Running Head: PRBQ Psychometrics

Abstract

The Posttrauma Risky Behaviors Questionnaire (PRBQ) assesses extent of engagement in posttrauma reckless and self-destructive behaviors (RSDBs). Given PRBQ's recent development with limited psychometric investigations, we used Item Response Theory to examine (1) item analysis; (2) person fit; and (3) differential item functioning (DIF) across gender-based groups and two different samples. One sample included 464 participants reporting potentially traumatic experiences (Mechanical Turk [MTurk], recruited online), and the other sample included 171 trauma-exposed women reporting current intimate partner violence (IPV) and substance use (recruited in-person). All PRBQ items contributed to the RSDB scale, and all PRBQ items and the PRBQ scale provided maximum information for high levels of the RSDB latent trait. Seven and 11 items were conceptualized as low information (LI) items in the MTurk and IPV samples, respectively. Eight MTurk participants' responses did not fit the overall pattern of responses as expected. Seven items were flagged for DIF between the two samples, and eight items were flagged for DIF between men and women in the MTurk sample. However, all effect sizes were < 8%. Conclusively, results suggest good psychometric properties for the PRBQ and support its use to compare RSDBs across different samples and gender-based groups.

Keywords. Posttrauma Risky Behaviors Questionnaire; psychometrics; item response theory; differential item functioning; gender.

Introduction

Evidence suggests that trauma-exposed individuals are more likely to engage in reckless and self-destructive behaviors (RSDBs) such as gambling (Roberts et al., 2017), problematic media/technology use (Contractor, Frankfurt, et al., 2017), disordered eating behaviors (Brewerton, 2007), substance use (Clark et al., 2001), aggressive behaviors (Lusk et al., 2017), and self-injurious/suicidal acts (Spitzer et al., 2020; Stein et al., 2010). Individuals who engage in posttrauma RSDBs report functional impairment and detrimental health such as more depression severity (Contractor, Weiss, et al., 2017; Sommer et al., 2020). Further, individuals who report posttrauma RSDBs (e.g., substance use) and posttrauma distress are more difficult to treat with clinical interventions and demonstrate poorer treatment outcomes/adherence (McCauley et al., 2012; Schäfer & Najavits, 2007). Indeed, clinical interventions for posttrauma symptoms have incorporated risk reduction/management techniques to mitigate RSDBs (Becker & Zayfert, 2001; Danielson et al., 2020; Senn et al., 2017).

Theoretical viewpoints help explain links between trauma and posttrauma RSDBs. For instance, the *emotion dysregulation* viewpoint suggests that trauma-exposed individuals may engage in RSDBs to reduce negative affect and increase positive affect (Ben-Zur & Zeidner, 2009; Cooper et al., 2000; Marshall-Berenz et al., 2011). Further, from a *cognitive* viewpoint, trauma-exposed individuals may experience limited attention, cognitive resources, and information processing abilities, which may reduce their capacity for adaptive decisions, and thus, increase RSDBs (Ben-Zur & Zeidner, 2009). From a *disinhibition* viewpoint, trauma-exposed individuals, especially those reporting chronic childhood traumas and posttrauma psychopathology, may engage in impulsive RSDBs (Braquehais et al., 2010) when encountering situations perceived as rewarding and/or distress-reducing (Casada & Roache, 2005). Finally, the

impaired risk recognition viewpoint suggests that the experience of trauma makes it difficult to accurately perceive and recognize risks in situations (Gidycz et al., 2006); such impaired judgement may contribute to RSDBs.

