*Original Research Article*

# Seven Symptoms Over Nearly 4,000 Days: Item-Level Variability in the Psychometric Properties of Daily Alcohol Use Disorder Symptoms in Young Adult Drinkers

Assessment
1–15

DOI: 10.1177/10731911251408577
journals.sagepub.com/home/asm
Sage

**Kevin M. King[1], Dahyeon Kang[1], Megan E. Schultz[1], Christine M. Lee[1], Jonas Dora[1], Cassandra L. Boness[2], and Ashley L. Watts[3]**

## Abstract

People experience symptoms of alcohol use disorders (AUD) in their daily lives, including more impairment-based symptoms (e.g., hazardous use, interpersonal problems) and symptoms based on the intensity or frequency of consumption (e.g., cravings or limit violations). We estimated the psychometric properties of seven manifestations of daily AUD symptoms in a high-risk sample of regularly drinking young adults and identified the optimal operationalization and thresholds for those symptoms. We estimated item response theory (IRT) models in ecological momentary assessment data ($n$ = 527, age 18–22, 45% female) assessed over 3,963 alcohol use days. Symptoms were relatively common on drinking days, especially symptoms related to consumption (such as time spent drinking or consuming larger among than intended). The specific threshold or item used to define each AUD symptom could have a substantial impact on item parameters. Tolerance was best loaded onto a factor of daily AUD symptoms when operationalized as *sensitivity* to the effects of alcohol, while larger/longer was best reflected as drinking *much more* than intended (e.g., 3+ drinks). Daily life research focusing only on alcohol-related consequences misses important information about common experiences of AUD symptoms in daily life. Refinement of daily measures of AUD symptoms could help researchers understand how the disorder develops over time.

## Keywords

alcohol use disorder, ecological momentary assessment, item response theory

Alcohol use disorder (AUD) is highly prevalent and rapidly changing during the transition to young adulthood. Nine percent of 18- to 20-year-olds meet past-year criteria for AUD, and prevalence rates peak at 19% when young adults are aged 24 to 25 (Substance Abuse and Mental Health Services Administration, 2021). At least some people who experience AUD in young adulthood experience a long-term persistent course of AUD, but desistence from even the most severe forms of AUD to milder forms of disorder (or to no disorder at all) also peaks during young adulthood (M. R. Lee et al., 2018; Vergés et al., 2012). These changes coincide with lifetime peaks in alcohol use: 83.5% of U.S.-based young adults report use, and 30.5% report binge drinking (Substance Abuse and Mental Health Services Administration, 2021). Thus, understanding the experience of AUD symptoms in young adulthood may provide fundamental insights into how AUD develops during this vulnerable developmental period.

Many theories of AUD describe the development of AUD in terms of how AUD symptoms change over long and short periods of time. For example, repeated episodes of heavy drinking are thought to promote increases in tolerance, which eventually lead to physical and psychological withdrawal following periods of abstinence (Koob & Le Moal, 2001). Experiencing craving is thought to increase the probability of use and lapses after abstinence, which is then thought to further increase cravings and further use (Koob & Le Moal, 2001, 2008). Some AUD symptoms are characterized by the combination of impairment from drinking behaviors (such as hazardous use or interpersonal problems) *and* by the continued drinking despite that impairment, reflecting increasing loss of control over use. Other AUD symptoms are defined by recurrent patterns of or changes in use-related behaviors (such as persistently drinking in larger amounts or

[1]University of Washington, Seattle, USA
[2]The University of New Mexico, Albuquerque, USA
[3]Vanderbilt University, Nashville, TN, USA

**Corresponding Author:**
Kevin M. King, Department of Psychology, University of Washington, Box 351525, Seattle, WA 98195-1525, USA.
Email: kingkm@uw.edu

for longer periods of time or than intended or increasing tolerance). In other words, AUD symptoms are defined as experiences related to drinking that persist, recur, or change in intensity or impact over time.

Although contemporary theoretical models articulate how people experience AUD symptoms in their daily lives (Berridge & Robinson, 2016; Koob & Le Moal, 2008), few studies have modeled how symptoms unfold and change over time in people's daily lives. For example, few studies have attempted to model how instances of alcohol-related impairment (such as getting into a fight while drunk) in daily life influence future drinking (Fairlie et al., 2016) and in turn future impairment, or how tolerance emerges as a function of drinking behaviors. In part, the disconnect between theories and models has been caused by a lack of validated measures that capture symptoms close in time to when they occur and at timescales where people can report them most accurately. Existing "gold standard" measures of AUD symptoms (e.g., Structured Clinical Interview for Diagnostic and Statistical Manual of Mental Disorders [SCID], Alcohol Use Disorders Identification Test (AUDIT), Alcohol Symptom Checklist; Allen et al., 1997; Babor & Higgins-Biddle, 2001; First, 2015; Hallgren et al., 2022) require people to retrospect about their experience of symptoms, as well as their frequency, persistence, recurrence, and change, often over prolonged periods like the past 12 months or one's lifetime. Thus, to bridge theoretical and statistical models, it is urgent to develop and validate reliable and informative measures of people's experience of AUD symptoms in their daily lives, which include more impairment-based symptoms (e.g., hazardous use, interpersonal problems) as well as symptoms based on the intensity or frequency of consumption (e.g., tolerance, cravings, or limit violations). The goal of the present study was to describe the prevalence of ad hoc measures of daily AUD symptoms in a high-risk sample of regularly drinking young adults, to estimate the psychometric properties of those symptoms, and to identify the optimal operationalization and thresholds for AUD symptoms in daily life.

## *Limitations of Existing Self-Reports of AUD*

Because retrospective self-reports of AUD ask people to report on their experiences across long stretches of time, and all possible contexts, they cannot be used to model how symptoms unfold or emerge over time. It is possible that retrospective reports of AUD may be reliable for screening or identifying AUD (Hallgren et al., 2022), but they also do not provide reliable information about how AUD is experienced from day to day. Retrospective reports require people to aggregate massive amounts of information, and the process of retrospective recall is thought to reflect an active reconstruction process that is influenced by factors related to the encoding and recall of memories, the emotional salience of a memory, and heuristics involved in judging a response to an item (Conner & Barrett, 2012; Shiffman et al., 2008). For instance, around half of the variance in global self-reports of personality is shared with daily measures of personality (Finnigan & Vazire, 2018), which indicates that people are relatively accurate at recalling what they *generally* do in daily life. Still, these reports can be biased by peak-end and other retrospective biases (Alaybek et al., 2022; Scharbert et al., 2024), which limit the inferences one can draw about daily life experiences from global reports.

Accurate recall is further complicated by other features. Global self-reports not only require participants to report the occurrence of a thought, feeling, or behavior but also its distal and proximal antecedent(s) and consequence(s), as well as its specific timing. Thus, these measures cannot be used to test hypotheses about how AUD symptoms change, recur, or persist over time, or what proximal factors might shape or be shaped by their occurrence (Cullum et al., 2012; Stone et al., 2007; Trela et al., 2016). In contrast, retrospection over shorter and more clearly defined periods of time (as in ecological momentary assessment [EMA] or daily diary studies) allows people to more accurately recall episodic information instead of heuristic beliefs about themselves (Robinson & Clore, 2002). However, a measure of AUD symptoms suitable for use in people's daily lives would allow researchers to study temporally dynamic processes related to AUD. For example, how do the characteristics or contexts of drinking episodes themselves (like the number of drinks, the rate of drinking, or the level of intoxication) influence the likelihood of experiencing AUD symptoms (Dvorak et al., 2014; Moore et al., 2011; Simons et al., 2010), and does this vary across different AUD symptoms? How do persistence, changes, or recurrence in AUD symptoms over time in people's daily life relate to the development of those symptoms in the future and to the emergence of AUD diagnoses? A validated measure of daily AUD symptoms would allow the interrogation of questions about *how and when* AUD symptoms are experienced and how they persist, change, and recur over time.

## *Existing Daily Measures of AUD*

C. M. Lee and colleagues (2017) validated measures of alcohol-related consequences that can be used to measure select AUD symptoms reflecting daily impairment from use, including hazardous use, social/interpersonal consequences, failure to fulfill role obligations, and time spent drinking or getting over the effects of drinking. A few other studies adapted global self-report measures of AUD symptoms reflecting compulsion and loss of control over drinking to create checklist of "acute dependence symptoms" (Dvorak et al., 2014; Simons et al., 2010, 2014), including loss of control, tolerance, and withdrawal. Existing studies

generally support the finding that consuming higher levels of alcohol in a given drinking episode was associated with more alcohol-related consequences or acute dependence symptoms. The present study is an attempt to build on this important prior research by conducting a formal psychometric analysis of item properties, such as discrimination or difficulty of daily AUD symptoms, to inform how well different items can be used to represent the experience of daily AUD symptoms.

However, in several of these assessments, individual symptoms are assessed with single indicators, despite robust evidence that a symptom's base rate, or severity, varies as a function of the particular phrasing of that item (Boness et al., 2019; Lane et al., 2016). For example, the prevalence of tolerance was 9% when students were asked if they ever consumed as much as a fifth of liquor in a day (20 drinks/3 bottles of wine/3 six packs), but rose to 22% when students were asked if they could drink a lot more before they got drunk than they had previously (Boness et al., 2019). A critical step toward developing robust and valid measures of AUD symptoms suitable for daily life research is to document how the use of different items that intend to capture AUD symptom impacts the psychometric properties of the assessment.

