Published in final edited form as:

Psychol Addict Behav. 2014 September ; 28(3): 852–862. doi:10.1037/a0035877.

Application of Item Response Theory to Tests of Substance-related Associative Memory

Yusuke Shono,

School of Community and Global Health, Claremont Graduate University.

Jerry L. Grenard,

School of Community and Global Health, Claremont Graduate University.

Susan L. Ames, and

School of Community and Global Health, Claremont Graduate University.

Alan W. Stacy

School of Community and Global Health, Claremont Graduate University.

Abstract

A substance-related word association test (WAT) is one of the commonly used indirect tests of substance-related implicit associative memory and has been shown to predict substance use. This study applied an item response theory (IRT) modeling approach to evaluate psychometric properties of the alcohol- and marijuana-related WATs and their items among 775 ethnically diverse at-risk adolescents. After examining the IRT assumptions, item fit, and differential item functioning (DIF) across gender and age groups, the original 18 WAT items were reduced to 14- and 15-items in the alcohol- and marijuana-related WAT, respectively. Thereafter, unidimensional one- and two-parameter logistic models (1PL and 2PL models) were fitted to the revised WAT items. The results demonstrated that both alcohol- and marijuana-related WATs have good psychometric properties. These results were discussed in light of the framework of a unified concept of construct validity (Messick, 1975, 1989, 1995).

Keywords

item response theory; construct validity; word association test; substance use; adolescents

In the past two decades, implicit memory and cognition approaches have gained substantial popularity in addiction and health behavior research. Focusing on the role of spontaneously activated cognitions on behavior (Stacy & Wiers, 2010, Wiers & Stacy, 2006), researchers have examined automatic/implicit cognitive processes at different levels of analysis, ranging from attention (e.g., Bradley, Field, Mogg, & De Houwer, 2004, Mogg & Bradley, 2002) to memory (e.g., Krank & Goldstein, 2006, Stacy, 1995, 1997) and attitude (e.g., Chassin, Presson, Sherman, Seo, & Macy, 2010, Huijding, de Jong, Wiers, & Verkooijen, 2005,

Houben, Havermans, & Wiers, 2010) by using various indirect tests of implicit cognitive processes related to addictive and health behaviors (see Ames, Grenard, Thush, Sussman, Wiers & Stacy, 2007, Stacy, Ames, & Grenard, 2006, Stacy & Wiers, 2010, for review). A recent meta-analysis examining the relationship between substance-related implicit cognition and substance use revealed that the substance-related implicit word association test (WAT) was the best predictor of substance use with the largest effect size (mean $r = .38$) among other implicit measures (Rooke, Hine, & Thorsteinsson, 2008). Although studies have reported good reliability of WAT (Ames et al., 2007, Preece, 1978), to the best of our knowledge, no *comprehensive* psychometric evaluations of WAT have been conducted in research on addiction, cognition, or memory. The current study extends previous research by applying a comprehensive item response theory framework to understand and improve the psychometric properties of WAT items and estimation of underlying latent traits of alcohol- and marijuana-related associative memory.

Word Association Tests (WATs) in Addiction and Health Behavior Research

The WAT is one of the most commonly used indirect memory tests for assessing the retrieval of preexisting substance-related associations in memory (Stacy, 1995, 1997). In substance-related WAT, a series of substance-related cue words or phrases are presented one by one visually or auditorily, and participants are asked to generate the first word or short phrase that comes to mind when they think of the cue. It is assumed that an association of a cue-target pair gets strengthened with repetitive encounters with a substance-related cue (e.g., “feeling good”) and target behavior (e.g., marijuana use). Therefore, those who frequently engage in substance use are more likely than those who do not to spontaneously think of substance-use behavior in response to substance-related cues in WAT.

Accumulated evidence has shown that substance- or risky behavior-related implicit associative memory, measured by WAT, has strong predictive power for substance use, including alcohol (Ames & Stacy, 1998, Kelly, Masterman, & Marlatt, 2005, Stacy, 1997), marijuana (Ames et al., 2007, Ames & Stacy, 1998, Stacy, 1997) and cigarette use (Grenard et al., 2008, Kelly, Haynes, & Marlatt, 2008), as well as risky sexual behavior (Ames, Grenard, & Stacy, 2013, Grenard, Ames, & Stacy, 2013, Stacy, Ames, Ullman, Zogg, & Leigh, 2006). Given the successful application of WAT to a wide range of issues in health and cognition, it is important to fully understand the psychometric characteristics and construct validity of the measure. The IRT modeling framework provides one of the most comprehensive strategies available to accomplish these goals and has a number of advantages over the traditional classical test theory (CTT) approach (Reise, Ainsworth, & Haviland, 2005).

Item Response Theory (IRT) Applied to Substance-related Word Association Test (WAT)

IRT consists of a series of statistical models specified to describe the probability of endorsing an item as a function of an underlying latent trait (θ). In the context of the alcohol-related WAT, IRT describes the association between the probability that a participant generates an alcohol-related response to a given WAT item and his/her level of the latent

alcohol-related implicit associative memory. The use of IRT in psychometric evaluation has several advantages over classical test theory (CTT). First, IRT allows for detailed investigation of WAT items in relation to the latent alcohol-related associative memory. It provides parameter estimates of item difficulty (b) and item discrimination (a). The b parameter indicates (1) how difficult a given WAT item is and (2) what level of latent memory association is needed so that 50% of participants would endorse an alcohol-related response to a given WAT item. Participants whose trait level (i.e., alcohol-related associative memory) is higher are likely to generate an alcohol-related response to a WAT item with a higher b parameter value. The a parameter tells how effectively a WAT item differentiates among individuals with different levels of latent implicit alcohol-related associative memory. The item with a higher value of a is a good item because such an item discriminates effectively between individuals of slightly different levels of latent alcohol-related implicit associations in memory.

A critical advantage of IRT over CTT is that IRT is sample-invariant whereas CTT is sample-dependent (Hambleton & Jones, 1993). Under the situation in which an IRT model fits the data, the item parameter estimates (i.e., the a and b parameters) can be interpreted independent of the study sample (item-parameter invariance; Lord, 1980). Similarly, a latent trait can be estimated independent of a set of test items used in a study (person-parameter invariance; Lord, 1980). These sample invariant characteristics are not true in CTT. In CTT, item discrimination (i.e., item-total correlation), item difficulty (i.e., proportion of correct) and scale scores (i.e., the summed score) are completely dependent on a sample. Thus, an estimate of a latent trait score in CTT is largely affected by the characteristics of a study sample (Hambleton & Jones, 1993).

