



How well do DSM-5 criteria measure alcohol use disorder in the general population of older Swedish adolescents? An item response theory analysis

Patrik Karlsson^{a,*}, Sarah Callinan^b, Gerhard Gmel^{c,d,e,f}, Jonas Raninen^{b,g}

^a Department of Social Work, Stockholm University, Stockholm, Sweden

^b Centre for Alcohol Policy Research, La Trobe University, Melbourne, Australia

^c Addiction Medicine, Lausanne University Hospital and University of Lausanne, Rue du Bugnon 23A, 1011 Lausanne, Switzerland

^d Research Department, Addiction Switzerland, Avenue Louis-Ruchonnet 14, 1003 Lausanne, Switzerland

^e Centre for Addiction and Mental Health, Institute for Mental Health Policy Research, 250 College Street, Toronto, ON M5T 1R8, Canada

^f Alcohol and Research Unit, University of the West of England, Frenchay Campus, Coldharbour Lane, Bristol BS16 1QY, UK

^g Karolinska Institutet, Department of Clinical Neuroscience, Stockholm, Sweden

ARTICLE INFO

Keywords:

Alcohol use disorder
Adolescents
Psychometric
Sweden

ABSTRACT

Background: This study assesses the psychometric properties of DSM-5 criteria of AUD in older Swedish adolescents using item response theory models, focusing specifically on the precision of the scale at the cut-offs for mild, moderate, and severe AUD.

Methods: Data from the second wave of Futura01 was used. Futura01 is a nationally representative cohort study of Swedish people born 2001 and data for the second wave was collected when participants were 17/18 years old. This study included only participants who had consumed alcohol during the past 12 months ($n = 2648$). AUD was measured with 11 binary items. A 2-parameter logistic item response theory model (2PL) estimated the items' difficulty and discrimination parameters.

Results: 31.8% of the participants met criteria for AUD. Among these, 75.6% had mild AUD, 18.3% had moderate, and 6.1% had severe AUD. A unidimensional AUD model had a good fit and 2PL models showed that the scale measured AUD over all three cut-offs for AUD severity. Although discrimination parameters ranged from moderate (1.24) to very high (2.38), the more commonly endorsed items discriminated less well than the more difficult items, as also reflected in less precision of the estimates at lower levels of AUD severity. The diagnostic uncertainty was pronounced at the cut-off for mild AUD.

Conclusion: DSM-5 criteria measure AUD with better precision at higher levels of AUD severity than at lower levels. As most older adolescents who fulfil an AUD diagnosis are in the mild category, notable uncertainties are involved when an AUD diagnosis is set in this group.

1. Introduction

This study explores the psychometric properties of DSM-5 criteria of alcohol use disorder (AUD) in a large and nationally representative sample of older Swedish adolescents. There are few recent studies on how well DSM-5 criteria measure AUD in older adolescents and updated studies on representative samples of the general population are needed. Most psychometric work on DSM-defined AUD has been carried out on adult samples. While some psychometric research has focused on adolescents or young adults (Beseler, Taylor, & Leeman, 2010; Boness, Lane, & Sher, 2019; Hagman and Cohn, 2011; Watts, Boness, Loeffelman, Steinley, & Sher, 2021), most of it is based on US college samples.

AUD is associated with substantial harm (Rehm & Shield, 2019). The two main systems for diagnosing AUD are the Diagnostic and Statistical Manual of Mental Disorders (DSM) and the International Statistical Classification of Diseases and Related Health Problems (ICD). DSM-5, launched in 2013, refers to the condition as AUD whereas the current ICD-system refers to either "alcohol dependence" or a "harmful pattern of use", with the latter being less severe (see e.g., Carvalho, Heilig, Perez, Probst, & Rehm, 2019). Regardless of which system is used, accurate measurements are required for adequate monitoring of prevalence, and treatment planning related to problems with alcohol use.

The change from DSM-IV to DSM-5 yielded important changes in how AUD was conceptualized. Whereas DSM-IV consisted of two

* Corresponding author at: Department of Social Work, Stockholm University, SE 106 91 Stockholm, Sweden.

E-mail address: patrik.karlsson@socarb.su.se (P. Karlsson).

<https://doi.org/10.1016/j.addbeh.2024.108007>

Received 8 November 2023; Received in revised form 4 March 2024; Accepted 5 March 2024

Available online 6 March 2024

0306-4603/© 2024 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

disorders – alcohol abuse and alcohol dependence - these were combined into a single AUD in DSM-5 (Hasin, 2014). The legal problems criterion for alcohol abuse was dropped, and a craving criterion was added. DSM-5 measures AUD with 11 criteria, and the same criteria are used regardless of age. AUD is thought to be unidimensional, where higher scores imply more severe AUD. The DSM-5 also includes a severity criterion, distinguishing between mild, moderate, and severe AUD based on the number of criteria fulfilled (Hasin, 2014). Diagnostically, a mild form of AUD is present if 2-3 items are fulfilled, a moderate if 4–5 are fulfilled, and severe if 6 or more symptoms are fulfilled. In adults, different measures of alcohol consumption – including harmful use, craving, binge drinking frequency and self-perceived problematic use – are associated with AUD severity group in a dose–response way, whereas other psychiatric disorders have been shown to be associated with severe AUD only (Mannes, Shmulewitz, Livne, Stohl, & Hasin, 2021).

