



Published in final edited form as:

*J Abnorm Psychol.* 2021 July ; 130(5): 512–524. doi:10.1037/abn0000678.

## Does crude measurement contribute to observed unidimensionality of psychological constructs? A demonstration with DSM-5 alcohol use disorder

Ashley L. Watts<sup>1</sup>, Cassandra L. Boness<sup>1</sup>, Jordan E. Loeffelman<sup>1,2</sup>, Douglas Steinley<sup>1</sup>, Kenneth J. Sher<sup>1</sup>

<sup>1</sup>Department of Psychological Sciences, University of Missouri, Columbia, MO

<sup>2</sup>Department of Psychology, University of Oklahoma, Norman, OK

### Abstract

Mental disorders are complex, multifaceted phenomena that are associated with profound heterogeneity and comorbidity. Despite the heterogeneity of mental disorder, most are generally considered unitary dimensions. We argue that certain measurement practices, especially using too few indicators per construct, preclude the detection of meaningful multidimensionality. We demonstrate the implications of crude measurement for detecting construct multidimensionality with alcohol use disorder (AUD). To do so, we used a large sample of college heavy drinkers (N=909) for whom AUD symptomology was thoroughly assessed (87 items) and a blend of confirmatory factor analysis, exploratory factor analysis, and hierarchical clustering. A unidimensional AUD model with one item per symptom criterion fit the data well, whereas a unidimensional model with all items fit the data poorly. Starting with an 11-item AUD model, model fit decreased and the variability in factor loadings increased as additional items were added to the model. Additionally, multidimensional models outperformed unidimensional ones in terms of variance explained in theoretically-relevant external criteria. All told, we converged on a hierarchically-organized model of AUD with three broad, transcriterial dimensions that reflected Tolerance, Withdrawal, and Loss of Control. In addition to introducing a hierarchical model of AUD, we propose that thorough assessment of psychological constructs paired with serious consideration of alternative, multidimensional structures can move past the deadlock of their unidimensional representations.

### General Scientific Summary

We show how crude measurement can essentially determine the inferences we draw about the dimensionality of diagnostic constructs. We use alcohol use disorder as an example of this general

---

Author correspondence: Correspondence should be addressed to Ashley L. Watts, 200 South Seventh Street, Columbia, MO, 65211. ashlewlwatts@gmail.com.

#### Author Contribution Statement

We define author contribution using the Contributor Roles Taxonomy (CRediT). Watts was responsible for conceptualization, formal analysis, methodology, software, validation, visualization, and writing (original draft preparation, lead). Boness was responsible for data curation, investigation, methodology, project administration, resources, validation, and writing (original draft preparation, supporting). Loeffelman was responsible for formal analysis, software, and writing (original draft preparation, supporting). Steinley was responsible for writing (review and editing). Sher was responsible for conceptualization, data curation, investigation, methodology, project administration, resources, supervision, and writing (review and editing).

argument because it is a heterogeneous construct that is assumed to reflect a unitary dimension. With thorough assessment, we propose a hierarchically-organized, multidimensional account of alcohol use disorder.

### Keywords

psychiatric classification; heterogeneity; measurement; factor analysis; hierarchical clustering; alcohol use disorder

---

Mental disorders are often described as unitary dimensions despite the fact that they contain diverse criteria that manifest profoundly heterogeneously. These conclusions often occur as researchers attempt to validate diagnostic criteria sets and observe that a single factor solution with a single item per criterion fits the data well. In turn, researchers conclude that the theoretical construct *per se* is unidimensional in nature, with little consideration of measurement practices that essentially guarantee such observed unidimensionality. Often, the role of measurement is underappreciated in these undertakings. In this paper, we demonstrate how crude measurement of our favored constructs can preclude the detection of their multidimensionality, and how mistakenly treating a multidimensional construct as unidimensional can obscure insights into a construct's causes, correlates, sequelae, and potential intervention targets.

We use alcohol use disorder (AUD) as a demonstration because it is a construct that exemplifies the tension between crude measurement and observed unidimensionality (Hasin et al., 2013). Along the way, we bring to bear data from a large sample of heavy drinkers ( $N=909$ ) for whom AUD was assessed thoroughly (87 items) and use a combination of confirmatory factor analysis, exploratory factor analysis, and hierarchical clustering. We suspected that AUD might best be represented multidimensionally, as a set of distinct dimensions that are potentially sufficiently correlated to have a meaningful higher-order factor.

### AUD and its Heterogeneity

Drinking that becomes severe and impairing is broadly defined as AUD. Currently, DSM-5 requires endorsement of 2 of 11 symptoms (Table 1) to meet criteria for AUD, meaning that there are over 2,000 possible combinations of symptoms and 55 possible combinations of two DSM-5 AUD symptoms alone that are sufficient for a diagnosis (Lane & Sher, 2015). It is therefore likely not surprising that AUD is associated with profound heterogeneity, inasmuch as those diagnosed with AUD vary dramatically in terms of symptom presentation and patterns of consumption, alcohol-related consequences, age of onset, profiles of comorbid psychopathology, family history of alcohol problems, reasons for drinking, and so on (Litten et al., 2015). Thus, an AUD diagnosis in isolation generally affords clinicians little information regarding its treatment and prognosis, severity, course and outcomes, and etiology.

Current consensus promotes a unitary conceptualization of AUD. That is, AUD diagnostic criteria are assumed to indicate a single, coherent latent dimension. A number of variable-

centered (i.e., factor analytic) studies found that AUD symptoms load highly onto a well-fitting single factor (see Hasin et al., 2013, for review). Those using person-centered approaches (e.g., cluster analysis, latent class analysis) tend to find that separate latent classes underlying AUD reflect different levels of severity as opposed to distinct subgroups of individuals who differ in terms of the types of AUD symptoms they experience (Lynskey et al., 2005).

Nevertheless, the decision to conceptualize AUD as unidimensional has received its fair share of criticisms (e.g., MacCoun, 2013; Wakefield, 2015). First, compared with DSM-IV, which included separate alcohol abuse and dependence diagnoses, DSM-5 AUD does not result in improved predictive validity of important external criteria (e.g., frequent heavy drinking, number of symptoms, treatment use; Wakefield & Schmitz, 2015). Second, the DSM has conventionally taken a “theory-free” approach, often ignoring etiology as an important component of a disorder in favor of prioritizing clinical presentation, reliability, and common language (Miller & Rockstroh, 2013). Because of this, unidimensional conceptualizations generally do not accommodate competing etiologic mechanisms described in modern models of addiction, such as the Alcohol Addiction Research Domain Criteria (AARDoC; Litten et al., 2015). More generally, unidimensional conceptualizations are not consistent with the way DSM-5 or other popular models organize substance use disorder symptoms. DSM-5 describes AUD as consisting of four components, impaired control, social impairment, risky use, and pharmacological criteria (APA, 2013; pp. 483-484), and AARDoC focuses on three functional domains, executive function, incentive salience, and negative emotionality. It is unclear how AUD can be both unidimensional and have distinct components. Thus, a tacit implication of these organizational frameworks is that there are homogeneous groupings of AUD features.

Several practices likely either contribute to observed unidimensionality within AUD, or, put another way, preclude the detection of its multidimensionality. First, researchers tend to use single indicators for each AUD symptom, typically diagnostic criteria. Statistically speaking, one cannot extract a factor without a sufficient number of indicators to identify it. To identify a factor in a confirmatory factor analytic (CFA) framework, a factor requires at least three indicators, ideally four or more (Bollen & Lennox, 1991). Additionally, if we theorize that certain complex constructs (e.g., tolerance, withdrawal) are distinct from other aspects of AUD but we model them with a single item, it is nearly impossible to detect separate factors for them. Complicating matters more, a number of diverse symptoms or behaviors can be collapsed or lumped into a single indicator, and such aggregation can result in important loss of information. When modeling depression, many researchers collapse increased or decreased appetite (or sleep) into a single indicator reflecting change in appetite, but these two manifestations of appetite change are likely to have different causes. In short, using single indicators per criterion makes it much more difficult to detect multidimensional solutions and results in an overly simplistic conceptualization of the construct. Second, one can observe unidimensionality within an item set even if the underlying structure is multidimensional (Vanderweele & Batty, 2020). Third, a number of conceptually non-equivalent but equally well-fitting models can be observed in the same data (Tomarken & Waller, 2003), so we cannot draw firm conclusions regarding construct

unidimensionality without thorough assessment and before systematically testing alternative models. This task was not explicitly undertaken by DSM-5 (Wakefield, 2015).