Another advantage of IRT is that reliability can be estimated with great flexibility. In CTT, a reliability estimate (e.g., Kuder-Richardson Formula 20, Cronbach's coefficient alpha) is a fixed constant for all items. In contrast, reliability in IRT can be estimated at any point in the range of an underlying latent trait. Moreover, reliability estimates can be computed at both the item and test levels, using the item information and test information functions (IIF and TIF), respectively. In our substance-related WAT, we determined the extent to which each WAT item and WAT as a test accurately estimated a specific level of implicit substance-related associative memory.

Lastly, the IRT framework allows for the investigation of differential item functioning (DIF). The DIF analysis assesses whether or not a test item functions equivalently across subgroups of a study sample while controlling for the overall difference in the latent trait levels. For example, if the a or b parameter of a given WAT item is different between male and female participants with the same level of the latent implicit alcohol-related associative memory, the item is considered having DIF and could be a threat to the construct validity of the alcohol-related WAT (Kristjansson, Aylesworth, McDowell, & Zumbo, 2005).

Current Study

The current study evaluated psychometric properties of two forms of a substance-related WAT, marijuana- and alcohol-related WAT, using a unidimensional IRT modeling

approach. The data were collected as part of a large-scale longitudinal study of dual-process theory and drug use in adolescents and consists of 775 ethnically diverse, at-risk high school students in Southern California. The adolescent sample was chosen because of sufficient variability in alcohol, marijuana, and other drug use and the importance of this age group for the study of drug use progression.

The aims of the study were to 1) evaluate parameters of substance-related WAT items including item difficulty and item discrimination; 2) examine the precision of WAT at the item and test levels, 3) estimate the latent trait scores (i.e., the level of substance-related implicit associative memory) for each participant, and 4) evaluate criterion validity through the association between WAT scores and substance-use measures. Results of comprehensive psychometric validation of substance-related WAT will be discussed in light of the framework of a unified concept of construct validity (Messick, 1975, 1989, 1995). The comprehensive IRT approach illustrated in this article is applicable to a wide variety of measurement issues in associative memory and other areas of addiction and health behavior research.

Method

Participants

The participants were 775 continuation high school (CHS) students (340 females) in the greater Los Angeles area. Their participation in this study did not require current or past history of substance use. The participants' ages ranged from 14 to 20, of which 94% were between the ages of 15 and 18. The study sample¹ comprised Hispanic (62.5%), Non-Hispanic White (12.5%), mixed race/ethnicity (18.7%), Black (3.2%), and other race/ethnicity that included Asian, Native American, and 'other' (3.1%). They were recruited from classes from 42 CHSs, which were selected from over 100 CHSs in the region. The schools sampled did not provide any drug education programs to their students.

Measures

Word Association Test (WAT)—As described in the introduction, the substance-related WAT is an indirect memory test designed to assess the spontaneous retrieval of preexisting substance-related associations in memory (e.g., Stacy, 1997). The current study used two formats of WAT, an outcome-behavior association task (OBAT) and a compound-cue version of WAT. In OBAT, all cues are phrases that are related to affective outcomes of drug use (e.g., “feeling good”). In the compound-cue WAT, cues consist of either a combination of location and affective outcome phrases (two compound cues; e.g., “my bedroom, feeling good”) or a combination of situation, location and affective outcome phrases (three compound cues; e.g., “weekend, friend’s house, having fun”). Fillers are cues not related to substance use (e.g., “doing homework”). Each of the three cue types had 6 target cues and 2 filler cues, totaling 18 target and 6 filler cues. Each trial started by visually

¹In several recent studies by other investigators on continuation high schools (CHSs) in the greater Los Angeles area (e.g., Barnett, et al., 2013; Sussman, et al., 2012), sample characteristics (including the male-to-female ratio, the mean age, racial/ethnic profile, and past alcohol and marijuana use) were very similar to those in the current study. Although demographic information was not available on all CHSs in the region, the general consistency across diverse studies in the region suggests that the present sample is at least similar to other samples previously drawn from the population.

presenting a cue phrase in the center of a computer display, and participants were instructed to respond with the first behavior or action that came to mind as quickly as possible. Responses were typed in a text box that appeared right below where the cue was presented. The next trial was generated by participants' clicking a text button that reads "click here to continue" or after 21 seconds had elapsed since the presentation of a cue, whichever came first.

The self-coding procedure (Frigon & Krank, 2009; Krank, Schoenfeld, & Frigon, 2010) was employed to code the WAT responses upon completion of the WAT session. In this procedure, participants were presented with a WAT cue and their typed response on the computer display, along with a list of 12 behavior categories (e.g., alcohol, marijuana, tobacco, exercise, etc.). They were asked to check one or more categories that were related to their response. A checked response was coded 1 and unchecked response was coded 0, and these scores were summed to yield a total WAT score for each category. In the current study, the scores for alcohol- and marijuana-related responses were examined separately.

Drug Use: Marijuana and Alcohol—Frequency of drug use was measured by a self-report drug use questionnaire (Stacy et al., 1990, Stacy, 1997) that asked participants to indicate how many times they had used each drug in the past year and the past 30 days. The questionnaire was an 11-point rating scale, with frequency response options ranging from 1 (None) to 11 (91+ times). The reliability and validity of these self-reported drug use measures were demonstrated elsewhere (e.g., Stacy et al., 1990).

Other variables—Participants' demographics (age, gender, and language use), scores on the Rutgers Alcohol Problem Index (RAPI; White & Labouvie, 1989), and frequencies of simultaneous polydrug use (Collins, Ellickson, & Bell, 1998) were also assessed. These measures were used as predictors in missing data analyses reported below (see Data Analysis Plan for more details).

Procedures

We contacted each continuation high school (CHS) to arrange recruitment and obtained both written assent from eligible students and consent from their parents. Assent and consent forms explained that the study was to investigate teenagers' health behaviors and they would need to participate in three assessments over the course of two years to complete their participation. Computer-based assessments were administered during regular school hours in groups of up to 20 ($M = 11.67$ participants per session) in a classroom that was provided by each CHS. Data collectors set up a mobile computer laboratory in each classroom that included 20 laptop computers supplied by the research project. Upon arrival to the laboratory, participants were randomly assigned to a computer. After the instructions were given, the assessments began by participants' pressing any number key on the keyboard. The rest of the assessments were self-directed by the computer program. A session lasted an average of about 60 minutes. Participants received a \$10 movie ticket in exchange for their participation during Wave 1, the data reported in this article.

Data Analysis Plan

Evaluation of the psychometric properties of marijuana- and alcohol-related WAT items consisted of three steps, and the analyses were conducted separately for each type of WAT².