The prevalence of AUD is usually found to be the highest among young adults (Grant et al., 2017). For instance, a Swiss study found that >30 % of a sample of young men in their early twenties met the criteria for AUD diagnosis (Marmet et al., 2019). These high prevalence rates have raised questions regarding the validity of AUD estimates among young people, and it has been questioned based on work on DSM-IV if we can measure AUD adequately among young people or if we need to use a “adolescent alcohol dependence” to get accurate estimates (Caetano & Babor, 2006). Even if the change in the diagnosis of AUD in DSM-5 was preceded by extensive psychometric research (Hasin, 2014), the diagnosis may not work as well in younger drinkers. For instance, there has been concern that some criteria may work differently in different age groups (Schuckit & Smith, 2021; Vergés et al., 2021). Recent research, using more items to measure each criterion, also suggest that the scale may not be unidimensional (Watts et al., 2021), a finding that runs contrary to much prior work in adolescent and adult samples (Beseler et al., 2010; Hagman & Cohn, 2011; Saha et al., 2020). The severity of specific AUD items is also not consistent across studies, which in part is attributable to varying age of participants and the instrument used to assess AUD criteria (Lane, Steinley, & Sher, 2016). Nonetheless, tolerance and larger/longer seems to be less severe criterion and withdrawal symptoms a more severe criterion among young people (e.g., Beseler et al., 2010). Other research suggests that people may interpret AUD criteria differently at different ages (Marmet et al., 2019) and that some criteria are not understood by younger people as they are intended to do (Slade, Teesson, Mewton, Memedovic, & Krueger, 2013). Thus, it is obvious that we cannot infer how useful current DSM-5 criteria are for diagnosing AUD among older adolescents from psychometric research in adult samples.

The aim of this study is to assess the psychometric properties of DSM-5 criteria of AUD in a large and nationally representative sample of older Swedish adolescents (17/18 years old) using item response theory models (IRT), focusing specifically at how the scale works at the cut-offs for mild, moderate, and severe AUD. As most people who meet an AUD diagnosis in general population surveys endorse relatively few items (Caetano & Babor, 2006), how the scale performs at the cut-off for a mild AUD is of specific interest. As this is where the line is drawn between having an AUD and not having an AUD, it is crucial that the scale measures as precisely as possible there (cf. Baker, 2001). Given the shortage of prior Swedish studies on the topic, and the low reliability in how severe different AUD criteria are estimated to be across studies (Lane et al., 2016), there is not enough information to formulate specific hypotheses and the study has more of an exploratory nature. However, given what is known in prior work, we do expect both tolerance and larger/longer to be among the more commonly endorsed criteria, and withdrawal to be among the least endorsed criteria.

2. Material and method

2.1. Sample

The sample comprises adolescents who report having used alcohol during the past 12 months and who participated in the second wave of the Futura01 study (see e.g., Sjödin, Larm, Karlsson, Livingston, & Raninen, 2021). Futura01 is a prospective cohort study that includes a nationally representative sample of adolescents in Sweden. Baseline data were collected during school-hours in 2017 when participants attended 9th grade (15/16 years old) and data for the first follow-up (wave 2) were collected between March and June in 2019 when participants were 17 or 18 years old, depending on what month they are born. The overwhelming majority were born in 2001 (96 %) and at wave 2 the mean age was 17.3 years (Std = 0.49). At baseline, 343 of 500 invited schools chose to participate in the study (68.8 %) and the number of individual participants were 5541. Of these, 3999 participated in wave 2 (72 %). Participants could either fill in the questionnaire on paper (17 %) or online (83 %). The current analysis includes past 12 months alcohol users who had valid responses to all questions measuring DSM-5 criteria of AUD (n = 2648). These comprise 98 % of all those who reported having used alcohol during the past year (n = 2715). There was no information about AUD in the baseline data, so the analysis is restricted to data from the second wave.

The data collection and overall study design was approved by the regional ethical review board of Stockholm (2017/103–31/5).

2.2. Variables

Participants who reported having ever used alcohol were asked to respond ‘yes’ or ‘no’ to 12 DSM criteria of AUD referring to the past 12 months. Legal problems were excluded from analysis as it is not included in DSM-5, leaving 11 binary DSM-5 criteria for AUD. The participants answered questions in Swedish with the questions being identical to those used to measure dependence and AUD by the Swedish Council of Information on Alcohol and Other Drugs (CAN, 2022). Table 1 lists the wording for each of the criteria in the questionnaire (excluding legal problems) a translated from Swedish to English in this paper by us. This translation may have altered the original meaning of the original

Table 1
Wording for each of the AUD criteria (translated from Swedish to English).

| Criterion | Wording |
|---------------------------------|---|
| Tolerance | During the past 12 months... ...did you drink more in order to get the same effect that you got when you first started drinking? |
| Withdrawal | ...has it happened that your hands shake, you start sweating or feel agitated when you cut down on drinking? |
| Larger/longer | ...during the times when you drank alcohol, did you end up drinking more than you planned when you started? |
| Quit/control | ...have you tried to reduce or stop drinking alcohol but failed? |
| Time spent | ...on the days that you drank, did you spend substantial time obtaining alcohol, drinking, or recovering from the effects of alcohol? |
| Given up activities | ...did you spend less time working, enjoying hobbies, or being with others because of your drinking? |
| Physical/psychological problems | ...have you continued to drink even though you knew that the drinking caused you physical or psychological problems? |
| Role neglect | ...has your drinking on repeated occasions caused you not to fulfil your obligations at work, in school or at home? |
| Hazardous use | ...did your drinking on repeated occasions caused great risk for physical harm (e.g., in traffic) |
| Interpersonal problems | ...did you continue drinking even though this causes continual and recurrent problems with other people? |
| Craving | ...have you had a strong desire or urge to drink alcohol? |

Swedish questions somewhat, so for full transparency we also present the original questions in [Supplementary Table 1](#).