To comprehensively screen posttrauma RSDBs, a 16-item Posttrauma Risky Behaviors Questionnaire (PRBQ) was developed following recommended guidelines (e.g., Furr, 2011; Germain, 2006). The PRBQ examines posttrauma RSDBs covered by the Diagnostic and Statistical Manual of Mental Disorders 5th edition (*DSM-5*) E2 criterion for posttraumatic stress disorder (PTSD); examples include substance use, dangerous driving, and suicidal behavior (American Psychiatric Association, 2013). Notably, the *DSM-5* does not outline all posttrauma RSDBs intended to be captured by E2. Thus, in developing the PRBQ, the authors reviewed the literature and consulted with experts to identify additional RSDBs that are prevalent among individuals reporting trauma experiences such as disordered eating and gambling (see Contractor et al., 2020). Hereby, the PRBQ is not limiting to a symptom criterion of PTSD, which has changed across *DSM* versions; and the PRBQ can be used to examine posttrauma RSDBs spanning diverse research questions, contexts, and diagnostic constructs (not specific to PTSD). Further, existing RSDB measures, such as the Risky, Impulsive, and Self-Destructive Measure (Sadeh & Baskin-Sommers, 2017) or the Risky Behaviors Questionnaire (Weiss et al., 2016) could be administered with a trauma measure to examine RSDBs in the context of trauma; however, such RSDB measures were not targeted to and do not measure extent of engaging in posttrauma RSDBs (e.g., some measures examine functionality of RSDBs), and are not brief. Notably, RSDBs in the context of trauma are different in functionality, outcomes, and characteristics (beyond unique reasons linking trauma and posttrauma RSDBs as outlined earlier; Kerig, 2019). For instance, in the short-run, engaging in posttrauma RSDBs may serve to avoid,

escape from, or distract from trauma-related aversive emotional/cognitive states (Cooper et al., 2000; Pat-Horenczyk et al., 2007); however, such engagement in posttrauma RSDBs may have negative consequences in the long-run (Ben-Zur & Zeidner, 2009). Further, posttrauma symptoms such as arousal or dissociation, may interfere with one's ability to avoid or escape from risky situations (Noll & Grych, 2011), increasing engagement in posttrauma RSDBs. Also, engaging in posttrauma RSDBs may help individuals reproduce the arousal linked to their trauma (Van der Kolk et al., 1985), which may explain elevated impulsivity and sensation-seeking among trauma-exposed individuals (Contractor et al., 2016; Roley et al., 2017).

Scores derived from the PRBQ have been subjected to preliminary psychometric investigations (Contractor et al., 2021; Contractor et al., 2019; Contractor et al., 2020). To-date, the PRBQ has demonstrated measurement of a unitary RSDB construct, high internal consistency (e.g., Cronbach's alpha = .94), good convergent validity with distinct measures of RSDBs (including with PTSD's E2 criterion scores assessed by the PTSD Checklist for DSM-5 [PCL-5]), and good incremental validity (e.g., PRBQ score predicted PTSD's E2 criterion score assessed by the PCL-5 above and beyond the Risky Behaviors Questionnaire; Contractor et al., 2020). Despite these initial strong psychometric indicators for PRBQ scores, there are a few unexplored areas. For instance, it would be helpful to examine if the existing response categories for the PRBQ items are optimal and useful (e.g., few participants in the samples used for the PRBQ validation paper endorsed "frequently and "very frequently" responses; Contractor et al., 2020); to identify groups with levels of RSDB latent trait that are most appropriate for PRBQ administration; to determine how much information each item adds to the PRBQ; and to examine differential item functioning (DIF) of PRBQ items across groups differentiated by demographic, clinical, or social characteristics.

To this end, the current study examined PRBQ's item parameter estimates based on Item Response Theory (IRT) which provides information on item performance in relation to an underlying latent construct assessed by a scale (Lord, 1980). IRT is an especially important tool for new measures such as the PRBQ because such IRT-based statistics help measure (a) amount of information contributed by each item, (b) extent to which the items distinguish individuals with vs. without clinically significant levels of RSDBs, (c) extent to which the various categories on a given scale of measurement are being endorsed for a given item, (d) whether some response patterns are misfitting the expected response pattern (i.e., whether participants respond as expected or if the PRBQ triggers unexpected response patterns), and (e) whether two individuals from two different groups who have the same level of RSDB latent trait have significantly different probabilities of endorsing a particular PRBQ item (i.e., if PRBQ items are performing differently towards a focal group relative to a reference group; Camilli & Shepard, 1994; Cole, 1993). IRT analyses of dichotomous/polytomous response data provide item- and individual-level information such as item analysis, person fit, and differential item functioning (DIF).