IRT assumption checking—We tested two assumptions of the item response theory (IRT) model, unidimensionality and local independence, using both categorical confirmatory factor analysis (CCFA) and IRT methods. In assessing unidimensionality, we fit a one-factor CCFA model, with weighted least-squares with mean and variance adjustment (WLSMV; Muthén, du Toit, & Spisic, 1997), by constraining all WAT items to load onto a single latent factor of implicit alcohol-related (or marijuana-related) associative memory. The model fit was evaluated according to the guidelines of Hu and Bentler (1999). A unidimensional two-parameter logistic (2PL) model was also fitted to the data to evaluate unidimensionality. We evaluated overall model fit by examining Maydeu-Olivares-Joe's M_2 (Maydeu-Olivares & Joe, 2006), a limited-information overall fit statistic, as well as the item-level model fit by assessing the $S-X^2$, an item misfit index (Orlando & Thissen, 2000).

The local independence (LI) assumption was evaluated by checking modification indices (MI) of residual covariances in the one-factor CCFA model and the LD $S-X^2$ statistic (Chen & Thissen, 1997) in the 2PL model. Potential local dependence (LD) is suspected when an excess correlation between a pair of items is observed after controlling for a single latent construct (Thissen & Steinberg, 2009). This suggests a violation of LI, implying to some investigators that the two items ask the exact same questions twice (Varni et al., 2010).

Differential item functioning (DIF)—DIF tests were conducted to test for item invariance across genders and age groups. Two types of DIF were examined: Uniform-DIF implies that the item exhibits a difference in the b parameter between two groups. Non-uniform DIF reflects a difference in the a parameter between two groups. Note that a group difference in DIF is examined while controlling for the overall group difference in the levels of the latent trait. The current study used a one-step Wald test (Cai, Thissen, & du Toit, 2011; Woods, Cai, & Wang, 2013), in which designated anchor items were used to link the latent trait metric for two groups. Anchor items are the items invariant across the groups. We identified the anchored items from a two-steps Wald test (see Langer, 2008, for more detail) before conducting the one-step Wald test. In the one-step Wald test, a model fit was conducted in the following one-step manner: while the mean and standard deviation (SD) of the reference group were fixed to 0 and 1, respectively, the mean and SD of the focal group and the item parameters (the a and b parameters) were estimated at the same time. The item parameters for the designated anchor items were constrained to be equal between the two groups whereas those for the candidate items were free to vary between the two groups. flexMIRT software (Cai, 2012) produces results of the Wald X^2 test for the comparisons of

²Although it may appear that the alcohol- and marijuana-related WATs should be analyzed by a multidimensional IRT approach, we used a unidimensional approach for the following reasons. A noncompensatory multidimensional model was not applicable as we used the same set of items for both substances. The use of overlapping items was inevitable to take into account individual differences in substance-related associative memory, as evidenced in previous studies (see Stacy, Galaif, Sussman, & Dent, 2006; Sussman, Stacy, Ames, & Freedman, 1998). A compensatory multidimensional model was also not relevant since a WAT response related to one substance should not be compensated by one's level on the construct of the second substance. Thus, the two forms of WATs were analyzed separately.

the candidate item(s) between the two groups. In comparisons between male and female participants, we used male participants as the reference group. In comparisons between younger (14–16 years old) and older participants (17 years old and above), the reference group was the young group.

IRT: Item parameter estimation—We evaluated item parameter estimates of any alternative sets of WAT items suggested by the preceding analysis. Both 1PL and 2PL models were fitted to examine if the a parameter should be fixed or varied across the WAT items. Further investigated was the amount of information each WAT item and the total WAT scale provided with respect to the latent trait. The item with a larger amount of information at a given level of the latent trait is considered more reliable. Latent trait scores for alcohol- and marijuana-related implicit associative memory were estimated separately as a function of various WAT item scores, using expected a posteriori (EAP) estimation.

Criterion-related validity—Criterion-related validity coefficients of substance-related WATs were calculated by separately correlating marijuana- and alcohol-related WAT scores with respective drug use frequencies in the past 30 days and 1 year. A nonparametric bootstrap method (Efron, 1979, 1987, Efron & Tibshirani, 1985) was used to estimate the Pearson correlation coefficients (termed r^*) and their confidence intervals, as the assumption of bivariate normality was violated. We used a bias-corrected and accelerated (BCa) procedure (Efron, 1987) to construct confidence intervals for r^* between the following pairs of variables: Alcohol WAT scores and alcohol use in the past 30 days, alcohol WAT scores and alcohol use in the past 1 year, marijuana WAT scores and marijuana use in the past 30 days, and marijuana WAT scores and marijuana use in the past year.

Missing data

The missing data rates on the WAT ranged from 2% to 15% across 18 items, which was not unexpected with open ended item formats. In the IRT analyses, list-wise deletion (LWD) of missing data was implemented. The use of LWD in IRT analyses is supported by several IRT simulation studies that have demonstrated acceptable-to-good parameter estimates of item discrimination and difficulty (Finch, 2008), no bias of uniform DIF detection with missing at random (MAR) data (Robitzsch & Rupp, 2009), and very close results to a complete data set (i.e., a data set with no missing data) in terms of power, type I error rate, and effect sizes in the detection of non-uniform DIF (Finch, 2011).

In the criterion-related validity analysis that was conducted with psychometrically validated WAT items, multiple imputation (MI; Rubin, 1987) was used for missing data in order to obtain unbiased estimates of parameters. Multivariate imputation by chained equations (Buuren, Boshuizen, & Knook, 1999; van Buuren & Oudshoorn, 2000) was used as the specific form of multiple imputation, applying the mice package (van Buuren & Groothuis-Oudshoorn, 2011) in the R statistical environment (R Development Core Team, 2012). This technique has recently gained popularity (Azur, Stuart, Frangakis, & Leaf, 2011) due to its ability to model each variable with missing data, regardless of its distribution (see van Buuren & Groothuis-Oudshoorn, 2011, for detailed procedures).

Results

Participants' demographic variables and their alcohol and marijuana use are summarized in Table 1. To determine whether or not the school cluster variables should be taken into account in subsequent analyses, we computed the design effect (DE) and intra-class correlation (ICC) for alcohol and marijuana use among the average of 42 CHSs. A DE of 2.0 was used as a cut-off (see Muthen & Satorra, 1995). The DE (ICC in parentheses) for alcohol and marijuana use was 1.8 (.016) and 1.5 (.012), respectively. Thus, the school cluster variable was not included in our analyses.

