riding a bicycle, driving, operating machinery, or anything else? | AUDADIS-IV; AUDADIS-5 | 14.87% | 1.27 | 1 | |
| Have you ever been under the influence of alcohol and put yourself in a risky situation where it increased your chances of being victimized - e.g. being physically or sexually assaulted, or robbed? | Authors | 7.81% | 1.75 | 2 | |
| Driven a car when you knew you had too much to drink to drive safely? | AUDADIS-5 | 3.54% | 2.25 | 3 | |
| *Mean* | | 8.74 % | 1.76 | | 0.64(0.64) |
| **Quit/Cut Down** | | | | | |
| Have you thought that you should quit or cut down on your drinking, whether you tried to or not? | Authors | 10.24% | 1.56 | 1 | |
| Have you ever tried to stop or cut down on your drinking? | SSAGA-IV; AUDADIS-IV | 7.42% | 1.79 | 2 | |
| Have you tried to stop or cut down on drinking but were unsuccessful? | AUDADIS-IV; AUDADIS-5 | 2.59% | 2.44 | 3 | |
| *Mean* | | 6.75% | 1.93 | | 0.71(0.71) |
| **Time Spent** | | | | | |
| Did you ever have a period where you spent a lot of time drinking? | AUDADIS-IV; AUDADIS-5 | 28.94% | 0.72 | 1 | |
| Did you have a period when you spent a lot of time being sick or getting over the bad after-effects of alcohol? | AUDADIS-IV; AUDADIS-5 | 21.30% | 1.03 | 2 | |
| Has there ever been a period of several days or more when you spent so much time drinking or recovering from the effects of alcohol that you had little time for anything else? | SSAGA-IV; CIDI | 12.43% | 1.49 | 3 | |
| *Mean* | | 20.89% | 1.08 | | 0.60(0.58) |
| **Social/Interpersonal** | | | | | |
| When drinking, have you said or done things in the presence of others that was regrettable or embarrassing? | Authors | 28.65% | 0.76 | 1 | |
| Have you ever been told that you become rude or obnoxious when drinking? | Authors | 5.74% | 2.15 | 2 | |
| Have you ever gotten into physical fights when drinking? | MAST; SSAGA-IV; AUDADIS-IV; AUDADIS-5 | 0.92% | 3.29 | 3 | |
| *Mean* | | 11.77% | 2.07 | | 0.37(0.30) |
| **Larger/Longer** | | | | | |
| Have you gone out drinking and ended up drinking more than you intended? | AUDADIS-IV; AUDADIS-5 | 38.93% | 0.34 | 1 | |
| Have you gotten drunk when you didn't mean to? | Authors | 17.86% | 1.12 | 2 | |
| After taking one or two drinks, have you been unable to stop? | ADS | 13.65% | 1.32 | 3 | |
| *Mean* | | 23.48% | 0.93 | | 0.59(0.58) |

