Objective

Implicit and explicit drinking self-identity appear to be useful in predicting alcohol-related outcomes. However, there are several different implicit and explicit measures which can be used to assess drinking self-identity. Some of these implicit measures can also capture relational information (e.g., I *am* a drinker, I *should be* a drinker), which might provide unique advantages. Despite the importance of having good measures of drinking self-identity, to date there has been little direct comparison of these measures.

Method

This study (N = 358) systematically compared two commonly-used measures of drinking self-identity (one implicit and one explicit: the IAT and the ASCS) with three relational measures of implicit self-identity (the aIAT, the RRT, and the pCIT) on a range of criteria relevant to experimental and clinical alcohol researchers.

Results

Overall, we found mixed performances on the implicit measures. Interestingly, the aIAT which probed should-based drinking identity performed better than the standard IAT. However, the explicit measure exhibited superior performance to all other measures across all criteria.

Conclusions

Our results suggest that researchers who wish to assess drinking-related self-identity and to predict alcohol-related outcomes cross-sectionally should set their focus primarily on the use (and further development) of the ASCS, rather than any of the implicit measures. Future research focusing on the ASCS should seek to investigate the generalisability of our findings to patient populations, and incorporate relational information within that procedure in order to further improve upon its already-strong utility.

Keywords: Implicit measures; implicit beliefs; drinking self-identity; Alcohol

Public Health Significance Statement. Psychologists increasingly recognise that the extent to which people self-identify as drinkers can provide unique and meaningful prediction of alcohol-related outcomes. Psychologists typically measure this using both self-report measures and measures of automatic behaviour, AKA implicit measures. In this study, we systematically compare a number of different measures of drinking self-identity based on a number of criteria of relevance to experimental and clinical psychologists.

On the role of (implicit) drinking self-identity in alcohol use and problematic drinking: A comparison of five measures

Drinking self-identity (Ramirez, Olin, & Lindgren, 2017) is an emergent, effective predictor of problematic consumption, with some work suggesting that self-identity may mediate changes in drinking behaviour over time (Blevins et al., 2018). Simultaneously, researchers have utilized *implicit measures* (i.e., measures which capture responses made quickly, without awareness, or unintentionally; see De Houwer, 2006) due to findings suggesting they may account for variance in alcohol consumption beyond that accounted for by traditional self-report measures (AKA explicit measures; Gray et al., 2011; Davies et al., 2017; but see Schimmack, 2019, for criticisms of implicit measures). Recent studies have investigated the utility of implicit measures of drinking self-identity to predict alcohol use and misuse. Findings have been promising: multiple studies have now demonstrated that implicit drinking self-identity predicts a variety of alcohol-related outcomes (Blevins et al., 2018; Lindgren et al., 2016a; Caudwell & Hagger, 2014). In one key study, Lindgren and colleagues (2013) compared five different variants of the Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998). They found that an IAT measuring drinking self-identity correlated best with self-report measures of drinking self-identification, exhibited superior psychometric properties, and was the best predictor of alcohol-related outcomes, including predicting those outcomes beyond an explicit measure of drinking self-identity.

Despite the predictive utility of drinking self-identity, important questions remain. In particular, published studies to date have measured implicit drinking self-identity only with the IAT[1]. However, the IAT has a critical limitation: it cannot assess *how* drinking is related to the

[1] Or its briefer alternative, the Brief IAT (BIAT; Werntz et al., 2016), for which the same limitations apply.

self; it can only assess whether or not a relationship exists (De Houwer, 2002). For example, if an individual has a strong effect in (i.e., a high score on) a drinking self-identity IAT, this effect may be driven by the belief that "alcohol is part of my identity". However, other beliefs might cause this effect, such as "alcohol *should be* part of my identity", "I *want* alcohol to be part of my identity", or "alcohol *has been* a part of my identity". Ultimately, the IAT (and all other similar *associative* implicit measures) are, by their very nature, unable to identify which implicit beliefs are responsible for producing observed effects. At the same time, different implicit beliefs about the same topic can have very different impacts on behavior (Remue et al., 2013; De Houwer, Van Dessel, & Moran, 2020), and *discrepancies* between different beliefs can be critical in predicting behavior (Higgins et al., 1985; Remue et al., 2014). The standard IAT cannot investigate these issues.

Relational implicit measures

Consequently, some researchers have moved away from *associative* implicit measures, and instead developed *relational* implicit measures. These relational implicit measures specify exactly *how* stimuli are related within their procedure, and by extension specify which implicit beliefs are assessed. These measures have utility in assessing and predicting phenomena such as depression, self-esteem, body dissatisfaction, and smoking (De Houwer et al., 2015; Heider et al., 2015, 2018; Remue et al., 2014; Tibboel et al., 2017). Notably, the *discrepancy* between different beliefs can provide unique insights into behaviour. For example, body dissatisfaction is associated with low am-based implicit identity with thinness, and high want-based implicit identity with thinness (Heider et al., 2018), and depression can be predicted by a combination of low am-based implicit self-esteem and high should-based implicit self-esteem (Remue et al., 2014). Critically, relational implicit measures provide us with a novel set of tools to explore implicit drinking self-identity findings implicit drinking self-identity by specifying the exact ways in which drinking is related to participants' self-identity.

To illustrate the distinction between a non-relational implicit measure and a relational implicit measure, let us consider the standard IAT and its relational variant, the autobiographical IAT (aIAT; Sartori et al., 2008). In the standard drinking-identity IAT, a word stimulus is presented on-screen in each trial. Participants might be asked to press one computer key (e.g., 'E') when they see a stimulus relating to the category "Me" or the category "Drinker", and to press another key (e.g., 'I') when they see a stimulus relating to the category "Not Me" or the category "Abstainer". This configuration would switch between blocks, such that on another block "Not Me" and "Drinker" would share a response key, and "Me" and "Abstainer" would share a response key. In the aIAT, sentences (rather than individual words) are presented on each trial. Participants might be required to press one key when they see a *true* sentence (e.g., "I am a human being") or when they see a sentence relating to drinking (e.g., "I am a drinker"), and another key when they see a *false* sentence (e.g., "I am a fish") or a sentence relating to abstaining (e.g., "I am an abstainer"). On another block of trials, participants might then be required to respond in an opposite manner where true sentences and sentences relating to abstaining share a response key, and false sentences and sentences relating to drinking share a response key. By comparing how quickly participants categorise sentences when they share a response key with true sentences vs. with false sentences, the aIAT can provide an insight to the extent which participants endorse drinking- and abstaining-related sentences as true or false. Importantly, between different aIATs, the relational information the drinker/abstainer sentences can be varied (e.g., "I am a drinker" in an am-aIAT, "I should be an abstainer" in a should-aIAT). By harnessing compatibility of *sentences* with the concepts true and false, measures like the aIAT can achieve precision over relational information (since sentences can contain such information) which the standard IAT cannot.

Research on relational implicit measures is emergent. Although different relational implicit measures demonstrated utility in different contexts, there is no systematic comparison of measure performances in the same domain. Thus, applied researchers have little guidance about which measure might be best suited to their goals. Since relational information could play an influential role in self-identity, and implicit drinking self-identity is important in the context of alcohol use, drinking self-identity is an ideal context to assess and compare the utility of different relational implicit measures.

Comparing measures of drinking self-identity

Ultimately, drinking self-identity generally demonstrates utility in predicting drinking-related outcomes, but there are multiple different measurement procedures which can be used to measure this (e.g., implicit vs. explicit; relational implicit measures vs. non-relational implicit measures). Thus, our goal was to systematically compare these different measurement procedures on criteria that have been of regular interest to basic and applied alcohol researchers. We also borrowed heavily from the criteria employed by Lindgren and colleagues (2013). Specifically, researchers generally wish for their measures to have high reliability (Flake et al., 2017; Revelle & Condon, 2019) and to correlate with explicit measures of the same construct (Muschalik et al., 2019; Nosek, 2007; but see Schimmack, 2019). Implicit measures and explicit measures may also be compared directly, and researchers frequently decide which measure to use on the basis of the measure's brevity (Millner et al., 2018) and its effectiveness in predicting alcohol use (Reich et al., 2010) and misuse (Lindgren et al., 2014). Importantly, researchers are also often interested in whether implicit measures can predict a criterion *over and above* explicit measures (De Houwer, 2006; Kurdi et al., 2019; Müller & Rothermund, 2019; Nosek et al., 2011).

We compared five measures of drinking self-identification: one associative implicit measure (the IAT), one explicit measure (the Alcohol Self-Concept Scale), and three relational implicit measures (the Autobiographical IAT, the Relational Responding Task, and the Propositional Concealed Information Test). For the relational implicit measures, we investigated two different kinds of beliefs involving drinking self-identity: descriptive beliefs (e.g., "I *am* a drinker") and prescriptive beliefs (e.g., "I *should be* a drinker"). We chose these specific beliefs because discrepancies between descriptive and prescriptive beliefs are well-documented in multiple contexts (Bear & Knobe, 2017), and would likely maximize the potential utility of the relational implicit measures. Indeed, should-based beliefs have provided additive predictive utility to am-based beliefs in a number of different implicit and explicit contexts (Heider et al., 2018; Rudman & Glick, 2001; Table 1 provides an overview of the measures used in this study and their rationale for inclusion).