There is limited evidence of AUD's multidimensionality. In a factor analysis of 21 DSM-III-R alcohol abuse and dependence criteria, Muthén and colleagues (1993) identified separable factors reflecting Tolerance, Withdrawal, and Impaired Control. These factors appear remarkably similar to those reported by Kendler and colleagues (2012), who found that alcohol dependence was undergirded by three genetic factors reflecting Tolerance and Excess Consumption, Withdrawal and Continued Use, and Loss of Control/Social Dysfunction. Kendler and colleagues' (2012) factors displayed divergent associations with external criteria. Tolerance/Excess Consumption was a relatively weak marker of alcohol dependence and other psychopathology (e.g., conduct disorder, cannabis dependence, depression). In contrast, Withdrawal and Continued Use and Loss of Control/Social Dysfunction factors were relatively poor markers of alcohol consumption and strong markers of pathology (e.g., treatment seeking, number of alcohol dependence episodes), with the latter associations being most pronounced for Loss of Control/Social Dysfunction. These findings strongly suggest that there are important sources of multidimensionality within AUD that appear to have distinct etiologies and distinguishable relations with external criteria.

Admittedly, we view multidimensional conceptualizations of AUD as especially alluring because they have the potential to: delineate critical sources of heterogeneity within psychiatric diagnoses, refine our ability to construct meaningful symptom profiles, predict course and related outcomes, create evidence-based triage protocols, and develop more personalized treatments. The latter possibility is based on the proposition that the mechanisms underlying different types of symptom profiles directly implicate different targeted treatments (Kendler et al., 2012; Kwako et al., 2016). The first steps towards parsing heterogeneity in psychiatric conditions requires the adoption of procedures to detect multidimensionality, should it exist.

## Present Study

We used a dataset of college drinkers who completed a large item set of AUD indicators (87 items) and substance use-relevant external criteria for two important reasons. First, college drinkers have among the highest rates of AUD (Grant et al., 2015), are in a critical developmental period for AUD onset (Kessler et al., 2005), and may progress to dependence more rapidly than adults (Martin et al., 1995). Thus, college heavy drinkers represent an important population to consider in the context of examining the latent structure of AUD. Studying college heavy drinkers effectively oversamples for tolerance (and bingeing), which is relatively less common in the general population.

Second, these data were collected with the aim of grading indicators of AUD symptoms along severity continua to provide more adequate coverage of the full breadth of latent criterion dimensions. Lane and colleagues (2016) showed in a recent meta-analysis of AUD assessments that within-symptom severities differed dramatically depending on the assessment. So, a withdrawal indicator can probe low-to-middling AUD severity on one

instrument and high severity on another. Thus, Boness and colleagues (2019) included multiple indicators per symptom criterion so as not to confound severity with specific criteria. Ultimately, they derived an abbreviated, 33-item unidimensional assessment of AUD *per se* for clinical settings. We extend this work by focusing on the complete 87-item set to detect multidimensionality within AUD.

## Aims

This study is organized into two broad aims. First, we examined the influence of crude measurement on observed unidimensionality of single, latent AUD factor by systematically adding items to a model that began with 11 indicators, one per criterion. Second, we examined alternatives to the unidimensional AUD structure by means of a thorough combination of confirmatory and exploratory factor analysis and hierarchical clustering. To start, we tested three confirmatory models put forth in the literature (Table 1). Anticipating complex, alternative structures, we ended with exploratory factor analysis and agglomerative hierarchical clustering. We anticipated a hierarchical solution given that psychopathology – and many other psychological constructs – is increasingly described as such (Kotov et al., 2017).

To determine the incremental utility of a multidimensional AUD structure above and beyond a unidimensional one, we compared the performance of our candidate alternative structures with that of two 1-factor conceptualizations of AUD, one with 11 items (1 per criterion) and one with 87 (all AUD items), in terms of variance explained in theoretically-relevant external criteria (i.e., alcohol and cannabis consumption, alcohol-related consequences, overall functioning/health). We also examined the candidate structure's relations with those same external criteria to consider convergent and discriminant validity among its factors.

## Method

### Participants and Procedure

We recruited 909 undergraduates ( $M_{\text{age}}=18.64$ ) enrolled in introduction to psychology courses at the University of Missouri. Participants completed the entire survey battery online and course credit was awarded upon completion. Methods and procedures were approved by the University of Missouri's Institutional Review Board (IRB #1203637). The majority of participants were female (56%) and White (91%; Table S1 includes all demographics). Although participants were simply required to have consumed alcohol on at least 12 separate occasions in the past year, they reported heavy past year drinking with an average of 1.25 weekly drinking days and 4.86 drinks per occasion.

### Measures

**AUD Symptom Assessment**—Boness and colleagues (2019) derived and/or adapted 87 AUD symptom indicators for the 11 DSM-5 criteria from 9 well-validated diagnostic interviews and self-report scales (Table 1 includes the number of items per criterion and Table S2 contains descriptive statistics). Items were rated on a five-point Likert scale (0 = Never/Not in the Past 12 Months, 4 = Yes, 4+ Times). We assessed past-year occurrences given concerns with lifetime report reliability (Haeny et al., 2014).

**Theoretically-relevant External Criteria**—Participants reported on age of first drinking experience, age of first heavy drinking experience, consumption, hangover, blackout, self- and other-identified problem use, treatment use, legal problems, cannabis use frequency, and three composites of functioning: general health, mental health, and physical functioning. To maintain consistency with AUD assessment, participants reported past year occurrences (see Table S3 for descriptive statistics and the Method Supplement for a thorough description of each measure). Details of these measures and their descriptive statistics can be found in the Supplemental Material.

## Data Analysis

All data, code, and Supplemental Materials are available on OSF (<https://osf.io/2vn58/>).

**Factor Analyses**—We conducted factor analyses using Mplus version 7.2 (Muthén & Muthén, 2014) and R version 3.5.1 with the psych (Revelle, 2019) package. We specified indicators as categorical and estimated CFAs with using WLSMV and exploratory structural equation models (ESEM) with target rotations. Latent factor means and variances were fixed at 0 and 1, respectively, as opposed to fixing the first factor loading to 1. We extracted EFAs with geomin rotation and used Horn’s parallel analysis (Horn, 1965) and Velicer’s MAP test (Velicer, 1976) to determine the number of factors in the data.

We relied on a combination of the  $\chi^2$  test statistic, the Comparative Fit Index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) to adjudicate our models given that no consensus exists regarding the use of a single fit index to evaluate the adequacy of model fit. We used the following model fit adequacy guidelines: CFI and TLI > .95 for reasonably good fit (Hu & Bentler, 1999), and RMSEA .08 for adequate fit and .05 for close fit (MacCallum et al., 1996). We did not use  $\chi^2$  and its associated *p* value as an index of absolute model fit given that it is overly sensitive to peripheral factors like sample size but report them in the interest of completeness and transparency. For exploratory solutions, we considered the model fit, solution interpretability, *and* each solution’s incremental utility over its CFA counterpart, given that ESEM solutions can yield preferential fit over CFA solutions due to model flexibility as opposed to increased validity (Hermann & Pfister, 2013).

We report coefficient omega ( $\Omega$ ; Revelle & Zinbarg, 2009), an index of factor reliability that reflects the proportion of variance in the observed total score attributable to all modeled sources of common variance. Given that we conduct CFA, ESEM, and EFA to test the same a priori structures, we report Tucker’s phi ( $\Phi$ ; Tucker, 1951) to examine factor similarity across models (e.g., CFA-ESEM). Like a correlation coefficient,  $\Phi$  ranges from  $-1$  to  $+1$ . For external validity analyses, factor models’ parameters were fixed to their unstandardized estimates prior to examining all factors’ associations with external criteria to prevent shifting of the structural models’ parameters as additional covariates were added to the model. All regression parameters were freely estimated. Age and sex were included as covariates in all regression models.

**Hierarchical Clustering**—We used hierarchical agglomerative cluster analysis, specifically Ward’s method,<sup>1</sup> to map the hierarchical structure of AUD using Matlab R2020a

and the ‘linkage’ function. Hierarchical clustering sorts indicators into a tree structure from the bottom up, into mutually exclusive item clusters based on their similarity, and into progressively broader clusters. We transposed the original  $N \times P$  dataset so that variables were the subject of the clustering and allowed for 2 to 16 clusters. To determine the number of clusters in the data, we considered consensus across three stopping rules (i.e., Calinski-Harabasz, Davies-Bouldin, and Silhouette) available in the ‘evalclusters’ function. A maximum value on the Calinski-Harabasz and Silhouette indices and a minimum on the Davies-Bouldin index determines the recommended number of clusters (Sharma & Batra, 2019). To assess the fit of each solution, we considered the cophenetic correlation (Sokal & Rohlf, 1962), which assesses how well the dendrogram induced by the hierarchical clustering procedure preserves the original pairwise distances between the original variables.