### *Present Study*

We are attempting to make progress toward validating measures of daily AUD symptoms in two ways at present. First, we have worked to test the convergence of daily AUD symptoms with retrospective measures of AUD symptoms (Kang et al., in press). Second, in the current study, we used item response theory (IRT) to estimate the psychometric properties of daily AUD symptoms. IRT models the probability of item endorsement given the level of an unobserved latent variable and provides information about both item difficulty and discrimination and (in the case of binary data) is identical to a confirmatory factor analysis of discrete variables (Takane & De Leeuw, 1987). Discrimination is akin to a factor loading, in that it reflects how the probability of item endorsement changes as a function of the level of latent daily AUD symptoms. A higher discrimination value means that an item does a better job differentiating between participants that are higher or lower than a given level of daily AUD symptoms. Item difficulty reflects the expected level of the latent daily AUD symptoms variable at which the item has a 50% probability of endorsement and is akin to an item intercept or threshold from a latent factor model (when multiplied by the discrimination, it reflects the expected probability of endorsing the item at the mean of the latent variable). For example, tolerance may be a low-difficulty item, such that reporting tolerance only may reflect a relatively low severity of daily AUD symptoms, while endorsing hazardous use may reflect a relatively high severity of daily AUD symptoms.

We used data from a high-risk sample of young adults, combined across two different EMA studies, who reported drinking alcohol at least weekly. Young adults who drink weekly exhibit twice the prevalence of AUD in epidemiological studies relative to the general population of young adults (Substance Abuse and Mental Health Services Administration, 2021). At least some people who experience AUD in young adulthood experience a long-term persistent course of AUD (M. R. Lee et al., 2018; Vergés et al., 2012), and this age reflects lifetime peaks in alcohol use: 83.5% of U.S.-based young adults report use, and 30.5% report binge drinking (Substance Abuse and Mental Health Services Administration, 2021). Thus (similar to prior studies, Dvorak et al., 2014; Simons et al., 2010), it is critical to understand the psychometric characteristics of daily AUD symptoms in this group of young adults at high risk for the development of AUD. To capture AUD symptoms, we identified items from existing measures of daily alcohol-related consequences (M. R. Lee et al., 2018), as well as measures of daily alcohol use, drinking intentions, and cravings that mapped onto different AUD symptoms. Seven of 11 AUD symptoms were well covered in existing EMA assessments: hazardous use, social consequences, failure to fulfill role obligations, time spent, tolerance, larger/longer, and craving. We aimed to estimate the prevalence of these AUD symptoms at the daily level, estimate their psychometric properties, and understand how different expressions of symptoms impacted those psychometric properties to identify items that optimally measured daily AUD symptoms.

## Method

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study.

### *Data*

We combined data from two samples that collected daily reports of the prior day's alcohol use and six AUD symptoms. In the second sample, we also collected EMA reports of a seventh symptom (craving). Thus, we combined data across studies for the analysis of six symptoms and report the degree to which parameter estimates changed when including craving. This study was exploratory, and the analyses were not pre-registered. We focused on estimating effect sizes and uncertainty in our estimates rather than focusing on hypothesis testing. All model syntax and supplemental figures and tables are found on OSF (https://osf.io/6m74u/). Data are available on request.

*Sample 1.* Participants were undergraduate students from a single university in the Pacific Northwest ($n$ = 134, 62%

female, 36% male, $M_{age}$ = 18.58, $SD$ = .67) who were recruited as part of a larger study via the university's Psychology Subject Pool and completed a general screening survey. Eligibility was limited to participants who reported alcohol or cannabis use at least weekly, owned a smartphone, and for whom English was their primary language. Participants reported 86% heterosexual orientation, with 8% bisexual and 6% gay, lesbian, or questioning identity. Participants were largely White (73%), with 8% identifying as Hispanic/Latino, 27% as Asian, Asian American, or Pacific Islander, and 6% as African or African American (participants could endorse more than one race/ethnicity).

*Sample 2.* Participants were recruited for a larger study on the development of alcohol and cannabis use problems during young adulthood. Participants were recruited from King, Pierce, and Snohomish Counties in Washington State from both college and non-college sources to ensure a representative sample of young adults in Washington State. We recruited using internet (Facebook, Instagram, TikTok, YouTube, Twitter, Craigslist, and Reddit) and non-internet (newspaper advertisements, flyers, and university registrar lists) sources. Participants were required to be between the ages of 18 and 22 at study screening, own a smartphone, be fluent in English, and report drinking *or* using cannabis "about once per week" or more over the past 3 months. Participants were excluded if they were not fluent in English or if they moved to the United States after age 12.

Participants were young adults at baseline ($n$ = 496, age 18–22, $M_{age}$ = 20.3, $SD$ = 1.3, 45% cisgender women, 42% cisgender men; 8.5% nonbinary/gender queer/gender nonconforming; 4.0% transgender; 0.2% nongendered). Approximately 67% of the sample attended a 4-year college at the time of recruitment. Participants endorsed a variety of racial and ethnic identities, and respective proportions broadly reflected Washington census data from the counties from which participants were recruited: 54% solely non-Hispanic White, 28.5% Asian, 6.6% Black, 8.4% Hispanic/Latino, and 22.7% who endorsed more than one ethnicity. Finally, 9.8% of the sample was born outside the United States.

## Sample Size Justification

This is a secondary data analysis study, so the sample size for the current study was based on different objectives (e.g., power, resource, and time limitations) than the present study. The sample size for Sample 1, which was a pilot study for Sample 2, was the number of participants that could be recruited in the 1-year period that the study was active. The sample size for Sample 2 was determined by the power to detect effects for the main aims of the larger study (e.g., testing moderation of the within-person association between momentary affect and urgency and predicting daily substance use from affect and urgency; King et al., 2024). We used all available data for the current study.

*Procedure.* Samples 1 and 2 were collected using similar study procedures except as noted and were approved by the university Institutional Review Board. Participants completed web-based baseline surveys assessing participant demographics, emotion regulation, substance use, mental health, and other variables not reported here. Participants were then oriented to the EMA portion of the study in a university lab setting (Sample 1) or over Zoom (Sample 2). First, participants completed a practice EMA survey in order for them to be familiarized with the study protocol. In addition, participants were given a handout providing details regarding the study and a study calendar on which they marked the days when their study session would start and which days the surveys would be sent out. Following the in-lab baseline and EMA training, participants completed 8 days of EMAs over the course of 2 weekends (Thursday through Monday morning; Sample 1) or 8 weekends (Thursday through Monday morning; Sample 2). EMAs were sent within five 3-hour blocks in between 9 am and 11 pm, with at least 1 hour between surveys. Participants had 1 hour to respond and were sent up to two reminders within the hour if they had not completed their survey. In Sample 1, participants were compensated with course credit for participation. In Sample 2, participants received $50 for the baseline survey and were paid $1 per EMA, with a $5 bonus for completing 80% (i.e., 17/21) of EMAs for a given weekend ($258 total possible).

## Measures

Each morning of the study, participants reported on their alcohol use and completed a measure of alcohol-related consequences. In the afternoon, participants reported on their intentions to drink later in the day. Participants also reported on their alcohol use and alcohol-related consequences in the afternoon survey if they missed the morning survey. At each EMA, participants reported on alcohol craving. From these data (retrospective reports of alcohol use and consequences, intentions, and cravings), we were able to construct measures of six (Sample 1) or seven (Sample 2) daily-level symptoms that aligned well with a *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.; DSM-5; American Psychiatric Association, 2013) AUD symptoms: larger/longer, time spent, craving (Sample 2 only), failure to fulfill role obligations, social/interpersonal problems, hazardous use, and tolerance. We could not proxy desire to quit/cut down on drinking, giving up important activities because of drinking, drinking despite physical/psychological problems, or withdrawal. All items have direct parallels to criterion described in manuals for existing diagnostic instruments, such as the Structured Clinical

Interview for the DSM (First et al., 2016) and the Alcohol Use Disorder and Associated Disabilities Interview Schedule-5 (AUDADIS-5) (Grant et al., 2015).

*Hazardous Use.* We used two items ("I hurt or injured myself by accident" and "I couldn't remember what I did") from a measure of alcohol-related consequences adapted from existing measures of AUD and extensively validated in young adults (C. M. Lee et al., 2017). These items map onto the AUD criteria of hazardous use because they indicate behaviors related to use that put the participant or others at risk of harm (First et al., 2016). Although they do not directly reflect hazardous behaviors in the context of drinking, blackouts have been directly linked to hazardous behaviors (Hingson et al., 2016). At the morning assessment (or the 2nd assessment of the day if they missed the morning assessment), participants reported which of the consequences they experienced as a result of their alcohol use the previous day (0 = *No*, 1 = *Yes*), or indicated they did not use, or that they did use but experienced no consequences of use. This AUD symptom was scored with an if/any criteria: if participants endorsed either symptom, the hazardous use symptom was scored as present.

*Social/Interpersonal Problems.* We used four items from C. M. Lee et al. (2017): "I became aggressive," "I did or said something that embarrassed me," "I was rude or obnoxious," and "I got into a fight/argument." These items reflect impairment in social relationships associated with alcohol use (First et al., 2016). Again, participants reported which of the consequences they experienced as a result of their alcohol use the previous day (0 = *No*, 1 = *Yes*). This AUD symptom was scored with an if/any criteria: if participants endorsed any symptom, the hazardous use symptom was scored as present.