[Table 1 here]

We aimed to compare these five different measures of drinking self-identity in terms of (i) the reliabilities of each of the measures, (iii) the implicit-explicit correlations of the implicit measures, (iv) how well they predicted drinking patterns (measured by the Daily Drinking Questionnaire), (v) how well they predicted problematic alcohol consumption patterns (measured by the Alcohol Use Disorders Identification Test), and (vi) how well they separated participants into high and low drinking groups. While we had no firm hypotheses, we expected that the standard IAT and the ASCS would be generally effective based on the above criteria, based on their performances in previous research.

Method

This experiment was not preregistered. All experimental materials, data, and processing and analysis scripts can be found on the Open Science Framework (https://osf.io/7df69/?view_only=aa6910c0f88142bfb004ceb9731707e5). Ethical approval for this project was provided by the Ethical Committee of the Faculty of Psychology and Educational Sciences at Ghent University (approval numbers 2015/13, 2016/63, and 2016/80).

Participants

Data were collected online via Prolific Academic (https://prolific.ac). Research suggests that data collected from such online data collection sites tend to be as valid as data collected from laboratory research, and also tend to more demographically diverse than lab-based studies (Palan & Schitter, 2018; Buhrmester et al., 2016; Casler et al., 2013; Levay et al., 2016). Individuals sign up to Prolific Academic on a voluntary basis, typically through word-of-mouth. Of the 100,000 individuals in the Prolific participant pool, approximately 30,000 of these people are of US nationality. Within this, approximately 10,000 participants fit the category of very low quantity drinkers and approximately 1,000 participants fit the category of very high quantity drinkers. All screening criteria were employed using Prolific Academic's internal screening tool. Individuals were eligible to participate if they were between the ages of 18 and 65, spoke fluent English, were of US nationality, and had not completed any previous similar studies from our research group. Participants were also only eligible for participation in this study if they consumed on average very low (less than 4 US standard drinks per week) or very high (more than 14 US standard drinks per week) quantities of alcohol. Participants were recruited on a first-come, first-served basis; that is, the study was available to all eligible participants on the Prolific site, who then manually signed up to participate in the study. 172 low drinkers (~2 % of eligible low drinkers) and 183 high drinkers (~19% of eligible high drinkers) completed the study. We do not have access to information about how many eligible participants opted not to sign up for our study, which in principle could result in non-inclusion bias.

We originally intended to use the aforementioned consumption criteria to distinguish between low and high groups for problematic alcohol consumption. Using such an “extreme groups” design can afford greater power to statistical tests (Preacher, 2015). Utilising such a design thus allowed us to maximise our power given constraints on the number of participants we planned to collect. However, we realised that the initial screening criteria failed to account for the fact that risk criteria are frequently differentiated by gender. Because participants’ responses to Prolific’s screening criteria were unavailable, we could not reformulate our drinking groups using this information. To address these issues, we deviated from our original strategy, and instead assigned participants into low or high drinking groups on the basis of their responses to the typical-quantity subscale of the Daily Drinking Questionnaire (DDQ). We categorized participants as low or high groups based on the criteria of the National Institute on Alcohol Abuse and Alcoholism (NIAAA): no more than 3 drinks on a single day or 7 drinks per week for women, and no more than 4 drinks on a single day or 14 drinks per week for men (NIAAA, 2017). However, we only employ this categorisation in our descriptive analyses and our final inference analysis. For the remainder of our analyses, we analyse DDQ/AUDIT scores as count data. Given that the DDQ and AUDIT responses followed a negative binomial distribution, we accordingly apply negative binomial link functions to regression analyses where applicable.

Given that we control for gender in our analyses, participants who did not identify their gender as male or female (3 participants: 2 non-binary, 1 no gender given) were excluded from analyses. All participants were paid at a rate of $6.28 per hour. After exclusions, our final sample consisted of 355 participants (157 women and 198 men), with a mean age of 36.58 years (SD = 10.81). Of these participants, 83 completed the standard IAT, 92 completed the RRT, 87 completed the pCIT, and 93 completed the aIAT.

Materials

All materials were programmed in Inquisit 5 and administered using the Inquisit Web Player via Prolific. The specific stimuli used in all procedures can be found in the Supplementary Materials.

IAT. Our drinking self-identity IAT (herein referred to as the standard IAT) was highly similar to the drinking-identity IAT used by Lindgren et al. (2013), including the fact that we used the same stimuli. The IAT followed the typically used seven-block structure: Blocks 3, 4, 6, and 7 served as the "critical" blocks, whereas Blocks 1, 2, and 5 served as practice blocks. On every trial, participants were required to respond to a stimulus using either the 'E' (left) or the 'I' (right) computer key based on categories on the top-left and -right sides of the screen. For example, if "me" and "not me" were respectively at the top-left and top-right of the screen, and the word "myself" appeared on-screen, participants would be required to press the left key. On the first block (20 trials), this was exactly the task of the participant: press one key for "me" stimuli, and another key for "not me" stimuli. On the second block (20 trials), participants were required to categorise stimuli as either drinker-related (e.g., "drunk") or abstainer-related (e.g., "sober"). On the third (20 trials) and fourth block (40 trials), participants were required to perform both categorisations simultaneously (i.e., respond 'E' for "me" stimuli and "drinker" stimuli, respond 'I' for "not me" stimuli and "abstainer" stimuli). The fifth block (20 trials) was identical to the first block, with one exception: the response keys required for the "me" and "not me" categorisations were in opposite positions. Finally, the sixth block (20 trials) and seventh block (40 trials) required participants to again perform the categorisations simultaneously, but now the overlapping responses for the different categorisations changed (i.e., press ‘E’ for “not me” and “drinker”, and press ‘I’ for “me” and “abstainer”).

If the participant responded incorrectly, then a red ‘X’ was presented in the center of the screen and remained there until the participant corrected their response. The primary outcome of interest in the IAT relates to whether there are differences in the response latencies to categorizing stimuli between response configurations on the critical blocks (i.e., between the configuration used for Blocks 3 and 4 and the configuration used for Blocks 6 and 7). The first blocks were considered as “drinking-consistent” blocks, because they involved response configurations which were consistent with identifying as a drinker. By contrast, the latter blocks were considered as “drinking-inconsistent” blocks.

aIAT. We administered two aIATs: an “am” aIAT, and a “should” aIAT. Both aIATs were almost identical to the IAT, with some critical distinctions. The third and fourth blocks were combined into a single block of sixty trials, and so too were the sixth and seventh blocks. In the first block, rather than responding to the categories “me” and “not me”, participants instead responded “true” or “false” to stimuli which related to personal information about the participant. Additionally, these stimuli consisted of full sentences, rather than individual words. For example, if the participant was presented with the sentence “I am at the beach”, then they would respond “false”. In addition, instead of responding based on the categories of “drinker” and “abstainer”, participants responded based on sentences of the category “I am [should be] a drinker” and “I am [should be] an abstainer” in the am [should] aIAT. Stimuli presented again consisted of whole sentences rather than individual words. For example, in the am-aIAT, if the participant was presented with the sentence “Drinking is part of who I am”, they should respond with the key which corresponds to the category “I am a drinker”. The order of presentation of the aIATs was counterbalanced across participants.

RRT. Similar to the aIAT measures, we administered am- and should-RRTs. Both RRTs were highly similar in layout to the aIAT, with some distinctions. Firstly, rather than categorise personally-relevant statements as true or false, participants were required to simply categorise individual words as true or false (for example, if the participant was presented with the word “correct”, then they should respond “true”). Secondly, participants also categorised statements about drinking identity as “true” or “false” (rather than as “I am/should be a drinker” or “I am/should be an abstainer”). To do this, they were instructed at the beginning of the relevant block to respond *as if* specific statements were true/false. For example, in the am-RRT, when responding in drinking-consistent ways, participants were instructed to respond *as if* “I am a drinker/I am not a drinker” is true, and *as if* “I am not a drinker/I am an abstainer” is false. The required *as if* response patterns varied between the two critical blocks, similarly to the aIATs. Like the aIATs, the am- and should-RRTs each consisted of 5 blocks. However, the third and fifth blocks now consisted of 80 trials, rather than 60 trials as in the aIATs.

pCIT.

The pCIT (Cummins, Verschuere, & De Houwer, 2020) is a recently developed variant on the standard Concealed Information Test (CIT; Agosta & Sartori, 2013), which is typically used in lie detection paradigms. In each trial, participants are required to respond ‘true’ or ‘false’ to sentence stimuli. There are three types of sentences in the task: irrelevants, probes, and targets, which are presented in a ratio of 4:1:1 respectively. Irrelevant stimuli consist of personal information sentences which are false (e.g., for a male participant, the sentence “I am a woman”), and target stimuli consist of sentences which are true (e.g., for the male participant, “I am a man”). Probe sentences consist of sentences which pertain to the belief which the experimenter intends to assess. In our case, probe stimuli consisted of sentences relating to drinking identity (e.g., “I am a drinker”, “I should be an abstainer”).