## Results

### Aim 1: Crude Measurement & Unidimensionality

We took two steps to address the influence of crude measurement on observed unidimensionality. First, we compared two 1-factor models of AUD in terms of fit, one with 11 items, one per symptom criterion, and one with 87; for the 11-item model, we counted a symptom criterion as present if any item assessing it was endorsed. The 11-indicator model fit well ( $\chi^2=128$ ; RMSEA=.046, CFI=.962, TLI=.952) but the 87-indicator model did not ( $\chi^2=18225$ ; RMSEA=.066, CFI=.893, TLI=.893; Tables S4-5).

Second, we started with an 11-item model and added randomly-selected items one at a time to examine how unidimensionality – operationalized in terms of model fit, variability in standardized factor loadings, and coefficient omega – varied with the inclusion of additional indicators. We fixed the random selection to draw one item per criterion per 11 items sampled (the only caveat to this is that there were an unequal number of items per criterion, so the models were not equally represented by criteria at later rounds of sampling, at 45+ indicators). Given that items were added to the initial model in a random order, we repeated this procedure 500 times to minimize the influence of item ordering on our inferences. We began with an 11-item model composed of a random mix of low, medium, and high IRT item difficulties from the 33-item set reported by Boness and colleagues (2019; this resulted in 38501 unique models tested [77 models \* 500 repetitions + 1 repetition for the 87-item model]; note that there were 500 initial 11-item solutions, so median fit from those models does not correspond directly to the earlier-described 11-item model). We plot the medians (and standard deviations as error ribbons) of the 500 repetitions for  $n$ -item each model (i.e., 11 items, 12 items).

Figure 1 displays model fit, range in factor loadings, and omega of the 1-factor solution as a function of the number of indicators included in the model. On average, the starting 11-item set fit the data well (median CFI=.937, TLI=.921, SRMR=.075, RMSEA=.066),

<sup>1</sup>We considered two other hierarchical clustering procedures (complete linkage, average linkage) and compared their performances in terms of cross-method agreement. Cophenetic correlations were good for all three methods ( $r_{\text{Ward's}}=.82$ ,  $r_{\text{complete linkage}}=.93$ ,  $r_{\text{average linkage}}=.93$ ). Numerous simulation studies have demonstrated the superiority of Ward’s method over others with respect to producing stable and interpretable clusters (Sharma & Batra, 2019) and Ward’s method performed better when conducting cross-validation for under a split-half design.

but there was variability in terms of fit that depended on which items were included in the starting set (range: CFI=.860-.982, TLI=.825-.977, SRMR=.037-.112, RMSEA=.046-.131). As indicators were added to the model, model fit decreased well below good or acceptable fit, as deemed by decreases in CFI and TLI and increases in SRMR and RMSEA. Effects for certain model fit indices were not linear, however, suggesting that the model appeared more unidimensional at a certain point. In addition, factor loadings became more variable with the inclusion of more indicators. Finally, in contrast, coefficient omega increased linearly as items were added to the model, consistent with the Spearman-Brown prophecy (e.g., Spearman, 1910).

## Aim 2: Alternative, Multidimensional Accounts of AUD

We began with confirmatory analyses that tested a set of *a priori* models<sup>2</sup> (Table 1): correlated (1) *DSM-IV Abuse and Dependence* factors<sup>3</sup> (Muthén, 1995); (2) *Tolerance, Withdrawal, and Loss of Control* factors (Kendler et al., 2012; Muthén et al., 1993), and (3) *Pharmacological, Risky Use, Impaired Control, and Social/Interpersonal* factors (APA, 2013). We then tested these same models using ESEM with target rotation, whereby the number of factors and hypothesized (target) loadings of items onto factors were specified, but all possible cross-loadings of items onto factors are allowed. ESEM with target rotation therefore tests whether the same general *a priori* structures tested using CFA can be retained but reveals potentially important deviations from simple structure not identified with CFA. We then considered whether we could retain the same structures tested with CFA and ESEM with target rotation using a method that imposes no *a priori* restrictions on the data, so we extracted 2-, 3-, and 4-factor EFA solutions (Table S8). Finally, we ended with purely exploratory EFAs and used parallel analysis and the MAP test to determine the number of factors in the data (Figure S1 contains the scree plot and Tables S9-10 contain the factor loadings).

**2-factor Model.**—The 2-factor CFA model fit poorly ( $\chi^2=14304$ ,  $df=2773$ ; RMSEA=.068; CFI=.918; TLI=.915), though its ESEM counterpart fit well ( $\chi^2=9276$ ,  $df=2699$ ; RMSEA=.052; CFI=.953; TLI=.950). The 2-factor ESEM Abuse factor nonetheless bore little resemblance to its CFA counterpart ( $\Phi_{Abuse}=.25$ ,  $\Phi_{Dependence}=.84$ ; Table S6). Similarly, the 2-factor EFA solution contained factors that appeared to reflect (1) Tolerance and Consumption and (2) Urge and Difficulty Controlling Use (Table S4). Thus, the 2-factor confirmatory model was not upheld in more exploratory contexts.

**3-factor Model.**—The 3-factor CFA fit well ( $\chi^2=10589$ ,  $df=3651$ ; RMSEA=.046; CFI=.949; TLI=.948). Its factors were reliable ( $\Omega_s$ .98-.99), well-represented by their indicators (median  $\lambda$ s were .95, .87, .65 for Tolerance, Withdrawal, and Loss of Control, respectively), and significantly intercorrelated, with Withdrawal and Loss of Control

<sup>2</sup>We also tested higher-order versions of the three and four correlated factors models, which included a second-order AUD factor that describes the shared variance among first-order factors. A higher-order model with 2 lower-order factors is not identified, so it could not be tested; the higher-order model with 3 lower-order factors did not converge; and fit for the model with 4 lower-order factors is reported in Table S4. Given that we could not report higher-order models for two of the three conceptual models of interest, we do not focus on higher-order models.

<sup>3</sup>Because the legal problems criterion exhibits differential item functioning with respect to sex and race (Harford et al., 2009) and was removed from DSM-5, we did not include it in our model with the abuse factor, or in our modeling more generally.



being especially highly correlated ( $r_{\text{Tolerance-Withdrawal}}=.44$ ;  $r_{\text{Tolerance-Loss of Control}}=.63$ ;  $r_{\text{Withdrawal-Loss of Control}}=.78$ ). The **3-factor** ESEM also fit well ( $\chi^2=9683$ ,  $df=3483$ ;  $RMSEA=.044$ ;  $CFI=.955$ ;  $TLI=.951$ ) and its factors were correlated with each other to a lesser extent than they were in the CFA model ( $r_{\text{Tolerance-Withdrawal}}=.18$ ;  $r_{\text{Tolerance-Loss of Control}}=.49$ ;  $r_{\text{Withdrawal-Loss of Control}}=.52$ ). Convergent congruence between CFA and ESEM factors was generally high ( $\Phi_{\text{Tolerance}}=.91$ ,  $\Phi_{\text{Withdrawal}}=.78$ ,  $\Phi_{\text{Loss of Control}}=.87$ ), as was the congruence between the CFA and EFA factors ( $\Phi_{\text{Tolerance}}=.71$ ,  $\Phi_{\text{Withdrawal}}=.76$ ,  $\Phi_{\text{Loss of Control}}=.91$ ).

Generally, the 3-factor EFA solution contained factors that appeared to reflect Tolerance, Withdrawal, and Loss of Control, though there were some potentially interesting cross-loadings. There was little evidence of cross-loadings onto the Tolerance factor (defined as  $\lambda = .30$ ), but there was far more evidence for cross-loadings of items between Withdrawal and Loss of Control factors. Craving and continued use items scattered across Withdrawal and Loss of Control factors. Finally, items reflecting giving up activities due to drinking loaded fairly highly ( $\lambda_s \approx .45$ ) onto Withdrawal.