*Failure to fulfill role obligations.* We used a single item from Lee et al. (2017): "I was unable to do work/schoolwork," (0 = No, 1 = Yes). This item reflects impairment in role obligations associated with use. (First et al., 2016)

*Time Spent (Obtaining Alcohol, Drinking, or Getting Over the Effects of Alcohol).* We used two items from C. M. Lee et al. (2017) and a third item reflecting the number of hours spent drinking. First, participants reported which of the consequences they experienced as a result of their alcohol use the previous day (0 = *No*, 1 = *Yes*). Time spent was measured with "I had a hangover" and "I experienced nausea or vomiting," to encompass broader aspects of time allocation associated with alcohol consumption (and specifically, time spent getting over the effects of alcohol); these items are found in existing clinical interviews for AUD (First et al., 2016; Grant et al., 2015). An important limitation here is that morning reports of nausea/vomiting could reference such symptoms that occurred either during OR after the drinking episode, so there may be some confounding with vomiting that occurred only in the context of heavy drinking episodes. We operationalized these items as time spent rather than withdrawal because we measured them only in the context of drinking episodes (e.g., the acute effects of alcohol wearing off the morning after drinking; Prat et al., 2009), whereas withdrawal is operationalized as symptoms that arise *after people cut down or quit* heavy and prolonged drinking. We also measured time spent drinking using a single item where participants reported how many hours they drank alcohol the previous day, with response options ranging from 0 to 24 hours, and explored four thresholds: at least 4, 8, 12, or 16 hours. For the final models, this AUD symptom was scored if participants reported experiencing either consequence (hangover or nausea/vomiting) or met the given threshold of the number of hours spent drinking.

*Tolerance.* We measured tolerance in two ways. First, we asked if participants reported consuming 20 or more drinks the previous day; the ability to consume more than 20 drinks in a day implies the presence of tolerance and has been used in prior measures of AUD (such as the AUDADIS) as a tolerance indicator (Grant et al., 2015). At the morning report, participants reported the number of drinks they consumed the previous night using a visual analogue scale ranging from 0 to 30 or more drinks, how many hours they drank (0–24 hours), and rated their level of intoxication on a visual analogue scale ranging from 0 ("Not at all/I didn't use alcohol") to 100 ("Very drunk/high") that was accompanied by a standard definition of an alcoholic drink. When participants reported consuming 20+ drinks on the previous day, we counted tolerance as present (Lejoyeux et al., 2001).

We also operationalized tolerance, initially, as the ratio of the rate of drinking (e.g., the number of drinks divided by the number of hours spent drinking divided by the level of intoxication reported), such that higher scores reflected more drinks per hour per unit of intoxication. This effort sought to capture the acquired feature of tolerance, feeling less intoxicated over time as you drink the same amount. However, this item was negatively correlated with all other daily AUD symptoms ($r = -.04$ to $-.17$), which produced negative factor loadings on a latent AUD factor in the initial multilevel factor analyses (factor loading = $-.33$), poor model fit, and difficulty with model estimation and convergence. In contrast, the *inverse* of that score (e.g., the ratio of intoxication over the rate of drinking, or drinks per hour, where *lower* scores reflected higher tolerance) had moderate correlations with other AUD symptoms and reasonable factor loadings on a latent AUD factor (factor loading = .58). We tested four thresholds for this ratio, coded at the mean, 66th, 75th, and 90th percentiles of the sample, and discuss the implications of this choice and these findings extensively in the discussion section. For the final models, this AUD symptom was scored if participants reported experiencing either

drinking 20+ drinks or met the given thresholds of the ratio of intoxication to the rate of drinking.

*Larger/Longer Drinking Than Intended.* We assessed larger/longer using a difference score computed from two items. The goal was to capture the AUD symptom reflecting loss of control over alcohol use and indexed by the difference between a participant's intended versus actual number of drinks consumed (First et al., 2016), although it is important to acknowledge that this approach only captures drinking in larger amounts, and not for longer periods of time, than intended. At the afternoon assessment, participants reported on their drinking intentions or plans for later that day using a single item "How many drinks do you intend to have later today" using a slider bar from 0 to "15 or more." We subtracted that item from the next day reported of drinking quantity, such that positive numbers represented how many more drinks participants consumed than they had planned to drink. We tested four thresholds for this difference: 1+, 3+, 4+, or 5+ drinks more than intended. This AUD symptom was scored as positive if participants reported meeting the given threshold.

*Craving.* At every EMA, we administered two of four possible craving items adapted from prior EMA research (Kaysen et al., 2014) and directly reflecting DSM criteria (First et al., 2016). Participants were instructed to record how they had felt in the past hour. Responses were recorded on a slider bar ranging from 0 to 100. Items included "How much have you wanted to drink alcohol?" and "How much have you been thinking about drinking?" (with anchors at 0 ["Not at all"] and 100 ["Very much"]), and "I felt like I could really use a drink" and "If I'd had some alcohol, I would probably have drunk it" (with anchors at 0 ["Strongly disagree"] and 100 ["Strongly agree"]). We took the mean of all items across all EMAs prior to the onset of daily alcohol use. We tested nine thresholds for this symptom: average cravings rated at >0, the 50%, 66%, 75%, and 90% threshold of the mean, as well as the 66% threshold for each specific item. This AUD symptom was scored as positive if participants reported meeting the given threshold.

## Analytic Plan

First, we used multilevel structural equation modeling in Mplus (Muthén & Muthén, 2017) to test the dimensionality of daily AUD symptoms and to estimate the multilevel reliability (omega) of daily AUD symptoms (Geldhof et al., 2014). We used the maximum likelihood with robust standard errors (ML) estimator with a logistic link function, estimated with Monte Carlo integration and random starting values to facilitate model convergence. We fixed the factor variances at the between- and within-person levels to 1 to allow free estimation of all item loadings for identification, and factor loadings were set to be equal (within each item) across the within- and between-person levels.

Next, we used IRT to estimate item parameters (difficulty and discrimination) of daily AUD symptoms. We estimated both Rasch and two-parameter (e.g., 2PL) IRT models. Generally, 2PL models are preferred because they avoid the unrealistic assumption that all slopes are equal to 1. However, it is common practice in both research and practice to use mean or sum score scales, which are equivalent to a latent factor model with slopes equal to 1 (McNeish & Wolf, 2020). In other words, it is common to assume a Rasch model. Thus, reporting both Rasch and 2PL results can provide information about what information mean or sum scoring can provide about daily AUD symptoms, as well as how much information might be lost by ignoring non-equivalence in slopes.

We implemented a Bayesian multilevel modeling approach to IRT (Bürkner, 2020) with a logit link function, implemented in the package *brms* (Bürkner, 2017) in R (R Core Team, 2017). This multilevel approach to IRT estimates a cross-classified multilevel model, where parameters are allowed to vary as a function of both items and people. We allowed item difficulties to vary across items and people, but discriminations were allowed to vary across only items, as models including variation in discriminations across people had substantial convergence issues. We fixed the estimate of variation in difficulties across participants to 1 to facilitate model identification in the 2PL.[1] In this implementation of IRT, item difficulties are modeled as $\theta_p + \xi_i$, where $\theta_p$ reflects a person's ability, and $\xi_i$ is the "easiness" of a given item. We transformed these estimates of easiness to the more traditional difficulty by multiplying by −1. In this case, then, $-\xi_i$ is equal to the level of $\theta_p$ at which an average participant has a 50% probability of endorsing the symptom. Item discriminations were estimated as log(discrimination) to prevent the estimation of negative slopes (Bürkner, 2020). We used weakly informative priors for all parameters. We then plotted item characteristic and item information curves, modeling the syntax from Kurz (2021).

We estimated these IRT models 30 times, varying the specific items or thresholds that could be used to define the experience of each AUD symptom, and report how item difficulties and discriminations varied depending on which specific item or threshold was used to represent each symptom, holding other symptom constant at their initially specified threshold.

Based on those analyses, we then estimated an "optimal" model, selecting the best-performing thresholds for the *time spent, larger/longer, tolerance*, and *craving* symptoms. For the final model, with all symptoms except failure to fulfill role obligations, we had at least two items or thresholds that could be used to represent a given AUD symptom. Each symptom was scored a 1 if any of the items used to represent them were endorsed or the threshold for that item was met. For these models, the *time spent* symptom was considered to be met if a participant reported spending at least 8 hours drinking or reported

a hangover or vomiting. The *larger/longer* symptom was considered to be met if a participant reported drinking at least one drink more than intended. The *tolerance* symptom was considered to be met if a participant reported drinking 20 drinks or more *or* reported a ratio of quantity of drinks/intoxication greater than the 75th percentile. The *craving* symptom was considered to be met if a participant reported daily average cravings greater than the 75th percentile for the sample. We report the Rasch and 2PL properties for this final model below.

*Missing Data.* We focused solely on AUD symptoms reported on days when participants reported using alcohol, because all symptoms except craving were only measured on those days. We lagged craving so that we used daily reports of craving on the same days that alcohol was used (and thus other AUD symptoms were reported). Because of missing observations for intentions and craving due to study design (e.g., we obtained morning reports of drinking for Wednesday through Sunday evenings, but intentions were only obtained from Thursday through Sunday for those potential drinking days), as well as the presence of craving only in Study 2, we obtained data for all AUD symptoms on 2,705 of 3,963 (68%) of all possible drinking days. However, because the analyses focused on the experience of symptoms at the item level (e.g., daily symptoms nested within participants *and* items), we were able to leverage data from all drinking days, whether or not a symptom was measured (or even present).