Initially, participants are presented with the target stimuli and told that they should respond ‘true’ to only these targets and ‘false’ to all other sentences, *even if* they believe some of those other sentences are true. For target stimuli and irrelevant stimuli, participants’ beliefs and the required responses are congruent: target stimuli are true for the participant and they must respond true, and irrelevant stimuli are false for the participant and they must respond false. The congruence of probe stimuli, however, can vary between participants. For non-drinkers, responding false to “I am a drinker” will be relatively easy. For drinkers, this will be more difficult: they must respond false even though the sentence is true for them. For the probe stimulus, “I am an abstainer”, it will be more difficult to respond with “false” for non-drinkers, while for drinkers it should be relatively easy. A pCIT score for ‘am’ drinking identity is calculated via the difference in response times to “I am a drinker” probes and “I am an abstainer” probes. The probe with the shorter response time may be inferred to be more consistent with the implicit self-identity of the participant. We used four probe stimuli: two assessing descriptive “am” beliefs (i.e., “I am a drinker” and “I am an abstainer”), and two assessing prescriptive “should” beliefs (i.e., “I should be a drinker” and “I should be an abstainer”).

The pCIT began with memorization phase where participants learned the target sentences. After this memorization phase, participants completed three 24-trial practice phases of the pCIT. In the first practice phase participants were given no time limit for their responses and were told simply to respond as accurately as they could. The second practice phase then introduced a time limit of 1500ms for each trial. If participants failed to respond within this time limit, then they were moved on to the next trial. The third practice phase consisted of this time limit plus a warning of 'TOO SLOW' which was presented on the screen if participants had not responded within 1200ms. If participants failed to respond correctly to at least 50% of the targets presented on any of the practice blocks, then they were required to repeat that block again until they achieved 50% accuracy. Additionally, participants in the second and third practice blocks were required to have a mean response latency on the overall block of less than 800ms. After completing all of the practice blocks, participants then completed 570 main trials of the pCIT.

ASCS. We measured explicit drinking identity via the Alcohol Self-Concept Scale (ASCS; Stein & Corte, 2007; Lindgren et al., 2013; adapted from Shadel & Mermelstein, 1996). The ASCS consists of seven 5-point Likert-scale items designed to assess the extent to which drinking is an important part of an individual's self-identity.

DDQ. The Daily Drinking Questionnaire (DDQ; Collins et al., 1985) assessed alcohol consumption. Participants reported the quantity of alcohol consumed on each day of a typical week in the past three months. We also administered a variant which asked participants to report the quantity of alcohol consumed during a particularly heavy drinking week in the past three months.

AUDIT. The Alcohol Use Disorders Identification Test (AUDIT; Saunders et al., 1993) measured problematic alcohol use. The AUDIT consists of ten items assessing the frequency and intensity of an individual's alcohol consumption, as well as symptoms of alcohol dependence and negative consequences resulting from drinking.

Procedure

Upon commencement of the experiment, participants firstly provided basic demographic information (age and gender). Then, participants completed one of the four implicit measures (i.e., the IAT, the aIAT, the RRT, or the pCIT). Participants assigned to complete either the aIAT or the RRT completed both the 'am' and 'should' variants. Following completion of the measure, all participants then completed the explicit measure of alcohol identification (i.e., the ASCS), and the measures of alcohol consumption (DDQ) and problematic alcohol-related behaviors (AUDIT). Finally, participants were asked whether they believed their data should be included in our analyses. They were told they would be paid regardless of their answer to this question. If the participant indicated that they believed we should exclude their data, then participants were prompted to provide a reason for why they believed they should be excluded. We excluded participants who indicated that their data should be excluded and provided a reason which might suggest that the integrity of their data might be suspect (for example, mentioning that they attended to their child midway through completing the implicit measure, n= 8).

Results

Analytic strategy

We firstly compared the four implicit measures based on (i) their split-half reliability and (ii) their correlations with the ASCS. Next, we compared all measures (including the ASCS) on the basis of (i) their average duration for completion, (ii) their prediction of alcohol use based on DDQ scores, (iii) their prediction of problematic drinking as measured by the AUDIT, and (iv) their success at classifying individuals as low- or high-drinkers. For analyses (ii) through (iv), we also investigated whether differences between 'am' and 'should' beliefs provided predictive utility. For analyses (ii) and (iii), we also evaluated whether the implicit measures of drinking identity predict self-reported drinking behavior over and above self-reported drinking identity.

Data preparation

To maximise the comparability of our different implicit measures, we used a consistent scoring method across all of the implicit measures (see Payne et al., 2008; Cummins & De Houwer, 2019). IAT and RRT effects are commonly quantified using the *D* score (Greenwald et al., 2003; De Houwer et al., 2015), and CIT effects are typically based on raw response times (Agosta & Sartori, 2013). However, in terms of psychometric soundness, neither of these scoring methods is the most optimal approach (see Blanton et al., 2015; Whelan, 2008). The use of a Probabilistic Index is more optimal (PI; Thas et al., 2012). A PI in the context of reaction-time-based implicit measures represents the probability that a randomly selected response time from one block/trial type will be larger than a randomly selected response time from another block/trial type. For example, in the IAT, a PI of .75 would suggest that, in 75% of cases, a randomly-selected drinking-inconsistent trial would have a longer RT than a randomly-selected drinking-consistent trial. This, in turn, would indicate that participants responded generally more quickly to drinking-consistent trials, and thus indicate that they identify more strongly with drinking than abstaining. The PI has been demonstrated to be psychometrically superior to both the *D* score and the use of raw response time data due to its normal distribution of effect sizes across participants, and its relative insensitivity to outliers (De Schryver et al., 2018; De Schryver & De Neve, 2019). For the relational implicit measures, we calculated two PI scores: one for “am” relations, and one for “should be” relations. RRT and aIAT PIs were calculated by comparing response times on the drinking-consistent block with response times on the drinking-inconsistent block. PIs for the pCIT were calculated by comparing response times from drinking-consistent probe trials with response times from drinking-inconsistent probe trials. For the IAT, only one PI was calculated (by comparing response times on the drinking-consistent block with response times on the drinking-inconsistent block).

For the ASCS, we computed a global sum score based on participants' responses to all item. For the DDQ, we calculated the sum total of units of alcohol consumed per week. For the AUDIT, we created a total summary score.

Sample Characteristics

Table 2 provides a breakdown of the characteristics of our sample between the two difference drinking conditions (i.e., low- and high-drinkers). Additionally, Figure 1 illustrates the distribution of effects for the DDQ-typical, DDQ-heavy, and AUDIT scores (all of which followed a negative binomial distribution).

[Table 2 here]

[Figure 1 here]

Hypothesis Testing: Comparing Implicit Measures

We compared the implicit measures on the basis of their Spearman-Brown-corrected split-half reliabilities and their correlation with the ASCS. Cronbach's alpha for the ASCS was extremely high, $\alpha = .95$, 95% CIs [.95, .96]. For split-half reliabilities, we bootstrapped the estimates for each variant of each procedure in order to facilitate direct comparisons of reliabilities between the procedures. Thus, the procedures' reliabilities can be compared directly by determining whether the median estimate of one procedure falls outside of the upper-/lower-bound confidence intervals of another. In general, the IAT outperformed all other measures in terms of split-half reliability. The aIAT also performed well, with its lower-bound confidence intervals excluding .7 (the typical lowest acceptable value for reliability; McDonald, 1999, chapter 6). The RRT's median R*sb* estimate also exceeded this value, though its lower-bound confidence interval includes it for both the 'am' and 'should' RRTs. The pCIT demonstrated extremely poor reliability, with all other procedures far exceeding it (see Table 3).

[Table 3 here]

We next evaluated the correlations between the implicit measures and the ASCS (the explicit measure of drinking self-identification). Both relational variants of the aIAT and RRT correlated significantly with the ASCS. The IAT and the pCIT did not. Whereas the point-estimates of the IAT and am-pCIT's implicit-explicit correlations were contained within the confidence intervals of the aIAT and RRT's estimates (implying that they did not significantly differ), the should pCIT's estimate fell well below most of the other procedures (see Table 4).

[Table 4 here]

Hypothesis Testing: Comparing All Measures

Finally, we compared the implicit measures *and* the explicit measure of drinking self-identity in terms of their average duration for completion, their prediction of self-reported drinking behavior, their prediction of problematic drinking patterns, and their rates of success at identifying individuals as low or high drinkers. In terms of duration for completion, the ASCS drastically outperformed all other measures, taking just over a minute to complete on average. By contrast, the IAT took about two minutes to complete, the aIATs and RRTs took 3-4 minutes to complete, and the pCIT took almost 15 minutes to complete.

The brevity of a task is only advantageous when it is accompanied by utility. We therefore assessed the utility of each measure by comparing how well each procedure could predict self-reported drinking behavior as assessed by the DDQ. Internal consistencies were high for both the typical quantity α = .91, 95% CI [.90, .93] and heavy quantity α = .92, 95% CI [.93, .94] DDQ sub-scales. Generalized linear models with a negative binomial link were used to evaluate how well the self-identity measures predicted drinking behavior. Models controlled for participant gender.