2.3. Statistical analysis

Confirmatory factor analysis (CFA) was first used to determine the dimensionality of the AUD scale. The CFA was performed in Mplus version 8.5 (Muthén & Muthén, 1998–2017) on a tetrachoric correlation matrix using a robust weighted least squares (WLSMV) estimator. Model fit was assessed by the Tucker-Lewis Index (TLI), the Comparative fit Index (CFI), the Root Mean Square Residual (RMSE) and Standardized Root Mean Square Residual (SRMR). According to Hu and Bentler (1999), for good model fit the TLI and the CFI should exceed 0.95, the RMSEA should be lower than 0.06, and the SRMR should be below 0.08.

Next, a two-parameter logistic (2PL) IRT model was fitted to the 11 AUD criteria using Stata vers.16.1 (StataCorp, 2019). Similar to other latent variable models, in an IRT model, the responses to the AUD criteria are assumed to be caused by an unobserved (i.e., latent) variable (Reise & Waller, 2008; Thomas, 2011). This latent variable is typically referred to as *theta*. Each item is assumed to provide information about the underlying trait being measured (i.e., AUD in this case) but tests are typically developed so that different items measure different parts of the trait (Baker & Kim, 2017). ‘Easy’ items are used to measure the lower part of the trait continuum (e.g., mild AUD) and ‘hard’ items the upper part (e.g., severe AUD). How well an entire test measures the entire trait continuum can then be assessed.

The 2PL estimates two kinds of parameters: a ‘difficulty’ parameter (also known as ‘severity’ or ‘location’) and a ‘discrimination’ parameter. Higher values on the difficulty parameter here imply that a given AUD criterion is more difficult to fulfil, and vice versa for lower difficulty values. More commonly endorsed criteria are thus considered less difficult in the context of IRT. The discrimination parameter measures the extent to which a given criterion can distinguish among people lying below and above a given criterion’s difficulty (Baker & Kim, 2017, p. 3). Items with higher discrimination values distinguish better. If the aim is to differentiate well between people at a given location – such as between those who are close to the cut-off for an AUD diagnosis – an item’s discrimination value is an important “index of quality” (Raykov & Marcoulides, 2018, p. 75). Though it would be hard to specify exact values of what may be considered good discrimination values, we follow the suggestions of Baker and Kim (2017, p. 26) (0.01–0.34 = very low, 0.35–0.64 = low, 0.65–1.34 = moderate, 1.35–1.69 = high, >1.70 = very high).

We also present the amount of information carried by the items individually (Item Information Functions, IIFs) and collectively (Test Information Function, TIF). The information value shows the precision of the estimates at given levels of *theta* (Baker & Kim, 2017). It can be seen as a sort of reliability estimate that is allowed to vary across *theta* (Thomas, 2011). This latter property is highly valuable here as it allows for an assessment of how precisely the items (individually and collectively) measure different parts of the AUD continuum. IIFs show the information provided by different items individually, and the TIF shows how much information the items provide jointly along *theta* (Baker & Kim, 2017; Thomas, 2011). TIF at a specific level of *theta* is the sum of the IIFs at the same level of *theta* (Baker, 2021, p. 106). For both IIFs and TIF, higher values mean higher precision. As to the TIF, we specifically illustrate how much information DSM-5 provides at *theta* values corresponding to the cut-off for mild (score of 2), moderate (score of 4), and severe AUD (score of 6). The location of the cut-offs on *theta* were in turn based on the Test Characteristic Curve (TCC). The TCC refers to the expected total score on the scale for each value of *theta* (Raykov & Marcoulides, 2018, p. 150) and *theta* values corresponding to specific sum scores are easily obtained using Stata’s `irttgraph tcc` command. We highlight these values in [Fig. 3](#).

3. Results

3.1. Descriptive statistics

[Table 2](#) presents the prevalence of endorsement of the different AUD criteria. There was a substantial variability in the endorsement prevalence of the AUD criteria, ranging from 33.27 % for having used more alcohol than intended (larger/longer) to 1.13 % for withdrawal symptoms. Tolerance and craving were other quite commonly endorsed criteria, having a prevalence of about 20 % each, whereas quit/control and hazardous use were other less commonly endorsed criteria. In total, 31.8 % of the participants met criteria for AUD. Of the 842 participants fulfilling criteria for AUD, 75.6 % were in the mild category, 38.3% in the moderate and 6.1% in the severe category. Individual fulfilling two criteria constituted 47 % of all participants meeting criteria for AUD (14.95/31.8). The internal consistency of the AUD scale was acceptable, though not quite satisfactory ($\alpha = 0.64$), and there would only be slight improvements if specific items were omitted.

3.2. Fit indices and factor loadings

The factor loadings were substantial for all items, ranging from 0.56 for larger/longer to 0.75 for hazardous use ([Table 2](#)). The factor model in which AUD was treated as unidimensional provided a good fit to the data (RMSEA = 0.030, CFI = 0.959, TLI = 0.948, SRMR = 0.074).