## Results

### Descriptive Statistics

We collected data from 527 participants on 3,963 drinking days from 630 total possible participants over 19,840 days of observation. On drinking days, participants reported a mean of 4.03 drinks ($SD = 3.48$), and an intoxication score of 37.38 (out of 100, $SD = 27.06$). In Sample 2, we also had information on hazardous/risky drinking patterns (AUDIT; (Saunders et al., 1993)) and self-reported AUD symptoms (Boness et al., 2019). The mean AUDIT score for participants who reported any drinking during the EMA period was 7.91 (where >8 is considered a threshold for harmful/hazardous use). Forty percent of the sample met criteria for past-year AUD, with 21.6% meeting criteria for mild AUD (2–3 symptoms) and 18.4% meeting criteria for moderate (4–5 symptoms) or severe (6+ symptoms) AUD.

At the day level, the rates of endorsing each AUD symptom varied across drinking days and depended on the symptom at hand. Hazardous use or failure to fulfill role obligations was reported on 5% of all drinking days, while people reported drinking 3 or more drinks than intended in 32% of drinking occasions. Endorsement rates and drinking behaviors did not vary across the full and analytic sample of drinking days. Table 1 presents the descriptive statistics for the sample. Note that the various cutoffs for AUD symptoms derived from continuous items are taken from the final model results.

**Table 1.** Descriptive Statistics.

| Alcohol use and AUD symptom variables | *N* | *M*/% | *SD* |
|---|---|---|---|
| Number of drinks | 3,963 | 4.03 | 3.48 |
| Intoxication | 3,934 | 37.12 | 26.84 |
| Hazardous use | 3,963 | 5% | 0.22 |
| Social/Interpersonal consequences | 3,963 | 10% | 0.30 |
| Failure to fulfill role obligations | 3,963 | 5% | 0.22 |
| Time spent (8+hours drinking) | 3,963 | 18% | 0.38 |
| Tolerance (75% cutoff or 20+ drinks) | 3,963 | 24% | 0.43 |
| Larger/longer (3+ more than intended) | 2,705 | 32% | 0.47 |
| Craving (75% cutoff) | 3,681 | 30% | 0.46 |

### Dimensionality

First, we tested the plausibility of a unidimensional structure of our daily AUD symptoms using exploratory multilevel factor analysis in MPlus 8 with the WLSMV estimator (Muthén & Muthén, 2017). Models capturing more than two factors failed to converge. Fit indices for a two-factor solution ($\chi^2 = 26$, $df = 16$, $p = .05$) in the exploratory factor analysis were better than those for a one-factor solution ($\chi^2 = 78$, $df = 28$, $p < .001$). However, the one-factor solution produced equivalent results to a two-factor solution (with one factor reflecting time spent, tolerance, larger/longer, and craving, and a second factor reflecting hazardous use, interpersonal consequences, and interference with role obligation) in terms of factor loadings. Moreover, the two factors in the two-factor solution were very highly correlated ($r = .90$), and as a result, we did not observe large differences in factor loadings across solutions. Thus, we consider the two solutions to be equivalent and present results from the one-factor solution only. Output from the EFAs and the results of 2PL IRTs for one- and two-factor solutions are available in Supplementary Materials on the OSF (https://osf.io/6m74u/).

### IRT

We then used multilevel IRT to estimate the relative difficulty and discrimination of each AUD symptom. Across 60 models (30 each under the Rasch and 2PL models), we explored how item-level parameter estimates changed depending on which specific item or item thresholds were used to represent a given symptom. Table 2 presents them for the 2PL model; Supplemental Table 1 presents the same results from Rasch models. Figure 1 illustrates how difficulties varied across symptoms and items for the 2PL models.

*Item Difficulties.* Item difficulties, or the estimated level of daily AUD symptoms where a participant had a 50% probability of endorsing a given symptom, varied substantially

**Table 2.** 2PL Difficulties and Discriminations.

| Symptom | Item/threshold | Difficulty | LCL | UCL | Discrimination | LCL | UCL |
|---|---|---|---|---|---|---|---|
| Hazardous | **Aggregate** | **2.67** | **3.04** | **2.34** | **1.41** | **1.19** | **1.66** |
| | Blackout | 2.56 | 2.92 | 2.24 | 1.49 | 1.25 | 1.76 |
| | Hurt | 2.83 | 3.23 | 2.48 | 1.95 | 1.59 | 2.39 |
| Social/ interpersonal consequences | **Aggregate** | **2.23** | **2.51** | **1.98** | **1.28** | **1.11** | **1.48** |
| | Aggressive | 3.50 | 4.25 | 2.92 | 1.35 | 1.04 | 1.72 |
| | Embarrassed | 3.41 | 4.09 | 2.87 | 1.02 | 0.82 | 1.26 |
| | Fight | 3.57 | 4.32 | 2.99 | 1.00 | 0.80 | 1.24 |
| | Rude | 3.58 | 4.33 | 3.00 | 1.22 | 0.96 | 1.53 |
| Role obligations | **Aggregate** | **2.95** | **3.43** | **2.54** | **1.23** | **1.01** | **1.48** |
| Time spent | **Aggregate** | 1.82 | 2.06 | 1.60 | 1.03 | 0.90 | 1.18 |
| | 4 hours drinking | 1.61 | 1.80 | 1.43 | 1.30 | 1.15 | 1.47 |
| | **8 hours drinking** | **2.90** | **3.28** | **2.55** | **1.61** | **1.35** | **1.93** |
| | 12 hours drinking | 3.66 | 4.76 | 2.91 | 1.78 | 1.20 | 2.58 |
| | 16 hours drinking | 4.82 | 7.21 | 3.43 | 1.56 | 0.92 | 2.52 |
| | **Hangover** | **3.97** | **4.92** | **3.26** | **0.72** | **0.57** | **0.90** |
| | **Vomit** | **2.71** | **3.20** | **2.31** | **0.82** | **0.67** | **0.98** |
| Tolerance | **Aggregate/a priori threshold** | 1.28 | 1.53 | 1.06 | 0.57 | 0.48 | 0.67 |
| | **20+ drinks** | **4.56** | **6.40** | **3.47** | **1.42** | **0.92** | **2.05** |
| | Ratio 50% | −0.05 | 0.08 | −0.18 | 0.76 | 0.66 | 0.87 |
| | **Ratio 66%** | **1.26** | **1.51** | **1.04** | **0.59** | **0.49** | **0.69** |
| | Ratio 75% | 2.39 | 2.91 | 1.98 | 0.52 | 0.42 | 0.63 |
| | Ratio 90% | 7.24 | 10.68 | 5.11 | 0.33 | 0.22 | 0.47 |
| Larger/longer | **Aggregate/a priori threshold** | 1.18 | 1.42 | 0.97 | 0.73 | 0.62 | 0.87 |
| | 1+ More drinks | −5.76 | −3.57 | −9.62 | 0.16 | 0.10 | 0.26 |
| | **3+ More drinks** | **1.17** | **1.41** | **0.97** | **0.74** | **0.62** | **0.88** |
| | 4+ More drinks | 1.89 | 2.23 | 1.60 | 0.76 | 0.64 | 0.90 |
| | 5+ More drinks | 2.46 | 2.90 | 2.10 | 0.81 | 0.67 | 0.97 |
| Craving | **Aggregate/a priori threshold** | 0.99 | 1.15 | 0.83 | 1.10 | 0.96 | 1.26 |
| | Any craving | −2.32 | −1.95 | −2.78 | 0.97 | 0.78 | 1.19 |
| | Craving 50% | −1.03 | −0.86 | −1.22 | 1.24 | 1.05 | 1.47 |
| | **Craving 66%** | **−0.04** | **0.06** | **−0.15** | **1.75** | **1.51** | **2.03** |
| | Craving 75% | 0.43 | 0.55 | 0.32 | 1.96 | 1.71 | 2.26 |
| | Craving 90% | 1.32 | 1.48 | 1.16 | 3.26 | 2.75 | 3.88 |
| | Craving could use 66% | 0.08 | 0.19 | −0.04 | 1.64 | 1.41 | 1.91 |
| | Craving thinking 66% | 0.04 | 0.16 | −0.07 | 1.30 | 1.12 | 1.52 |
| | Craving want 66% | 0.01 | 0.13 | −0.12 | 1.11 | 0.95 | 1.30 |
| | Craving would 66% | −0.01 | 0.11 | −0.14 | 1.03 | 0.86 | 1.22 |

*Note*. Ratio or craving % refers to the threshold of the continuous ratio or craving scale (or item) that was used as a cutoff to identify the presence/absence of that symptom. Bolded items and coefficients were selected for the final model.

across items within symptoms. Across both Rasch and 2PL models, the specific item or threshold chosen to determine the presence versus absence of a symptom sometimes had a substantial impact on a given item's difficulty. The extent to which the items or thresholds yielded differential item difficulties depended on the symptom.

For some symptoms, the varied items reflecting the symptom had minimal impact on the difficulty (e.g., hazardous use, social/interpersonal consequences). These findings indicate, at least in part, that these items had relatively similar prevalence in the sample. This finding held across both the Rasch and 2PL models.

As expected, the difficulty of items using a threshold also increased as the threshold increased. Reporting drinking 1 or more drinks than intended, for example, had a difficulty of −5.76, but its difficulty increased as we raised the threshold, with an unexpected binge (or near binge; 4+ drinks more than intended) having a difficulty of 1.89 (Figure 1).