| | | | | | |
|---|---|---|---|---|---|
| **Craving** | | | | | |
| Have you ever had times when you thought about drinking or how good a drink would make you feel? | Authors | 39.89% | 0.30 | 1 | |
| Have you ever gotten impatient when you were ready to have a drink/felt like having a drink and you couldn't (i.e., situational factors prevented you from drinking at that time)? | Authors | 7.62% | 1.66 | 2 | |
| Have you ever wanted a drink so badly that you couldn't think of anything else? | AUDADIS-5; SSAGA-IV | 2.60% | 2.31 | 3 | |
| *Mean* | | 16.70% | 1.42 | | 0.55(0.40) |
| **Give up** | | | | | |
| Have you ever felt that activities you used to enjoy without drinking alcohol are now only enjoyable if drinking alcohol is involved? | Authors | 9.06% | 1.52 | 1 | |
| Have you ever given up or cut down on activities that you were interested in or that gave your pleasure in order to drink? | AUDADIS-IV; AUDADIS-5 | 3.14% | 2.14 | 2 | |
| Have you ever had a period of time when alcohol was such a focus that you didn't even think about the things you usually enjoy doing? | Authors | 1.85% | 2.43 | 3 | |
| *Mean* | | 4.68% | 2.03 | | 0.65(0.54) |
| **Failure to Fulfill** | | | | | |
| Was there ever a period when being hungover or sick from drinking interfered with your responsibilities at work, school, or at home? | CIDI | 10.91% | 1.31 | 1 | |
| Was there ever a period when your drinking or being intoxicated interfered with your responsibilities at work, school, or at home? | CIDI; SSAGA-IV; AUDADIS-IV; AUDADIS-5 | 5.38% | 1.68 | 2 | |
| Have you ever gotten into trouble at work, home, or school because of being hungover or being sick from drinking? | Authors | 3.34% | 1.90 | 3 | |
| *Mean* | | 6.54% | 1.63 | | 0.59(0.58) |
| **Physical/Psychological** | | | | | |
| Have you ever continued to drink when feeling overwhelmed or burnt out even though you knew drinking would make you feel worse? | Authors | 7.61% | 1.60 | 1 | |
| Have you ever continued to drink alcohol despite having a physical illness or psychological problem that gets worse with drinking? | CIDI | 5.19% | 1.81 | 2 | |
| Have you ever continued to drink when you knew you had a serious physical illness that might be made worse by drinking? | SSAGA-IV; CIDI | 3.33% | 2.02 | 3 | |
| *Mean* | | 5.38% | 1.81 | | 0.65(0.62) |
| **Withdrawal** | | | | | |
| After drinking alcohol, did you EVER take a drink or use any drug or medicine, OTHER THAN aspirin, Advil or Tylenol, to GET OVER any bad after effects of drinking? | AUDADIS-IV; AUDADIS-5 | 11.38% | 1.40 | 1 | |
| Were you unable to sleep? | SSAGA-IV | 2.25% | 2.27 | 2 | |
| Find yourself shaking (when the effects of alcohol were wearing off)? | AUDADIS-IV | 1.12% | 2.62 | 3 | |
| *Mean* | | 4.92% | 2.10 | | 0.54(0.32) |

*Note.* 1PL = one-parameter logistic. Relative rank: 1 = low, 2= medium, and 3 = high severity. Larger/Longer = "Alcohol is taken in larger amounts or over longer periods than was intended;" Time Spent = "A great deal of time is spent in activities necessary to obtain alcohol, use alcohol, or recover from its effects;" Quit/Cut Down = "There is a persistent desire or unsuccessful efforts to cut down or control alcohol use;" Craving = "Craving, or a strong desire or urge to use alcohol;" Failure to Fulfill = "Recurrent alcohol use resulting in a failure to fulfill major role obligations at work, school, or home;" Give Up = "Important social, occupational, or recreational activities given up or reduced because of alcohol use;" Hazardous Use = "Recurrent alcohol use in situations in which it is physically hazardous." Physical/Psychological = "Alcohol use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by alcohol." Tolerance = "Tolerance, as defined by either of the following: (a) A need for markedly increased amounts of alcohol to achieve intoxication or desired effect (b) A markedly diminished effect with continued use of the same amount of alcohol." Withdrawal = "Withdrawal, as manifested by either of the following: (a) The characteristic withdrawal syndrome for alcohol, (b) Alcohol (or a closely related substance, such as benzodiazepine) is taken to relieve or avoid withdrawal symptoms." Social/Interpersonal = "Continued alcohol use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of alcohol." AUDADIS-IV = Alcohol Use Disorder and Associated Disabilities Interview Schedule-IV (Grant, Dawson, & Hasin, 2001); AUDADIS-5 = Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (Grant et al., 2011); SSAGA-IV = Semi-Structured Assessment for the Genetics of Alcoholism-IV (Reich, Herjanic, Welner, & Gandhy, 1982); CIDI = Composite International Diagnostic Interview (Robins et al., 1989; Cottler, 2000); MAST = Michigan Alcohol Screening Test (Selzer, 1971); ADS = Alcohol Dependence Scale (Skinner & Allen, 1982); OCDS = Obsessive-Compulsive Drinking Scale (Kranzler, Mulgrew, Modesto-Lowe & Burleson, 1999); AUQ = Alcohol Urge Questionnaire (Bohn, Krahn, & Staehler, 1995); SADD = Short Alcohol Dependence Data Questionnaire (Raistrick, Dunbar, & Davidson, 1983). Authors = created by authors.