The *relational* measures may contribute to predicting drinking one other way – i.e., the *discrepancy* between "am" and "should" beliefs might also predict DDQ scores. To test this, we calculated a PI discrepancy score for each procedure, which consisted of subtracting participants' 'am' PI scores from their 'should' PI scores. As such, positive discrepancy scores indicate greater endorsement of "am" than "should" beliefs, and a negative discrepancy score indicates the opposite. Only the aIAT (both "am" and "should" aIATs for heavy-drinking periods, but only the "should" aIAT for typical-drinking periods) and the ASCS significantly predicted DDQ scores. Estimates for the ASCS also exceeded the upper-bound confidence intervals for the aIAT, suggesting that the ASCS was also more effective at predicting DDQ scores (see Table 5).

[Table 5 here]

The utility of implicit measures can also be touted in terms of their ability to account for unique variance in a model over-and-above an explicit measure. Thus, we next examined whether any of the implicit measures were capable of predicting either of the DDQ scores after controlling for ASCS scores. To do so, we conducted a series of negative binomial generalized linear regressions using either DDQ score as the dependent variable of interest, ASCS sum scores as one independent variable, and the PI scores from each measure as the second independent variable (models also controlled for participant gender). The critical test of interest is whether there is a significant main effect of each measure after controlling for the ASCS. Whereas none of the measures predicted heavy-quantity DDQ scores above and beyond the ASCS[2], the should-RRT and the standard IAT both significantly predicted beyond the ASCS for the typical-quantity DDQ scores. However, the direction of these effects was counterintuitive,

[2] The ASCS remained a significant predictor in all models.

particularly for the IAT: for both the should-RRT and the standard IAT, greater identification with drinking predicted *less* quantities of alcohol consumed (see Table 6).

[Table 6 here]

We ran six additional negative binomial generalized linear regressions (one for each measure, using the two different DDQ outcomes) in order to investigate whether PI discrepancy scores could predict beyond the ASCS. The interaction between PI scores for 'am' and 'should' relations also failed to provide any added predictive utility beyond ASCS scores for predicting any of the DDQ scores[3] (all ps > .098).

We next compared the procedures in terms of their ability to predict *problematic* alcohol usage (as measured by the AUDIT). The internal consistency of the AUDIT in our sample was high, $\alpha = .90$, 95% CI [.88, .91]. We followed an identical approach to our analyses of the DDQ data. That is, we firstly conducted negative binomial generalized linear regressions for each of the measures, with AUDIT sum score as the dependent variable, and the PI score from one of the measures, or the PI discrepancy score of that procedure, as a dependent variable (controlling for gender in all models). As with the DDQ, the ASCS was the most effective measure at predicting AUDIT scores. The should aIAT was also a significant predictor as with the DDQ; however, the am aIAT was no longer a significant predictor (see Table 7).

[Table 7 here]

As previously, we next investigated whether any of the measures were capable of providing accounting for variance in predicting AUDIT scores above and beyond the ASCS. We again conducted a series of negative binomial generalized linear regressions, using AUDIT scores as the dependent variable, ASCS scores as one independent variable, and the PI score of

[3] In all models, the main effect of ASCS remained a significant predictor of both DDQ scores.

the corresponding procedure as the second independent variable (all controlling for gender). As Table 8 demonstrates, only the am and should RRTs predicted significantly beyond the ASCS[4]. As with the DDQ, this prediction was in the opposite direction to what might typically be expected: greater identification with drinking in the should RRT predicted less problematic drinking (i.e., lower scores in the AUDIT).

[Table 8 here]

We again investigated whether the discrepancy between 'am' and 'should' PI scores in each measure could predict scores in the AUDIT above and beyond ASCS scores. The PI discrepancy scores again failed to provide any additional utility beyond the ASCS in predicting problematic drinking[5] (all *p*s > .28).

We lastly investigated the extent to which all of these procedures could accurately classify participants as either low or high drinkers. To do so, we firstly conducted a series of logistic regressions[6], each using drinking status (low or high) as the dependent variable, and the score from each measure (PI score for implicit measures, sum score for the ASCS) as the independent variable in each separate model. We conducted three additional logistic regressions: this time using the PI discrepancy score for each of the three relational implicit measures as the independent variable. We then calculated the predicted probabilities for each participant based on the parameter estimates from each model, and then calculated bootstrapped area-under-curve (AUC) coefficients for each model, which serve as an index of the rate of correct classification of participants as high or low drinkers. AUC statistics vary in range from 0 to 1: an AUC of 0 would indicate all cases were classified incorrectly, and an AUC of 1 would indicate that all

[4] In all models, the main effect of ASCS remained a significant predictor of AUDIT scores.
[5] In all models, the main effect of ASCS remained a significant predictor of AUDIT scores.

cases were classified correctly. An AUC score of .5 would imply that half of the cases were classified correctly: given that the classifier is binary in nature, this would amount to classification at a chance level. As such, each measure can be considered to classify at a greater-than-chance level when its lower-bound 95% confidence interval excludes .5. Consistent with the trend of results found across all of our analyses, the ASCS vastly outperformed all other measures in classifying participants. Among the implicit measures, only the aIAT classified participants at an above-chance level (see Table 9).

[Table 9 here]

Discussion

Summary of findings

We compared different implicit and explicit measures of drinking self-identity. Our results paint a clear picture: for alcohol researchers seeking a single measure which is optimal in terms of duration and cross-sectional predictive utility of AUDIT or DDQ scores, and accuracy at classifying individuals into low and high drinking groups, the ASCS is far-and-above the best option. The measure consistently outperformed all of the implicit measures on every criterion by which they were compared. For many researchers, especially researchers who are familiar with implicit measures, this may come as no surprise: explicit measures very often outperform implicit measures in outcome prediction (Genschow et al., 2017; Greenwald et al., 2009). However, what may be more surprising to researchers familiar with implicit measures was the inability of most of the implicit measures to provide any added predictive utility over-and-above the explicit measure. Whether it be in terms of predicting alcohol use on typical- or heavy-drinking weeks or predicting hazardous drinking patterns, the implicit measures provided added utility in only three instances (which will be discussed in greater depth below). Whereas the standard IAT, aIAT, and RRT varied in terms of their performances in different analyses, the pCIT was ineffective by every metric.

Practical implications for previous findings

One of our most surprising findings was that the standard IAT failed to predict AUDIT scores over-and-above ASCS scores, which is in contrast to previous findings (Lindgren et al., 2013; Lindgren et al., 2016b; Werntz et al., 2016)[7]. This discrepancy, however, may be explained by the fact that previous studies which found these effects typically have larger sample sizes for this analysis than we did here (N = 300 in Lindgren et al., 2013; N = 506 in Lindgren et al., 2016a; N = 12,387 in Werntz et al., 2016; compared to our N = 83). Indeed, a brief[8] power calculation using our sample size and the effect size for the standard IAT's prediction of drinking related outcomes when controlling for ASCS scores from Lindgren et al. (Cohen's d = 0.29) suggests that we would have approximately 17% power to detect a true effect of the standard IAT's incremental prediction beyond the ASCS. By contrast, when conducting such a power analysis based on the ASCS's prediction of drinking outcomes using our sample size and the effect size for the ASCS found by Lindgren et al. (Cohen's d = 0.74), we would have approximately 85% power to detect a true effect.

We want to be clear that we are not suggesting that the IAT cannot provide incremental utility over and above the ASCS. Indeed, the literature on this to date (including our results here)

[7] Researchers familiar with the drinking-identity IAT may see our counterintuitive results and wonder: if we had used the IAT D score, would our effects conform more to extant findings? We considered this possibility and ran post-hoc analyses using the D score. IAT D scores produced an almost identical pattern of results. Only two results differed: the split-half reliability of the IAT is significantly lower using D (vs. the PI) and predicting typical-quantity DDQ scores above and beyond the ASCS becomes non-significant for the D score (p = 0.06). Importantly, the direction of all effects remained in the same direction across analyses, including those that produce counterintuitive findings. We invite readers to explore and reproduce these results (both for the PI score and the D score) using our openly available processing and analysis scripts.

[8] We conduct this power analysis using G*Power based on these extracted coefficients primarily to illustrate the differences in the meaningfulness of these effect sizes. However, more accurate values could be calculated through a direct simulation study.

tells a consistent story: the IAT *is* incrementally predictive beyond the ASCS, but its true incremental effect size is likely small. The question is not whether such an incremental effect exists, but rather in which contexts is it useful or meaningful to try to find, or take advantage of, this small incremental effect. In this sense, the above power analyses have stark implications for researchers without access to large sample sizes, suggesting that the IAT in its current form will very likely only be incrementally useful to researchers when they have access to larger populations (i.e., at least an N of 300). By contrast, the ASCS' true effect size appears to be much larger as a direct predictor of alcohol-related outcomes, which suggests it may have more utility to researchers who do not have access to such large populations.