First, item characteristics derived from item analysis using the graded response model (GRM) include thresholds (b) and discrimination (a) parameters. Thresholds indicate if each PRBQ item is measured by appropriate response categories; they identify items where the respondents are not choosing certain response categories or are choosing only extreme response categories. Information relating the probability of endorsing items and item parameters is portrayed figuratively with category response curves (CRCs). Further, item discrimination parameters are the counterpart of slopes (steepness) in the general linear model (Baker & Kim, 2004); they indicate an item's precision in measuring across varying RSDB latent trait levels and an item's contribution to the RSDB latent trait. For PRBQ items with a higher discrimination

parameter, the probability of choosing a response category will change more rapidly corresponding to changes in RSDB latent trait levels; these items also contribute more to the RSDB latent trait. Item and test information curves (IIC and TIC) indicate the amount of information given by each item and the overall test, respectively.

Second, through person fit estimates, we can identify atypical/aberrant response patterns indicating misfit. These person fit estimates can detect whether a person with a given RSDB latent trait level responds appropriately/atypically to a PRBQ item. This idea is also referred to as traitedness - magnitude of consistency between a person's expected pattern of responses (i.e., indication of behavior) and theoretically anticipated response patterns on a particular dimension (Reise & Waller, 1993). Atypical responses may be attributed to psychological processes characterized by guessing, cheating, or different item interpretations (Tendeiro et al., 2016).

Finally, DIF analysis (Raju, 1988, 1990) identifies item bias, i.e., whether two individuals with the same RSDB latent trait level belonging to different groups have similar or different probabilities of choosing a response category on PRBQ items. Hereby, DIF analysis can help to statistically/visually determine what specific aspects of the PRBQ item are biased (i.e., items or response categories). For the current study, we first examined DIF between men and women. Indeed, research indicates that engagement in the different types of RSDBs differs across gender; men report more substance use (Brady & Randall, 1999; Nolen-Hoeksema, 2004), illegal behaviors, risky driving, and physical aggression (Pat-Horenczyk et al., 2007), while women report more dysfunctional eating (Pat-Horenczyk et al., 2007) and non-suicidal self-injury (Bakken & Gunter, 2012). Broadly, men have been found to engage in more RSDBs (Archer, 2004; Contractor & Weiss, 2019; Grant et al., 2015). For the current study, we also examined DIF between two samples differing on some socio-demographic characteristics and method of

recruitment. In this regard, engagement in RSDBs has been shown to differ across samples differing on socio-demographic and clinical characteristics: race-ethnicity defined groups differ in engagement in risky health behaviors (Daw et al., 2017); groups reporting interpersonal traumas report more RSDBs (Contractor et al., 2018); and lower socioeconomic status is associated with engagement in multiple RSDBs (Meader et al., 2016). For the current study, one sample included participants recruited from Amazon's Mechanical Turk (hereby referenced as MTurk sample) and the other sample included women recruited from the community who reported experiencing intimate partner violence (IPV) and substance use (hereby referenced as the IPV sample). The reference group, i.e., the MTurk sample, had an almost uniform distribution of income, were mostly employed, and recruited online from anywhere in the US. The focal group, i.e., the IPV sample, was recruited in-person and locally from Providence County in Rhode Island, comprised of only women, and had a higher percentage of individuals reporting low income and unemployment. DIF analyses would indicate if valid comparisons of RSDBs across the two different samples can be made; RSDB latent trait would be compared across the groups only if PRBQ items did not exhibit DIF across the compared groups (i.e., no bias towards a group). This is especially important given concerns of data quality for MTurk (Chmielewski & Kucker, 2020).

In summary, applying IRT and DIF, we examined and compared psychometric properties of PRBQ scores in and across two different samples. Specifically, we conducted item analysis (threshold and discrimination parameters) of PRBQ items in each sample; examined person fit estimates in each sample; and conducted DIF analysis to investigate whether the PRBQ items measured the RSDB latent trait identically across men and women (in the MTurk sample) and across the two different samples (i.e., MTurk and IPV). Broadly, stronger psychometric

properties are established for an instrument that can measure the latent trait identically across different samples.