Table 3

*Factor Loadings for 3-Factor Exploratory Factor Analysis with a BI-CF-Quartimax Rotation (N = 536)*

| Variable | General Factor | Residual 1 | Residual 2 |
|---|---|---|---|
| After taking one or two drinks, have you been unable to stop? | 0.476* | 0.211* | -0.125 |
| Have you gotten drunk when you didn't mean to? | 0.668* | 0.282* | 0.003 |
| Have you gone out drinking and ended up drinking more than you intended? | 0.723* | 0.243* | -0.047 |
| Did you ever have a period where you spent a lot of time drinking? | 0.663* | 0.252* | 0.207* |
| Has there ever been a period of several days or more when you spent so much time drinking or recovering from the effects of alcohol that you had little time for anything else? | 0.664* | 0.267* | 0.062 |
| Did you have a period when you spent a lot of time being sick or getting over the bad after-effects of alcohol? | 0.712* | 0.046 | 0.018 |
| Have you thought that you should quit or cut down on your drinking, whether you tried to or not? | 0.414* | 0.756* | -0.055 |
| Have you ever tried to stop or cut down on your drinking? | 0.330* | 0.926* | 0.040 |
| Some people find it hard to go several days without drinking. Have you ever found it difficult to go without drinking for an extended period of time? [a] | 0.518* | -0.023 | 0.449* |
| Have you ever had times when you thought about drinking or how good a drink would make you feel? | 0.650* | -0.099 | 0.162* |
| Have you ever wanted a drink so badly that you couldn't think of anything else? | 0.604* | 0.006 | 0.538* |
| Have you ever gotten impatient when you were ready to have a drink/felt like having a drink and you couldn't (i.e., situational factors prevented you from drinking at that time)? | 0.611* | -0.253* | 0.478* |
| Was there ever a period when your drinking or being intoxicated interfered with your responsibilities at work, school, or at home? | 0.792* | -0.138 | -0.021 |
| Have you ever gotten into trouble at work, home, or school because of being hungover or being sick from drinking? | 0.797* | 0.103 | -0.003 |
| Was there ever a period when being hungover or sick from drinking interfered with your responsibilities at work, school, or at home? | 0.833* | -0.180 | 0.023 |
| Have you ever given up or cut down on activities that you were interested in or that gave your pleasure in order to drink? | 0.418* | -0.014 | 0.433* |
| Have you ever felt that activities you used to enjoy without drinking alcohol are now only enjoyable if drinking alcohol is involved? | 0.490* | -0.209 | 0.258* |
| Have you ever given up or greatly reduced important activities because of drinking, like participating in sports, work or associating with friends or relatives? [a] | 0.348 | 0.072 | 0.479* |
| Have you ever been under the influence of alcohol in a situation where it increased your chances of getting hurt - e.g., when riding a bicycle, driving, operating machinery, or anything else? | 0.598* | -0.547* | -0.054 |
| Have you ever been under the influence of alcohol and put yourself in a risky situation where it increased your chances of being victimized - e.g. being physically or sexually assaulted, or robbed? | 0.670* | -0.195 | -0.003 |
| Driven a car when you knew you had too much to drink to drive safely? | 0.653* | -0.666* | -0.029 |
| Have you ever continued to drink when you knew you had a serious physical illness that might be made worse by drinking? | 0.726* | 0.138 | -0.268 |
| Have you ever continued to drink alcohol despite having a physical illness or psychological problem that gets worse with drinking? | 0.683* | 0.298* | -0.304* |
| Have you ever continued to drink when feeling overwhelmed or burnt out even though you knew drinking would make you feel worse? | 0.686* | 0.127 | -0.060 |
| Have you ever drank as much as a fifth of liquor in one day, that would be about 20 drinks, or 3 bottles of wine, or as much as 3 six packs of beer in a single day? | 0.616* | -0.002 | 0.394* |
| Did you ever find that you could drink a lot more before you got drunk? | 0.504* | 0.049 | 0.763* |
| Did you ever find that you could drink a lot more before you passed out? | 0.467* | -0.029 | 0.787* |
| Have you ever gotten into physical fights when drinking? | 0.334* | -0.126* | 0.393* |
| When drinking, have you said or done things in the presence of others that was regrettable or embarrassing? | 0.642* | -0.001 | 0.083 |
| Have you ever been told that you become rude or obnoxious when drinking? | 0.491* | -0.292* | -0.034 |
| Find yourself shaking (when the effects of alcohol were wearing off)? | 0.219 | 0.007 | 0.983* |
| Were you unable to sleep? | 0.210 | 0.106 | 0.734* |
| After drinking alcohol, did you EVER take a drink or use any drug or medicine, OTHER THAN aspirin, Advil or Tylenol, to GET OVER any bad after effects of drinking? | 0.479* | -0.19 | 0.301* |

*Note.* $\chi^2$ [432] = 540.69*; CFI = 0.96; TLI = 0.95; RMSEA = 0.02. * $p < .05$. [a] Denotes cases in which an alternative variable (i.e., the item with the next highest 1PL IRT difficulty; see Supplemental Table 1) had to be chosen because the original IRT-indicated item used in the other diagnostic scores had too low of a base rate to be included in the model. For items that significantly loaded on Residual 1 and Residual 2, these items were only included in Residual 2 when used to generate the bi-factor scores.