*Item Discriminations.* Item discriminations, or how well a given daily AUD symptom discriminated days that were higher or lower in AUD symptoms, also varied depending on which item was used to represent a given AUD symptom. The specific pattern varied across items. Generally, for some

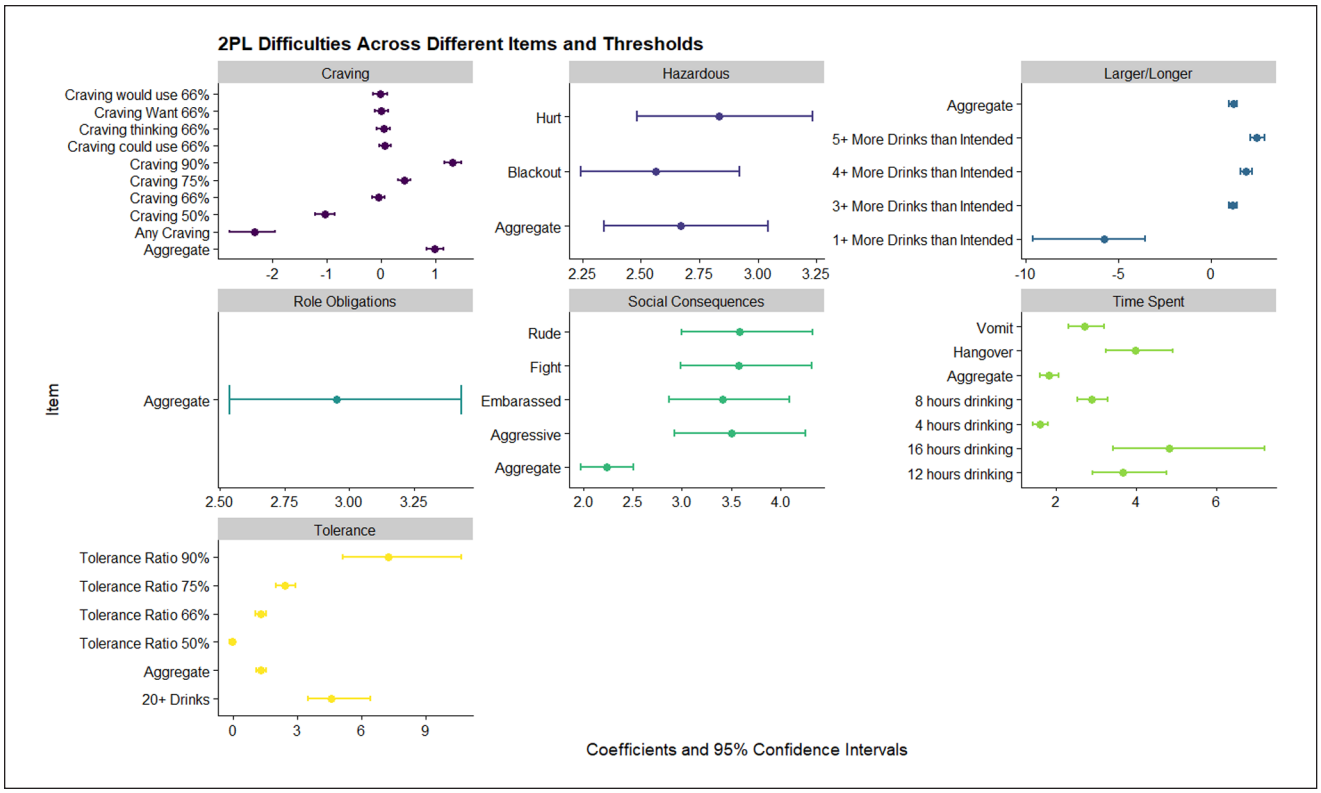


**Figure 1.** Variation in Item Difficulty Within Each AUD Symptom Across Items/Thresholds.
*Note.* X scales are freed across AUD symptoms to highlight within-symptom variation in difficulties.

items, discrimination decreased among the most difficult items, while for others, discrimination increased. The most difficult items (such as the 90% cutoff for craving, getting hurt, or drinking for 16 or more hours) also had the most uncertainty in their estimates, despite having relatively higher discrimination. The discrimination for aggregate symptoms scored with an "if/any" algorithm was somewhere in between the discriminations of the individual items.

## Final Models

Finally, we re-estimated Rasch and 2PL models with thresholds that provided the best balance of higher discrimination and minimizing uncertainty. We used the following post hoc thresholds based on the Rasch and 2PL sensitivity analysis reported in Table 2: Larger/longer: 3+ drinks; Craving: 66%; Tolerance: 66%; Time spent: 8+ hours or reporting a hangover or vomiting. This model exhibited good evidence of convergence of the Markov Chain Monte Carlo chains (all Rhat = 1.00), and effective sample sizes were all high (>5,000).

*Rasch Modeling.* As expected, difficulty estimates closely mirrored the relative frequency of the AUD criteria in the sample. Craving was the least difficult item, such that a person at −0.04 logits below the mean of daily AUD symptoms had a 50% probability of endorsing the cravings symptom, while they only reached a 50% probability of reporting hazardous use on days when their daily AUD symptoms were 3.35 logits above the mean. Put another way, when theta was at zero (e.g., at the mean of daily AUD symptom severity), participants had a $1/(1+\exp(-1.01)) = 51\%$ probability of meeting the criteria for craving (defined as being at or above the 66%ile of the sample, or a rating of approximately 18 on a 0–100 scale), and a 3.4% chance of endorsing hazardous use. Table 3 presents the results of the final Rasch model.

*2PL IRT Modeling.* The 2PL IRT model improved model fit relative to the Rasch models (ΔWAIC = −173). Table 4 presents these results. In the 2PL models, where item slopes were allowed to vary, results were relatively similar with two exceptions. Relative to the symptoms reflecting the impairment from alcohol use (i.e., hazardous use, social consequences, role obligations), daily AUD symptoms that more directly reflected aspects of alcohol consumption (i.e., time spent, tolerance, larger/longer, craving) had lower difficulties and poorer discrimination. Tolerance and larger/longer had especially weak discrimination estimates relative to all other symptoms and thus provided poorer discrimination between days characterized as higher versus lower levels of latent AUD symptoms.

**Table 3.** Item Difficulties From Rasch Modeling of Daily AUD Symptoms.

| Item | Difficulty | LCL | UCL | Probability of endorsement at the mean of daily AUD symptoms (e.g., $\theta = 0$) (%) |
|---|---|---|---|---|
| Hazardous use | 3.35 | 3.18 | 3.52 | 3.4 |
| Social/interpersonal consequences | 2.61 | 2.47 | 2.76 | 6.8 |
| Failure to fulfill role obligations | 3.40 | 3.22 | 3.57 | 3.2 |
| Time spent | 3.84 | 3.64 | 4.04 | 2.1 |
| Tolerance | 0.82 | 0.71 | 0.93 | 30.5 |
| Larger/longer | 0.92 | 0.80 | 1.04 | 28.5 |
| Craving | −0.04 | −0.15 | 0.08 | 51.0 |

**Table 4.** Item Difficulties From the Final 2PL Modeling of Daily AUD Symptoms.

| Item | Difficulty | LCL | UCL | Discrimination | LCL | UCL |
|---|---|---|---|---|---|---|
| Hazardous use | 2.67 | 2.34 | 3.04 | 1.41 | 1.19 | 1.66 |
| Social/interpersonal consequences | 2.23 | 1.98 | 2.51 | 1.28 | 1.11 | 1.48 |
| Failure to fulfill role obligations | 2.95 | 2.54 | 3.43 | 1.23 | 1.01 | 1.48 |
| Time spent | 1.82 | 1.60 | 2.06 | 1.03 | 0.90 | 1.18 |
| Tolerance | 1.28 | 1.06 | 1.53 | 0.57 | 0.48 | 0.67 |
| Larger/longer | 1.18 | 0.97 | 1.42 | 0.74 | 0.62 | 0.87 |
| Craving | 0.99 | 0.84 | 1.15 | 1.10 | 0.96 | 1.26 |

Relative to the Rasch model, allowing item discriminations to vary in the 2PL model produced lower estimates of relative difficulty of most items, meaning that at the mean of theta, participants were somewhat more likely to endorse that item at the daily level.

Figure 2 illustrates the item-level discrimination and information curves from the final model, while Table 4 presents the final parameter estimates. In this final model, craving, larger/longer, and tolerance indicated relatively less severe AUD problems and provided somewhat more information about daily AUD symptoms. Item information curves suggested that alcohol-related impairment provided information about relatively more severe daily AUD symptoms, whereas consumption-related items (i.e., craving, tolerance, larger/longer, time spent) provided information about relatively less severe daily AUD symptoms.

## Discussion

Symptoms of AUD involve experiences at the daily or momentary level, such as getting into a fight while drinking, violating self-imposed drinking limits, or experiencing cravings. Gold standard measures of AUD symptoms (e.g., SCID, AUDIT, Alcohol Symptom Checklist; Allen et al., 1997; Babor & Higgins-Biddle, 2001; First, 2015; Hallgren et al., 2022) require people to retrospect about their experiences with those symptoms and factor in temporal features, such as symptom frequency, persistence, and recurrence. Validating measures that can capture AUD symptoms as they are experienced, or soon thereafter, would allow researchers to directly model the temporal dynamics among AUD symptoms, unlocking new lines of research about the development of AUD. To our knowledge, only one study has validated daily-level measures of alcohol-related consequences that capture the experience of some AUD symptoms (C. M. Lee et al., 2017), while other studies have used ad hoc checklists of "acute dependence" symptoms (Dvorak et al., 2014; Simons et al., 2010). The present study built upon this prior work by estimating the psychometric properties of seven AUD symptoms using IRT and testing the impact of item choices on those psychometric properties.

