# Deconstructing the Architecture of Alcohol Abuse and Dependence Symptoms in a Community Sample of Late Adolescent and Emerging Adult Women: An Item Response Approach

**Alexis E. Duncan**[1,2], **Arpana Agrawal**[2], **Kathleen K. Bucholz**[2], **Carolyn E. Sartor**[2], **Pamela A.F. Madden**[2], and **Andrew C. Heath**[2]

[1]George Warren Brown School of Social Work, Washington University, St. Louis, MO

[2]Midwest Alcoholism Research Center, Department of Psychiatry, Washington University School of Medicine, St. Louis, MO

## Abstract

The objective of this study was to examine the underlying factorial architecture of lifetime DSM-IV alcohol use disorder (AUD) criteria in a population-based sample of adolescent and emerging adult female twins who had ever used alcohol (n=2832; aged 18-25 years), and to determine whether thresholds and factor loadings differed by age. Item response modeling was applied to DSM-IV AUD criteria. Compound criteria (e.g., persistent desire or unsuccessful attempts to quit or cut down) were included as separate items. Of the remaining 16 items, tolerance and use despite physical problems were the most and least commonly endorsed items, respectively. Underlying the items was a single factor representing liability to AUDs. Factor loadings ranged from 0.67 for blackouts to 0.90 for time spent using/recovering from effects. Some items assessing different DSM-IV criteria had very similar measurement characteristics, while others assessing the same criterion showed markedly different thresholds and factor loadings. Compared to that of women aged 21-25 years, the threshold for hazardous use was higher in women aged 18-20 years, but lower for used longer than intended and persistent desire to cut down. After accounting for threshold differences, no variations in discrimination across age groups were observed. In agreement with the extant literature, our findings indicate that the factorial structure of AUD is unidimensional, with no support for the abuse/dependence distinction. Individual components of compound criteria may differ in measurement properties; therefore pooling information from such divergent items will reduce information about the AUD construct.

*Correspondence to: Alexis E. Duncan, Ph.D., Department of Psychiatry, Washington University School of Medicine, 660 S. Euclid, Box 8134, St. Louis, MO 63110, USA, 314-286-2210 (phone), 314-286-2213 (fax), aduncan@wustl.edu.

---

## 1. Introduction

Although the Diagnostic and Statistical Manual 4th edition (DSM-IV; American Psychiatric Association, 2000) contains two alcohol use disorder diagnoses, abuse and dependence, the preponderance of evidence indicates that a single factor underlies the eleven alcohol abuse and dependence criteria (Dawson et al., 2010; Gelhorn et al., 2008; Harford et al., 2009; Kahler et al., 2006; Saha et al., 2006). Many investigators have thus used item response theory (IRT) modeling, which requires that a single factor underlies a group of items, to study the psychometric properties of these alcohol use and dependence criteria. Using IRT, one can estimate the difficulty and discrimination of each item along the continuum of a latent trait, where <u>discrimination</u> refers to an item's ability to distinguish individuals with high vs. low liability to AUDs, and <u>difficulty</u> reflects the location along the liability distribution where the item functions. Therefore, how an item relates to the underlying liability to AUD is characterized by the location along the liability continuum where the item operates (i.e. how difficult it is), and, importantly how well the item discriminates those at high versus low risk for AUDs, given a particular liability to AUD. Alcohol abuse (AA) was originally conceptualized as a less severe form of AUD compared to alcohol dependence (AD; American Psychiatric Association, 2000); however, many investigators have used IRT models to demonstrate that some AA criteria are at the more severe end of the AUD continuum (e.g., failure to fulfill role obligations), while some AD criteria are at the less severe end (e.g., drank larger amounts/longer than intended) (Langenbucher et al., 2004; Saha et al., 2006).

Many of these IRT analyses have been conducted using one item per criterion. As such, results from these analyses may not provide a complete picture of the architecture of AUD symptoms because the majority of the AD criteria are "compound items," consisting of two components. For example, AD criterion 4 is endorsed if an individual (1) has a persistent desire or (2) has made unsuccessful efforts to cut down or control substance use, implying that persistent desire and unsuccessful efforts are equal to one another in severity and discrimination. It is, however, quite possible that each component has different psychometric properties, and that for research purposes collapsing them into a single item results in loss of valuable information.

AUD criteria may also have different psychometric properties in different groups, referred to as differential item functioning (DIF) (Muthen et al., 1985a). For example, Martin, et al. (2006) found that the AA criteria hazardous use and legal problems displayed greater difficulty in female than in male adolescents recruited from addiction treatment programs. Other studies have found DIF by age group; however, these studies included emerging adults (aged 18-24/25) as a single age group (Harford et al., 2009; Saha et al., 2006). Since the emerging adulthood age range spans the legal drinking age, when there is a marked increase in binge drinking (Grucza et al., 2009), further information may be gained by investigating the possibility of differential item function in individuals in emerging adulthood based on whether or not an individual is of legal drinking age. For example, Kahler et al., (2009) found differential item functioning between 18-20 and 21-24 year olds for drinking frequency and for being drunk at school or work using a Rasch model to examine alcohol consumption and problems across adolescence and young adulthood.