Utility of the implicit measures

In which contexts *did* the implicit measures provide incremental utility? For the prediction of alcohol consumption on typical weeks, both the 'should' RRT and the standard IAT predicted beyond the ASCS: greater implicit drinking self-identity counterintuitively predicted *less* quantities of alcohol consumed on a typical week. This finding is unusual for the standard IAT, but may be related to poor estimation due to the smaller sample size in this study compared to others. Indeed, the IAT did not follow this pattern for the AUDIT scores, suggesting it may be an artefact of poor estimation. However, findings for the 'should' RRT *were* replicated with AUDIT scores. This consistent pattern with the 'should' RRT may indicate that should-based drinking-identity beliefs can play a role in predicting alcohol-related outcomes. Additionally, the should-aIAT generally performed best out of all of the implicit measures in terms of the direct prediction of alcohol-related outcomes (though not in over-and-above ASCS predictions), As such, the broad pattern of our results suggests that implicit should-based drinking-identity beliefs specifically may provide a valuable contribution to the prediction of drinking behaviors.

Why might greater should-based drinking self-identity predict less consumption of alcohol? Although counterintuitive at first glance, such an effect might stem from one of two sources: the first conceptual, the second statistical. First, participants who drink in large quantities might be less likely to endorse the belief that they should be drinkers (particularly if they consider intensive drinking problematic), as was the case for the RRT data (at the descriptive level). While the opposite pattern was seen for the aIAT, it is notable that the SDs around these estimates were substantially larger for the high group compared to the low group for both the RRT and aIAT, suggesting at the very least that there is greater heterogeneity in the should-based drinking identity of high drinkers. However, these findings can also easily be explained through statistical phenomena. First, because of our small sample size, poor estimation may be the source of these counterintuitive findings. In addition, our results might be explained as suppressor effects (Watson et al., 2013). That is, when two variables that have oppositional relationships to a dependent variable are entered into an interactive model as predictors, one variable (e.g., the ASCS) can artificially increase the predictive power of another variable (e.g., the standard IAT or the should-RRT). Indeed, as seen in Table 5, the standard IAT and should-RRT both had negative relationships with DDQ scores, while the ASCS had a positive relationship. As such, although potentially interesting, replication of these effects is minimally required before substantial claims can be made and to ensure that they are not by-products of poor estimation or suppressor effects.

Despite their overall procedural similarity, results from the RRT and the aIAT diverged substantially. It is unclear why this was the case. The aIAT outperformed the RRT in all analyses with the exception of those which involved prediction over and above the ASCS. If suppressor effects are ultimately shown to account for this discrepancy, then a clearer picture emerges: the aIAT in the context is simply a better measure than the RRT. However, even if this explanation proves to be inaccurate, it is unclear what specific features of the aIAT lead to its superiority to the RRT. The most notable distinction between the measures is the fact that the aIAT continuously presents the required responses on-screen, whereas the RRT presents only the words “true” and “false”. Having the additional information provided by the aIAT may in turn make the task clearer for participants, and by extension result in more meaningful performances than those in the RRT.

Though the discrepancy between am and should variants of the three relational implicit measures had utility in predicting outcomes in other domains, there was little evidence for their utility in the alcohol domain. This absence may stem from the high correlations between the ‘am’ and ‘should’ scores on all three measures (particularly on the RRT and aIAT). This, however, begs a further question: are relational implicit measures useful? The findings above might suggest not. On the other hand, specifying should-based beliefs in relational implicit measures appeared to be broadly useful in our analyses. Most critically, in a number of instances the should-aIAT outperformed the standard IAT. The aIAT is effectively identical to the standard IAT with the exception that the aIAT specifies relational information. Thus, although it seems that discrepancies between different implicit beliefs are not useful for predicting alcohol-related behavior, targeting ‘should’-based implicit beliefs relating to drinking identity might very well be of greater value than simply not specifying those relations.

Implications for drinking self-identity

The predictive utility of the ASCS, as well as the individual predictive utility of most of the implicit measures, represents further evidence for the valuable role which alcohol-related self-identity plays in predicting alcohol-related use and misuse (Gray et al., 2011; Lindgren et al., 2013, 2017; Werntz et al., 2016). In our work, only should-based implicit self-identity consistently predicted alcohol-related behaviors over-and-above the ASCS. Lindgren and colleagues (2013) suggested that implicit drinking self-identity could be a viable target for interventions focused on reducing drinking behaviors. Our findings provide a deeper insight into this possiblity, and suggest that should-based implicit drinking self-identity beliefs could be particularly viable targets of interest for such interventions.

In addition to the potential promise of should-based implicit drinking self-identity, researchers should take one other central message from these results: the ASCS is far-and-above the most practical and predictive measure of drinking self-identity and offers a very useful balance of brevity and predictive utility[9]. In addition, the ASCS can be administered with far greater ease than the implicit measures: while the implicit measures typically require specialist timed software, the ASCS can be administered through any basic forms website, as well as via pencil-and-paper. Our results may well be idiosyncratic to our sampling strategy – i.e., our sampling of low- and high-frequency drinkers using Prolific Academic's screening criteria and then subsequent use of continuous statistical analyses. The performance of the ASCS might decrease if a more continuous sampling strategy were applied from the beginning. Nonetheless, the ASCS still performed well, and has performed well in other contexts (Blevins et al., 2018; Lindgren et al., 2017).

For researchers whose primary goal is to predict typical or hazardous alcohol use, we recommend efforts aimed at improving and developing the ASCS. Indeed, the ASCS has not yet

[9] It might be suggested that the ASCS's superior performance could be attributed to the fact that all participants completed the ASCS, which meant that the ASCS had a sample size four times larger than any individual implicit measure. However, the results of the ASCS also replicate when analysing each group of participants separately. This, combined with the large effect size of the ASCS across our analyses, suggests that our results cannot be entirely attributed to this comparably larger sample size.

received much attention in the context of its psychometrics. Given recent findings that many commonly-used measures in psychology are psychometrically-defective (Hussey & Hughes, 2019), it is imperative for basic researchers to ensure that the ASCS can achieve a high standard of validity. Although some findings, including our own, suggest the ASCS has very high internal consistency (Blevins et al., 2018; Lindgren et al., 2013, 2017), there is less research investigating its test-retest reliability, underlying factor structure, or whether it achieves measurement invariance. The ASCS might be further improved by learning lessons from our findings on implicit should-based drinking self-identity. The ASCS items focus only on descriptive (i.e., "am") beliefs, and the ASCS might benefit from it the addition of items assessing prescriptive (i.e., "should") beliefs.

Limitations

Although our findings have important implications, our study comes with a number of limitations. As the divergence between our findings and previous findings with the IAT illustrate, smaller sample sizes reduce the possibility of detecting true effects. Combining this with the fact that ours is the first study of its kind to use multiple relational implicit measures in the context of alcohol-related outcomes, the inferences that we draw based on the relational implicit measures are necessarily preliminary. Future research, preferably with a much larger sample size, is required to attempt to replicate these findings. Additionally, our results speak only to the cross-sectional use of these measures, but evidence suggests that scores in the IAT may be of particular use in predicting changes in drinking outcomes across time (Lindgren et al., 2016a; 2018). Future work should therefore seek to establish whether our findings of the effectiveness of should-based drinking-identity also hold up in cross-temporal measurement contexts.

Our findings are in the context of initially sampling from high and low drinking groups, and then conducting continuous analyses. Such a strategy represents an optimal context to detect drinking-identity-based effects (Preacher, 2015). However, most individuals likely fall in the category of moderate drinkers. For this group, implicit drinking-identity may be less stable, and therefore reduce the predictive utility of all of the measures here. Indeed, this is a limitation inherent in any similar such extreme-groups experimental design (Preacher et al., 2005). Indeed, one particular issue resides in the fact that the negative binomial distribution of the effects seen here may be an artefact of this extreme groups sampling strategy. If the distribution of these scores would differ when sampling continuously along the spectrum of drinkers, then the findings from our statistical models here which assume a negative-binomial distribution would not be generalisable to this more comprehensive sample. Ultimately, this will also need to be investigated in order to determine, which (if any), measures of drinking self-identity can be most effective in predicting drinking-related outcomes in the population at-large (some studies have already begun to address this question; Werntz et al., 2016; Lindgren et al., 2016b). Indeed, in terms of the population at-large, one further limitation relates to our use of an online sample: results from such a sample may differ from what could be found amongst a population recruited in-person or among a clinical population. Future research should therefore also seek to assess the generalisability of our findings to other populations.

Further, although the explicit measure outperformed all of the implicit measures here, that may not be the case in all assessment contexts. Since implicit measures capture behaviour putatively under more automatic conditions, they may provide insight beyond self-reports into clinically-relevant behaviour which may occur automatically. For example, in the context of predicting adherence to a treatment program, implicit measures may provide utility in the fact that they might predict reasons for relapsing which occur outside of the awareness or intention of patients (Van Dessel et al., 2020; but see Cummins et al., 2019; Schimmack, 2019).

Conclusions

We compared five measures of drinking self-identity on criteria relevant to alcohol researchers. Our results suggest that the RRT and aIAT in particular show some promise as measures of drinking self-identity, with the should-aIAT the best-performing individual predictor, and the should-RRT being the only measure which consistently predicted above and beyond the ASCS. The pCIT, by contrast, had virtually no promise as a measure. Our findings also demonstrated that the ASCS, far and above, out-performed all of the implicit measures. The ASCS might be further improved by addition of items that assess prescriptive drinking self-identity.