First, our results indicate that AUD symptoms are indeed commonly experienced in conjunction with drinking episodes, with prevalence estimates from our final model ranging from 5% to 30% of drinking days, depending on how symptoms or thresholds were defined. This variation in prevalence is reflected in the range of item difficulties estimated by the IRT models. Some AUD symptoms (i.e., hazardous use, social/interpersonal consequences, failure to fulfill role obligations) occurred infrequently in this sample of regularly drinking young adults, many of whom met criteria for at least mild AUD. Importantly, the majority of the low base rate AUD symptoms were derived from existing measures of alcohol-related consequences (C. M. Lee et al., 2017), which have been commonly used in EMA research. In contrast, other symptoms (e.g., craving, consuming larger amounts than intended) occurred more frequently but have also been measured only rarely. It is notable that these symptoms more directly reflect alcohol consumption. This finding implies that studies that *only* focus on alcohol-related

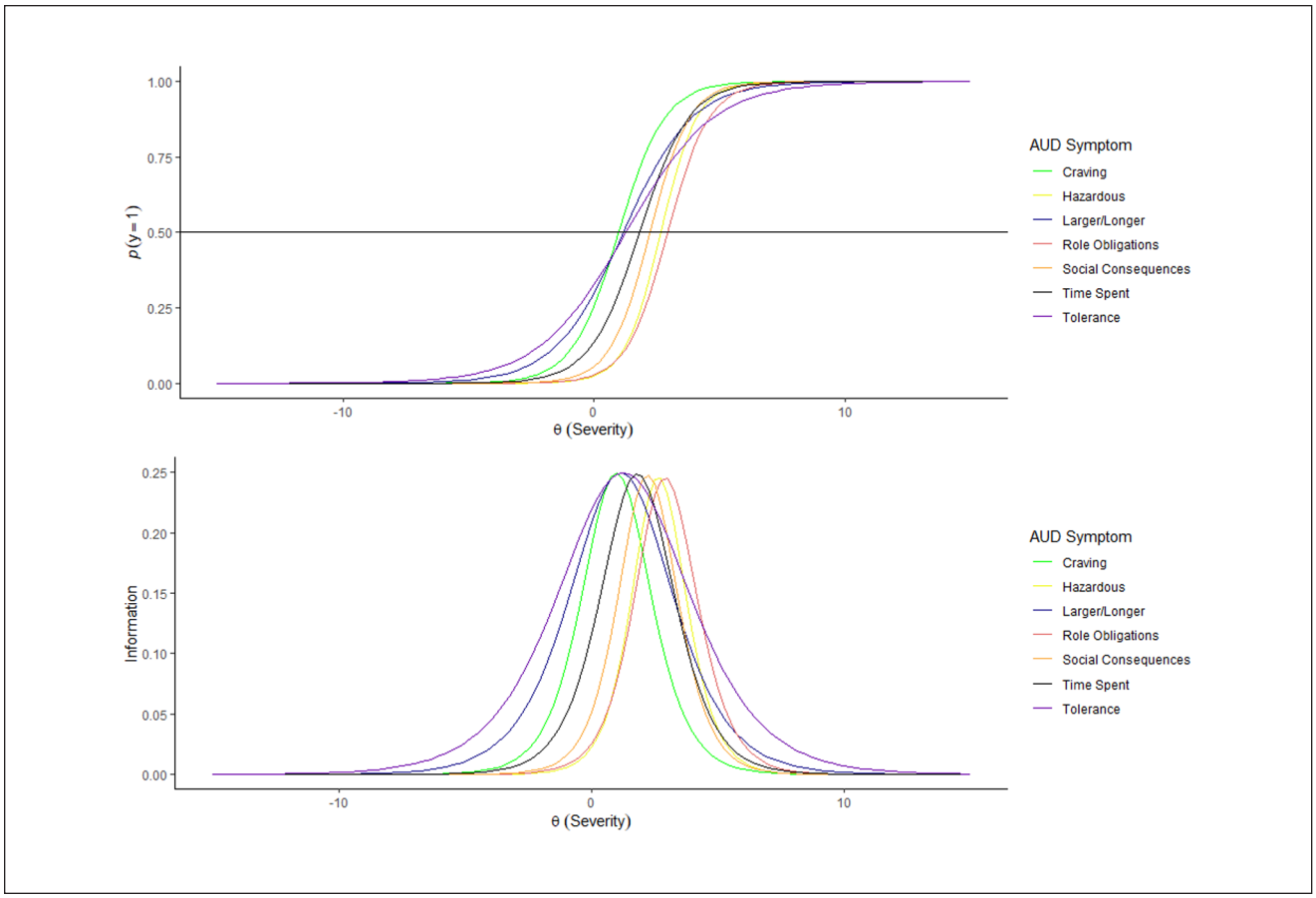


**Figure 2.** Item-Level Discrimination and Item Information Curves of the Final Model of Daily AUD Symptoms.

consequences as a way to understand the development of AUD are missing important information about common experiences of AUD symptoms in daily life.

Moreover, the vast majority of items had at least modest associations with the latent factor reflecting the difficulty of their latent daily AUD symptoms and provided a great deal of information about a relatively specific region of severity of the underlying spectrum of daily AUD symptoms. Importantly, although a 2PL parameterization fit the data better than a Rasch model (which fixed all discriminations to 1), the Rasch and 2PL produced relatively similar estimates of item difficulties across models and provided (relatively) similar amounts of information about the latent AUD construct. This finding suggests that, at least in the present study, sum scores of daily AUD symptoms would be expected to perform relatively similarly to scores that accounted for differential symptom weighting based on the 2PL model. Taken together, relatively simpler models of daily AUD symptoms sacrifice some assumptions at a relatively minor cost to inferences about the data. At the same time, it will be important to replicate and extend this work with a broader pool of items and coverage of all AUD symptoms.

Two items, tolerance and larger/longer, as we a priori operationalized, had relatively weaker associations with the latent factor. When larger/longer was defined as drinking at least one more drink than intended, it had very poor discrimination (close to 0). However, discrimination estimates improved when we set the threshold to 2+, 3+, or 4+ drinks than intended, which suggests that more substantial limit violations are better indicators of AUD. Structured clinical interviews measuring AUD queries drinking *much more* than intended (First, 2015), and Boness and colleagues reported that drinking "more than intended" was relatively common (38%). In other work, this threshold had relatively weak discriminations in the 2PL model, whereas getting drunk when not intending to was less common and also had a relatively better discrimination (Boness et al., 2019). It may also be that people do not necessarily think about their drinking intentions in terms of specific and strict limits and instead set fuzzier intentions for their drinking such as whether or not they intend to get drunk (Chung & Martin, 2005; Slade et al., 2013). For example, other studies of drinking intentions use multiple items to capture more heuristic measures of intentions, such as drinking one to two or four to five drinks (Lewis et al., 2020), which may better reflect how people

formulate their drinking intentions than specific numbers of drinks. Relatedly, it may also be that the larger/longer criterion is better understood as regret following drinking, so items that capture people's post-episode evaluation of their alcohol use may better reflect the nature of this AUD symptom. Future studies could use cognitive interviewing (Boness & Sher, 2020) and other methods to understand how people think about their drinking intentions to better understand how to operationalize and measure limit violations.

Tolerance is even more complicated to assess. First, regardless of how we operationalized tolerance, it was a relatively weak indicator of daily AUD symptoms compared with all other symptoms except larger/longer. In fact, when we defined tolerance literally, to capture people being less intoxicated for a given rate of drinking, it was *inversely* related to the experience of other AUD symptoms. Thus, in the present study, we operationalized "tolerance" as the *inverse* of how it is typically defined: higher scores reflected *higher* intoxication for a given rate of drinking, or people's *sensitivity* to the effects of alcohol. Although tolerance may be *broadly* defined as insensitivity to alcohol's effects at the person level (e.g., people with more tolerance feel less intoxication for a given number of drinks), our data make it clear that this definition of tolerance does not coincide with people's experience of AUD symptoms at the daily level. When people were less intoxicated at a given number of drinks, they experienced *fewer* other symptoms of AUD, and it was only when people reported higher intoxication given their drinking that they experienced other daily AUD symptoms. This aligns with prior work that both identified challenges measuring tolerance in retrospective surveys and classical definitions of tolerance, which is described as the ability to consume larger amounts of alcohol without significant impairment over time (Caetano & Babor, 2006; Lejoyeux et al., 2001).

As with larger/longer, this finding for tolerance highlights the importance of examining different operationalizations of AUD symptoms to understand how they develop over time and are experienced in people's daily lives. However, it is likely that tolerance, defined as insensitivity, *does* drive the development of AUD and the experience of other AUD symptoms as they unfold over time. As people progressively consume more and more alcohol to obtain the same effects, they will also be likely to consume alcohol in excess and, in those episodes, experience more AUD symptoms. Indeed, prior studies had operationalized tolerance in terms of people's perceptions of their drinking episode (e.g., "had to drink more to feel same effects" Dvorak et al., 2014), which likely captures episodes more characterized by overconsumption than insensitivity to alcohol's effects. That is, other assessments of tolerance may query the construct in a way that conflates it with excess consumption and not the construct of interest (Caetano & Babor, 2006). Future studies should test this hypothesis to understand how tolerance, as characterized by insensitivity to the effects of alcohol and episodes of overconsumption, might drive the development of drinking episodes and the experience of other AUD symptoms over time.