**4-factor Model.**—The 4-factor CFA ( $\chi^2=12254$ ,  $df=3399$ ;  $RMSEA=.051$ ;  $CFI=.937$ ;  $TLI=.935$ ) fit relatively poorly, but the 4-factor ESEM ( $\chi^2=8425$ ,  $df=3648$ ;  $RMSEA=.040$ ;  $CFI=.963$ ;  $TLI=.960$ ) fit well. Generally, the theoretical model was not well-supported (Tables S5-6). In the ESEM, the Pharmacological factor was not well-represented by withdrawal indicators (median  $\lambda_{\text{Withdrawal}}=.32$ ,  $\lambda_{\text{Tolerance}}=.99$ ), and the Social/Interpersonal Problems and Risky Use factors were poorly represented by their indicators (median  $\lambda_s = -.06$ ,  $.36$ , respectively;  $\Omega_s = .08$ ,  $.80$ ). Convergent congruence between CFA and ESEM factors was relatively high for Pharmacological and Impaired Control factors ( $\Phi_s = .77$ ,  $.70$ ), but quite low for Risky Use and Social/Interpersonal factors ( $\Phi_s = -.01$ ,  $.51$ ). The 4-factor EFA solution contained factors that appeared to reflect *Tolerance*, *Withdrawal*, *Loss of Control*, and *Compulsive Use*. This factor solution generated Tolerance and Withdrawal factors that were highly similar to those from the 3-factor solution ( $\Phi_s = .96$ ,  $.88$ ) and partitioned the Loss of Control factor into two factors ( $\Phi_s = .77$ ,  $.66$ ). Although the 4-factor ESEM fit better than the 3-factor ESEM, this was likely due to increased flexibility in the model as opposed to model validity because the ESEM and EFA factors often bore limited resemblance to their CFA counterparts.

**Other alternative solutions.**—Parallel analysis suggested that there were 25 factors in the data and the MAP test suggested 12. Both of these EFA solutions yielded symptom-specific factors (e.g., tolerance) that were moderately correlated with certain other symptom-specific (e.g., larger/longer, time spent) and relatively uncorrelated with others (e.g., craving), indicating the presence of multiple higher-order factors in AUD. Other times, these solutions yielded both broad and narrow factors of the same symptom (e.g., tolerance), suggesting a hierarchical organization within the item set. Thus, we next used hierarchical agglomerative clustering.

The Calinski-Harabasz and Silhouette stopping rules supported a 2-cluster solution whereas the Davies-Bouldin stopping rule supported a 3-cluster solution.<sup>4</sup> We elected to move forward with a 2-cluster solution because the majority of stopping rules indicated a

preference for it, seminal work argues that the Calinski-Harabasz is the preferred method for choosing the number of clusters (Milligan & Cooper, 1985), and the Calinski-Harabasz has a functional form for minimizing the within-cluster variances that is consistent with Ward's method. This solution produced a good cophenetic correlation ( $r = .82$ ), which indicates a strong hierarchical structure. Taken together, stopping rules indicate that there may be at least two distinct clusters of AUD indicators within the data (see Hubert & Steinley [2005], for a method to formally compare hierarchical clustering and unidimensional structures): The variance accounted for by the unidimensional structure was close to 0% (.0017), whereas the variance accounted for by the hierarchical structure was 67% (the squared cophenetic correlation,  $.82^2 = .67$ ).

Because hierarchical agglomerative clustering can produce illusory clusters, especially for subclusters composed of a small number of indicators (Pavan & Pelillo, 2003), we focused on clusters that were composed of at least 4 items and appeared substantive as opposed to methodological (Figure 2, see Figure S2 for the full dendrogram). By methodological, we mean that several clusters split into subclusters based on item severity (e.g., low, high; Boness et al., 2019). Pruning cluster trees can lead to more stable and interpretable solutions (Kpotufe & Von Luxburg, 2011).

Atop the hierarchy were *Urge and Difficulty Controlling Use* (A1) and *Tolerance and Consumption* (A2). *Urge and Difficulty Controlling Use* (A1) split into *Withdrawal and Craving* (B1) and *Loss of Control* (B2). *Withdrawal and Craving* (B1) split into *Withdrawal* (C1) and *Preoccupation* (C2). *Loss of Control* (B2) split into *Risky and Impulsive Use* (C3) and *Compulsive Use* (C4), the latter of which split into *Desire to Control Use* (D1) and *Unintentional Intoxication* (D2). *Tolerance and Consumption* (A2) split into *Tolerance* (B3) and *Consumption and Recovery from It* (B4). *Consumption and Recovery from It* (B4) further split into *Consumption* (C7) and *Time Spent Obtaining, Using, and Recovering from Alcohol* (C8), and *Consumption* (C7) further split into *Exceeds Personal Limits* (D3) and *Low Severity Craving* (D4).

**Summary.:** Of the alternative models tested, only the correlated Tolerance, Withdrawal, and Loss of Control model fit well and was upheld in confirmatory, semi-confirmatory, and exploratory contexts. The EFAs and hierarchical clustering revealed a complex, hierarchically-organized structure, with broad dimensions spanning multiple criteria. In the hierarchical clustering, independent clusters emerged for: (1) Tolerance and Excess Consumption and (2) Urge and Difficulty Controlling Use, which bifurcated into Withdrawal and Loss of Control. Withdrawal and Loss of Control appear intertwined at a superordinate level of the hierarchy.

### Aim 3: External Validity of Alternative AUD Structures

**Incremental Validity—**To determine the utility of multidimensional structures of AUD, we compared the amount of explained variance in theoretically-relevant external criteria for each of our multidimensional solutions (3-factor CFA and ESEM, 2-cluster) with two latent

---

<sup>4</sup>The 3-cluster solution broke the Tolerance/Excess Consumption cluster from the 2-cluster solution into Tolerance and Excess Consumption (see Supplemental Materials).

1-factor solutions, with 11 and 87 items respectively. The 3-factor CFA and ESEM solutions tended to outperform both the 1-factor models in terms of variance explained in external criteria, though this was not always the case (Figure S3). The 2-cluster solution tended to outperform the 3-factor CFA and ESEM solutions, as well. On average, the 3-factor CFA and ESEM solutions explained  $\approx 15\text{-}16\%$  of the variance in external criteria, the 2-cluster solution explained 21%, and the 1-factor CFAs with 11 and 87 items explained  $\approx 12$  and 14%, respectively. The absolute differences in explained variance correspond to a relative increment of 1.1 and 1.4 times the amount of explained by the 11- and 87-item 1-factor CFAs by the 3-factor CFA and ESEM solutions, and 1.8 and 1.5 times for the 2-cluster solution.

**Differential Validity of AUD Factors**—We present only on the 3-factor CFA solution's external validity to prevent overinterpreting findings from the exploratory models (Figure 3; see Table S10 for ESEM and clustering results). External validity includes regressions in which each external criterion was regressed onto each AUD CFA factor in isolation (zero-order associations) and onto all factors simultaneously (partial associations).

On balance, when entered into the model in isolation, AUD factors did not relate differentially with external criteria, although the Tolerance factor was more strongly associated with consumption, blackout, other-identified problem use, and treatment use than were Withdrawal and Loss of Control. In contrast, when AUD factors were entered into the model simultaneously, they more often than not displayed diverging relations with external criteria. Loss of Control tended to emerge as the strongest correlate of age of first drink, consumption, hangover, blackout, cannabis use frequency, sexual consequences, legal problems, and poor general health. Tolerance was most strongly positively associated with consumption and blackout, albeit to a lesser extent than Loss of Control, decreased legal problems, and increased physical functioning. Withdrawal was most strongly associated with later age of first drunk and first time being drunk, and hangover, whereas it was negatively associated with consumption, blackout, and cannabis use frequency.

## Discussion

With thorough AUD assessment and a systematic set of tests of alternative structural models using confirmatory and exploratory analysis, we converged on a multidimensional, hierarchically-organized structure of AUD composed of three broad, transcriterial dimensions or facets. By all accounts, a unidimensional account of AUD was not supported in factor or cluster analyses. Factor analyses indicated that a unidimensional structure fit the data poorly in both an absolute and relative sense, and hierarchical clustering further indicated that there may be at least two mutually exclusive clusters of items at the top of the hierarchy. We also showed that researchers can artificially create unidimensional models of constructs by choosing to include few items in their models. As symptom criteria are modeled with more than one indicator, our models may become more multidimensional. Indeed, a multidimensional, three-factor solution fit well, and outperformed the unidimensional solution in terms of variance explained in external criteria. In our view, there was clear multidimensionality within the data.

## Alternative Structures of AUD

What is slightly less clear is what accounts for the meaningful sources of multidimensionality within AUD. We can say with relative confidence that the two-factor Abuse and Dependence model was not at all supported by the data: the confirmatory model fit poorly and yielded highly correlated Abuse and Dependence factors, neither of which were evidenced in exploratory endeavors. These findings are consistent with considerable research. Although DSM-IV distinguished between Abuse and Dependence, they were merged in DSM-5 because they are highly correlated and are not differentiable in terms of severity (Proudfoot et al., 2006).