While the above studies have reported several instances of DIF attributable to varying item difficulty, there have been fewer demonstrated instances of DIF contributing to variations in discrimination. There are two possible explanations for this: first, the computation of discrimination and difficulty in the item response framework are inextricably interconnected such that one parameter can "compensate" for the other (http://www.statmodel.com/download/MplusIRT2.pdf). Second, and perhaps more importantly, variations in difficulty may reflect expected changes (e.g. higher versus lower endorsement of certain items in specific groups), and are thus noted more commonly, but the ability of the item to separate those at high versus low risk at a particular liability level may remain unaltered after accounting for varying difficulties. This would suggest that while some items are more typically observed or endorsed in certain sub-groups, the inherent architecture created by these items that underlies the composite liability to AUDs (i.e. the factor) remains fairly stable.

The objective of the current study was to examine the psychometric functioning of the DSM-IV AUD criteria, breaking the compound criteria into their component parts, and to evaluate the presence of differential item functioning by age in data from a sample of young women in late adolescence and emerging adulthood (ages 18-25 years).

## 2. Method

### 2.1 Study Sample

The Missouri Adolescent Female Twin Study (MOAFTS) is a study of female twin pairs born between 1975 and 1985 to parents residing in the state of Missouri. Parents of twins were identified and traced through birth records and contacted regarding participation (target $n = 1,999$ European-American and $n = 370$ African-American pairs representing all live-born pairs of Missouri resident parents; 95.9% of families located, $n = 2,279$). Participants reflect statewide demographics, including individuals of both African and European ancestry coming from rural and urban areas. Data collection from the twins began with the baseline interview in 1995-1999, when twins were of median age 15 (mean [SD] = 15.52 [2.42]; range: 12-23 years). Data for the current study come from the Wave 4 assessment conducted between 2000 and 2005 when the twins were median age 22 (mean = 21.69 [2.76]; range: 18-29 years; on average 5 years after the baseline assessment). Since minors for whom parental consent was denied at baseline were of legal age of consent at Wave 4, all women originally targeted at baseline were recontacted, unless they had previously withdrawn from the study. Eighty-three percent of baseline participants also participated at Wave 4, and there were an additional 930 new participants (25% of total). All research protocols were approved by the institutional review board at the Washington University School of Medicine. Additional details regarding the sample are available elsewhere (Heath et al., 1999; Heath et al., 2002).

### 2.2 Assessment

Study participants were interviewed with a telephone adaptation of the Semi-Structured Assessment for the Genetics of Alcoholism (SSAGA), a comprehensive structured psychiatric diagnostic instrument (Bucholz et al., 1994). Reliability data for individual AUD criteria from the original SSAGA, have indicated good to excellent reliability, with κ exceeding .60 for all of the nine DSM-III-R criteria studied (Bucholz et al., 1994; Bucholz et al., 1995).

Analyses for the current study were conducted only on women in the MOAFTS sample who were in emerging adulthood (aged 18-25 years) and had ever consumed a full drink of alcohol (n=2838; 74.94% of sample). Women who had consumed alcohol on <7 separate

days (lifetime) were skipped from the alcohol diagnostic section and were assumed to have negative responses to all AUD symptoms. While all of the remaining women who had ever had a drink on > 6 days (79.16% of ever drinkers) were asked questions about tolerance and AA symptoms, questions regarding the other six AD criteria were only asked of those who had ever consumed >3 drinks in a 24 hour period or of those who had consumed ≤3 drinks in 24 hours but indicated that they had ever felt that they were an excessive drinker. Women who were not asked about these AD symptoms were assigned negative responses (41.79% of ever drinkers).

The two elements of four of the "compound" AD lifetime criteria (larger/longer, persistent desire/unsuccessful efforts to quit, continued to drink despite physical/psychological problems caused/worsened by drinking, withdrawal/relief) were included in the analyses as separate items (see Table 1). Tolerance was included as a single item, collapsing the increased amounts and diminished effect elements because inclusion of the separate elements resulted in the identification of a third factor in the exploratory factor analysis that included only the tolerance items, one of which had a factor loading >1. Since only eight women endorsed the AA recurrent legal problems criterion, it was not included in the analyses.

### 2.3 Statistical analyses

All data analyses were conducted using Mplus, version 5 (Muthen et al., 2007). Exploratory and Confirmatory Factor analyses were conducted using the maximum likelihood estimator in Mplus.