Method

Procedure and Participants

MTurk Sample. Study procedures were approved by the [redacted] Institutional Review Board (IRB). We recruited adult participants (≥18 years) from Amazon's Mechanical Turk (MTurk), which is a crowd-sourcing platform (Paolacci et al., 2010). MTurk's subject pool (1) is diverse compared to traditional internet-recruited samples; (2) represents the US population in demographic characteristics such as gender distribution and mean age; (3) generates reliable data (Buhrmester et al., 2011; Mischra & Carleton, 2017; Shapiro et al., 2013; U.S. Census Bureau, 2016a, 2016b); and (4) has demonstrated utility for trauma research by capturing individuals with posttrauma symptoms in a cost and time-effective manner (Engle et al., 2020; van Stolk-Cooke et al., 2018). The current study was described as a 45-60-minute survey to develop a measure of posttrauma RSDBs. Eligible adult participants (1) lived in North America, (2) reported English fluency, and (3) endorsed trauma screened with the Primary Care Posttraumatic Stress Disorder Screen for *DSM-5* (PC-PTSD-5; Prins et al., 2015). Eligible participants who provided informed consent and completed the Qualtrics survey validly were given $1.25.

Of the obtained 891 responses, 47 responses of 18 participants who attempted the survey multiple times were excluded (remainder *n* = 844). We further excluded 150 participants who did not meet eligibility criteria, 122 participants who did not pass all four validity checks to ensure attentive responding and comprehension (e.g., "please click on the little blue circle on the bottom of the screen; "I am paid paid biweekly by leprechauns;" Meade & Craig, 2012; Thomas & Clifford, 2017), 97 participants who missed data on all measures, and 11 participants who did not

endorse ==one/worst potentially traumatic event== on the Life Events Checklist for *DSM-5* (LEC-5;

Weathers et al., 2013). This resulted in a sample size of 464 participants. For PRBQ items, 10%

of the responses were missing completely and were therefore discarded. Therefore, the final

sample included 418 participants. See Table 1 for detailed demographic information.

　　　　IPV Sample. Study procedures were approved by the [redacted] IRB. Data were collected

as part of a larger ongoing study examining day-level relations among PTSD, emotion

dysregulation, substance use, and HIV/sexual risk in a sample of community women currently

experiencing IPV and substance use. Participants were recruited from Providence County in

Rhode Island using posters, brochures, and flyers posted in community establishments and

internet forums. Inclusion criteria included: (a) female gender, (b) ≥18 years; (c) English

speaking; (d) involvement in a heterosexual intimate relationship with the presence of physical

and/or sexual victimization; and (e) and use of drugs/alcohol ==(required to endorse at least one==

==episode of alcohol use, illicit drug use, or licit drug misuse in the past month).== Exclusion criteria

included: (a) current mania/psychosis; (b) self-reported pregnancy; (c) color-blindness; (d)

cardiovascular disease; and (e) residence in a shelter/group home. Eligible participants were

provided with study information, following which written informed consent was obtained. The

study was completed in four parts: (a) an initial session, (b) an experimental session, (c) 30 days

of thrice daily experience sampling methodology, and (d) a follow-up session. Data from the

initial session were used in the current study; in this session, structured diagnostic interviews,

self-report measures, and a protocol for developing emotion induction scripts were administered

(Lang & Cuthbert, 1984; Lang et al., 1983). Remuneration was $40 for participation in the initial

session.

The sample of 171 individuals reported a Criterion A trauma as determined by the Structured Clinical Diagnostic Interview for *DSM-5* (Brodey et al., 2016) and affirmative responses to either one of the two screening questions ("In the past 6 months, did your partner do anything to physically hurt you, such as push or shove you, grab you, or punch or hit you?" or "In the past 6 months, did your partner make you do something sexual that you didn't want to do, such as pressuring or forcing you to do something sexual when you didn't want to?"). See Table 2 for detailed demographic information. Data were not missing for the PRBQ items.