3.3. IRT analysis

The estimated difficulty and discrimination parameters from the IRT analysis of the 11 AUD criteria are presented in [Table 2](#) and [Fig. 1](#). The item characteristic curves (ICCs) in [Fig. 1](#) are based on both difficulty and discrimination estimates from a 2PL IRT. As to difficulty, the ICCs show where along *theta* (the latent AUD trait) people have a probability of 0.5 of endorsing a particular criterion. The steepness of the curves at this point illustrates the criteria’s discrimination. The AUD criteria measured AUD at difficulties from about +0.7 to +3.15. The least severe criterion was larger/longer (0.73), followed by tolerance (1.32) and time spent (1.37). Withdrawal was by far the most severe item (3.16), followed by quit/control (2.61) and hazardous use (2.56).

The discrimination parameters ranged from moderate to very high. The least severe criterion, larger/longer, also had the lowest discrimination value (1.24), followed by tolerance and craving (1.30 and 1.34, respectively) which also had relatively low difficulty. Role neglect had the highest discrimination (2.55), followed by hazardous use (2.38). Thus, the more difficult items overall had higher discrimination values than the less difficult items.

[Fig. 2](#) presents item information functions (IIFs) whereas the test information function (TIF) is shown in [Fig. 3](#). For individual criteria, the highest level of information (i.e. precision) was provided by role neglect, closely followed by hazardous use and interpersonal problems (see [Fig. 2](#)). The relatively mild criteria larger/longer, tolerance and craving provided least information. The TIF in [Fig. 3](#) shows that the joint amount of information in the AUD scale was highest in the more severe continuum, and then declined at the top end of the scale. The information value of the entire scale peaked at a *theta* value of about + 2.5, i.e., slightly above the cut-off for severe AUD. Conversely, the information value was much smaller at the cut-off for mild AUD (*theta* = 0.915, or 2 endorsed criteria). For moderate AUD, the information value was higher than for mild AUD and lower than for severe AUD.

4. Discussion

This study assessed the psychometric properties of self-reported DSM-5 criteria of alcohol use disorder (AUD) in a nationally representative sample of older Swedish adolescents. The findings show that DSM-5 measures severe AUD relatively well, but that the mild form of AUD is

Table 2

Prevalence, factor loadings and IRT (2PL) parameters of DSM-5 criteria of alcohol use disorder during the past 12 months among past 12 months drinkers (n = 2648).

| DSM-5 criteria of AUD during the past 12 months | N | % | Chronbach's alpha if item is deleted | CFA ^a | IRT (2 PL) ^a | |
|---|-------------|--------------|--------------------------------------|------------------------------|-------------------------|----------------|
| | | | | Factor loadings ^b | Difficulty | Discrimination |
| Larger/longer | 881 | 33.27 | 0.63 | 0.56 | 0.73 | 1.24 |
| Tolerance | 560 | 21.15 | 0.61 | 0.59 | 1.32 | 1.30 |
| Craving | 504 | 19.03 | 0.60 | 0.59 | 1.42 | 1.34 |
| Time spent | 459 | 17.33 | 0.60 | 0.65 | 1.37 | 1.63 |
| Physical/psychological problems | 354 | 13.37 | 0.62 | 0.68 | 1.58 | 1.71 |
| Role neglect | 79 | 2.98 | 0.63 | 0.78 | 2.29 | 2.55 |
| Given up activities | 78 | 2.95 | 0.63 | 0.73 | 2.52 | 2.02 |
| Interpersonal problems | 78 | 2.95 | 0.63 | 0.71 | 2.47 | 2.13 |
| Quit/control | 72 | 2.72 | 0.63 | 0.71 | 2.61 | 1.96 |
| Hazardous use | 54 | 2.04 | 0.63 | 0.75 | 2.56 | 2.38 |
| Withdrawal | 30 | 1.13 | 0.64 | 0.69 | 3.16 | 1.92 |
| Cronbach's alpha | | | 0.64 | | | |
| Fit-indices | | | | | | |
| RMSEA (90 % CI) | | | | 0.030 (0.025 0.036) | | |
| CFI | | | | 0.959 | | |
| TLI | | | | 0.948 | | |
| SRMR | | | | 0.074 | | |
| AUD | | | | | | |
| No | 1806 | 68.20 | | | | |
| 0 | 1187 | 44.83 | | | | |
| 1 | 619 | 23.38 | | | | |
| Mild | 637 | 24.05 | | | | |
| 2 | 396 | 14.95 | | | | |
| 3 | 241 | 9.10 | | | | |
| Moderate | 154 | 5.82 | | | | |
| 4 | 108 | 4.08 | | | | |
| 5 | 46 | 1.74 | | | | |
| Severe | 51 | 1.93 | | | | |
| 6 | 27 | 1.02 | | | | |
| 7 | 12 | 0.45 | | | | |
| 8 | 7 | 0.26 | | | | |
| 9 | 1 | 0.04 | | | | |
| 10 | 2 | 0.08 | | | | |
| 11 | 2 | 0.08 | | | | |

^a All estimates are significant at $p < 0.001$.

^b Completely standardized factor loadings.