In item response modeling (IRM), a single factor confirmatory factor model can also be interpreted as a 2-parameter logistic IRM, where the threshold, when divided by the factor loading, refers to criterion difficulty (or 'b') and the (unstandardized) factor loading, when divided by 1.7 to approximate the probit scale, is reflective of criterion discrimination (or 'a') (Birnbaum, 1968; Bock, 1997; Muthen, 1985). Item difficulty and discrimination, which are key to the conceptualization of item characteristic curves (ICC), were computed using a 2 parameter (2P) logistic model with logit function $L = 1.7*a(\theta-b)$, where $\theta$ = the liability distribution. In general, <u>discrimination</u> refers to the ability of an item to distinguish individuals with high vs. low liability to AUDs at corresponding thresholds across the AUD continuum. Discrimination is indexed by the steepness of the ICC, with steeper curves representing greater discrimination. In contrast, item <u>difficulty</u> reflects the location along the liability distribution where the item functions – i.e., items falling closer to the y-axis of ICC represent less difficult items, or those that are more commonly endorsed, given their discrimination.

We tested for the presence of differential item functioning (DIF) (Muthen et al., 1985b), or differences in the difficulty or discrimination of each item, across age groups. After accounting for mean differences in the underlying latent factor, we first fit a two group model in which factor means were freely estimated in each group but thresholds and factor loadings for each item were constrained across age groups (18-20 years and 21-25 years), Second, we fit a series of models in which the item thresholds and factor loadings were freely estimated in each group – because an omnibus test of free thresholds and factor loadings is not identified, tests for individual items were conducted. Each of these models was compared to the fully constrained model using likelihood-based fit statistics with two degrees of freedom. In order to determine whether it was the factor loading, threshold or both that differed across groups, another series of models was fit testing the thresholds for each item separately. Next, a series of models in which factor loadings were freely estimated across groups were fit and compared to a model which allowed significantly different thresholds to be freely estimated (i.e., from the previous step) using a one degree of freedom

test. Standard errors were adjusted for non-independence using the maximum likelihood robust (MLR) estimator which allows for clustered observations, to correct for the non-independence of observations in twin pairs.

To view the information afforded by items in the item response model, both individual item characteristic curves (ICC) and individual item and total information curves (TIC) were calculated. The TIC is the sum of the information from each item and represents measurement precision afforded to the liability being assessed. The height or area under the curve represents measurement precision while the kurtosis represents the range of liability across which the items distinguish. If, for instance, an item with poor measurement characteristics (e.g., low factor loading) is excluded from the IRM, the drop in TIC should be negligible (or may improve, if eliminating this item increases the coherence amongst the remaining items). On the other hand, if an item makes a considerable contribution to the measurement of the liability distribution, its exclusion should result in a dramatic lowering of the TIC height. There is no statistical test for differences in the area under the TICs; however, the TICS were visually compared for a series of models in which each item was deleted.

## 3. Results

### 3.1 Symptom endorsement

The proportions of women endorsing each item overall and by age category are presented in Table 2. For both age categories, *tolerance* was the most frequently endorsed criterion (38.31% and 40.88% for age 18-20 and 21-25, respectively), and *legal problems* was the least frequently endorsed criterion (0.26% - 0.30%). As noted in section 2.2, since only eight women in the sample endorsed this criterion, it was not included in further analyses. A significantly higher proportion of women in the older age group reported *hazardous use* and *blackouts* compared to the younger group (p<0.05 for both comparisons).

### 3.2 Factor Analysis

Eigenvalues from an exploratory factor analysis suggested a two-factor solution, with eigenvalues for the 1, 2 and 3 factor solutions as 9.74, 1.14 and 0.93 respectively; however, visual inspection of the scree plot suggested that the one class solution might also be a viable solution. In the two factor solution, the items *hazardous use, tolerance, larger amounts, longer duration, time spent*, and *blackouts* loaded on one factor and the rest of the items loaded on the other, but these factors were significantly and highly correlated (r=0.89) after adjusting for age. Furthermore, the 1- and 2- factor CFA models had similar fit statistics (1-factor CFI=0.96 and RMSEA=0.04; 2-factor CFI = 0.99 and RMSEA=0.02). We therefore concluded that the high factor correlation and comparable fit statistics, in combination with multiple previously published findings that a single factor underlies the 11 alcohol abuse and dependence symptoms, provided sufficient evidence for a single factor, and thus we proceeded with the IRT modeling.

### 3.3 Item Response Modeling

Results from the IRT models are displayed in Figure 1. Item discrimination ranged from 0.89 (*physical problems*) to 2.25 (*time spent*) (corresponding factor loadings ranged from .64-.90). *Physical problems* (2.49) and *tolerance* (0.31) had the highest and lowest thresholds, respectively. After accounting for threshold differences between age groups, means and variances did not differ. There were significant differences between models in which thresholds and factor loadings were freely estimated for each item compared to the fully constrained model for the items *hazardous use* ($\chi^2$=36.30, df=2, p<0.001), *larger amount* ($\chi^2$=6.93, df=2, p=0.031) *and persistent desire* ($\chi^2$=14.83, df=2, p=001). Upon

further investigation, these differences appeared to be driven by differing thresholds rather than factor loadings. The younger group had higher thresholds (i.e. less common endorsement) for *hazardous use*, while the older group had higher thresholds for *larger amounts* and *persistent desire*. After accounting for these threshold differences, which reflect DIF-induced changes in difficulty parameters, discriminations (and corresponding factor loadings) appeared to be stable across age groups.