Table 4

*Correlation Table of Diagnostic Scores (N = 546)*

| | | 1 | 2 | 3 | Bi-Factor 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| | 1 DSM-5 AUD (0-11) | - | 0.89* | 0.88* | 0.86* | 0.20* | 0.10* |
| | 2 Criterion Count (0-11) | | - | 0.93* | 0.86* | 0.22* | 0.11* |
| | 3 Symptom Count (0-33) | | | - | 0.85* | 0.20* | 0.11* |
| Bi-Factor | 4 General AUD Factor | | | | - | 0.08 | 0.07 |
| | 5 Residual Factor 1: Compulsive Use | | | | | - | 0.15* |
| | 6 Residual Factor 2: Physiological Dependence | | | | | | - |

*Note.* Correlates among DSM-5 Alcohol Use Disorder (AUD) and the five new scores. DSM-5 AUD is based on the AUDADIS-5 (Grant et al., 2011) algorithm. Criterion Count is based on the 33 IRT-derived items whereby if any of the three items for each of the 11 criteria were endorsed, the participant received a '1' for that AUD criterion (range: 0-11). Symptom Count uses the 33 IRT-derived items whereby the total number of items endorsed were summed to create a symptom count ranging from 0 to 33. The Bi-Factor scores include three separate factor scores based on a bi-factor model (see Table 3).

Table 5

*Incremental Validity of New AUD Diagnostic Scores Compared to DSM-5 AUD (N = 546)*