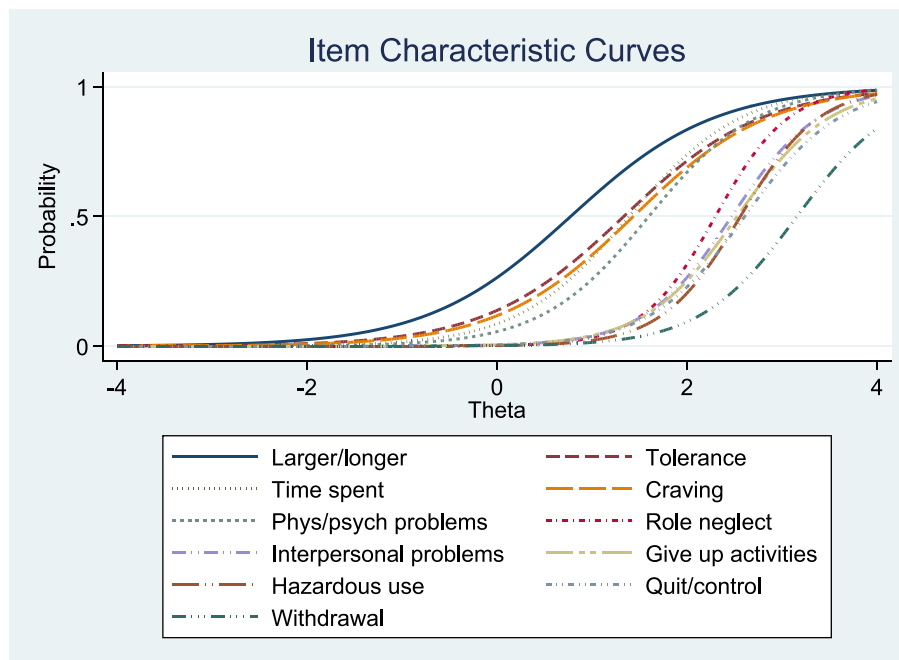


Fig. 1. Item characteristic curves from a two-parametric logistic IRT model of AUD criteria.

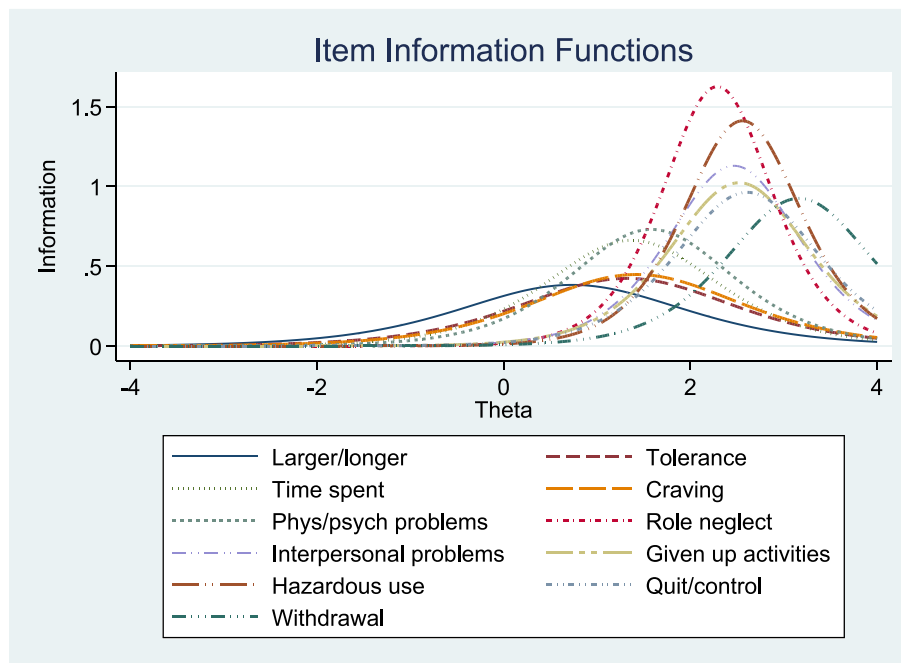


Fig. 2. Item information functions for AUD criteria. Higher values on the Y-axis means that a criterion carries more information.

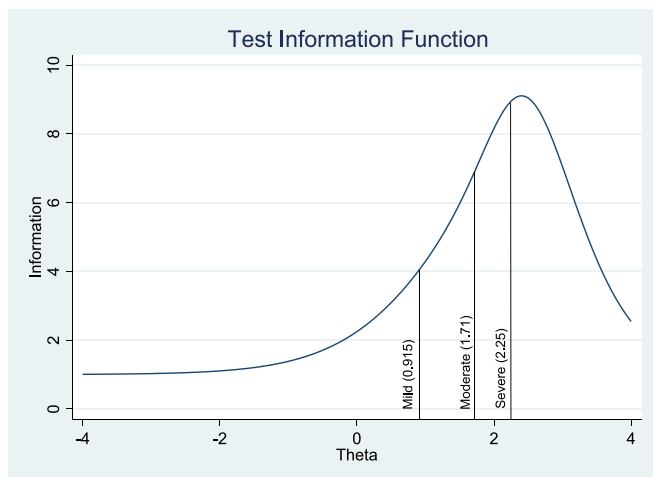


Fig. 3. Test information function for the AUD scale. Higher values on the Y-axis means that the scale carries more information. Theta values shown in the figure corresponds to the cut-off values for mild (2 endorsed criteria), moderate (4 endorsed criteria) and severe (6 endorsed criteria) AUD.

measured with less precision. In this study, most participants (76 %) who fulfilled AUD were in the mild category. The instrument appeared to be particularly unprecise at the point of the scale corresponding to the lowest cut-off for AUD (a score of 2) where almost half (47 %) of those who qualified for an AUD diagnosis found themselves. These findings highlight potential problems with defining mild AUD in the general population of older adolescents according to DSM-5 criteria. IRT analyses in adults, using DSM-5 criteria, corroborates that the information value is lowest in the less severe part of the AUD continuum (Saha, Chou, & Grant, 2020).