Of note, in both age groups several pairs of items had very similar psychometric properties (see Figure 1). *Give up activities, psychological problems, social problems* and *longer duration* had similar thresholds and discrimination in both age groups. The pairs of items *hazardous use* and *larger amounts* in the younger group and *blackouts* and *persistent desire* in the older group had similar properties. With the exception of the two components of the *withdrawal* criterion, the different items constituting each compound criterion had very different psychometric properties. For example, for the DSM-IV compound criterion *persistent desire/unsuccessful attempts to quit, unsuccessful attempts* had a lower discrimination and a much higher threshold than *persistent desire*. Likewise, *physical problems* (part of the compound criterion *of physical/emotional problems*) had the highest threshold and the lowest discrimination of the four items used to assess this AD criterion as well as all other items in the analysis.

### 3.4 Total Information Curves

The TICs for each age group almost completely overlapped. Since it was difficult to distinguish the two visually, a single TIC is displayed in Figure 2. While the peaks of the curves were high, the TICs indicated that the AUD items provided information over a somewhat narrow mid to upper range of the AUD continuum. For individual items, *withdrawal syndrome* and *relief* displayed greatest item information; however, they appeared to discriminate over a narrow range of liability in the middle of the AUD continuum. The removal of these withdrawal items from the model resulted in the largest drop in the height of the TIC compared to removal of other items. Removal of the items *physical problems, larger amounts* (in the younger group only), and *tolerance* did not appear to affect the height of the TIC when removed from the model. Removal of the remaining items resulted in an intermediate reduction in curve height.

## 4. Discussion

We confirmed the presence of a single factor underlying the DSM-IV AUD criteria, as reported in numerous previous studies (Dawson et al., 2010; Gelhorn et al., 2008; Harford et al., 2009; Kahler et al., 2006; Saha et al., 2006). Also similar to findings from previous studies, AA and AD items were mixed in terms of severity (Harford et al., 2009; Langenbucher et al., 2004; Martin et al., 2006; Saha et al., 2006), with the AD *tolerance* criterion as the least severe item, and the *social problems* and *role interference* AA criteria toward the more severe end of the AUD continuum. This is contrary to the DSM-IV concept of alcohol dependence and abuse representing more and less severe disorders, respectively (American Psychiatric Association, 2000).

While several studies have used population-based, community and clinical samples to study the psychometric properties of AUD symptoms with IRT models (e.g., Harford et al., 2009; Langenbucher et al., 2004; Martin et al., 2006; Saha et al., 2006), to our knowledge, none to our knowledge have done so to examine the psychometric properties of different elements of compound AUD items in a population-based sample. Our results show that, in addition to varying thresholds across the AUD continuum, the different elements of several compound items also varied in their ability to distinguish between individuals above and below those thresholds. Although there were no differences in factor loadings or thresholds between the

two *withdrawal* elements *(syndrome* and *relief)*, the two elements of *larger amounts/longer duration* and *persistent desire/unsuccessful attempts* had significantly different factor loadings and thresholds as evidenced by non-overlapping confidence intervals (ICCs are shown in figure 1; 95% CI's for individual items are available upon request). *Larger amounts* and *persistent desire* had lower factor loadings and thresholds than their corresponding elements *longer duration* and *unsuccessful attempts*, respectively. In addition, many of the four elements *of the physical/psychological* problems criterion also had thresholds and factor loadings that differed significantly from one another, with *physical problems*, in particular, having a very high threshold but a low factor loading (i.e. distinguishing poorly across the higher range of AUD liability). Therefore, the elements of these compound items likely represent facets of a single diagnostic criterion, with each facet having unique measurement characteristics. This issue has previously been identified in latent class analyses of adult samples, both community-ascertained (Heath et al., 1994) and proband-ascertained (Bucholz et al., 1996), so it is unlikely to be specific to the age group of the current sample. Intriguingly, our analyses revealed that some facets of compound items are more superior indicators of AUD liability than others. For example, the *psychological problems* component of the continued use despite physical or psychological problems criterion has a much higher factor loading than the *physical problems* component, implying that compound criteria, comprised of items with varying measurement properties, can reduce net information in the measurement of AUD liability. Of note, however, is that the measurement properties of the *physical problems* item might be influenced by the youthful age of our sample which is likely comprised of women early in their drinking careers when *physical problems* may be less common.