Measures

Life Event Checklist for DSM-5 (LEC-5; Weathers et al., 2013)*. Administered to both samples, the LEC-5 is a 17-item self-report measure assessing exposure to potentially traumatic experiences. Participants respond to each item with these 6 response options: "Happened to me," "Witnessed it," "Learned about it," "Part of my job," "Not sure," or "Doesn't apply." An 18th item asked about the most potentially traumatic event. A positive response to one of the first four response options for Items 1-16 was considered a potential trauma endorsement.

Structured Clinical Interview for DSM-5 (SCID-5). Administered to the IPV sample, a computerized version of the SCID-5 was used to ascertain trauma exposure and establish current psychiatric diagnoses (First & Williams, 2016). The SCID-5 severity scales have demonstrated good psychometric properties (Shankman et al., 2018) and there is evidence of moderate to excellent inter-rater reliability across major diagnostic categories (Osório et al., 2019).

Posttrauma Risky Behaviors Questionnaire (PRBQ; Contractor et al., 2020). Administered to both samples, the PRBQ is a 16-item self-report measure of posttrauma RSDBs referencing the past month. Fourteen items measure the extent of engaging in specific posttrauma RSDBs on a response scale ranging from 0 (*never*) to 4 (*very frequently*). Two supplemental

items measure functional impairment and associations between posttrauma RSDB frequency and onset of the worst trauma. The PRBQ scores have good psychometric properties (Contractor et al., 2019; Contractor et al., 2020). In the present study, ordinal Cronbach alphas were 0.97 and 0.93 for the MTurk and IPV samples, respectively; and ordinal omegas were 0.98 and 0.94 for the MTurk and IPV samples, respectively.

Statistical Analyses

All analyses were conducted using R (R Core Team, 2019). Principal components analysis was conducted on the 14 PRBQ items using the R package *factoextra* (Kassambara & Mundt, 2020). According to Reckase (1979) and Reeve et al. (2007), the first factor should explain at least 20% of the variance in order to meet the unidimensionality assumption. Additionally, if the ratio of the first factor eigenvalue to the second factor eigenvalue < 3, such an estimate also indicates unidimensionality (Morizot et al., 2007). Monotonicity in PRBQ data was examined using the R package *Mokken* (van der Ark, 2007) based on Mokken scale analysis (Mokken, 1971). The monotonicity plots were visually examined to ensure that the items were increasing monotonically. Local item dependence was tested by examining whether model fit of a single factor confirmatory factor analysis (CFA) would increase by adding error correlations across items. This error correlation has to be < 0.20 to establish local item independence.

For IRT analyses, we used GRM (Samejima, 1969) for the PRBQ data because the item categories have a meaningful increasing order. Both constrained and unconstrained GRMs were fitted to the data using the R package *ltm* (Rizopoulos, 2006). We classified items as being low informative (LI) when the CRC peak for any response category in the PRBQ item was under another CRC throughout the entire latent trait space; this meant that not all five response category options were utilized by participants. Items were also classified as LI when the

discrimination parameter of the item was < 0.75 (Baker, 2001). Items for which participants utilized all response categories and with discrimination parameters > 0.75 were categorized as being moderate to highly informative (MHI). Baker (2001) classified item discrimination parameters < 0.65 as moderate to low because item discrimination parameters are analogous to factor coefficients where items with higher discrimination parameters are able to better differentiate between individuals (Reiss et al., 2011). Erring on the more conservative side, we considered PRBQ items that had discrimination parameter estimates < 0.75 (also lower factor coefficients) and/or unutilized response categories as providing 'lower' information than the other items. This being said, we acknowledge the potential arbitrary nature of the existing classifications and the lack of consensus on what defines "low" while inferring results and outlining recommendations.