## Clinical Implications

Current retrospective measures of AUD (e.g., the SCID; First, 2015) are well suited for gaining a snapshot of a person's current level of AUD severity, but they may be less well suited for determining what the most appropriate clinical intervention may be. For instance, current measures conflate the actual behavior being measured (e.g., tolerance, hazardous use, consequences) with the timing of these behaviors (e.g., recurrence, persistence). By developing a measure suitable for daily assessment, researchers and clinicians can examine how these symptoms unfold over time and understand what relevant situational factors precede drinking behaviors and the resulting impairment. For instance, one individual may consistently drink for larger/longer when they are with a particular group of people, while another individual may experience more social/interpersonal consequences if they drink when they are frustrated. However, these patterns may not become clear without closely monitoring the behaviors themselves as opposed to a person's overall impressions of their drinking.

Moreover, daily measures may also be used to aid in prevention efforts. Specifically, although certain AUD symptoms are more common among individuals who drink regularly, individuals who begin endorsing these symptoms may benefit from intervention to avoid further escalating the problems they experience as a result of drinking. In fact, this study has demonstrated that item parameters vary depending on what item or cutoff is used and we can further optimize these measures for different samples and settings, meaning that we may develop cutoffs specific to college or young adult populations, which may differ from the cutoffs used in adult populations.

## Limitations and Future Directions

To date, our study is the largest to examine people's experience of (selected) AUD symptoms in their daily lives and the first to use IRT to describe item-level characteristics in those experiences. At the same time, there are important limitations to consider. Moreover, although we captured nearly 4,000 drinking occasions, the present study was designed around time intervals (e.g., 10 days or 2 months), and some daily AUD symptoms were experienced relatively rarely. It may be that capturing more drinking episodes could improve the precision of the measurement of daily AUD symptoms. Future studies could consider, for example, a design built around intensively measuring a certain number of drinking episodes per participant, rather than continuously measuring participants while hoping to capture sufficient drinking occasions. In addition, we

studied AUD symptoms in regularly drinking young adults. Although they drank heavily and often and thus were at high risk for AUD, fewer than half met criteria for AUD and (because of their age) could not have long histories of AUD symptoms. It would be critical to extend this research to older populations and those with more severe and longer histories of AUD to understand how daily AUD symptoms might be experienced differently.

Future research should build on and broaden the psychometric work presented here. First, and most notably, we did not measure all AUD symptoms that people might experience in their daily lives. Future studies should seek to include all symptoms, including desire to cut down or quit, giving up important activities to drink alcohol, continued use despite knowledge of a physical or psychological problem, and withdrawal. More comprehensive assessment of AUD symptoms, paired with repeated assessments over time, would allow researchers to model both how people experience AUD symptoms in their daily lives and how they change, recur, or persist over time. It would also facilitate theory-based tests of temporal dynamics among AUD symptoms, the development and maintenance of AUD over time, and the involvement of micro-level processes in AUD's time course and experience.

Moreover, future research should seek to more broadly measure each AUD symptoms in more ways they might manifest in daily life and to consider how different manifestations might differently represent the experience of that AUD symptom. Future studies should also measure a wider range of experiences that could represent daily AUD symptoms in the same way previous studies have accomplished this for retrospective measures of AUD (Boness et al., 2019). Importantly, these expanded assessments should also measure AUD symptoms outside of drinking episodes, as several AUD symptoms (cravings, desires to cut down or quit, and withdrawal) can be (or must be, in the case of withdrawal) experienced on non-drinking days. Given how narrowly we measured AUD symptoms in the current study, it is highly likely that a broader measurement would produce different inferences about the psychometric properties of daily AUD symptoms. Refined measures that have psychometric information for a broad range of AUD symptom items could also build on this work to select item sets that could be used to provide high resolution of a specific range of AUD symptom severity. Although this approach might come at the cost of overall information about daily AUD symptoms, it could be used to improve the specificity of information obtained in a given sample. For example, a more detailed assessment of the daily experiences of withdrawal symptoms would be expected to provide a great deal of information in a sample of people seeking treatment for AUD but not very much information in a sample of adolescents experimenting with alcohol use. Future studies should seek to validate a measure that captures the experience of all 11 symptoms of AUD and to understand how people's experience of symptoms varies across a range of severities of AUD. Such a measure could be used to measure the daily life manifestation of AUD symptoms, how they persist, recur, or change over time, and how they are associated with contextual factors (such as depressive or post-traumatic stress disorder [PTSD] symptoms or social contexts).

## Conclusion

Young adults who drink experience at least some AUD symptoms in their daily lives. Our findings highlight the importance of measuring the daily experiences of AUD and how critical it is to consider a broad range of operationalizing how specific AUD symptoms are experienced. Developing and validating daily measures of AUD symptoms could provide fresh insights into the development and maintenance of AUD over time because they would allow researchers to directly model the emergence of individual symptoms of AUD in the contexts in which people experience AUD. In turn, researchers can consider: How does the social or physical context of drinking occasions influence the experience of different AUD symptoms? How are AUD symptoms linked to the daily experience of other psychological disorders, such as PTSD or depression? Are there protective behavioral strategies that people can use to limit the experience of specific AUD symptoms? How does experiencing AUD symptoms in daily life influence future drinking behaviors, and how, when, and for whom do those daily symptoms recur or persist over time?

### Declaration of Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Preparation of this article was supported by grants DA047247 and K02AA0288327 to K.M.K. C.L.B. is funded by K08AA030301. A.L.W. is funded by R00AA028306.

### ORCID iDs

Kevin M. King https://orcid.org/0000-0001-8358-9946
Dahyeon Kang https://orcid.org/0000-0002-7299-8557
Jonas Dora https://orcid.org/0000-0003-0651-8479
Cassandra L. Boness https://orcid.org/0000-0001-8734-193X

### Supplemental Material

Supplemental material for this article is available online.

### Note

1. We attempted to freely estimate variation in item difficulties across participants in the 2PL models, as we did with the Rasch models. Initial modeling attempts indicated that all 2PL models had issues with the MCMC algorithm, with substantially high numbers of divergent transitions (which means that), despite having Rhat values of 1.00 and sufficient

effective sample sizes. To solve this issue, we introduced more informative priors (e.g., priors that used the means of the aggregate models) and increased adaptive delta and the maximum tree depth to improve MCMC convergence. This approach was able to eliminate divergent transitions for the majority of items, but a few models still produced one or two divergent transitions, and we were concerned that the parameter estimates from those models were untrustworthy. Thus, we fixed the participant-level variance estimate to 1 to facilitate model convergence; this choice effectively standardized the estimate of variation across participants while still allowing variation. The inferences from these models were nearly identical to those for models (with and without divergences) that freely estimated this parameter, suggesting that at *least in that case*, this modeling choice did not substantially impact our conclusions.

## References


Alaybek, B., Dalal, R. S., Fyffe, S., Aitken, J. A., Zhou, Y., Qu, X., Roman, A., & Baines, J. I. (2022). All's well that ends (and peaks) well? A meta-analysis of the peak-end rule and duration neglect. *Organizational Behavior and Human Decision Processes*, *170*, 104149. https://doi.org/10.1016/j.obhdp.2022.104149

Allen, J., Litten, R., Fertig, J., & Babor, T. (1997). A review of research on the Alcohol Use Disorders Identification Test (AUDIT). *Alcoholism: Clinical and Experimental Research*, *21*, 613–619. https://doi.org/10.1111/j.1530-0277.1997.tb03811.x

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). American Psychiatric Publishing.

Babor, T., & Higgins-Biddle, J. (2001). *The Alcohol Use Disorders Identification Test: Guidelines for use in primary care*. World Health Organization, Department of Mental Health and Substance Dependence.

Berridge, K. C., & Robinson, T. E. (2016). Liking, wanting, and the incentive-sensitization theory of addiction. *American Psychologist*, *71*(8), 670–679. https://doi.org/10.1037/amp0000059

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*(1), 35–49. https://doi.org/10.1037/adb0000443

Boness, C. L., & Sher, K. J. (2020). The case for cognitive interviewing in survey item validation: A useful approach for improving the measurement and assessment of substance use disorders. *Journal of Studies on Alcohol and Drugs*, *81*(4), 401–404. https://doi.org/10.15288/jsad.2020.81.401

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, *80*(1). https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C. (2020). *Bayesian Item Response Modeling in R with brms and Stan* (arXiv:1905.09501). arXiv. http://arxiv.org/abs/1905.09501

Caetano, R., & Babor, T. F. (2006). Diagnosis of alcohol dependence in epidemiological surveys: An epidemic of youthful alcohol dependence or a case of measurement error? *Addiction*, *101*(s1), 111–114. https://doi.org/10.1111/j.1360-0443.2006.01599.x

Chung, T., & Martin, C. S. (2005). What were they thinking? *Drug and Alcohol Dependence*, *80*(2), 191–200. https://doi.org/10.1016/j.drugalcdep.2005.03.023

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, *74*(4), 327–337. https://doi.org/10.1097/PSY.0b013e3182546f18

Cullum, J., O'Grady, M., Armeli, S., & Tennen, H. (2012). The role of context-specific norms and group size in alcohol consumption and compliance drinking during natural drinking events. *Basic and Applied Social Psychology*, *34*(4), 304–312. https://doi.org/10.1080/01973533.2012.693341

Dvorak, R. D., Pearson, M. R., & Day, A. M. (2014). Ecological momentary assessment of acute alcohol use disorder symptoms: Associations with mood, motives, and use on planned drinking days. *Experimental and Clinical Psychopharmacology*, *22*(4), 285–297. https://doi.org/10/gnp3dd

Fairlie, A. M., Ramirez, J. J., Patrick, M. E., & Lee, C. M. (2016). When do college students have less favorable views of drinking? Evaluations of alcohol experiences and positive and negative consequences. *Psychology of Addictive Behaviors*, *30*(5), 555–565. https://doi.org/10.1037/adb0000190

Finnigan, K. M., & Vazire, S. (2018). The incremental validity of average state self-reports over global self-reports of personality. *Journal of Personality and Social Psychology*, *115*(2), 321–337. https://doi.org/10.1037/pspp0000136

First, M. B. (2015). Structured Clinical Interview for the DSM (SCID). In R. L. Cautin, & S. O. Lilienfeld (Eds.), *The encyclopedia of clinical psychology* (pp. 1–6). Wiley. https://doi.org/10.1002/9781118625392.wbecp351

First, M. B., Williams, J. B. W., Spitzer, R. L., Karg, R. S., Williams, J. B. W., & Karg, R. S. (2016). *User's guide for the SCID-5-CV structured clinical interview for DSM-5 disorders: Clinician version*. American Psychiatric Association.