We found that these AUD items only provide information in a relatively narrow mid-range of the AUD severity continuum, indicating that the current AUD items do not discriminate well at either end of the continuum. While some previous IRT studies reported similar findings (Langenbucher et al., 2004; Martin et al., 2006), others reported poor discrimination only at the less severe end of the continuum (Harford et al., 2009; Krueger et al., 2004; Saha et al., 2006). This has been partly ameliorated in one study with the inclusion of a measure of alcohol consumption (Saha et al., 2006). While this poor resolution at the extremes of the liability continuum may not be problematic with the dichotomous alcohol abuse and dependence diagnoses of the current DSM-IV diagnostic system, it is an important issue to consider in the context of the dimensional approach proposed for DSM-V (American Psychiatric Association, 2010). Apart from *tolerance* at the less severe end and *physical problems* at the more severe end, the rest of the AUD items cluster together, with several pairs of items having nearly identical thresholds and factor loadings. In each of these pairs, at least one item is a component of a compound item. While it is valid to view these items as potentially redundant, they may also be seen as opportunities to capture individuals who may have been false negatives on the other item with similar properties, improving quantification of symptom severity.

The findings for the hazardous use variable are of particular interest given that the revision of the DSM criteria is currently being discussed. In a recently published study using data from the National Epidemiologic Survey on Alcohol and Related Conditions, Agrawal et al (2010) found that endorsement of this criterion alone was associated with less severe outcomes than the endorsement of other AA criteria. Additional studies using IRT models in treatment and national probability samples of adolescents and young adults have indicated that the hazardous use criterion has relatively poor discrimination and shows DIF by gender, with higher thresholds for adolescent and young adult women than men (e.g., Martin et al, 2006; Harford et al., 2009). Our study adds to this body of literature, demonstrating that in this sample of young adult women, hazardous use had relatively low discrimination and DIF

by age, such that the threshold was higher among women below legal drinking age than among those who could drink legally.

This study extends the literature on the performance of alcohol use disorder criteria by focusing on differential item functioning in emerging adult women by whether or not they were of legal drinking age. In addition to differing thresholds for hazardous use, thresholds differed by age for *larger amounts* and *persistent desire to quit/cut down*, although factor loadings appeared to remain invariant across the two age groups. In contrast to hazardous use, the items *larger amounts* and *persistent desire to quit/cut down* had higher thresholds among women over age 21. It is important to note, however, that this heterogeneity in the effect of age has little impact on the overall latent trait of alcohol use disorder, which is evidenced by total information curves for each age group that almost entirely overlap as well as the modest magnitude of the statistically significant differences in the thresholds of these items between age groups.

### 4.1 Limitations

This study has several potential limitations. First, the study sample was limited to women aged 18-25 years from a Midwestern state. Therefore, results may not be generalizable to men or to women of different ages and different regions. Second, since analyses were conducted using lifetime AUD symptoms, it is possible that results may differ from analyses using current (past 12 month) symptoms. However, our findings are very similar to those from studies in which past 12 month symptomatology was analyzed (Saha et al., 2006). Furthermore, as few studies have focused on the application of IRT modeling to lifetime criteria, this may also be viewed as a contribution of the present study. Finally, we relied on retrospective self-report of lifetime AUD symptoms. It is possible that women may not remember AUD symptoms occurring when they were younger; however, the relatively young age of the study sample minimizes this possibility.

### 4.2 Conclusions

This study adds to a growing body of literature reporting that a single continuum underlies AUD symptomatology. Elements of several DSM-IV compound symptoms had significantly different psychometric properties from one another, and several items had nearly identical psychometric properties. This suggests that each constituent of a compound items merits separate evaluation. Additionally, it may be possible to eliminate some symptoms without affecting the amount or quality of the information obtained from the remaining AUD symptoms. Future studies should attempt to replicate these results in additional study samples.

## Acknowledgments

## References

Agrawal A, Bucholz KK, Lynskey MT. DSM-IV alcohol abuse due to hazardous use: a less severe form of abuse? J Stud Alcohol Drugs. 2010; 71(6):857–63. [PubMed: 20946742]

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, DSM-IV-TR. American Psychiatric Association; Washington, DC: 2000.

American Psychiatric Association. DSM-V Development. American Psychiatric Association; 2010. www.dsm5.org

Birnbaum, A. Some Latent Trait Models. In: Lord, FM.; Norvick, MR., editors. Statistical Theory of Mental Test Scores. Addison-Wesley; Reading, MA: 1968. p. 397-472.

Bock RD. A brief history of item response theory. Educ Meas Iss Prac. 1997; 16:21–33.