Further, Emon's (2008) polytomous extension of van der Flier's (1980, 1982) U3 statistic was computed to detect inconsistent, aberrant, or misfitting patterns of PRBQ item scores using the package *PerFit* (Tendeiro et al., 2016). Lastly, a DIF model (DFIT; differential functioning of items and tests; Oshima & Morris, 2008) was fit to the data to compare PRBQ item functioning across men vs. women in the MTurk sample, and across the MTurk vs. IPV samples using the R package *lordif* (Choi et al., 2011). Uniform DIF is indicated when the latent trait is consistently lower for one group compared to another, while the magnitude/order of the difference between the latent traits differ across groups for non-uniform DIF (Walker, 2011). The magnitude of these DIF estimates was determined using the McFadden's R^2 estimate which was classified as low when the value <10%; the difference in the group membership accounted for <10% of variation in the response patterns for the same level of the latent trait.

Results

Unidimensionality, Monotonicity, and Local Item Dependence

Principal components analysis revealed that the first factor explained 57.14% of the variance and the subsequent factors explained < 7% of the variance. The ratio of the first factor eigenvalue to the second factor eigenvalue was 8.08. Thus, results suggested unidimensionality of the 14-item PRBQ factor. Further, none of the error correlations were greater than 0.20; there was no correlation between the items after accounting for variation in the items due to variation in the factor scores. Thus, local item independence was established. Lastly, all the items were monotonically increasing; choosing a higher response category on a PRBQ item indicated greater engagement in posttrauma RSDBs. Therefore, the assumptions of unidimensionality, monotonicity, and local item independence for IRT analyses were met.

MTurk Sample

Item Analysis. Both unconstrained and constrained GRMs had comparable fit as evidenced by similar Akaike information criterion (AIC) and Bayesian information criterion (BIC) values (-2[log likelihood] = 69.12, *p* <.001). Therefore, we retained the unconstrained GRM. The unconstrained GRM revealed the following decreasing order of item discriminations: 13, 11, 7, 9, 12, 3, 14, 5, 10, 2, 8, 4, 6, and 1 (see Table 3). The CRC patterns are given in Figure 1. All items had discrimination parameter values > 0.75. However, all response categories were utilised for items 3, 5, 8, 9, 10, 11, and 13 only; therefore, these items were classified as MHI. Items 1, 2, 4, 6, 7, 12, and 14 were classified as LI.

ICCs (see Figure 2, left panel) indicated that PRBQ Item 9 contained the least information of all the items. The information provided by all PRBQ items together was 79.73 whereas the information provided by all PRBQ items except Item 9 was 71.93 (this item

contributed 9.70% of the information to the PRBQ measure). Although this is less than the information provided by the other items, this value is not negligible. Therefore, it was deemed that all the PRBQ items contributed some information to the scale. TIC (see Figure 2, right panel) indicated that > 58% of the information was contained between latent trait RSDB values of 1 and 3. Additionally, 96.96% of the information was measured for latent trait level ≥ 0. Thus, the PRBQ provided more information on participants who endorsed a greater extent of engagement in RSDBs. In fact, all items provided maximum information between the RSDB latent trait levels of 1 and 3.

Person Fit Estimates. Person fit statistics (U3) revealed that 8 participants did not fit the pattern of responses as expected; of these participants, three identified as male, 3 identified as female, 1 identified as male-to-female transgender, 1 identified as female-to-male transgender, and their ages ranged from 23 to 72. Variables such as relationship status, ethnicity, employment, and income levels did not explain the reason for the misfit. As an explanation for the misfit, all 8 respondents endorsed all but one PRBQ item with a response category of 1; they endorsed PRBQ items 3, 5, 7, 9, 13, and 14 with a response category of 2 or 3 (but not higher). Thus, there was no particular reason discernible from the data for these response patterns to be labelled as misfit.

DIF Across Gender Groups. PRBQ items 3, 6, 8, and 13 were flagged as exhibiting DIF. The average RSDB latent trait for men and women was 1.57 and 1.40, respectively. The Cohen's *d* standardized mean difference effect size was 0.27 which is considered small (Cohen, 1988). Thus, the men were only slightly higher on the RDSB latent trait than the women. Even when the LR χ^2 test for non-uniform DIF [i.e., χ^2_{13}) and $\Pr(\chi^2_{23})$] or uniform DIF [$\Pr(\chi^2_{12})$] flagged items for possible DIF, McFadden's R^2 change effect size remained <8% for all the tests indicating a small effect size.