References

Agosta, S., & Sartori, G. (2013). The autobiographical IAT: A review. *Frontiers in Psychology*, *4*. https://doi.org/10.3389/fpsyg.2013.00519

Bear, A., & Knobe, J. (2017). Normality: Part descriptive, part prescriptive. *Cognition*, *167*, 25–37. https://doi.org/10.1016/j.cognition.2016.10.024

Blanton, H., Jaccard, J., Strauts, E., Mitchell, G., & Tetlock, P. E. (2015). Toward a meaningful metric of implicit prejudice. *The Journal of Applied Psychology*, *100*(5), 1468–1481. https://doi.org/10.1037/a0038379

Blevins, C. E., Abrantes, A. M., Anderson, B. J., Caviness, C. M., Herman, D. S., & Stein, M. D. (2018). A longitudinal evaluation of the role of alcohol self-concept in alcohol use, motives, negative affect, and alcohol-related problems among emerging adults. *The American Journal on Addictions*, *27*(6), 501–508. https://doi.org/10.1111/ajad.12768

Caudwell, K. M., & Hagger, M. S. (2014). Pre-drinking and alcohol-related harm in undergraduates: The influence of explicit motives and implicit alcohol identity. *Journal of Behavioral Medicine*, *37*(6), 1252–1262. https://doi.org/10.1007/s10865-014-9573-6

Cathelyn, F., Van Dessel, P., Cummins, J., Spruyt, A., & De Houwer, J. (2020). Driving Under the Influence of Alcohol. Retrieved from osf.io/b7wur.

Christoffersen, M. N., & Soothill, K. (2003). The long-term consequences of parental alcohol abuse: A cohort study of children in Denmark. *Journal of Substance Abuse Treatment*, *25*(2), 107–116. https://doi.org/10.1016/S0740-5472(03)00116-8

Collins, R. L., Parks, G. A., & Marlatt, G. A. (1985). Social determinants of alcohol consumption: The effects of social interaction and model status on the self-administration of alcohol. *Journal of Consulting and Clinical Psychology*, *53*(2), 189–200. https://doi.org/10.1037/0022-006X.53.2.189

Cummins, J., & De Houwer, J. (2019). An inkblot for beliefs: The Truth Misattribution Procedure. *PloS One*, *14*(6), e0218661. https://doi.org/10.1371/journal.pone.0218661

Cummins, J., Hussey, I., & Hughes, S. (2019). The AMPeror's New Clothes: Performance on the Affect Misattribution Procedure is Mainly Driven by Awareness of Influence of the Primes. https://doi.org/10.31234/osf.io/d5zn8

Cummins, J., Lindgren, K., & De Houwer, J. (2020). *Comparing (relational) implicit measures: Predicting drinking outcomes*. Retrieved from https://osf.io/7df69/.

Cummins, J., Verschuere, B., & De Houwer, J. (2020). The Propositional Concealed Information Test. *Manuscript in preparation.*

Davies, E. L., Paltoglou, A. E., & Foxcroft, D. R. (2017). Implicit alcohol attitudes predict drinking behaviour over and above intentions and willingness in young adults but willingness is more important in adolescents: Implications for the Prototype Willingness Model. *British Journal of Health Psychology*, *22*(2), 238–253. https://doi.org/10.1111/bjhp.12225

De Bruyn, S., Wouters, E., Ponnet, K., Van Damme, J., Maes, L., & Van Hal, G. (2018). Problem drinking among Flemish students: Beverage type, early drinking onset and negative personal & social consequences. *BMC Public Health*, *18*(1), 234. https://doi.org/10.1186/s12889-018-5120-7

De Houwer, J. (2002). The Implicit Association Test as a tool for studying dysfunctional associations in psychopathology: Strengths and limitations. *Journal of Behavior Therapy and Experimental Psychiatry*, *33*(2), 115–133. https://doi.org/10.1016/S0005-7916(02)00024-1

De Houwer, J. (2006). What Are Implicit Measures and Why Are We Using Them? In *Handbook of implicit cognition and addiction*. 11–28. https://doi.org/10.4135/9781412976237.n2

De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The relational responding task: Toward a new implicit measure of beliefs. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00319

De Schryver, M., & De Neve, J. (2019). A tutorial on probabilistic index models: Regression models for the effect size P(Y1 < Y2). *Psychological Methods*, *24*(4), 403–418. https://doi.org/10.1037/met0000194

De Schryver, M., Hussey, I., De Neve, J., Cartwright, A., & Barnes-Holmes, P. M. D. (2018). The PI IRAP: An alternative scoring algorithm for the IRAP using a probabilistic semiparametric effect size measure. *JOURNAL OF CONTEXTUAL BEHAVIORAL SCIENCE*, *7*, 97–103. http://dx.doi.org/10.1016/j.jcbs.2018.01.001

Degenhardt, L., Chiu, W. T., Sampson, N., Kessler, R. C., Anthony, J. C., Angermeyer, M., … Wells, J. E. (2008). Toward a Global View of Alcohol, Tobacco, Cannabis, and Cocaine Use: Findings from the WHO World Mental Health Surveys. *PLOS Medicine*, *5*(7), e141. https://doi.org/10.1371/journal.pmed.0050141

Flake, J. K., Pek, J., & Hehman, E. (2017). Construct Validation in Social and Personality Research: Current Practice and Recommendations. *Social Psychological and Personality Science*, *8*(4), 370–378. https://doi.org/10.1177/1948550617693063

Genschow, O., Demanet, J., Hersche, L., & Brass, M. (2017). An empirical comparison of different implicit measures to predict consumer choice. *PLoS ONE*, *12*(8). https://doi.org/10.1371/journal.pone.0183937

Gray, H. M., LaPlante, D. A., Bannon, B. L., Ambady, N., & Shaffer, H. J. (2011). Development and validation of the Alcohol Identity Implicit Associations Test (AI-IAT). *Addictive Behaviors*, *36*(9), 919–926. https://doi.org/10.1016/j.addbeh.2011.05.003

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480. https://doi.org/10.1037//0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the Implicit Association Test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216. https://doi.org/10.1037/0022-3514.85.2.197

Greenwald, A. G., Poehlman, T. A., Uhlmann, E. L., & Banaji, M. R. (2009). Understanding and using the Implicit Association Test: III. Meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, *97*(1), 17–41. https://doi.org/10.1037/a0015575

Heider, N., Spruyt, A., & De Houwer, J. (2015). Implicit beliefs about ideal body image predict body image dissatisfaction. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.01402

Heider, N., Spruyt, A., & De Houwer, J. (2018). Body Dissatisfaction Revisited: On the Importance of Implicit Beliefs about Actual and Ideal Body Image. *Psychologica Belgica*, *57*(4), 158–173. https://doi.org/10.5334/pb.362

Higgins, E. T., Klein, R., & Strauman, T. (1985). Self-Concept Discrepancy Theory: A Psychological Model for Distinguishing among Different Aspects of Depression and Anxiety. *Social Cognition*, *3*(1), 51–76. https://doi.org/10.1521/soco.1985.3.1.51

Hussey, I., & Hughes, S. (2019). Hidden invalidity among fifteen commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science*. https://psyarxiv.com/7rbfp/

Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., … Banaji, M. R. (2019). Relationship between the Implicit Association Test and intergroup behavior: A meta-analysis. *American Psychologist*, *74*(5), 569–586. https://doi.org/10.1037/amp0000364

Lechner, W. V., Day, A. M., Metrik, J., Leventhal, A. M., & Kahler, C. W. (2016). Effects of alcohol-induced working memory decline on alcohol consumption and adverse consequences of use. *Psychopharmacology*, *233*(1), 83–88. https://doi.org/10.1007/s00213-015-4090-z

Lindgren, K. P., Baldwin, S. A., Olin, C. C., Wiers, R. W., Teachman, B. A., Norris, J., Kaysen, D., & Neighbors, C. (2018). Evaluating Within-Person Change in Implicit Measures of Alcohol Associations: Increases in Alcohol Associations Predict Increases in Drinking Risk and Vice Versa. *Alcohol and Alcoholism (Oxford, Oxfordshire)*, *53*(4), 386–393. https://doi.org/10.1093/alcalc/agy012

Lindgren, K. P., Gasser, M. L., Werntz, A., Namaky, N., Baldwin, S. A., & Teachman, B. A. (2016b). Moderators of Implicit and Explicit Drinking Identity in a Large US Adult Sample. *Addictive Behaviors*, *60*, 177–183. https://doi.org/10.1016/j.addbeh.2016.04.022

Lindgren, K. P., Neighbors, C., Gasser, M. L., Ramirez, J. J., & Cvencek, D. (2017). A review of implicit and explicit substance self-concept as a predictor of alcohol and tobacco use and misuse. *The American Journal of Drug and Alcohol Abuse*, *43*(3), 237–246. https://doi.org/10.1080/00952990.2016.1229324

Lindgren, K. P., Neighbors, C., Teachman, B. A., Baldwin, S. A., Norris, J., Kaysen, D., … Wiers, R. W. (2016a). Implicit alcohol associations, especially drinking identity, predict drinking over time. *Health Psychology*, *35*(8), 908–918. https://doi.org/10.1037/hea0000396

Lindgren, K. P., Neighbors, C., Teachman, B. A., Wiers, R. W., Westgate, E., & Greenwald, A. G. (2013). I Drink Therefore I am: Validating Alcohol-related Implicit Association Tests. *Psychology of Addictive Behaviors : Journal of the Society of Psychologists in Addictive Behaviors*, *27*(1), 1–13. https://doi.org/10.1037/a0027640

Lindgren, K. P., Neighbors, C., Westgate, E., & Salemink, E. (2014). Self-Control and Implicit Drinking Identity as Predictors of Alcohol Consumption, Problems, and Cravings. *Journal of Studies on Alcohol and Drugs*, *75*(2), 290–298.