Alcohol-related word association

IRT assumption checking—Both CCFA and 2PL models showed a good fit to the data, indicating unidimensionality of the 18 alcohol-related WAT items. Results from CCFA, conducted using Mplus, version 6.11 (Muthen & Muthen, 2011), revealed fit indices as follows: comparative fit index (CFI) = .944, Tucker-Lewis Index (TLI) = .937, and RMSEA = .045, with a 90% confidence interval of .036 and .054. All of the 18 factor loadings were significant ($p < .01$), ranging from .46 to .75. A 2PL model was fitted using flexMIRT, version 1.0.4.3 (Cai, 2012) and indicated a good model fit (RMSEA = .04). Regarding the local independence (LI) assumption, there were three potential item pairs with local dependency (LD), implied by relatively large values of modification indices (MI) for residual covariances: (1) “friend’s house, feeling a rush” and “weekend, friend’s house, feeling a rush,” (2) “friend’s house, feeling a rush” and “feeling a rush,” and (3) “my bedroom, feeling good” and “my bedroom, feeling relaxed.” In the IRT analysis, no indication of LD item pairs ($LD\ S-X^2 > 10$) was obtained. Only one item (“weekend, party, feeling high”) showed a poor item fit ($p < .0001$). After examining the item contents, we set aside “friend’s house, feeling a rush” and “weekend, party, feeling high,” from a subsequent analysis. The model fit was slightly improved after removing these two items, CFI = .961, TLI = .955, and RMSEA = .037 (90% CI = .027 – .047).

DIF—The DIF test detected only one item exhibiting DIF across gender. The item “weekend, friend’s house, feeling a rush” discriminated more effectively for males ($a = 2.51$) than females ($a = 1.27$; $p < .02$). Thus, the item was excluded from the subsequent analysis. We also dropped “feeling high” since the discrimination parameter for males was substantially low ($a = .78$). With regard to age groups, the item “feeling more relaxed” was the only item with a significant uniform-DIF ($p < .02$), indicating that this item was easier for older participants ($b = 1.62$) than younger participants ($b = 2.31$). However, we expected that some items may be more difficult at younger ages while still being potentially applicable to later changes with increasing age. Thus, the item remained in the analysis.

IRT: Item parameter estimation—Our revised alcohol-related WAT, reduced to 14 items (coefficient alpha = .80), was fitted with both 1PL and 2PL models. Both models indicated a good fit (RMSEA = .04 and .03 for 1PL and 2PL, respectively), with no evidence of a violation of LI. A likelihood ratio test revealed a significant improvement in fit by the 2PL, relative to the 1PL, $G^2(17) = 42.27$, $p < .001$. These results indicated that alcohol-WAT data were reproduced by the model better when the a parameters were

estimated freely (2PL), rather than being constrained to be equal (1PL). Table 2 presents the estimated parameters for both models. The common a parameter in 1PL was 2.02. In 2PL, the a parameters ranged from 1.62 to 2.46, indicating that all 14 alcohol WAT items effectively differentiated the participants across different levels of the latent trait. The b parameters in both models were very similar for each item. For most of the WAT items, moderate to strong levels of implicit alcohol-related memory associations were needed to endorse alcohol-related responses. All of these parameter estimates are graphically represented in the item characteristics curves (ICC; Figure 1).

All ICCs show that the probability of endorsing alcohol-related responses was low for those participants whose latent trait levels were below 1.0. The slopes of most WAT items were steepest throughout the range of the latent levels from about 1.0 to 2.0. These items also provided most information about the latent trait (i.e., most reliable) in this range of the latent level (Figure 2). The amount of information provided by each WAT item was summed to create the TIC (Figure 3), which demonstrates that the alcohol-related WAT is most reliable at moderate-to-high levels of the latent trait. Estimated latent trait scores (Table 3) revealed that those who endorsed one alcohol-related response were estimated to possess an average level of the latent alcohol-related associative memory. As participants endorsed more alcohol-related responses, their latent score increased.

Marijuana-related word association

IRT assumption checking—CCFA showed a good fit of the marijuana-model, CFI = .978, TFI = .975, RMSEA = .043 (90% CI = .034 – .052). IRT analysis revealed an adequate model fit (RMSEA = .05) with no indication of a poor item fit. Thus, the marijuana-related WAT was determined unidimensional. With regard to LI, both CCFA and 2PL detected only one potential LD item pair, “feeling high” and “my bedroom, feeling high.” After reviewing the item contents, we removed the latter item from the analysis.

DIF—For genders, two items were detected as having uniform-DIF: “feeling a rush” ($p = .02$) and “forgetting problems” ($p = .03$). “Feeling a rush” was an easier item for females ($b = 1.59$) to endorse a marijuana-related response than for males ($b = 2.03$) after matching the two groups on the latent trait. Conversely, “forgetting problems” was easier for males ($b = 1.18$) than for females ($b = 1.44$). These two items were removed from the subsequent analyses. As for the age groups, no items exhibited a significant DIF.

IRT: Item parameter estimation—The revised marijuana-related WAT had a total of 15 items (coefficient alpha=.87). Both 1PL and 2PL models fit the data adequately (RMSEA = .05) with no sign of LD item pairs. A likelihood ratio test showed that 2PL had a significantly better fit than 1PL, $G^2(17) = 77.88, p < .001$. Estimated item parameters by both models and ICCs (2PL only) are presented in Table 2 and Figure 1, respectively. In 2PL, the a parameter varied from 1.37 to 3.62 and the b parameters ranged from $-.48$ to 2.05. As shown in Figure 2, IIFs show that “feeling high” was the only item most reliable at the below-average level of the latent trait. Still, TIF illustrates that the marijuana WAT was most reliable around the moderate-to-high levels of the latent trait continuum (Figure 3). Estimated latent trait scores showed that a total marijuana WAT score of three corresponded

to the average level of the latent marijuana-related associative memory. A monotonically increasing relationship was observed between the total WAT scores and the latent trait scores (Table 3).

Criterion-related validity

Table 4 shows the correlations between substance-related WAT scores and drug use frequencies. In both alcohol- and marijuana-related WATs, the participants who endorsed more substance-related responses in WAT tended to report higher frequencies of substance uses both in past year use ($r^* = .44$ [BCa CI = .37 – .51] and .56 [BCa CI = .50 – .61], for alcohol and marijuana, respectively) and past 30 days use ($r^* = .38$ [BCa CI = .30 – .47] and .48 [BCa CI = .42 – .54], for alcohol and marijuana, respectively).