Given this finding, the most straightforward approach would be to raise the bar of what is considered AUD in older adolescents. The low level for what is currently considered mild AUD in DSM-5 may also inflate prevalence rates of AUD (Martin, Steinley, Vergés, & Sher, 2011). However, the working group behind AUD in DSM-5 were cautious to set

the cut-off higher because this may fail to identify people with potential treatment needs (Hasin, 2014). To the extent that the cut-off for any AUD will continue to be two endorsed criteria, it is imperative to include additional items that primarily taps the mild form of AUD. Future work may consider extending the DSM-5 list of AUD criteria with items pertaining to drinking patterns of youth alcohol use, similar to what has been done in adult samples (e.g., Saha et al., 2020). This may help to capture mild AUD more precisely. Ultimately, as shown by our study, the DSM criterion seem to have better discrimination for more severe AUDs than for milder AUDs. Various measures of drinking have in IRT-analysis been found to measure less severe alcohol use and related problems, which suggests that questions from for example AUDIT may be included as well. Interestingly, different measures of consumption seem to predict AUD differently across ages. An Australian study has shown that frequency of drinking more strongly predicts scores on the dependence and harmful drinking subscales of AUDIT among young people than what number of drinks per occasion do, and that the reverse holds for people in older age ranges (Callinan, Livingston, Dietze, & Room, 2022). At the same time, younger people typically drank more per occasion and less frequent, whereas older drinkers drank more frequently but less per occasion.

Thus, how well different measures of consumption may complement the present AUD criteria may be related to what is more, and less, normative drinking behavior in a specific age group. As the massive decline in drinking in adolescents during the past two decades (e.g., Raitasalo et al., 2021) may have changed what is normative drinking behavior also among adolescent drinkers (Raninen et al., 2021), several different consumption measures may be tested to identify items that accurately measure different forms of AUD.

Similar to prior work on older adolescents and young adults (Hagman & Cohn, 2011; Marmet et al., 2019; Slade et al., 2021), larger/longer and tolerance were commonly endorsed criteria in this study, whereas withdrawal was not. Somewhat unexpectedly, craving was the third most endorsed item, and the IRT analysis also showed it to have relatively low difficulty and discrimination compared to several other items. This finding contrasts with the Marmet et al. (2019) study on young men, where craving was among the least endorsed items, thus having a relatively high difficulty value. Other more recent research also suggests craving to be relatively uncommon in similarly aged samples

(Slade et al., 2021). In part, the relatively high endorsement rate of craving in this study may be due to the phrasing of the question, which only asked about an urge or desire to drink, but without querying about whether this implied an inability to abstain from drinking. Hazardous use was another criterion that were relatively more difficult in this study compared to what have been found in some other work (Marmet et al., 2019). As the wording of this criteria in the survey explicitly mentioned traffic as an example, the fact that some participants either had not reached the minimum legal age for driving a car (18 years in Sweden) or had not yet received their driving license may have affected the low endorsement of this criterion. Future studies should consider other measures of hazardous use in samples of younger users where traffic may be a less relevant reference point.

It could also be that some items are simply not well-suited to younger drinkers. For instance, the item on increased *tolerance* to alcohol makes sense with an older adult – they have a baseline level in the preceding years to which they can compare their current tolerance. However, for a younger drinker, who only recently started drinking, this question is trickier – any consumption would presumably result in an increase in alcohol tolerance if said tolerance was at its lowest possible level (as a lifetime abstainer) recently. Qualitative work highlights that the high endorsement prevalence of the tolerance criterion among adolescents to a large extent may be considered false positives as also “normative tolerance development” seems to be considered by them when queried about tolerance (Chung & Martin, 2005, p. 198). Prior work has also shown that the most frequently endorsed criterion here – *larger/longer* – is typically not interpreted in terms of compulsion by young drinkers but rather according to social or other reasons (Slade et al., 2013). The assumption involved in this criterion that people actually set limits to their consumption level may not apply particularly well to adolescents (Chung & Martin, 2005). Our results are in line with these previous findings in that larger/longer had the lowest level of discrimination, as also reflected in its relatively poor information value. Thus, in younger populations, this criterion’s validity as an indicator of AUD may be questioned.

Finally, it can be noted that the most commonly endorsed criteria in this study all refer to non-consequences (cf. Martin, Langenbucher, Chung, & Sher, 2014b). Compared to some US studies on college samples (e.g., Hagman & Cohn, 2011), a criterion such as interpersonal problems were rarely endorsed, suggesting that at least some criteria are context dependent (Martin et al., 2014). Besides physical/psychological problems, only a few percent endorsed criteria related to negative consequences of drinking. An alternative way to diagnosing AUD may be to use the Harmful dysfunction approach (Wakefield & Schmitz, 2015). This imposes the requisite that for fulfilling a substance use disorder the individual needs to fulfil criteria that both relates to compulsion and to consequences. Estimating prevalence of substance use disorder according to this approach leads to substantially lower estimates (Wakefield & Schmitz, 2014). Future research should explore whether this approach works better than the DSM-5 approach at different cut-offs of AUD.