Bucholz KK, Cadoret R, Cloninger C, Dinwiddie SH, Hesselbrock VM, Nurnberger JI, Reich T, Schmidt I, Schuckit MA. A new, semi-structured psychiatric interview for use in genetic linkage studies: a report of the reliability of the SSAGA. J Stud Alcohol. 1994; 55:149–158. [PubMed: 8189735]

Bucholz KK, Hesselbrock VM, Shayka JJ, Nurnberger JI Jr, Schuckit MA, Schmidt I, Reich T. Reliability of individual diagnostic criterion items for psychoactive substance dependence and the impact on diagnosis. J Stud Alcohol. 1995; 56(5):500–505. [PubMed: 7475029]

Bucholz KK, Heath AC, Reich T, Hesselbrock BM, Kramer JR, Nurnberger JI, Schuckit MA. Can we subtype alcoholism: a latent class analysis of data from relatives of alcoholics in a multi-center family study of alcoholism. Alcohol Clin Exp Res. 1996; 20:1462–1471. [PubMed: 8947326]

Dawson DA, Saha TD, Grant BF. A multidimensional assessment of the validity and utility of alcohol use disorder severity as determined by item response theory models. Drug Alcohol Depend. 2010; 107:31–38. [PubMed: 19782481]

Gelhorn H, Hartman C, Sakai J, Stallings M, Young S, Rhee SH, Corley RP, Hewitt JK, Hopfer C, Crowley TJ. Toward DSM-V: an item response theory analysis of the diagnostic process for DSM-IV alcohol abuse and dependence in adolescents. J Am Acad Child Adolesc Psychiatry. 2008; 47:1329–1339. [PubMed: 18827724]

Grucza RA, Norberg KE, Bierut LJ. Binge drinking among youths and young adults in the United States: 1979-2006. J Am Acad Child Adolesc Psychiatry. 2009; 48:692–702. [PubMed: 19465879]

Harford TC, Yi HY, Chen CM. The Dimensionality of DSM-IV alcohol use disorders among adolescent and adult drinkers and symptom patterns by age, gender, and race/ethnicity. Alcohol Clin Exp Res. 2009; 33:868–878. [PubMed: 19320629]

Heath AC, Bucholz KK, Slutske WS, Madden PAF, Dinwiddie SH, Dunne MP, Statham DJ, Whitfield JB, Martin NG, Eaves LJ. The assessment of alcoholism surveys of the general community: what are we measuring? Some insights from the Australian twin panel survey. Int Rev Psychiatry. 1994; 6:295–307.

Heath AC, Howells W, Bucholz KK, Glowinski AL, Nelson EC, Madden PAF. Ascertainment of a Mid-Western US female adolescent twin cohort for alcohol studies: assessment of sample representativeness using birth record data. Twin Res. 2002; 5:107–112. [PubMed: 11931688]

Heath AC, Madden PAF, Grant JD, McLaughlin TL, Todorov AA, Bucholz KK. Resiliency factors protecting against teenage alcohol use and smoking: influences of religion, religious involvement and values, and ethnicity in the Missouri Adolescent Female Twin Study. Twin Res. 1999; 2:145–155. [PubMed: 10480749]

Hesselbrock MN, Mesa CE, Bucholz KK, Schuckit MA, Hesselbrock VM. A validity study of the SSAGA: a comparison with the SCAN. Addiction. 1999; 94:1361–1370. [PubMed: 10615721]

Kahler CW, Hoeppner BB, Jackson KM. A Rasch model analysis of alcohol consumption and problems across adolescence and young adulthood. Alcohol Clin Exp Res. 2009; 33:663–673. [PubMed: 19183135]

Kahler CW, Strong DR. A Rasch model analysis of DSM-IV alcohol abuse and dependence items in the National Epidemiological Survey on Alcohol and Related Conditions. Alcohol Clin Exp Res. 2006; 30:1165–1175. [PubMed: 16792564]

Krueger RF, Nichol PE, Kicks BM, Markon KE, Patrick CJ, Iacono WG, McGue M. Using latent trait modeling to conceptualize an alcohol problems continuum. Psychol Assess. 2004; 16:107–119. [PubMed: 15222807]

Langenbucher JW, Labouvie E, Martin CS, Sanjuan PM, Bavly L, Kirisci L, Chung T. An application of item response theory analysis to alcohol, cannabis, and cocaine criteria in DSM-IV. J Abnorm Psychol. 2004; 113:72–80. [PubMed: 14992659]

Martin CS, Chung T, Kirisci L, Langenbucher JW. Item response theory analysis of diagnostic criteria for alcohol and cannabis use disorders in adolescents: Implications for DSM-V. J Abnorm Psychol. 2006; 115:807–814. [PubMed: 17100538]

Muthen BO. A method for studying the homogeneity of test items with respect to other relevant variables. J Educ Stat. 1985; 10:121–132.

Muthen BO, Lehman J. Multiple group IRT modeling: Applications to item bias analysis. J Educ Stat. 1985a; 10:133–142.

Muthen BO, Lehman J. Multiple group IRT modeling: applications to item bias analysis. J Educ Stat. 1985b; 10:133–142.

Muthen, LK.; Muthen, BO. Mplus User's Guide. 5th. Muthen and Muthen; Los Angeles, CA: 2007.