Discussion

The present study was the first to apply a comprehensive psychometric framework, applying IRT approaches, to evaluate psychometric properties of alcohol- and marijuana-related word association tests (WATs) in a sample of ethnically diverse at-risk adolescents. Our results demonstrate that both forms of WAT have good psychometric properties when subjected to comprehensive latent variable and IRT analyses. The discussion below focuses on key findings regarding item and scale properties as well as evidence of construct validity (Messick, 1989, 1995).

Alcohol- and Marijuana-related WAT: Scale Properties

The original 18 WAT items were reduced to 14 and 15 items in alcohol- and marijuana-related WAT, respectively. Items were removed because they exhibited poor item fit (two items each in both WATs), local dependency (LD) issues (one item each in both WATs), or gender bias (one item in the alcohol WAT and two items in the marijuana WAT). Excluding these items improved the revised versions of the substance-related WAT. As expected, both forms of WAT were shown to be unidimensional and most reliable with individuals with moderate-to-high levels of latent alcohol- or marijuana-related associative memory (see Figure 3). A monotonically increasing relationship between the total WAT scores and estimated latent trait scores was observed in both WATs (see Table 3). These results confirmed that the substance-related WATs measure a single construct of substance-related associative memory as it purports to measure. Furthermore, the total alcohol- and marijuana-WAT scores were positively correlated with frequencies of respective past substance use behaviors, providing strong evidence of the criterion-related validity. This finding is in agreement with that of Krank et al. (2010), who reported that self-coded WAT scores were positively associated with past 30 days alcohol use among college students while adding to evidence supporting the use of self-coded scoring procedures (Frigon & Krank, 2009; Krank et al., 2010).

Alcohol- and Marijuana-related WAT: Item Properties

Item discrimination for all items in both WATs showed high discrimination parameters ($a > 1.35$). Among the items, some of the compound cues exhibited very high discrimination parameter values, especially in the marijuana-related WAT. Those compound cues included

“friend’s house, having fun” ($a = 3.62$), “friend’s house, hanging out, feeling good” ($a = 3.14$), and “Friday night, friend’s house, having fun” ($a = 3.20$). A possible explanation for this is that when a positive affective outcome cue was combined with a peer cue to create a compound cue, its item discrimination was further improved. This explanation is consistent with some theories of adolescent substance use that focus on peer influence as a pivotal risk factor (e.g., Hawkins, Catalano, & Miller, 1992; Petraitis, Flay, & Miller, 1995).

With respect to item difficulty, most of the WAT items were most reliable at the moderate-to-high levels of the latent construct. However, a slightly different pattern of results was observed across the two WATs. In the alcohol WAT, all but one item (“Friday night, friend’s house, having fun”) had difficulty parameter estimates greater than 1.0. This indicates that the probability of endorsing alcohol-related responses to the items was lower than .5 for the participants with below moderate levels of latent alcohol-related associative memory. In contrast, the marijuana-related WAT contained a mix of items with moderate-to-high difficulty parameters and items with lower difficulty parameters. This led the marijuana-WAT to cover a wider range of the latent trait continuum than the alcohol-WAT. For example, even among those participants with a lower level of latent marijuana associative memory, half of them endorsed marijuana-related responses to the cues, “feeling high” ($b = -.48$) and “weekend, party, feeling high” ($b = .20$). On the other hand, these two cues were not good ones for alcohol. Both items showed a poor model fit and hence were excluded from the revised alcohol-WAT. Further, the only item with the phrase “feeling high” in the revised alcohol-WAT had a high difficulty parameter estimate (“my bedroom, feeling high”; $b = 1.97$). Hence, we consider “feeling high” as a cue strongly associated with marijuana, particularly at a lower range of the latent trait. This suggests that inclusion of behavior-specific cues may further improve the psychometric properties of substance-related WAT.

Alcohol and Marijuana-related WAT: Unified Concept of Validity

Traditionally, construct validity has been examined by use of multitrait-multimethod matrix (MTMM matrix, Campbell & Fiske, 1959) or confirmatory factor analysis (CFA, Joreskog, 1969; Kenny & Kashy, 1992; Stacy, Widaman, Hays, & DiMatteo, 1985) procedures to gather evidence of convergent and discriminant validity. In contrast, Messick (1989, 1995) suggested six aspects of construct validity, arguing that construct validity of a measurement instrument should be justified by use of the available evidence for a wide variety of aspects of construct validity, including content, substantive, structural, generalizability, external, and consequential aspects. The current study showed that the substance-related WATs exhibited evidence of each of these aspects of construct validity. For example, the *content aspect* is evidenced by the fact that all substance-related WAT items were selected from, or created based on, past studies that reported the utility of WAT as a measure of substance-related implicit associative memory (e.g., Ames et al., 2007). A unidimensional structure of both forms of WATs supports the *structural aspect* of construct validity, indicating that a single construct of alcohol- or marijuana-related implicit associative memory is evaluated in the WAT. Regarding the *substantive aspect*, which requires empirical evidence of response consistencies from data, both forms of WAT revealed good internal consistency across a range of the latent trait. For example, the amount of information (I) exceeded 5.0, which is

equivalent to a reliability estimate of .80, at the underlying latent trait levels between 0 and 2.0 (see Figures 3). In terms of the *generalizability aspect* of construct validity, the DIF tests demonstrated that all items in the revised version of the alcohol- and marijuana-related WATs were invariant across gender and age groups. Finally, although we were not able to investigate any evidence of convergent and discriminant validity in the current study, the obtained evidence of criterion-related validity for both forms of WAT justifies the *external aspect* of construct validity. As reported above, a significant correlation was found between substance-related WAT scores and frequencies of substance use both in the past 30 days and past year. Overall, the current study revealed multiple lines of evidence for the construct validity of the alcohol- and marijuana-related WAT, in accord with the unified concept of construct validity (Messick, 1989, 1995).

Limitations

Several caveats in the present study need to be addressed. First, item invariance was examined only across gender and age groups due to the limited number of samples representing different subgroups (e.g., ethnicity). Thus, the current WAT items might have shown DIF across other subgroups. Future investigations that explore DIF of substance-related WAT items could be conducted across ethnicity and other defining characteristics. Second, because the data were cross-sectional, the direction of the possible causal relationship between WAT scores and past drug use was not inferred. Lastly, drug use behavior was measured via a self-report questionnaire and thus responses are sensitive to demand characteristics and/or social desirability bias. However, under circumstances where adolescents were assured that responses would be confidential, adolescent self-reports have been shown to be accurate (Dent, Sussman, & Stacy, 1997, Donohue, Hill, Azrin, Cross, & Strada, 2007).