| External Validator | DSM-5 (0-11) $r^2$ | $R^2$ | Criterion Count (0-11) $r^2$ | Std. Parameter Estimate | $\Delta R^2$ | Symptom Count (0-33) $r^2$ | Std. Parameter Estimate | $\Delta R^2$ | Bi-Factor $r^2$ General AUD Factor | Std. Parameter Estimate | $r^2$ Resid. 1: CU | Std. Parameter Estimate | $r^2$ Resid. 2: PD | Std. Parameter Estimate | $\Sigma r^2$ | $\Delta R^2$ | Low AUD (0-11) $r^2$ | Std. Parameter Estimate | $\Delta R^2$ | Moderate AUD (0-11) $r^2$ | Std. Parameter Estimate | $\Delta R^2$ | Severe AUD (0-11) $r^2$ | Std. Parameter Estimate | $\Delta R^2$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Consumption | 0.220 | 0.405 | 0.207 | 0.177 | 0.006 * | 0.213 | 0.218 | 0.011 * | 0.189 | 0.139 | 0.017 | 0.035 | 0.027 | 0.116 | 0.232 | 0.019 * | 0.206 | 0.197 | 0.010 * | 0.153 | 0.070 ₛ | 0.002 | 0.130 | 0.101 | 0.006 * |
| Hangover | 0.096 | 0.108 | 0.099 | 0.190 | 0.007 * | 0.098 | 0.182 | 0.008 * | 0.104 | 0.223 | 0.004 | 0.011 | 0.013 | 0.083 | 0.121 | 0.020 * | 0.082 | 0.076 | 0.002 | 0.096 | 0.173 | 0.012 * | 0.053 | 0.053 | 0.002 |
| Blackout | 0.379 | 0.380 | 0.421 | 0.492 | 0.048 * | 0.399 | 0.396 | 0.036 * | 0.347 | 0.239 | 0.021 | 0.025 ₛ | 0.046 | 0.155 | 0.415 | 0.039 * | 0.417 | 0.446 | 0.052 * | 0.283 | 0.137 ₛ | 0.007 * | 0.193 | 0.074 ₛ | 0.003 |
| Self-identified problem use | 0.189 | 0.192 | 0.152 | 0.008 ₛ | 0.000 | 0.238 | 0.460 | 0.049 * | 0.137 | 0.034 | 0.040 | 0.140 | 0.004 | -0.122 ₛ | 0.181 | 0.029 * | 0.126 | -0.070 ₛ | 0.001 | 0.246 | 0.397 ₗ | 0.063 * | 0.256 | 0.386 ₗ | 0.088 * |
| Other-identified problem use | 0.127 | 0.137 | 0.077 | -0.203 ₛ | 0.008 | 0.116 | 0.123 | 0.004 | 0.065 | -0.158 ₛ | 0.029 | 0.097 | 0.001 | -0.079 ₛ | 0.095 | 0.023 | 0.053 | -0.287 ₛ | 0.022 | 0.125 | 0.196 | 0.015 | 0.133 | 0.230 | 0.031 * |
| Treatment utilization | 0.053 | 0.057 | 0.021 | -0.303 ₛ | 0.018 * | 0.035 | -0.057 ₛ | 0.001 | 0.017 | -0.238 ₛ | 0.006 | 0.026 | 0.006 | -0.102 ₛ | 0.030 | 0.026 * | 0.009 | -0.392 ₛ | 0.041 * | 0.046 | 0.091 | 0.003 | 0.073 | 0.210 ₗ | 0.026 * |
| Legal problems | 0.108 | 0.109 | 0.056 | -0.285 ₛ | 0.016 * | 0.098 | 0.103 | 0.003 | 0.059 | -0.152 ₛ | 0.006 | 0.001 ₛ | 0.000 | -0.016 ₛ | 0.065 | 0.006 | 0.031 | -0.410 ₛ | 0.044 * | 0.103 | 0.166 | 0.011 * | 0.183 | 0.368 ₗ | 0.079 * |
| Sexual consequences | 0.280 | 0.282 | 0.267 | 0.272 | 0.010 * | 0.290 | 0.320 | 0.024 * | 0.181 | -0.098 ₛ | 0.020 | 0.024 ₛ | 0.005 | 0.016 ₛ | 0.207 | 0.004 | 0.223 | 0.067 | 0.001 * | 0.250 | 0.224 | 0.020 * | 0.224 | 0.225 | 0.029 * |
| SF-12† | 0.059 | 0.063 | 0.063 | 0.170 | 0.005 | 0.054 | 0.077 | 0.001 | 0.045 | 0.039 | 0.016 | 0.091 | 0.001 | -0.058 ₛ | 0.061 | 0.010 | 0.001 | 0.006 | 0.000 | 0.015 | -0.237 ₛ | 0.021 * | 0.008 | -0.107 | 0.007 |