Saha TD, Chou SP, Grant BF. Toward an alcohol use disorder continuum using item response theory: results from the National Epidemiologic Survey on Alcohol and Related Conditions. Psych Med. 2006; 36:931–941.
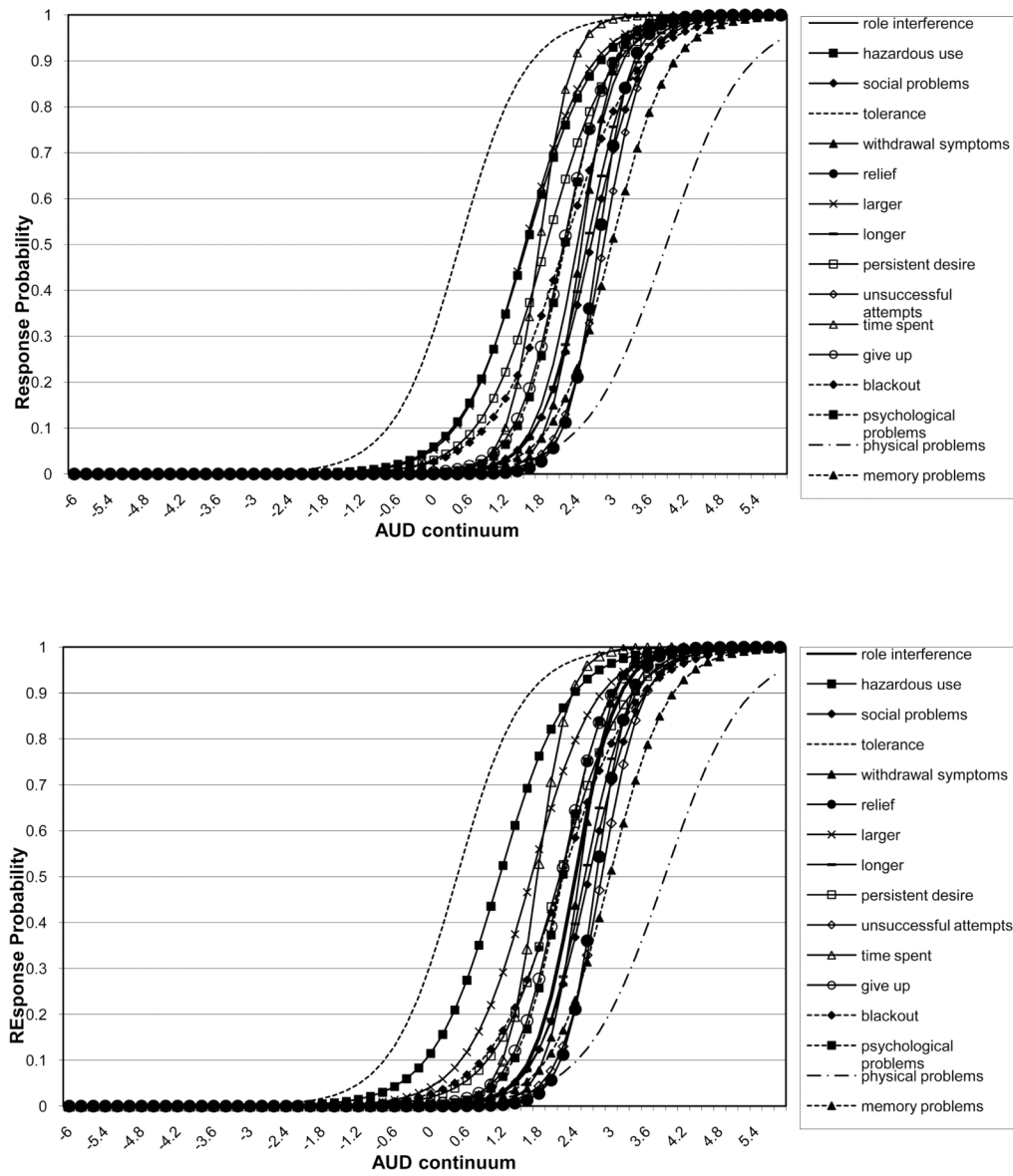
**Figure 1.**
Item characteristic curves from an IRT model of DSM-IV alcohol abuse and dependence symptoms in women aged (a) 18-20 and (b) 21-25.
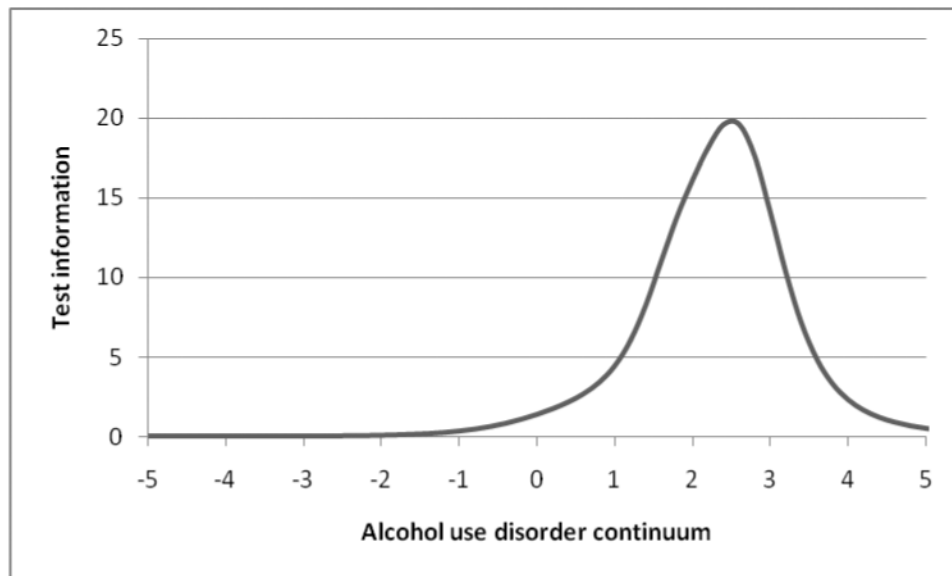
**Figure 2.**
Total information curve for (a) all DSM-IV alcohol abuse and dependence symptoms and for (b) all symptoms except withdrawal and relief in women aged 18-25.