Conclusion

Despite these limitations, the present study revealed sound psychometric properties of the alcohol- and marijuana-related WAT. Both forms of WATs were most reliable at moderate-to-high levels of the underlying implicit alcohol- or marijuana-related associative memory. Knowledge of the level of reliability at *different levels of the latent trait* is one of the several fundamental advantages of IRT over traditional psychometric evaluation (e.g., CTT), in addition to advantages of sample invariance, flexibility, and rigor in evaluating differential item functioning. The IRT and construct validation procedures shown here are useful for a wide range of research topics in addiction as well as basic cognitive research on WAT. Although the procedures can be applied to any presumed measures of an underlying trait, it may be surprising that these highly *quantitative* procedures can be effectively applied to responses that are self-generated and open-ended – the responses are essentially *qualitative* in origin. When such responses are amenable to numeric coding, they can be usefully integrated into formal and comprehensive tests of psychometrics and construct validity as revealed here.

Acknowledgments

This research was supported by two grants from the National Institute on Drug Abuse (DA024659-04 and DA023368-06). We wish to thank Amy Custer for her work on this project.

References

- Ames SL, Grenard JL, Stacy AW. Dual process interaction model of HIV-risk behaviors among drug offenders. *AIDS and Behavior*. 2013; 17(3):1–12. [PubMed: 23054037]
- Ames SL, Grenard JL, Thush C, Sussman S, Wiers RW, Stacy AW. Comparison of indirect assessments of association as predictors of marijuana use among at-risk adolescents. *Experimental and Clinical Psychopharmacology*. 2007; 15(2):204–218. [PubMed: 17469944]
- Ames SL, Stacy AW. Implicit cognition in the prediction of substance use among drug offenders. *Psychology of Addictive Behaviors*. 1998; 12(4):272–281.
- Azur MJ, Stuart EA, Frangakis C, Leaf PJ. Multiple imputation by chained equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*. 2011; 20(1):40–49. [PubMed: 21499542]
- Barnett E, Spruijt-Metz D, Unger JB, Rohrbach LA, Sun P, Sussman S. Bidirectional associations between future time perspective and substance use among continuation high-school students. *Substance Use & Misuse*. 2013; 48(8):574–580. [PubMed: 23750661]
- Bradley B, Field M, Mogg K, De Houwer J. Attentional and evaluative biases for smoking cues in nicotine dependence: Component processes of biases in visual orienting. *Behavioural Pharmacology*. 2004; 15(1):29–36. [PubMed: 15075624]
- Cai, L. flexMIRT (TM) version 1.88: A numerical engine for multilevel item factor analysis and test scoring [Computer software]. Seattle, WA: Vector Psychometric Group; 2012.
- Cai, L.; Thissen, D.; du Toit, SHC. IRTPRO: Flexible, multidimensional, multiple categorical IRT modeling [Computer software]. Lincolnwood, IL: Scientific Software International, Inc.; 2011.
- Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*. 1959; 56(2):81–105. [PubMed: 13634291]
- Chassin L, Presson CC, Sherman SJ, Seo D-C, Macy JT. Implicit and explicit attitudes predict smoking cessation: Moderating effects of experienced failure to control smoking and plans to quit. *Psychology of Addictive Behaviors*. 2010; 24(4):670–679. [PubMed: 21198227]
- Chen WH, Thissen D. Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics*. 1997; 22(3):265–289.
- Collins RL, Ellickson PL, Bell RM. Simultaneous polydrug use among teens: prevalence and predictors. *Journal of Substance Abuse*. 1998; 10(3):233–253. [PubMed: 10689657]
- Dent CW, Sussman SY, Stacy AW. The impact of a written parental consent policy on estimates from a school-based drug use survey. *Evaluation Review*. 1997; 21(6):698–712. [PubMed: 10183297]
- Donohue B, Hill HH, Azrin NH, Cross C, Strada MJ. Psychometric support for contemporaneous and retrospective youth and parent reports of adolescent marijuana use frequency in an adolescent outpatient treatment population. *Addictive Behaviors*. 2007; 32(9):1787–1797. [PubMed: 17261356]
- Efron B. Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*. 1979; 7(1):1–26.
- Efron B. Better bootstrap confidence intervals. *Journal of the American Statistical Association*. 1987; 82(397):171–185.
- Efron B, Tibshirani R. The bootstrap method for assessing statistical accuracy. *Behaviormetrika*. 1985; 17:1–35.
- Finch H. The impact of missing data on the detection of nonuniform differential item functioning. *Educational and Psychological Measurement*. 2011; 71(4):663–683.
- Finch H. Estimation of item response theory parameters in the presence of missing data. *Journal of Educational Measurement*. 2008; 45(3):225–245.
- Frigon AP, Krank MD. Self-coded indirect memory associations in a brief school-based intervention for substance use suspensions. *Psychology of Addictive Behaviors*. 2009; 23(4):736–742. [PubMed: 20025382]
- Grenard JL, Ames SL, Stacy AW. Deliberative and spontaneous cognitive processes associated with HIV risk behavior. *Journal of Behavioral Medicine*. 2013; 36(1):1–13. [PubMed: 22108762]