Also, DSM-5's conceptual organization of substance use disorder criteria into four groupings – Pharmacological, Impaired Control, Risky Use, Social/Interpersonal – was not supported. Factors from the four-factor CFA and ESEM did not resemble each other. Withdrawal and Tolerance did not cohere into a single factor in any solution, and neither Impaired Control nor Risky Use factors were retained in any exploratory solution. Based on these data, DSM-5's groupings appear more conceptual than empirical.

Our candidate alternative structure of AUD contained three factors: Tolerance and Excess Consumption, Withdrawal, and Loss of Control. Kendler and colleagues (2012) first identified this structure for alcohol dependence, and we attempted to conceptually extend it to DSM-5 AUD. Kendler and colleagues' model was generally corroborated, although our findings highlighted the intertwined nature of withdrawal and loss of control. Withdrawal and Loss of Control factors shared items that loaded onto both, remained highly correlated after allowing item cross-loadings, and formed a superordinate cluster in the hierarchical clustering that appeared to reflect Urge and Difficulty Controlling Use. Note, however, that Withdrawal and Loss of Control were distinct in the penultimate level of the hierarchy. Popular models of alcohol addiction (Koob & Le Moal, 2001) describe the important interplay between withdrawal and loss of control, although it is difficult to characterize the precise nature of their relationship without longitudinal data. The differentiation of Tolerance, Withdrawal, and Loss of Control factors across stages of the disease process remains high priority for our future work.

## Limitations and Future Directions

What we present here is necessarily preliminary. A notable limitation of this work includes the fact that our sample comprised college heavy drinkers, which potentially limits its generalizability to other samples with longer-established drinking histories and in terms of demographics. In particular, the current sample has limited racial, ethnic, and geographic diversity given that the majority of participants were White Midwesterners. Using a relatively homogenous population prioritizes examination of internal validity at the potential expense of external validity. As we mentioned earlier, a strength of our data is that we effectively oversampled for tolerance and binge drinking, which is relatively less common in the general population. It seems likely that we sampled many transient drinkers who exhibit high degrees of certain symptoms (tolerance) but are not necessarily at long-term risk for alcohol addiction (Wakefield & Schmitz, 2015). We view this as a strong starting, but not ending, point for future work that will examine the structure of AUD in more diverse

samples. Because of this, we view undergraduates as a useful starting, but not ending, place for examining the latent structure of AUD. We look forward to related follow-up research that uses a range of samples, from undergraduate to community to high-risk. In what follows, we focus on ways that using a college heavy drinking sample may result in a slightly different structure of AUD than would other samples (e.g., community, clinical).

Many conclude that undergraduate samples are not informative of the general population, but this issue is more complicated with respect to alcohol consumption. College drinkers have considerable experience with alcohol, have among the highest rates of AUD (Grant et al., 2015), are at increased risk for AUD onset at this age (Kessler et al., 2005), and are more likely to experience serious drinking consequences than their peers who do not attend college (Slutske, 2005). Indeed, we sampled fairly heavy drinkers who reported bingeing on a weekly basis (see Table S1).

Also, Caetano and Babor (2006) have argued that rates of alcohol dependence may be inflated in adolescents and young adults because they misattribute tolerance and withdrawal to acute effects of excessive consumption. Although this is certainly a possibility, we developed our assessment to avoid these types of misattribution (Boness et al., 2019). Along these lines, although some view hangover and withdrawal as distinct but conflated in self-reports (Karriker-Jaffe et al., 2015), they are also likely physiologically related processes (Prat et al., 2009). Further, Kendler and colleagues (2012) found that excess consumption and tolerance loaded onto the same genetic factor among adults, suggesting that they covary meaningfully in college *and* community drinkers, so this covariation cannot be explained alone by sample type.

We also look forward to follow-up research that examines other correlates, including personality and a broader swath of psychopathology. Based on the generality of the correlates we examined, we might not expect to find considerably divergent associations with external criteria among tolerance, withdrawal, and loss of control. We saw some evidence of differential validity among these factors, but only when the factors' shared variance was controlled.

Net of Tolerance and Loss of Control, Withdrawal was negatively associated with early alcohol initiation, blackout, and cannabis use frequency. Net of Tolerance and Withdrawal, Loss of Control was the strongest positive marker of earlier alcohol initiation and excess drinking, alcohol consumption, blackout, cannabis use frequency, and legal problems. Loss of Control's hint that it is most closely aligned with externalizing liability. We expect those with marked externalizing tendencies and related psychopathology to initiate alcohol and other substance use earlier, use most consistently across the lifespan, and experience more legal problems and other consequences from drinking (Tully & Iacono, 2016; Watts et al., in press). Based on theory and existing research, we expect that loss of control would bear the most robust associations with impulsivity and externalizing psychopathology (Kwako et al., 2019), withdrawal with negative emotionality and internalizing (Koob & Volkow, 2010), and tolerance with extraversion, novelty seeking, and the like (Park et al., 2009).

## A Note on the Importance of Exploratory Modeling

Given that current consensus promotes a unidimensional model of AUD, we approached our study with the spirit of establishing “dual phenomenology,” using varied technologies to identify the same structures to minimize the likelihood of spurious findings. To avoid false alarms generated by confirmatory analysis, we considered replication in AUD’s structure across multiple exploratory methods. To avoid false alarms generated by any single type of analysis, we considered replication in AUD’s structure across diverse methods, in this case factor and cluster analysis. This approach allowed for numerous alternative latent structures of AUD.

The joint use of confirmatory and exploratory analysis was critical to these efforts. We supplemented confirmatory modeling with exploratory modeling to put our confirmatory models to even stronger (riskier) tests. In so doing, we discovered that some confirmatory models produced solutions that generally resembled their hypothesized structures, but their structures could not be retained in a semi-confirmatory (ESEM with target rotation) or exploratory context (EFA), nor were their factors identified at any level of the hierarchical model. Such an observation highlights that one *can* generate hypothesized structure with some coercion, even when the data do not necessarily conform to such a structure on their own. In this way, our exploratory analysis provided a scientific foil to our theory-driven efforts.

## Moving Towards a Hierarchical, Multidimensional Conceptualization of AUD

The notion that AUD manifests dramatically differently across individuals is a virtual truism (Litten et al., 2005). Empirical investigations nonetheless support the traditional conceptualization of AUD as a homogenous, unidimensional construct (Hasin et al., 2013; cf. MacCoun, 2013). These countervailing conclusions – heterogeneity in presentation but homogeneity in construct – present somewhat of a paradox. This work was motivated by the opinion that heterogeneity in AUD requires closer examination.

To our knowledge, this was one of the first comprehensive examinations of the latent structure of AUD with thorough assessment and analysis (c.f., Krueger et al., 2004; Muthén et al., 1993). This work culminated in a hierarchically-organized, multidimensional model of AUD that describes it in terms of both narrow dimensions (or facets) that correspond with symptoms and increasingly broad, transcriterial dimensions that describe covariation in subsets of AUD symptoms and indicators. In broad brush, our findings challenge unidimensional conceptualizations of AUD. We advance three relatively novel higher-order dimensions of AUD – Tolerance/Excess Consumption, Withdrawal, and Loss of Control – that require replication in other samples. Tolerance/Excess Consumption appeared relatively distinct from other components of AUD, whereas Withdrawal and Loss of Control dimensions appear more closely entangled at uppermost levels of the hierarchy.

What is most clear from the present study is that tolerance appears relatively distinct from withdrawal. We observed this in the tests of the two- and three-factor models and in the hierarchical clustering. That tolerance and withdrawal did not form a single dimension reflecting neuroadaptation to alcohol consumption is not all that surprising, despite the fact

that they are grouped together as pharmacological AUD criteria in DSM-5 and ICD-11. First, other factor analyses find that tolerance separates from other AUD criteria (Grant et al., 2007; Muthén et al., 1993), so much so that others have dropped tolerance from their models of AUD due to poor fit (Langenbucher et al., 2004). Second, tolerance and withdrawal are genetically distinct in rodents (Crabbe et al., 2011) and, more preliminarily, in humans (Kendler et al., 2012). Relatedly, dissociation of tolerance from other aspects of dependence is well established in the tobacco literature (Perkins, 2002). Third, tolerance is not a strong marker of severity or dependence after controlling for heavy consumption (Martin et al., 2008; Vergés et al., 2018). Fourth and finally, tolerance does not substantially increment withdrawal in terms of defining physiological dependence (Schuckit et al., 1998). These findings suggest – but are certainly not conclusive – that tolerance is relatively distinct from the rest of the dependence syndrome.