4.1. Strengths and limitations

An important limitation is that the study relies on a reflective model of measurement. Items are in this model assumed to be caused by the latent AUD construct rather than causing it (Edwards & Bagozzi, 2000). Though the unidimensional AUD model had an acceptable fit to the data, this in itself does not establish that the reflective model is the correct one (MacCoun, 2013), or that all the included items are conceptually relevant regarding AUD (Martin, Langenbucher, Chung, & Sher, 2014a). Unidimensionality is compatible with several different models, including complex, direct causal relations between AUD items (MacCoun, 2013). The fact that this, and prior work (Hagman & Cohn, 2011), support a unidimensional model may also be related to our use of single rather than multiple items to measure the 11 AUD criteria (Watts et al., 2021). The actual instrument used to measure AUD contributes to

variability in both difficulty (Lane et al., 2016) and discrimination (Vize & Lane, 2022) parameters across studies, where apparently minor differences in how the questions are administrated or worded can have an important effect on the estimates (Lane et al., 2016). The items were administrated in Swedish, and it is possible that there exist some important nuances compared to how they are phrased in other languages, such as English, and that in turn may have impacted on the results and their generalizability to other countries. Our use of single items with binary responses to assess each criterion may also fail to capture a more nuanced picture. For instance, the measure on time spent only refer to actual drinking days, and as such do not capture time spent recovering from drinking the day after the drinking day. Also, all analyses were based on self-reported survey information, which could induce social desirability or recall bias, and this needs to be kept in mind when interpreting the results. The fact that we have not done any external validation analysis of the AUD scale should also be kept in mind. Finally, while we have shown that the scale measures particularly severe AUD more precisely than mild AUD, we lack any quantitative standards for how much better this measurement is.

Key-strengths of the study include a large and nationally representative sample of older adolescents in Sweden. The measure of AUD included all 11 criteria of AUD in DSM-5, rather than proxy measures derived from other instruments. Our use of IRT also allowed us to make a more fine-grained assessment of the psychometric properties of the AUD scale than what would be the case if relying solely on more traditional techniques such as factor analysis. Instead of providing an overall assessment of the AUD scale, IRT facilitates the assessment of the AUD scale’s psychometric properties at different levels of AUD severity. Using DSM-5 criteria provides updated information on AUD using the latest version of the DSM and the general population sample makes our results generalizable to the wider target population of Swedish older Swedish adolescents.

5. Conclusion

This large, general population study shows that DSM-5 measures severe AUD with higher precision than mild AUD in older adolescents. This, together with the fact that most older adolescents who meet criteria for AUD have mild AUD, raises challenges to how well the DSM-5 measures AUD in general population samples of older adolescents.

Funding

PK was funded by Systembolagets alkoholforskningsråd (SRA) 2021-0019. JR was funded by Systembolagets alkoholforskningsråd (SRA) 2021-0019 and Forte 2021-01725.

CRedit authorship contribution statement

Patrik Karlsson: Writing – review & editing, Writing – original draft, Formal analysis, Conceptualization. **Sarah Callinan:** Writing – review & editing, Conceptualization. **Gerhard Gmel:** Writing – review & editing, Conceptualization. **Jonas Raninen:** Writing – review & editing, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.addbeh.2024.108007>.