- Grenard JL, Ames SL, Wiers RW, Thush C, Sussman S, Stacy AW. Working memory capacity moderates the predictive effects of drug-related associations on substance use. *Psychology of Addictive Behaviors*. 2008; 22(3):426–432. [PubMed: 18778136]
- Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*. 1993; 12(3):38–47.
- Hawkins JD, Catalano RF, Miller JY. Risk and protective factors for alcohol and other drug problems in adolescence and early adulthood: Implications for substance abuse prevention. *Psychological Bulletin*. 1992; 112(1):64–105. [PubMed: 1529040]
- Houben K, Havermans RC, Wiers RW. Learning to dislike alcohol: Conditioning negative implicit attitudes toward alcohol and its effect on drinking behavior. *Psychopharmacology*. 2010; 211(1):79–86. [PubMed: 20431994]
- Hu L, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*. 1999; 6(1):1–55.
- Huijding J, de Jong PJ, Wiers RW, Verkooijen K. Implicit and explicit attitudes toward smoking in a smoking and a nonsmoking setting. *Addictive Behaviors*. 2005; 30(5):949–961. [PubMed: 15893091]
- Joreskog KG. A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*. 1969; 34(2):183–202.
- Kelly AB, Haynes MA, Marlatt GA. The impact of adolescent tobacco-related associative memory on smoking trajectory: An application of negative binomial regression to highly skewed. *Addictive Behaviors*. 2008; 33(5):640–650. [PubMed: 18222050]
- Kelly AB, Masterman PW, Marlatt GA. Alcohol-related associative strength and drinking behaviours: Concurrent and prospective relationships. *Drug and Alcohol Review*. 2005; 24(6):489–498. [PubMed: 16361205]
- Kenny DA, Kashy DA. Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin*. 1992; 112(1):165–172.
- Krank, MD.; Goldstein, AL. Adolescent changes in implicit cognitions and prevention of substance abuse. In: Wiers, RW.; Stacy, AW., editors. *Handbook of implicit cognition and addiction*. Thousand Oaks, CA: Sage Publications; 2006. p. 439–453.
- Krank MD, Schoenfeld T, Frigon AP. Self-coded indirect memory associations and alcohol and marijuana use in college students. *Behavior Research Methods*. 2010; 42(3):733–738. [PubMed: 20805595]
- Kristjansson E, Aylesworth R, McDowell I, Zumbo BD. A Comparison of Four Methods for Detecting Differential Item Functioning in Ordered Response Items. *Educational and Psychological Measurement*. 2005; 65(6):935–953.
- Langer, M. A reexamination of Lord's Wald test for differential item functioning using item response theory and modern error estimation. Unpublished doctoral dissertation. the University of North Carolina at Chapel Hill; 2008.
- Lord, FM. Applications of item response theory to practical testing problem. Hillsdale, NJ: Lawrence Erlbaum Associates; 1980.
- Maydeu-Olivares A, Joe H. Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*. 2006; 71(4):713–732.
- Messick, S. Validity. In: Linn, RL., editor. *Educational measurement*. New York, NY: Macmillan Publishing; 1989. p. 13–103.
- Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*. 1995; 50(9):741–49.
- Messick S. The standard problem: Meaning and values in measurement and evaluation. *American Psychologist*. 1975; 30(10):955–966.
- Mogg K, Bradley BP. Selective processing of smoking-related cues in smokers: Manipulation of deprivation level and comparison of three measures of processing bias. *Journal of Psychopharmacology*. 2002; 16(4):385–392. [PubMed: 12503841]

- Muthén, LK.; Muthén, BO. Mplus user's guide. Sixth Edition. Los Angeles, CA: Muthén & Muthén; 1998–2011.
- Muthen, B.; Satorra, A. Complex sample data in structural equation modeling. In: Marsden, PV., editor. Sociological methodology. Oxford, England: Blackwell; 1995. p. 267–316.
- Muthen B, du Toit SHC, Spisic D. Robust inference using weighted least squares and quadratic estimating equations in latent variable modeling with categorical and continuous outcomes. 1997 Unpublished technical report.
- Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. *Applied Psychological Measurement*. 2000; 24(1):50–64.
- Petratis J, Flay BR, Miller TQ. Reviewing theories of adolescent substance use: Organizing pieces in the puzzle. *Psychological Bulletin*. 1995; 117(1):67–86. [PubMed: 7870864]
- Preece PF. Three-year stability of certain word-association indices. *Psychological Reports*. 1978; 42(1):25–26.
- Reise SP, Ainsworth AT, Haviland MG. Item response theory: Fundamentals, applications, and promise in psychological research. *Current Directions in Psychological Science*. 2005; 14(2):95–101.
- R Development Core Team. R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing; 2012.
- Robitzsch A, Rupp AA. Impact of missing data on the detection of differential item functioning: The case of Mantel-Haenszel and logistic regression analysis. *Educational and Psychological Measurement*. 2009; 69(1):18–34.
- Rooke SE, Hine DW, Thorsteinsson EB. Implicit cognition and substance use: A meta-analysis. *Addictive Behaviors*. 2008; 33(10):1314–1328. [PubMed: 18640788]
- Rubin, DB. Multiple Imputation for Nonresponse in Surveys. New York, NY: Wiley; 1987.
- Stacy AW. Memory association and ambiguous cues in models of alcohol and marijuana use. *Experimental and Clinical Psychopharmacology*. 1995; 3(2):183–194.
- Stacy AW. Memory activation and expectancy as prospective predictors of alcohol and marijuana use. *Journal of Abnormal Psychology*. 1997; 106(1):61–73. [PubMed: 9103718]
- Stacy, AW.; Ames, SL.; Grenard, JL. Word association tests of associative memory and implicit processes: Theoretical and assessment issues. In: Wiers, RW.; Stacy, AW., editors. *Handbook of implicit cognition and addiction*. Thousand Oaks, CA: Sage Publications; 2006. p. 75–90.
- Stacy AW, Ames SL, Ullman JB, Zogg JB, Leigh BC. Spontaneous cognition and HIV risk behavior. *Psychology of Addictive Behaviors*. 2006; 20(2):196–206. [PubMed: 16784366]
- Stacy AW, Flay BR, Sussman S, Brown KS, Santi S, Best JA. Validity of alternative self-report indices of smoking among adolescents. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*. 1990; 2(4):442–446.
- Stacy AW, Galaif ER, Sussman S, Dent CW. Self-generated drug outcomes in high-risk adolescents. *Psychology of Addictive Behaviors*. 1996; 10(1):18–27.
- Stacy AW, Wiers RW. Implicit cognition and addiction: A tool for explaining paradoxical behavior. *Annual Review of Clinical Psychology*. 2010; 6:551–575.
- Stacy AW, Widaman KF, Hays R, DiMatteo MR. Validity of self-reports of alcohol and other drug use: A multitrait-multimethod assessment. *Journal of Personality and Social Psychology*. 1985; 49(1):219–232. [PubMed: 4020615]
- Sussman S, Stacy AW, Ames SL, Freedman LB. Self-reported high-risk locations of adolescent drug use. *Addictive Behaviors*. 1998; 23(3):405–411. [PubMed: 9668939]
- Thissen, D.; Steinberg, L. Item response theory. In: Millsap, R.; Maydeu-Olivares, A., editors. *The Sage handbook of quantitative methods in psychology*. London: Sage Publications Ltd; 2009. p. 148–177.
- van Buuren S, Groothuis-Oudshoorn K. MICE: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011; 45(3):1–67.
- van Buuren, S.; Oudshoorn, K. Multivariate imputation by chained equations: MICE V1.0 User's manual. Leiden: TNO Prevention and Health; 2000. (TNO Report PG/VGZ/00.038).