This raises the question of how we should think about tolerance if it is peripheral to alcohol dependence. This general idea bore out in the eternal validity analyses, where tolerance had the least unique associations with alcohol-relevant external criteria (Figure 3). Although we cannot offer a definitive answer at this point, it seems likely that tolerance is an intermediate phenotype (or non-specific prerequisite) that promotes the development of dependence by facilitating the ability to experience repeated, high-dose exposures. If AUD reflects a staged, developmental process (Koob & Le Moal, 2001) whereby tolerance is generally independent from but encourages other aspects of dependence, it further challenges the notion that AUD reflects a unitary dimension. This is especially true if AUD reflects a complex blend of acquired neuroadaptations (e.g., tolerance, withdrawal) and premorbid or dispositional features (e.g., impulsivity, reward sensitivity). By contrast, AUD might reflect a formative configuration of several reasonably independent, etiologically distinct features. Compared with a strict network model conceptualization, we believe that these independent features might themselves reflect latent variables composed of a number of symptoms that cut across diagnostic criteria, resulting in a hybrid formative-reflective conceptualization of AUD (MacCoun, 2013).

Notwithstanding AUD's ontology, we are optimistic that multidimensional conceptualizations of AUD can clarify sources of heterogeneity in symptom presentation, comorbid psychopathology, course, treatment targets, and the like. Multidimensional conceptualizations can also accommodate the sundry competing mechanisms thought to underlie AUD, including but not limited to reward sensitivity, impulse control deficits, and negative affectivity (Kwako et al., 2015). In addition to the fact that such characterizations are more consistent with modern models of addiction (i.e., ANA), they are also more consistent with contemporary models of other psychopathology that describe it in terms of broad and narrow transdiagnostic dimensions, such the Hierarchical Taxonomy of Psychopathology (Kotov et al., 2017) and the Research Domain Criteria (Insel et al., 2010). How the dimensions we advance in this work interface with those outlined in other frameworks remains high priority for future research. Research that parses critical sources of heterogeneity in AUD is an important starting place for such endeavors.

## Parsing Heterogeneity of Psychiatric Disorders

Despite the fact that most all psychiatric conditions are remarkably heterogeneous, diagnostic manuals (i.e., DSM-5, ICD-11) tend to conclude that they conform to a unidimensional structure. This conclusion is likely due to a number of conspiring factors – including but not limited to the use of too few indicators; the use of single indicators for each symptom; and the failure to rigorously examine alternative, multidimensional conceptualizations of psychiatric diagnoses – that contribute to a seemingly monolithic structure that can be difficult to pierce. Our proposed antidote is the adoption of methodological and analytic procedures that are more likely to detect multidimensionality, should it exist. Though there may be certain scenarios in which it is useful to extract a unidimensional factor of a given construct, this should not be confused with evidence that the *construct itself* is unidimensional (Watts et al., 2020).

We focused on AUD but it is likely that efforts to parse within-disorder heterogeneity could be applied to virtually any disorder. Indeed, most DSM constructs have been criticized for their heterogeneity, including major depression (Fried, 2017), autism spectrum disorder (Masi et al., 2017), attention-deficit/hyperactivity disorder (Roberts et al., 2017), and schizophrenia (Ahmed et al., 2018). Major depression, for instance, is associated with 10377 unique symptom profiles such that two patients may not share a single overlapping symptom. What's more, a recent empirical examination of 3703 depressed patients identified 1030 unique profiles exhibited by those patients alone (Fried & Nesse, 2015). Such profound heterogeneity might cause one to question the utility of a unidimensional phenotype, particularly when its features are associated with different risk factors (Fried, 2017; Fried & Nesse, 2015).

Parsing critical sources of heterogeneity in psychiatric phenomena has the potential to move past the deadlock of rampant comorbidity exhibited by disorders in our current and past diagnostic systems. Multidimensional, hierarchical conceptualizations acknowledge, rather than overlook, the remarkably heterogeneous nature of most psychiatric conditions (Kotov et al., 2017), and, in turn, focus on more homogeneous symptom clusters that are more likely to carve nature at its joints. We remain hopeful that such symptom clusters will serve as more fruitful targets in investigations into the etiology, distal and proximal correlates, sequelae, and treatment of psychiatric phenomena.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We thank Eric Feckzo for his programming assistance. Cassandra Boness is funded through F31AA026177 (Principal Investigator: Cassandra L. Boness); Jordan E. Loeffelman, Douglas Steinley, and Kenneth J. Sher are funded through R01AA024133 (Principal Investigator: Kenneth J. Sher). The computation for this work was performed on the high performance computing infrastructure provided by Research Computing Support Services and in part by the National Science Foundation under grant number CNS-1429294 at the University of Missouri, Columbia MO. DOI: <https://doi.org/10.32469/10355/69802>