**Table 1**

DSM-IV Alcohol Use Disorder items used in Factor Analytic and Item Response Modeling. Note: the recurrent legal problems criterion was eliminated from the analyses due to negligible prevalence)

| | |
|---|---|
| 1 | **Role failure:** Failure to fulfill major role obligations at work, school or home |
| 2 | **Hazardous use:** recurrent use in physically hazardous situations |
| 3 | **Social problems:** continued use despite social or interpersonal problems |
| 4 | **Tolerance:** need for markedly increased amounts to achieve intoxication/desired effect or markedly diminished effect with continued use of the same amount Withdrawal |
| 5 | **Withdrawal syndrome:** characteristic alcohol withdrawal syndrome |
| 6 | **Relief:** the same/closely related substances taken to relieve/avoid withdrawal symptoms |
| | Larger/longer |
| 7 | **Larger amounts:** drinking larger amounts than intended |
| 8 | **Longer duration:** drinking over a longer period than intended |
| | Cut down/control |
| 9 | **Persistent desire:** Persistent desire to cut down or quit drinking |
| 10 | **Unsuccessful efforts:** Unsuccessful efforts to cut down or quit drinking |
| 11 | **Time spent:** great deal of time is spent in activities necessary to obtain, use, or recover from effects |
| 12 | **Give up activities:** give up/reduce important social, occupational, or recreational activities |
| | Continued use despite problems |
| 13 | **Blackouts:** three or more blackouts from drinking |
| 14 | **Psychological problems:** continued use despite psychological problems caused/made worse by alcohol |
| 15 | **Physical problems:** continued use despite physical problems caused/made worse by alcohol |
| 16 | **Memory problems:** continued use despite memory problems caused/made worse by alcohol (does not include blackouts) |

**Table 2**

Prevalence of lifetime DSM-IV alcohol abuse and dependence criteria, stratified by age (American Psychiatric Association, 2000).

| | All (%) | 18-20 years (%) | 21-25 years (%) |
|---|---|---|---|
| *n* | 2835 | 1158 | 1677 |
| Role failure | 2.47 | 2.07 | 2.74 |
| Hazardous use[***] | 19.07 | 13.80 | 22.71 |
| Legal problems | 0.28 | 0.26 | 0.30 |
| Social problems[*] | 2.19 | 1.55 | 2.62 |
| Tolerance | 39.43 | 38.31 | 40.20 |
| Withdrawal | 1.94 | 1.73 | 2.08 |
|   Syndrome | 1.62 | 1.55 | 1.67 |
|   Relief | 0.81 | 0.52 | 1.01 |
| Larger/Longer | 13.67 | 14.32 | 13.22 |
|   Larger | 13.04 | 13.63 | 12.63 |
|   Longer | 2.15 | 2.16 | 2.14 |
| Cut down/control[*] | 8.10 | 14.32 | 13.22 |
|   Persistent desire[*] | 7.96 | 9.06 | 7.21 |
|   Unsuccessful efforts | 0.95 | 0.95 | 0.95 |
| Time spent | 6.34 | 5.87 | 6.67 |
| Give up activities | 4.33 | 3.97 | 4.59 |
| Use despite physical/psychological problems | 11.17 | 9.92 | 12.03 |
|   Blackouts[**] | 7.54 | 6.21 | 8.46 |
|   Psychological problems | 3.98 | 3.45 | 4.35 |
|   Physical problems | 0.92 | 0.95 | 0.89 |
|   Memory problems | 1.52 | 1.81 | 1.31 |

[***]
p<0.001

[**]
p<0.05

[*]
p<0.10