Mattiuzzi, C., & Lippi, G. (2019). Worldwide epidemiology of alcohol and drugs abuse. *European Journal of Internal Medicine*, *70*, e27–e28. https://doi.org/10.1016/j.ejim.2019.10.019

Millner, A. J., Coppersmith, D. D. L., Teachman, B. A., & Nock, M. K. (2018). The Brief Death Implicit Association Test: Scoring Recommendations, Reliability, Validity, and Comparisons with the Death Implicit Association Test. *Psychological Assessment*, *30*(10), 1356–1366. https://doi.org/10.1037/pas0000580

Müller, F., & Rothermund, K. (2019). The Propositional Evaluation Paradigm: Indirect Assessment of Personal Beliefs and Attitudes. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02385

Muschalik, C., Elfeddali, I., Candel, M. J. J. M., Crutzen, R., & de Vries, H. (2019). Does the discrepancy between implicit and explicit attitudes moderate the relationships between explicit attitude and (intention to) being physically active? *BMC Psychology*, *7*(1), 52. https://doi.org/10.1186/s40359-019-0322-z

National Institute on Alcohol Abuse and Alcoholism. (2017) Drinking Levels Defined. Retrieved from NIAAA website: https://www.niaaa.gov/alcohol-health/overview-alcohol-consumption/moderate-binge-drinking

Nosek, B. A. (2007). Implicit–Explicit Relations. *Current Directions in Psychological Science*, *16*(2), 65–69. https://doi.org/10.1111/j.1467-8721.2007.00477.x

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, *15*(4), 152–159. https://doi.org/10.1016/j.tics.2011.01.005

Payne, B. K., Burkley, M. A., & Stokes, M. B. (2008). Why do implicit and explicit attitude tests diverge? The role of structural fit. *Journal of Personality and Social Psychology*, 16–31.

Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the Extreme Groups Approach: A Critical Reexamination and New Recommendations. *Psychological Methods*, *10*(2), 178–192. https://doi.org/10.1037/1082-989X.10.2.178

Preacher, K. J. (2015). Extreme Groups Designs. In R. L. Cautin & S. O. Lilienfeld (Eds.), *The Encyclopedia of Clinical Psychology* (pp. 1–4). John Wiley & Sons, Inc. https://doi.org/10.1002/9781118625392.wbecp190

Rachdaoui, N., & Sarkar, D. K. (2017). Pathophysiology of the Effects of Alcohol Abuse on the Endocrine System. *Alcohol Research : Current Reviews*, *38*(2), 255–276.

Ramirez, J. J., Olin, C. C., & Lindgren, K. P. (2017). Resolving an identity crisis: Implicit drinking identity and implicit alcohol identity are related but not the same. *Addictive Behaviors*, *72*, 41–44. https://doi.org/10.1016/j.addbeh.2017.03.014

Rehm, J., Mathers, C., Popova, S., Thavorncharoensap, M., Teerawattananon, Y., & Patra, J. (2009). Global burden of disease and injury and economic cost attributable to alcohol use and alcohol-use disorders. *The Lancet*, *373*(9682), 2223–2233. https://doi.org/10.1016/S0140-6736(09)60746-7

Reich, R. R., Below, M. C., & Goldman, M. S. (2010). Explicit and Implicit Measures of Expectancy and Related Alcohol Cognitions: A Meta-Analytic Comparison. *Psychology of Addictive Behaviors : Journal of the Society of Psychologists in Addictive Behaviors*, *24*(1), 13–25. https://doi.org/10.1037/a0016556

Remue, J., Houwer, J. D., Barnes-Holmes, D., Vanderhasselt, M.-A., & Raedt, R. D. (2013). Self-esteem revisited: Performance on the implicit relational assessment procedure as a measure of self- versus ideal self-related cognitions in dysphoria. *Cognition and Emotion*, *27*(8), 1441–1449. https://doi.org/10.1080/02699931.2013.786681

Remue, J., Hughes, S., De Houwer, J., & De Raedt, R. (2014). To Be or Want to Be: Disentangling the Role of Actual versus Ideal Self in Implicit Self-Esteem. *PLoS ONE*, *9*(9). https://doi.org/10.1371/journal.pone.0108837

Revelle, W., & Condon, D. M. (2019). Reliability from α to ω: A tutorial. *Psychological Assessment*, *31*(12), 1395–1411. https://doi.org/10.1037/pas0000754

Rudman, L. A., & Glick, P. (2001). Prescriptive Gender Stereotypes and Backlash Toward Agentic Women. *Journal of Social Issues*, *57*(4), 743–762. https://doi.org/10.1111/0022-4537.00239

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. https://doi.org/10.1111/j.1360-0443.1993.tb02093.x

Schimmack, U. (2019). The Implicit Association Test: A Method in Search of a Construct. *Perspectives on Psychological Science*, 1745691619863798. https://doi.org/10.1177/1745691619863798

Shadel, W. G., & Mermelstein, R. (1996). Individual differences in self-concept among smokers attempting to quit: Validation and predictive utility of measures of the smoker self-concept and abstainer self-concept. *Annals of Behavioral Medicine: A Publication of the Society of Behavioral Medicine*, *18*(3), 151–156. https://doi.org/10.1007/BF02883391

Stein, K. F., & Corte, C. (2007). Identity impairment and the eating disorders: Content and organization of the self-concept in women with anorexia nervosa and bulimia nervosa. *European Eating Disorders Review: The Journal of the Eating Disorders Association*, *15*(1), 58–69. https://doi.org/10.1002/erv.726

Thas, O., Neve, J. D., Clement, L., & Ottoy, J.-P. (2012). Probabilistic index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *74*(4), 623–671. https://doi.org/10.1111/j.1467-9868.2011.01020.x

Tibboel, H., De Houwer, J., Dirix, N., & Spruyt, A. (2017). Beyond associations: Do implicit beliefs play a role in smoking addiction? *Journal of Psychopharmacology*, *31*(1), 43–53. https://doi.org/10.1177/0269881116665327

Van Dessel, P., Cummins, J., Hughes, S., Kasran, S., Cathelyn, F., & Moran, T. (2020). Reflecting on twenty-five years of research using implicit measures: Recommendations for their future use. *Social Cognition*.

Werntz, A. J., Steinman, S. A., Glenn, J. J., Nock, M. K., & Teachman, B. A. (2016). Characterizing Implicit Mental Health Associations across Clinical Domains. *Journal of Behavior Therapy and Experimental Psychiatry*, *52*, 17–28. https://doi.org/10.1016/j.jbtep.2016.02.004

Whelan, R. (2008). Effective Analysis of Reaction Time Data. *The Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630

Table 1. Differences between the different measures and rationale for inclusion. Words in square brackets indicate different relational clauses between different versions of each procedure.

Measure	Rationale	Response labels	Example stimuli		
Alcohol Self-Concept Scale	Most commonly-used explicit measure of drinking self-identity	5-point Likert scale responses	Drinking is a large part of my personality; drinking is part of my daily life		
Implicit Association Test	Already shown utility in measuring drinking self-identity in the past (e.g., Lindgren et al., 2013)	Me/Not Me (categories); Drinker/Abstainer (attributes)	Categories: me, my; they, them	Attributes: drunk, drink; sober, abstain	
Autobiographical Implicit Association Test	Relational version of the IAT; already some evidence for its effectiveness in predicting alcohol consumption (Cathelyn et al., 2020)	True/False (categories); I [am/should be] a drinker, I [am/should be] an abstainer (attributes)	Categories: I'm doing a psychology experiment; I'm eating at a downtown restaurant	Attributes: Drinking [is/should be] part of who I am; Abstaining [is/should be] part of who I am	
Relational Responding Task	Has shown efficacy in a range of domains (e.g.,	True/False (categories and attributes)	Categories: correct,	Attributes: I [am/should be] a drinker;	

	body dissatisfaction, anti-immigrant beliefs, and smoking behavior; De Houwer et al., 2015; Heider et al., 2015; Tibboel et al., 2017)		confirm; wrong, deny	I [am/should be] an abstainer	
Propositional Concealed Information Test	Newly developed relational implicit measure (Cummins et al., 2020); the measure it is based on (the CIT) shows very large effect sizes and excellent reliability (Agosta & Sartori, 2013)	True/False (targets, irrelevants, probes)	Targets (always respond true): I am a human, I should be a good person	Irrelevants (always respond false): I am a toddler, I should be a bad person	Probes (always respond false): I am/should be a drinker, I am/should be an abstainer

Table 2. Breakdown of sample characteristics.