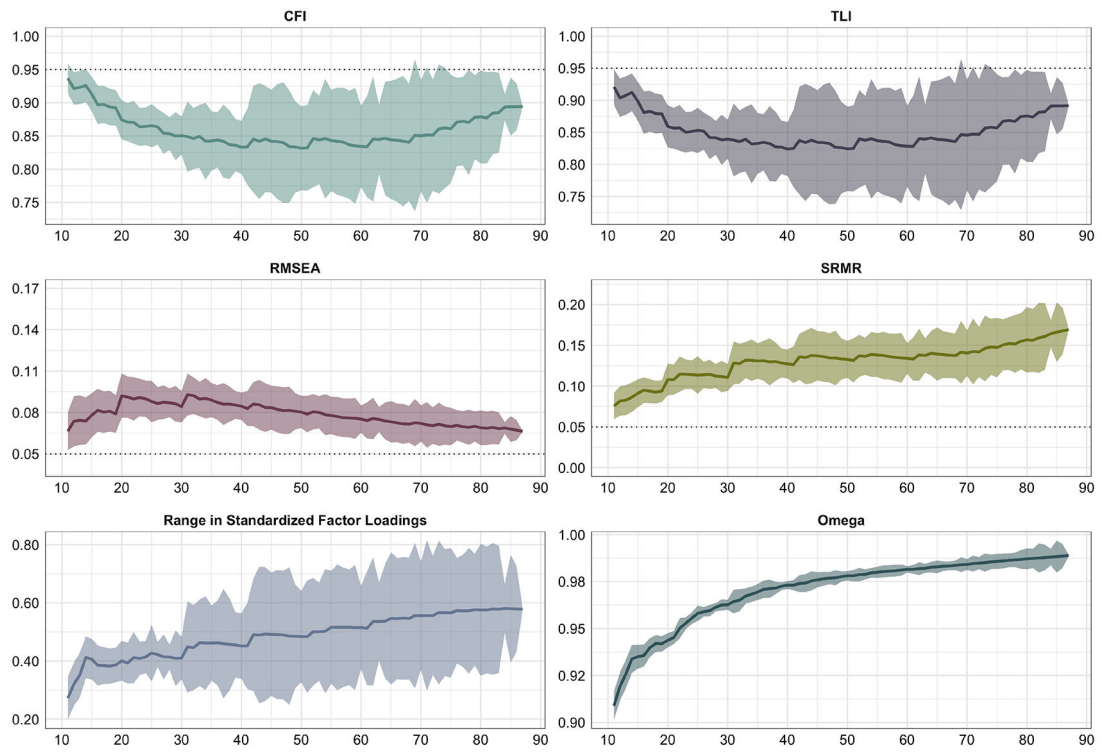
## References

- American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (DSM-5)*. Author: Washington, DC.
- Ahmed AO, Strauss GP, Buchanan RW, Kirkpatrick B, & Carpenter WT (2018). Schizophrenia heterogeneity revisited: Clinical, cognitive, and psychosocial correlates of statistically-derived negative symptoms subgroups. *Journal of Psychiatric Research*, 97, 8–15. [PubMed: 29156414]
- Bollen K, & Lennox R (1991). Conventional wisdom on measurement: A structural equation perspective. *Psychological Bulletin*, 110, 305–314.
- Boness CL, Lane SP, & Sher KJ (2019). Not all alcohol use disorder criteria are equally severe: Toward severity grading of individual criteria in college drinkers. *Psychology of Addictive Behaviors*, 33, 35–49. [PubMed: 30676037]
- Caetano R, & Babor TF (2006). Diagnosis of alcohol dependence in epidemiological surveys: An epidemic of youthful alcohol dependence or a case of measurement error?. *Addiction*, 101, 111–114. [PubMed: 16930167]
- Crabbe JC, Kendler KS, & Hitzemann RJ (2011). Modeling the diagnostic criteria for alcohol dependence with genetic animal models. In *Behavioral Neurobiology of Alcohol Addiction* (pp. 187–221). Springer, Berlin, Heidelberg.
- Fried EI (2017). Moving forward: how depression heterogeneity hinders progress in treatment and research. *Expert Review of Neurotherapeutics*, 17, 423–425. [PubMed: 28293960]
- Fried EI, & Nesse RM (2015). Depression sum-scores don't add up: why analyzing specific depression symptoms is essential. *BMC Medicine*, 13, 72. [PubMed: 25879936]
- Grant BF, Goldstein RB, Saha TD, Chou SP, Jung J, Zhang H, ... & Hasin DS (2015). Epidemiology of DSM-5 alcohol use disorder: results from the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry*, 72, 757–766. [PubMed: 26039070]
- Grant BF, Harford TC, Muthén BO, Yi HY, Hasin DS, & Stinson FS (2007). DSM-IV alcohol dependence and abuse: further evidence of validity in the general population. *Drug and Alcohol Dependence*, 86, 154–166. [PubMed: 16814489]
- Haeny AM, Littlefield AK, & Sher KJ (2014). Repeated diagnoses of lifetime alcohol use disorders in a prospective study: Insights into the extent and nature of the reliability and validity problem. *Alcoholism: Clinical and Experimental Research*, 38, 489–500.
- Harford TC, Yi HY, Faden VB, & Chen CM (2009). The dimensionality of DSM-IV alcohol use disorders among adolescent and adult drinkers and symptom patterns by age, gender, and race/ethnicity. *Alcoholism: Clinical and Experimental Research*, 33, 868–878.
- Hasin DS, O'Brien CP, Auriacombe M, Borges G, Bucholz K, Budney A, ... Grant BF (2013). DSM-5 criteria for substance use disorders: Recommendations and rationale. *American Journal of Psychiatry*, 170, 834–851.
- Herrmann A, & Pfister HR (2013). Simple measures and complex structures: Is it worth employing a more complex model of personality in Big Five inventories?. *Journal of Research in Personality*, 47, 599–608.
- Horn JL (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 32, 179–185.
- Hu LT, & Bentler PM (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55.
- Hubert L, & Steinley D (2005). Agreement among Supreme Court justices: Categorical vs. continuous representation. *SIAM News*, 38, 4–7.
- Insel T, Cuthbert B, Garvey M, Heinssen R, Pine DS, Quinn K, ... & Wang P (2010). Research domain criteria (RDoC): toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, 167, 748–751.
- Karriker-Jaffe KJ, Witbrodt J, & Greenfield TK (2015). Refining measures of alcohol problems for general population surveys. *Alcoholism: Clinical and Experimental Research*, 39, 363–370.

- Kendler KS, Aggen SH, Prescott CA, Crabbe J, & Neale MC (2012). Evidence for multiple genetic factors underlying the DSM-IV criteria for alcohol dependence. *Molecular Psychiatry*, 17, 1306–1315. [PubMed: 22105626]
- Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, & Walters EE (2005). Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Archives of General Psychiatry*, 62, 593–602. [PubMed: 15939837]
- Koob GF, & Le Moal M (2001). Drug Addiction, Dysregulation of Reward, and Allostasis. *Neuropsychopharmacology*, 24, 97–129. [PubMed: 11120394]
- Koob GF, & Volkow ND (2010). Neurocircuitry of addiction. *Neuropsychopharmacology*, 35, 217–238. [PubMed: 19710631]
- Kotov R, Krueger RF, Watson D, Achenbach TM, Althoff RR, Bagby RM, ... & Eaton NR (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): a dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, 126, 454–477. [PubMed: 28333488]
- Kpotufe S & Von Luxburg U (2011). Pruning nearest neighbor cluster trees. arXiv preprint arXiv:1105.0540
- Krueger RF, Nichol PE, Hicks BM, Markon KE, Patrick CJ, & McGue M (2004). Using latent trait modeling to conceptualize an alcohol problems continuum. *Psychological Assessment*, 16, 107–119. [PubMed: 15222807]
- Kwako LE, Momenan R, Litten RZ, Koob GF, & Goldman D (2016). Addictions neuroclinical assessment: a neuroscience-based framework for addictive disorders. *Biological Psychiatry*, 80, 179–189. [PubMed: 26772405]
- Kwako LE, Schwandt ML, Ramchandani VA, Diazgranados N, Koob GF, Volkow ND, ... & Goldman D (2019). Neurofunctional domains derived from deep behavioral phenotyping in alcohol use disorder. *American Journal of Psychiatry*, 176, 744–753.
- Lane SP, & Sher KJ (2015). Limits of current approaches to diagnosis severity based on criterion counts. *Clinical Psychological Science*, 3, 819–835. [PubMed: 26783505]
- Lane SP, Steinley D, & Sher KJ (2016). Meta-analysis of DSM alcohol use disorder criteria severities: Structural consistency is only ‘skin deep’. *Psychological Medicine*, 46, 1769–1784. [PubMed: 27019218]
- Langenbucher JW, Labouvie E, Martin CS, Sanjuan PM, Bavly L, Kirisci L, & Chung T (2004). An application of item response theory analysis to alcohol, cannabis, and cocaine criteria in DSM-IV. *Journal of Abnormal Psychology*, 113, 72–80. [PubMed: 14992659]
- Litten RZ, Ryan ML, Falk DE, Reilly M, Fertig JB, & Koob GF (2015). Heterogeneity of alcohol use disorder: Understanding mechanisms to advance personalized treatment. *Alcoholism: Clinical and Experimental Research*, 39, 579–584.
- Lynskey MT, Nelson EC, Neuman RJ, Bucholz KK, Madden PA, Knopik VS, ... & Heath AC (2005). Limitations of DSM-IV operationalizations of alcohol abuse and dependence in a sample of Australian twins. *Twin Research and Human Genetics*, 8, 574–584. [PubMed: 16354499]
- MacCallum RC, Browne MW, & Sugawara HM (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, 1, 130–149.
- MacCoun RJ (2013). The puzzling unidimensionality of DSM-5 substance use disorder diagnoses. *Frontiers in Psychiatry*, 4, 1–5. [PubMed: 23346060]
- Martin CS, Chung T, & Langenbucher JW (2008). How should we revise diagnostic criteria for substance use disorders in the DSM-V?. *Journal of Abnormal Psychology*, 117, 561–575. [PubMed: 18729609]
- Martin CS, Kaczynski NA, Maisto SA, Bukstein OM, & Moss HB (1995). Patterns of DSM-IV alcohol abuse and dependence symptoms in adolescent drinkers. *Journal of Studies on Alcohol*, 56, 672–680. [PubMed: 8558899]
- Masi A, DeMayo MM, Glozier N, & Guastella AJ (2017). An overview of autism spectrum disorder, heterogeneity and treatment options. *Neuroscience Bulletin*, 33, 183–193. [PubMed: 28213805]
- Miller GA, & Rockstroh B (2013). Endophenotypes in psychopathology research: where do we stand?. *Annual Review of Clinical Psychology*, 9, 177–213.
- Milligan GW, & Cooper MC (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50, 159–179.