References

- Baker, F. B., & Kim, S. H. (2017). *The basics of item response theory using R*. New York: Springer.
- Baker, F. B. 2001, *The basics of item response theory*. College Park, M.D, ERIC Clearinghouse on Assessment and Evaluation.
- Beseler, C. L., Taylor, L. A., & Leeman, R. F. (2010). An item-response theory analysis of DSM-IV alcohol-use disorder criteria and “binge” drinking in undergraduates. *Journal of Studies on Alcohol and Drugs*, *71*, 418–423.
- Boness, C. L., Lane, S. P., & Sher, K. J. (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.
- Caetano, R., & Babor, T. F. (2006). Diagnosis of alcohol dependence in epidemiological surveys: An epidemic of youthful alcohol dependence or a case of measurement error? *Addiction*, *101*, 111–114.
- Callinan, S., Livingston, M., Dietze, P., Gmel, G., & Room, R. (2022). Age-based differences in quantity and frequency of consumption when screening for harmful alcohol use. *Addiction*, *117*, 2431–2437.
- CAN, 2022. Användning och beroendeproblem av alkohol, narkotika och tobak. En studie med fokus på år 2021 i Sverige. [Use and dependence problems of alcohol, narcotics and tobacco. A study with focus on year 2021 in Sweden]. Centrum för Alkohol och narkotikaupplysning [The Swedish Council for Information about Alcohol and other Drugs], Stockholm.
- Carvalho, A. F., Heilig, M., Perez, A., Probst, C., & Rehm, J. (2019). Alcohol use disorders. *The Lancet*, *394*, 781–792.
- Chung, T., & Martin, C. S. (2005). What were they thinking? Adolescents' interpretations of DSM-IV alcohol dependence symptom queries and implications for diagnostic validity. *Drug and Alcohol Dependence*, *80*, 191–200.
- Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*, 155.
- Grant, B. F., Chou, S. P., Saha, T. D., Pickering, R. P., Kerridge, B. T., Ruan, W. J., & Hasin, D. S. (2017). Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001–2002 to 2012–2013: Results from the National Epidemiologic Survey on alcohol and related conditions. *JAMA Psychiatry*, *74*, 911–923.
- Hagman, B. T., & Cohn, A. M. (2011). Toward DSM-V: Mapping the alcohol use disorder continuum in college students. *Drug and Alcohol Dependence*, *118*, 202–208.
- Hasin, D. (2014). Truth (validity) and use despite consequences: The DSM-5 substance use disorder unidimensional syndrome. *Addiction*, *109*, 1781–1782.
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1–55.
- Lane, S. P., Steinley, D., & Sher, K. J. (2016). Meta-analysis of DSM alcohol use disorder criteria severities: Structural consistency is only ‘skin deep’. *Psychological Medicine*, *46*, 1769–1784.
- MacCoun, R. J. (2013). The puzzling unidimensionality of DSM-5 substance use disorder diagnosis. *Frontiers in Psychiatry*, *4*, 10.
- Mannes, Z. L., Shmulewitz, D., Livne, O., Stohl, M., & Hasin, D. S. (2021). Correlates of mild, moderate, and severe alcohol use disorder among adults with problem substance use: Validity implications for DSM-5. *Alcoholism: Clinical and Experimental Research*, *45*, 2118–2129.
- Marmet, S., Studer, J., Bertholet, N., Grazioli, V. S., Daeppen, J. B., & Gmel, G. (2019). Interpretation of DSM-5 alcohol use disorder criteria in self-report surveys may change with age. A longitudinal analysis of young Swiss men. *Addiction Research and Theory*, *27*, 489–497.
- Martin, C. S., Langenbucher, J. W., Chung, T., & Sher, K. J. (2014a). Response to commentaries. *Addiction*, *109*, 1784–1785.
- Martin, C. S., Langenbucher, J. W., Chung, T., & Sher, K. J. (2014b). Truth or consequences in the diagnosis of substance use disorders. *Addiction*, *109*, 1773–1778.
- Martin, C. S., Steinley, D. L., Vergés, A., & Sher, K. J. (2011). The proposed 2/11 symptom algorithm for DSM-5 substance-use disorders is too lenient. *Psychol Med.*, *41*, 2008–2010.
- Muthén, L. K., Muthén, B. O., 1998-2017. *Mplus User's Guide*. eight ed. Muthén & Muthén, Los Angeles.
- Raitasalo, K., Kraus, L., Bye, E. K., Karlsson, P., Tigerstedt, C., Törrönen, J., & Raninen, J. (2021). Similar countries, similar factors? Studying the decline of heavy episodic drinking in adolescents in Finland, Norway and Sweden. *Addiction*, *116*, 62–71.
- Raninen, J., Larm, P., Svensson, J., Livingston, M., Sjödin, L., & Karlsson, P. (2021). Normalization of non-drinking? Health, school situation and social relations among Swedish ninth graders that drink and do not drink alcohol. *International Journal of Environmental Research and Public Health*, *18*, 11201.
- Raykov, T., & Marcoulouides, G. A. (2018). *A course in item response theory and modeling with Stata*. College Station, Texas: StataPress.
- Rehm, J., & Shield, K. D. (2019). Global burden of disease and the impact of mental and addictive disorders. *Current Psychiatry Reports*, *21*, 1–7.
- Reise, S. P., & Waller, N. G. (2008). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*, 27–48.
- Saha, T. D., Chou, S. P., & Grant, B. F. (2020). The performance of DSM-5 alcohol use disorder and quantity-frequency of alcohol consumption criteria: An item response theory analysis. *Drug and Alcohol Dependence*, *216*, Article 108299.
- Schuckit, M. A., & Smith, T. L. (2021). Endorsement of specific alcohol use disorder criterion items changes with age in individuals with persistent alcohol use disorders in 2 generations of the San Diego prospective study. *Alcoholism: Clinical and Experimental Research*, *45*, 2059–2068.
- Sjödin, L., Larm, P., Karlsson, P., Livingston, M., & Raninen, J. (2021). Drinking motives and their associations with alcohol use among adolescents in Sweden. *Nord Stud Alcohol Drugs*, *38*, 256–269.
- Slade, T., Mewton, L., O'Dean, S., Tibbetts, J., Clay, P., Isik, A., & Swift, W. (2021). DSM-5 and ICD-11 alcohol use disorder criteria in young adult regular drinkers: Lifetime prevalence and age of onset. *Drug and Alcohol Dependence*, *229*, Article 109184.
- Slade, T., Teesson, M., Mewton, L., Memedovic, S., & Krueger, R. F. (2013). Do young adults interpret the DSM diagnostic criteria for alcohol use disorders as intended? A cognitive interviewing study. *Alcoholism: Clinical and Experimental Research*, *37*, 1001–1007.
- StataCorp. (2019). *Stata statistical software: Release 16*. College Station, TX: StataCorp.
- Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment*, *18*, 291–307.
- Vergés, A., Lee, M. R., Martin, C. S., Trull, T. J., Martens, M. P., Wood, P. K., & Sher, K. J. (2021). Not all symptoms of alcohol dependence are developmentally equivalent: Implications for the false-positives problem. *Psychology of Addictive Behaviors*, *35*, 444–456.
- Wakefield, J. C., & Schmitz, M. F. (2014). How many people have alcohol use disorders? Using the harmful dysfunction analysis to reconcile prevalence estimates in two community surveys. *Frontiers in Psychiatry*, *5*, 10.
- Wakefield, J. C., & Schmitz, M. F. (2015). The harmful dysfunction model of alcohol use disorder: Revised criteria to improve the validity of diagnosis and prevalence estimates. *Addiction*, *110*, 931–942.
- Watts, A. L., Boness, C. L., Loeffelman, J. E., Steinley, D., & Sher, K. J. (2021). Does crude measurement contribute to observed unidimensionality of psychological constructs? A demonstration with DSM-5 alcohol use disorder. *Journal of Abnormal Psychology*, *130*, 512.