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, *19*(1), 72–91. https://doi.org/10.1037/a0032138

Grant, B. F., Goldstein, R. B., Smith, S. M., Jung, J., Zhang, H., Chou, S. P., Pickering, R. P., Ruan, W. J., Huang, B., Saha, T. D., Aivadyan, C., Greenstein, E., & Hasin, D. S. (2015). The alcohol use disorder and associated disabilities interview schedule-5 (AUDADIS-5): Reliability of substance use and psychiatric disorder modules in a general population sample. *Drug and Alcohol Dependence*, *148*, 27–33. https://doi.org/10/gjrhxs

Hallgren, K. A., Matson, T. E., Oliver, M., Caldeiro, R. M., Kivlahan, D., & Bradley, K. A. (2022). Practical assessment of DSM-5 alcohol use disorder criteria in routine care: High test-retest reliability of an Alcohol Symptom Checklist. *Alcoholism, Clinical and Experimental Research*, *46*(3), 458–467. https://doi.org/10.1111/acer.14778

Hingson, R., Zha, W., Simons-Morton, B., & White, A. (2016). Alcohol-induced blackouts as predictors of other drinking related harms among emerging young adults. *Alcoholism: Clinical and Experimental Research*, *40*(4), 776–784. https://doi.org/10.1111/acer.13010

Kang, D., Watts, A. L., Boness, C.L., Schultz, M.E., Dora, J., Lee, C.M., King, K.M. (in press). Daily life assessment of seven alcohol use disorder symptoms. *Clinical Psychological Science*.

Kaysen, D., Atkins, D. C., Simpson, T. L., Stappenbeck, C. A., Blayney, J. A., Lee, C. M., & Larimer, M. E. (2014). Proximal relationships between PTSD symptoms and drinking among female college students: Results from a daily monitoring study. *Psychology of Addictive Behaviors*, *28*(1), 62–73. https://doi.org/10.1037/a0033588

King, K. M., Feil, M. C., Gomez Juarez, N., Moss, D., Halvorson, M. A., Dora, J., Upton, N. F., Bryson, M. A., Seldin, K., Shoda, Y., Lee, C. M., & Smith, G. T. (2024). Negative urgency as a state-level process. *Journal of Personality*, *93*(3), 529–552. https://doi.org/10.1111/jopy.12961

Koob, G. F., & Le Moal, M. (2001). Drug addiction, dysregulation of reward, and allostasis. *Neuropsychopharmacology*, *24*(2), 97–129. https://doi.org/10.1016/S0893-133X(00)00195-0

Koob, G. F., & Le Moal, M. (2008). Addiction and the brain antireward system. *Annual Review of Psychology*, *59*, 29–53. https://doi.org/10.1146/annurev.psych.59.103006.093548

Kurz, S. (2021). *Make ICC plots for your brms IRT models*. https://solomonkurz.netlify.app/blog/2021-06-29-make-icc-plots-for-your-brms-irt-models/

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*(8), 1769–1784. https://doi.org/10.1017/S0033291716000404

Lee, C. M., Cronce, J. M., Baldwin, S. A., Fairlie, A. M., Atkins, D. C., Patrick, M. E., Zimmerman, L., Larimer, M. E., & Leigh, B. C. (2017). Psychometric analysis and validity of the daily alcohol-related consequences and evaluations measure for young adults. *Psychological Assessment*, *29*(3), 253–263. https://doi.org/10.1037/pas0000320

Lee, M. R., Boness, C. L., McDowell, Y. E., Vergés, A., Steinley, D. L., & Sher, K. J. (2018). Desistance and severity of alcohol use disorder: A lifespan-developmental investigation. *Clinical Psychological Science*, *6*(1), 90–105. https://doi.org/10.1177/2167702617736852

Lejoyeux, M., Claudon, M., McLoughlin, M., & Adès, J. (2001). Comparison of alcohol-dependent patients with and without physiological dependence. *European Addiction Research*, *7*(4), 198–201. https://doi.org/10.1159/000050741

Lewis, M. A., Litt, D. M., King, K. M., Fairlie, A. M., Waldron, K. A., Garcia, T. A., LoParco, C., & Lee, C. M. (2020). Examining the ecological validity of the prototype willingness model for adolescent and young adult alcohol use. *Psychology of Addictive Behaviors*, *34*(2), 293–302. https://doi.org/10.1037/adb0000533

McNeish, D., & Wolf, M. G. (2020). Thinking twice about sum scores. *Behavior Research Methods*, *52*(6), 2287–2305. https://doi.org/10.3758/s13428-020-01398-0

Moore, T. M., Elkins, S. R., McNulty, J. K., Kivisto, A. J., & Handsel, V. A. (2011). Alcohol use and intimate partner violence perpetration among college students: Assessing the temporal association using electronic diary technology. *Psychology of Violence*, *1*(4), 315–328. https://doi.org/10.1037/a0025077

Muthén, L. K., & Muthén, B. O. (2017). *Mplus user's guide* (8th ed.). Muthén & Muthén.

Prat, G., Adan, A., & Sánchez-Turet, M. (2009). Alcohol hangover: A critical review of explanatory factors. *Human Psychopharmacology: Clinical and Experimental*, *24*(4), 259–267. https://doi.org/10.1002/hup.1023

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, *128*(6), 934–960. https://doi.org/10.1037/0033-2909.128.6.934

Saunders, J. B., Aasland, O. G., Babor, T. F., de la Fuente, J. R., & Grant, M. (1993). Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO collaborative project on early detection of persons with harmful alcohol consumption-II. *Addiction*, *88*(6), 791–804.

Scharbert, J., Utesch, K., Reiter, T., Ter Horst, J., Van Zalk, M., Back, M. D., & Rau, R. (2024). If you were happy and you know it, clap your hands! Testing the peak-end rule for retrospective judgments of well-being in everyday life. *European Journal of Personality*, *39*(1), 55–69. https://doi.org/10.1177/08902070241235969

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology*, *4*, 1–32.

Simons, J. S., Dvorak, R. D., Batien, B. D., & Wray, T. B. (2010). Event-level associations between affect, alcohol intoxication, and acute dependence symptoms: Effects of urgency, self-control, and drinking experience. *Addictive Behaviors*, *35*(12), 1045–1053. https://doi.org/10.1016/j.addbeh.2010.07.001

Simons, J. S., Wills, T. A., & Neal, D. J. (2014). The many faces of affect: A multilevel model of drinking frequency/quantity and alcohol dependence symptoms among young adults. *Journal of Abnormal Psychology*, *123*(3), 676–694. https://doi.org/10.1037/a0036926

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*(6), 1001–1007. https://doi.org/10/f4z2tx

Stone, A. A., Shiffman, S., Atienza, A. A., & Nebeling, A. (2007). Historical roots and rationale of ecological momentary assessment (EMA). In A. A. Stone, S. Shiffman, A. A. Atienza, & L. Nebeling (Eds.), *The science of real-time data capture: Self-reports in health research* (pp. 3–10). Oxford University Press.

Substance Abuse and Mental Health Services Administration. (2021). *Key substance use and mental health indicators in the United States: Results from the 2020 National Survey on Drug Use and Health*. Center for Behavioral Health Statistics and Quality, Substance Abuse and Mental Health Services Administration.

Takane, Y., & De Leeuw, J. (1987). On the relationship between iTem Response Theory and Factor Analysis of Discretized Variables. *Psychometrika*, *52*(3), 393–408. https://doi.org/10.1007/bf02294363

Trela, C. J., Piasecki, T. M., Bartholow, B. D., Heath, A. C., & Sher, K. J. (2016). The natural expression of individual differences in self-reported level of response to alcohol during ecologically assessed drinking episodes. *Psychopharmacology*, *233*(11), 2185–2195. https://doi.org/10.1007/s00213-016-4270-5

Vergés, A., Jackson, K. M., Bucholz, K. K., Grant, J. D., Trull, T. J., Wood, P. K., & Sher, K. J. (2012). Deconstructing the age-prevalence curve of alcohol dependence: Why "maturing out" is only a small piece of the puzzle. *Journal of Abnormal Psychology*, *121*(2), 511–523. https://doi.org/10/bx7m3d