- Muthén BO (1995). Factor analysis of alcohol abuse and dependence symptom items in the 1988 National Health Interview Survey. *Addiction*, 90, 637–645. [PubMed: 7795500]
- Muthén BO, Grant B, & Hasin D (1993). The dimensionality of alcohol abuse and dependence: factor analysis of DSM-III-R and proposed DSM-IV criteria in the 1988 National Health Interview Survey. *Addiction*, 88, 1079–1090. [PubMed: 8401162]
- Muthén LK and Muthén BO (1998–2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.
- Park A, Sher KJ, Wood PK, & Krull JL (2009). Dual mechanisms underlying accentuation of risky drinking via fraternity/sorority affiliation: The role of personality, peer norms, and alcohol availability. *Journal of Abnormal Psychology*, 118, 241–255. [PubMed: 19413401]
- Pavan M & Pelillo M (2003). Dominant sets and hierarchical clustering. *Proc. IEEE Int'l Conf. Computer Vision*, 1, 362–369.
- Perkins KA (2002). Chronic tolerance to nicotine in humans and its relationship to tobacco dependence. *Nicotine & Tobacco Research*, 4, 405–422. [PubMed: 12521400]
- Prat G, Adan A, & Sánchez-Turet M (2009). Alcohol hangover: a critical review of explanatory factors. *Human Psychopharmacology: Clinical and Experimental*, 24, 259–267. [PubMed: 19347842]
- Proudfoot H, Baillie AJ, & Teesson M (2006). The structure of alcohol dependence in the community. *Drug and Alcohol Dependence*, 81, 21–26. [PubMed: 16005578]
- Revelle W (2019). *psych: Procedures for Psychological, Psychometric, and Personality Research*. Northwestern University, Evanston, Illinois. R package version 1.9.12, <https://CRAN.R-project.org/package=psych>.
- Revelle W, & Zinbarg RE (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sjitsma. *Psychometrika*, 74, 145–154.
- Roberts BA, Martel MM, & Nigg JT (2017). Are there executive dysfunction subtypes within ADHD?. *Journal of Attention Disorders*, 21, 284–293. [PubMed: 24214969]
- Schuckit MA, Smith TL, Daepfen JB, Eng M, Li TK, Hesselbrock VM, ... & Bucholz KK (1998). Clinical relevance of the distinction between alcohol dependence with and without a physiological component. *American Journal of Psychiatry*, 155, 733–740.
- Sharma S, & Batra N (2019). Comparative Study of Single Linkage, Complete Linkage, and Ward Method of Agglomerative Clustering. In 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), IEEE, 568–573.
- Slutske WS (2005). Alcohol use disorders among U.S. college students and their non-college-attending peers. *Archives of General Psychiatry*, 62, 321–327. [PubMed: 15753245]
- Spearman C (1910). Correlation calculated from faulty data. *British Journal of Psychology*, 3, 271–295.
- Sokal RR, & Rohlf FJ (1962). The comparison of dendrograms by objective methods. *Taxon*, 11, 33–40.
- Tomarken AJ, & Waller NG (2003). Potential problems with "well fitting" models. *Journal of Abnormal Psychology*, 112, 578–598. [PubMed: 14674870]
- Tucker LR (1951). A method for synthesis of factor analysis studies (No. PRS-984). Educational Testing Service Princeton NJ.
- Tully EC, & Iacono WG (2016). An integrative common liabilities model for the comorbidity of substance use disorders with externalizing and internalizing disorders. In Sher KJ (Eds.), *The Oxford Handbook of Substance Use and Substance Use Disorders: Volume 2* (pp. 187–214). New York, New York: Oxford University Press.
- VanderWeele TJ, & Batty CJ (2020). On the Dimensional Indeterminacy of One-Wave Factor Analysis Under Causal Effects. *arXiv preprint arXiv:2001.10352*.
- Velicer WF (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 31, 321–327.
- Vergés A, Ellingson JM, Schroder SA, Slutske WS, & Sher KJ (2018). Intensity of daily drinking and its relation to alcohol use disorders. *Alcoholism: Clinical and Experimental Research*, 42, 1674–1683.

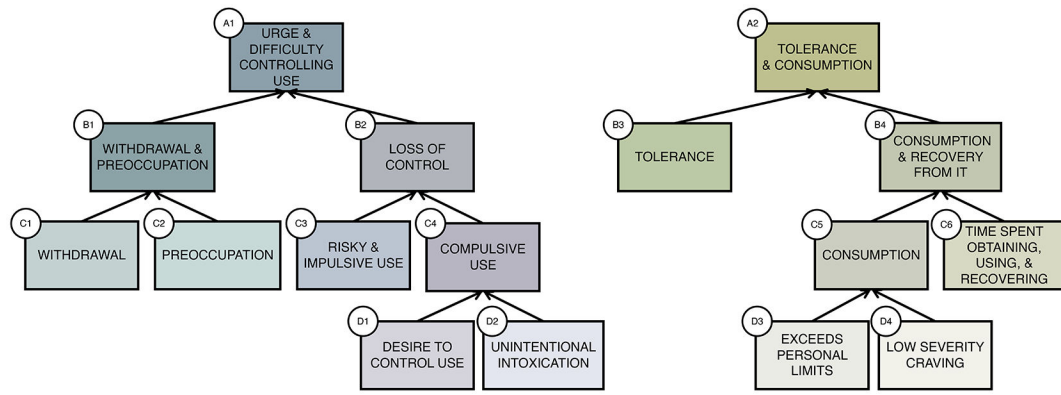
- Wakefield JC (2015). DSM-5 substance use disorder: How conceptual missteps weakened the foundations of the addictive disorders field. *Acta Psychiatrica Scandinavica*, 132, 327–334. [PubMed: 25968656]
- Wakefield JC, & Schmitz MF (2015). The harmful dysfunction model of alcohol use disorder: Revised criteria to improve the validity of diagnosis and prevalence estimates. *Addiction*, 110, 931–942. [PubMed: 25622535]
- Watts AL, Lane SP, Bonifay W, Steinley D, & Meyer FAC (2020). Building theories on top of, and not independent of, statistical models: The case of the p-factor. *Psychological Inquiry*.
- Watts AL, Wood PK, Jackson KM, Lisdahl KM, Heitzeg MM, ... & Sher KJ (*in press*). Incipient alcohol use in childhood: Early alcohol sipping and its relations with psychopathology and personality. *Development and Psychopathology*.



**Figure 1.**

Model fit of a 1-factor solution as a function of number of indicators in the model.

*Note.* Medians of the 500 repetitions are reported with standard deviations as error ribbons



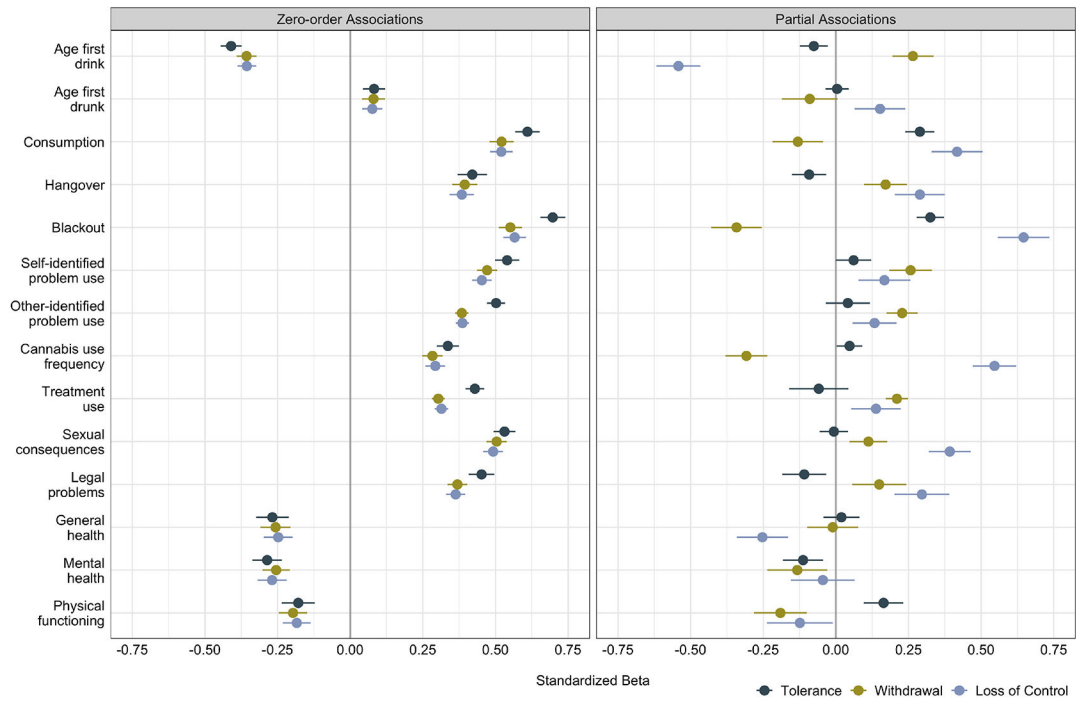
**Figure 2.**  
Pruned hierarchical structure yielded from agglomerative hierarchical clustering

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 3.**  
 Relations between 3-factor CFA factors and external criteria  
 Note: Betas are included with standard errors.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Table 1.**

Confirmatory models.

Diagnostic criterion	# items	2-factor model			3-factor model			4-factor model		
		Abuse	Dependence	Tolerance	Withdrawal	Loss of Control	Pharmacological	Risky Use	Impaired Control	Social/ Interpersonal
Continued use	7		X		X			X		
Craving	11	---	---			X			X	
Cut down	7		X			X			X	
Give up	4		X			X				X
Hazardous use	7	X				X				
Larger/longer	6		X			X			X	
Failure to fulfill obligations	6	X				X				X
Social/interpersonal problems	6	X				X				X
Time spent	7		X			X				X
Tolerance	12		X	X				X		
Withdrawal	14		X		X					X