*Note.* Models control for sex. Consumption is standardized. Std. = standardized. $\Delta R^2$ is the change in model fit when a given score is added to a model already including sex and DSM-5 alcohol use disorder. Resid. = residual factor. CU = Compulsive Use, PD = Physiological Dependence. ₛ = significantly smaller than standardized DSM-5 estimate. ₗ = significantly larger than standardized DSM-5 estimate. *$p$ < .05. † = Each of the 8 SF-12 subscales were also evaluated and the pattern of results were the same, therefore, an overall SF-12 composite is presented. Low AUD represents a criteria count whereby only those items indicative of the lowest severity (i.e., rank = 1) for each of the 11 criteria were counted. Moderate AUD only counts those of moderate severity (i.e., rank = 2), and Severe AUD only counts those with the highest severity (i.e., rank = 3).
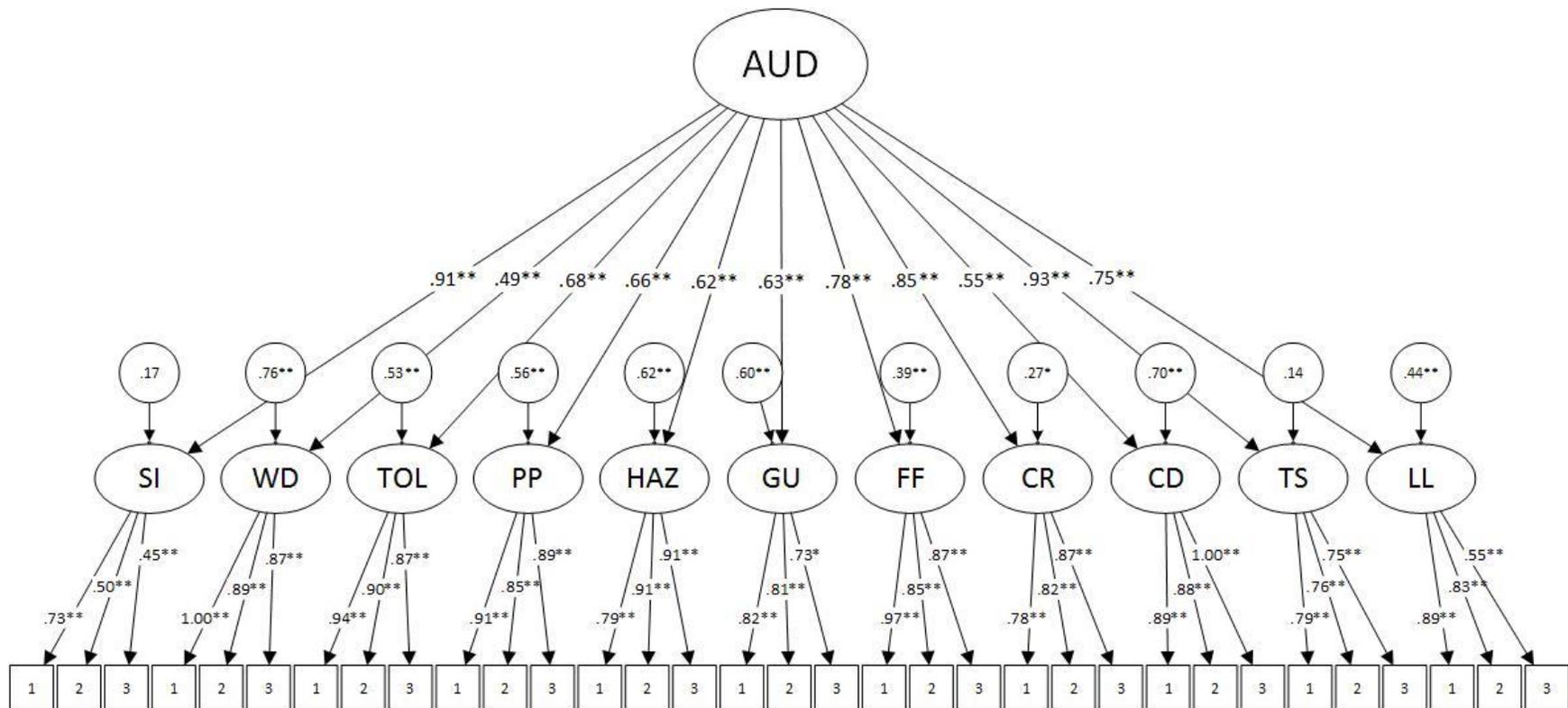
*Figure 1.* Higher-Order Confirmatory Factor Analysis of the 33 Alcohol Use Disorder Symptom Indicators. This figure demonstrates the higher-order model specified to demonstrate the unidimensionality of the 33 alcohol use disorder symptom indicators chosen from the IRT-models to indicate low, medium, and high severity. Table 2 lists full descriptions of the 33 items ranging from low to high severity. All estimates are standardized. Estimator = WLSMV. $\chi^2 [551] = 2959.35$, $p < .0001$; RMSEA = 0.09. $**p < .001$, $*p < .01$. LL = "Alcohol is taken in larger amounts or over longer periods than was intended;" TS = "A great deal of time is spent in activities necessary to obtain alcohol, use alcohol, or recover from its effects;" CD = "There is a persistent desire or unsuccessful efforts to cut down or control alcohol use;" CR = "Craving, or a strong desire or urge to use alcohol;" FF = "Recurrent alcohol use resulting in a failure to fulfill major role obligations at work, school, or home;" GU = "Important social, occupational, or recreational activities given up or reduced because of alcohol use;" HAZ = "Recurrent alcohol use in situations in which it is physically hazardous." PP = "Alcohol use is continued despite knowledge of having a persistent or recurrent physical or psychological problem that is likely to have been caused or exacerbated by alcohol." TOL = "Tolerance, as defined by either of the following: (a) A need for markedly increased amounts of alcohol to achieve intoxication or desired effect (b) A markedly diminished effect with continued use of the same amount of alcohol." WD = "Withdrawal, as manifested by either of the following: (a) The characteristic withdrawal syndrome for alcohol, (b) Alcohol (or a closely related substance, such as benzodiazepine) is taken to relieve or avoid withdrawal symptoms." SI = "Continued alcohol use despite having persistent or recurrent social or interpersonal problems caused or exacerbated by the effects of alcohol."