Problematic Drinking Group	Mean Age (SD)	Mean AUDIT Score (SD)	Mean DDQ Typical Quantity Score (SD)	Mean DDQ Heavy Quantity Score (SD)
Low men (n = 87)	36.91 (11.59)	5.76 (4.55)	4.57 (3.84)	7.33 (6.66)
Low women (n = 85)	38.31 (11.76)	4.00 (3.14)	2.44 (2.10)	5.01 (8.63)
High men (n = 111)	34.37 (9.36)	15.99 (7.34)	37.12 (23.02)	47.61 (28.73)
High women (n = 72)	37.56 (10.40)	14.83 (8.26)	24.39 (23.55)	33.79 (26.48)

Table 3. Median split-half reliability estimates for each implicit measure.

Measure	Relation type	Median R*sb* estimate	95% CIs
aIAT	Am	.853	.765, .928
	Should	.870	.766, .959
RRT	Am	.788	.607, .902
	Should	.805	.597, .901
pCIT	Am	.089	-.327, .394
	Should	.065	-.406, .378
IAT	NA	.934	.867, .987

Table 4. Implicit-explicit correlations for each of the implicit measures.

Measure	Relation type	Median *r*	95% CIs	p
aIAT	Am	.283	.085, .459	.006
	Should	.309	.113, .481	.002
RRT	Am	.269	.068, .449	.010
	Should	.217	.013, .404	.038
pCIT	Am	.129	-.081, .329	.228
	Should	.025	-.184, .232	.817
IAT	NA	.154	-.064, .357	.165

Table 5. Relationships between DDQ scores and scores on each measure (implicit and explicit).

Measure	Relation type	DDQ	Std. Beta	95% CIs	p
aIAT	Am	Typical	0.20	-0.03, 0.42	.083
		Heavy	0.24	0.04, 0.45	.022
	Should	Typical	0.37	0.14, 0.61	.004
		Heavy	0.37	0.13, 0.62	.002
	Discrepancy	Typical	-0.02	-0.26, 0.20	.895
		Heavy	-0.03	-0.27, 0.22	.807
RRT	Am	Typical	0.03	-0.19, 0.51	.886
		Heavy	0.16	-0.17, 0.45	.362
	Should	Typical	-0.09	-0.31, 0.33	.578
		Heavy	0.01	-0.32, 0.31	.934
	Discrepancy	Typical	0.10	-0.19, 0.44	.531
		Heavy	0.12	-0.18, 0.43	.444
pCIT	Am	Typical	0.06	-0.19, 0.28	.637
		Heavy	0.05	-0.18, 0.28	.709
	Should	Typical	0.08	-0.16, 0.32	.532
		Heavy	0.08	-0.16, 0.33	.503
	Discrepancy	Typical	-0.01	-0.21, 0.20	.952
		Heavy	-0.02	-0.22, 0.18	.847
IAT	NA	Typical	-0.18	-0.42, 0.06	.144
		Heavy	-0.09	-0.34, 0.16	.495
ASCS	NA	Typical	0.69	0.61, 0.78	<.001
		Heavy	0.74	0.66, 0.81	<.001

Table 6. Prediction of DDQ scores by the implicit measures over and above the ASCS.

Measure	Relation type	DDQ	Std. Beta	95% CIs	p
aIAT	Am	Typical	0.04	-0.15, 0.23	.694
		Heavy	0.05	-0.13, 0.23	.578

	Should	Typical	0.20	-0.02, 0.42	.072
		Heavy	0.19	-0.02, 0.40	.072
RRT	Am	Typical	-0.11	-0.41, 0.19	.457
		Heavy	0.02	-0.27, 0.30	.916
	Should	Typical	-0.39	-0.65, -0.13	.004
		Heavy	-0.25	-0.51, -0.00	.050
pCIT	Am	Typical	-0.06	-0.24, 0.12	.526
		Heavy	-0.06	-0.25, 0.13	.523
	Should	Typical	0.07	-0.12, 0.25	.470
		Heavy	0.04	-0.15, 0.23	.671
IAT	NA	Typical	-0.61	-0.85, -0.36	<.001
		Heavy	-0.16	-0.37, 0.06	.151

Table 7. Prediction of AUDIT scores by each of the measures.

Measure	Relation type	Std. Beta	95% CIs	p
aIAT	Am	0.12	-0.03, 0.27	.126
	Should	0.18	0.01, 0.35	.037
	Discrepancy	0.00	-0.16, 0.17	.957
RRT	Am	-0.03	-0.26, 0.20	.804
	Should	-0.10	-0.31, 0.10	.323
	Discrepancy	0.09	-0.12, 0.30	.390
pCIT	Am	0.12	-0.04, 0.28	.129
	Should	0.03	-0.14, 0.19	.750
	Discrepancy	0.06	-0.07, 0.20	.358
IAT	NA	0.06	-0.11, 0.22	.499
ASCS	NA	0.53	0.48, 0.57	<.001

Table 8. Prediction of AUDIT scores by the implicit measures over and above the ASCS.

Measure	Relation type	Std. Beta	95% CIs	p

aIAT	Am	0.03	-0.09 0.14	.648
	Should	0.04	-0.09, 0.18	.527
RRT	Am	-0.20	-0.39, -0.01	.037
	Should	-0.24	-0.41, -0.07	.006
pCIT	Am	0.07	-0.04, 0.18	.226
	Should	0.02	-0.09, 0.13	.731
IAT	NA	0.02	-0.09, 0.13	.692

Table 9. Accuracy of each of the measures in classifying participants as low or high drinkers.

Measure	Relation type	AUC	95% CIs
aIAT	Am	.67	.55, .77
	Should	.63	.51, .74
	Discrepancy	.53	.41, .64
RRT	Am	.50	.38, .61
	Should	.56	.44, .67
	Discrepancy	.57	.45, .69
pCIT	Am	.52	.39, .63
	Should	.57	.43, .69
	Discrepancy	.52	.40, .65
IAT	NA	.58	.43, .70
ASCS	NA	.82	.78, .85

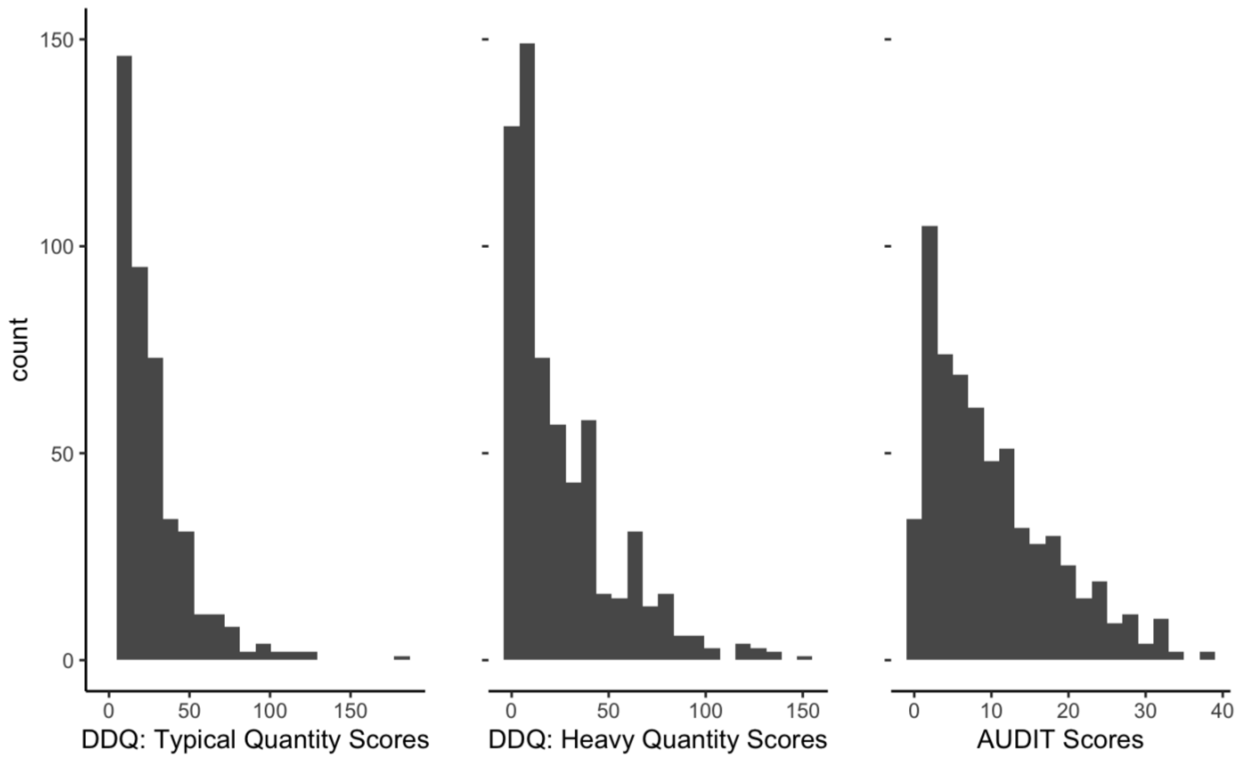

Figure 1. The distribution of DDQ (typical and heavy) and AUDIT scores across participants.