- van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Statistics in Medicine*. 1999; 18:681–694. [PubMed: 10204197]
- Varni JW, Stucky BD, Thissen D, DeWitt EM, Irwin DE, Lai J, Yeatts K, DeWalt DA. PROMIS pediatric pain interference scale: An item response theory analysis of the pediatric pain item bank. *The Journal of Pain*. 2010; 11(11):1109–1119. [PubMed: 20627819]
- White HR, Labouvie EW. Towards the assessment of adolescent problem drinking. *Journal of Studies on Alcohol*. 1989; 50(1):30–37. [PubMed: 2927120]
- Wiers RW, Stacy AW. Implicit cognition and addiction. *Current Directions in Psychological Science*. 2006; 15(6):292–296.
- Woods CM, Cai L, Wang M. The Langer-improved Wald test for DIF testing with multiple groups: Evaluation and comparison to two-group IRT. *Educational and Psychological Measurement*. 2013; 73(3):532–547.

Figure 1.

Item characteristic curves for the revised alcohol (solid) and marijuana (dashed) WAT items. The x-axes show the level of theta (latent implicit alcohol- or marijuana-related associative memory), with 0 representing the average level of theta. The y-axes show the proportion of alcohol- or marijuana-related responses generated to each WAT cue.

Figure 2.

Item information functions for the revised alcohol (solid) and marijuana (dashed) WAT items. The x-axes show the level of theta (latent implicit alcohol- or marijuana-related associative memory), with 0 representing the average level of theta. The y-axes show the amount of information (I), an index of how accurately each item contributes to estimate the latent trait at a given level of theta.

Figure 3. Test information functions (TIFs) for the revised alcohol (solid) and marijuana (dashed) WATs. The x-axis shows the level of theta (latent implicit alcohol- or marijuana-related associative memory), with 0 representing the average level of theta. The y-axis shows the amount of information (I), an index of how accurately each form of WAT estimates the latent trait at a given level of theta. I = 5, 10, and 20 is equivalent to a reliability estimate of .80, .90, and .95, respectively.

Table 1

Demographic variables and Substance Use

	n	%	Mean(SD)
Gender			
Female	340	43.9	
Male	435	56.1	
Age			16.6(1.1)
14	17	2.2	
15	79	10.2	
16	215	27.7	
17	278	35.9	
18	121	15.6	
18+	19	2.5	
Not reported	46	5.9	
Ethnicity			
Hispanic	484	62.5	
Mixed	145	18.7	
Non-Hispanic White	97	12.5	
Black	25	3.2	
Other	24	3.1	
Alcohol Use			
Past 30 days	421 ^a	54.8	2.01 ^b (1.6)
Past 1 year	501 ^a	80.0	2.65 ^b (2.7)
Marijuana Use			
Past 30 days	355 ^a	46.6	3.48 ^b (3.1)
Past 1 year	460 ^a	60.6	4.07 ^b (3.9)

Note.

^aThe number of people who reported they used alcohol or marijuana at least once in the past 30 days or past year.

^bPast use of alcohol and marijuana was assessed with an 11-point rating scale: 1 = 0 time, 2 = 1–10 times, 3 = 11–20 times, 11 = 91+ times.

Table 2

Item parameter estimates for 1PL and 2PL in the Alcohol- and Marijuana-related WAT

Item	Label	Alcohol-related WAT				Marijuana-related WAT			
		a		b		a		b	
		1PL	2PL	1PL	2PL	1PL	2PL	1PL	2PL
1	feeling hyper	2.02	1.62	1.48	1.64	2.17	1.37	1.63	2.05
2	feeling good	2.02	2.12	1.46	1.44	2.17	2.15	0.81	0.83
3	feeling high	-	-	-	-	2.17	1.77	-0.44	-0.48
4	feeling more relaxed	2.02	2.10	1.98	1.96	2.17	2.09	1.05	1.07
5	forgetting problems	2.02	1.54	1.31	1.50	2.17	1.87	0.90	0.96
6	feeling a rush	2.02	1.98	1.66	1.69	-	-	-	-
7	my bedroom, feeling good	2.02	2.08	2.15	2.13	2.17	2.39	1.42	1.37
8	my bedroom, feeling more relaxed	2.02	1.84	2.08	2.19	2.17	2.20	1.23	1.23
9	friend's house, feeling hyper	2.02	1.91	1.22	1.25	2.17	2.42	1.25	1.21
10	my bedroom, feeling high	2.02	1.92	1.91	1.97	-	-	-	-
11	friend's house, feeling a rush	-	-	-	-	2.17	1.89	1.31	1.40
12	friend's house, having fun	2.02	2.17	1.26	1.23	2.17	3.62	1.27	1.12
13	friend's house, hanging out, feeling good	2.02	2.46	1.16	1.09	2.17	3.14	0.98	0.91
14	Friday night, my bedroom, feeling hyper	2.02	2.23	1.38	1.34	2.17	2.51	1.37	1.31
15	Weekend, party, feeling high	-	-	-	-	2.17	1.93	0.19	0.20
16	Friday night, friend's house, having fun	2.02	2.45	0.92	0.87	2.17	3.20	0.98	0.90
17	Weekend, my bedroom, feeling more relaxed	2.02	1.75	2.02	2.17	-	-	-	-
18	Weekend, friend's house, feeling a rush	-	-	-	-	2.17	2.33	1.29	1.26

Note. a = item discrimination parameter, b = item difficulty parameter, 1PL = one-parameter logistic model, 2PL = two-parameter logistic model, and the dash (-) indicates the cue was dropped from the revised WAT.

Table 3

Alcohol- and Marijuana-related WAT Scores, Latent Trait Scores (EAP) and their standard deviations

WAT Score	Alcohol		Marijuana	
	EAP	SD	EAP	SD
0	−0.72	0.71	−1.12	0.65
1	0.00	0.51	−0.51	0.51
2	0.42	0.40	−0.10	0.41
3	0.71	0.35	0.20	0.34
4	0.93	0.32	0.43	0.30
5	1.12	0.30	0.62	0.27
6	1.29	0.29	0.78	0.25
7	1.46	0.29	0.92	0.24
8	1.62	0.29	1.06	0.24
9	1.79	0.29	1.20	0.24
10	1.96	0.30	1.33	0.25
11	2.14	0.32	1.48	0.26
12	2.35	0.34	1.65	0.29
13	2.60	0.38	1.86	0.33
14	2.93	0.45	2.12	0.39
15			2.45	0.47

Note. EAP = Expected A Posteriori.

Table 4

Bivariate Correlations (BCa 95% CI) between the Substance-related WAT Scores and Frequencies of Past Drug Use

	Alcohol	Marijuana
Past 30 days	.38 (.30 – .47)	.48 (.42 – .54)
Past 1 year	.44 (.37 – .51)	.56 (.50 – .61)

Note. BCa 95% CI = Bias-corrected and accelerated bootstrap confidence interval.