IPV Sample

Item Analysis. Both unconstrained and constrained GRMs had similar fit as evidenced by their similar AIC and BIC values (-2[log likelihood] = 68.89, p <.001). Therefore, we retained the unconstrained GRM. The unconstrained GRM revealed the following decreasing order of item discriminations: 11, 12, 9, 13, 10, 5, 7, 14, 8, 3, 6, 2, 4, and 1 (see Table 3). The CRC patterns are given in Figure 3. Although all item discrimination parameters were > 0.75, all response categories were utilized only for items 9, 10, and 11 (classified as MHI). The rest of the 10 items were classified as LI.

ICCs (see Figure 4, left panel) indicated that items 1, 4, 2, and 6 contained the least information of all the items. The information provided by all items together was 53.26 whereas the information provided by all PRBQ items except these items was 46.18. Items 1 and 4 contributed 3% each, and items 2 and 6 contributed 3.5% each indicating that each of these items contributed very little information to the entire PRBQ instrument. TIC (see Figure 4, right panel) indicated that > 68% of the information was contained between latent trait RDSB values of 0.50 and 3.50. Additionally, 86.50% of the information was measured for latent trait level ≥ 0. This indicated that the PRBQ provided more information on participants with a greater extent of engagement in RDSBs. For this sample, all items provided maximum information for RSDB values between 1 and 3, except items such as 9 and 10 (physically and verbally aggressive behaviors), which provided maximum information at lower levels of RSDB values such as 0 to 2.5 and -1 to 2, respectively.

Person Fit Estimates. Person fit statistics (U3) revealed that all participants fit the pattern of responses as expected.

DIF Analyses Across MTurk and IPV Samples

DIF analysis was conducted to compare the MTurk and IPV samples. Eight out of 14 PRBQ items were flagged for DIF (items 1, 2, 4, 5, 7, 9, 10, and 12). The average RSDB for the IPV and MTurk samples were 1.86 and 1.49, respectively. The Cohen's d standardized mean difference effect size was 0.53 which is considered medium (Cohen, 1988). McFadden's R^2 change was <3.5% for all tests. Therefore, the amount of DIF between the groups is considered low. In sum, although the two groups differed in their RSDBs latent trait levels (IPV vs. MTurk samples exhibited more RSDB latent trait levels), there was no substantial difference in how individuals with the same amount of RSDBs belonging to the IPV or MTurk samples answered the PRBQ items.

Discussion

The current study extends psychometric investigations of the PRBQ scores using an Item Response Theory (IRT) framework in two different samples (MTurk and IPV). Broadly, study results indicate that PRBQ items have strong psychometric properties across both samples. Considering item discrimination parameters, all PRBQ items across both samples contributed information to the understanding and measurement of the RSDB construct. In other words, all PRBQ items were able to distinguish between participants with varying RSDB latent trait levels; the probability of choosing a response category for all PRBQ items changed corresponding to changes in the RSDB latent trait levels. This being said, across samples, PRBQ items 9 (physically aggressive behaviors), 11 (property destruction), and 13 (deliberately injuring self without intent to kill oneself) were most discriminating. Perhaps, being indicative of mental health impairment and/or poorer functioning (Armour et al., 2020; Jacobson et al., 2008; McFall et al., 1999; Selby et al., 2012) and commonly having anger regulation/expression as underlying mechanisms (Chen et al., 2012; Connor et al., 2004; Klonsky, 2007) may explain why

endorsement of these items helps differentiate individuals with higher vs. lower RSDB traits. Also, prior factor-analytical research has indicated relatively higher factor loadings for these items on the RSDB latent trait (Contractor et al., 2020).

Further, across samples, PRBQ items 1 (problematic alcohol use) and 4 (problematic technology use) were the least discriminating. For PRBQ Item 1, perhaps the normative US culture on social drinking (Moore et al., 2005), including less stigma for alcohol use compared to other RSDBs such as drug use (because consuming alcohol does not constitute illegal behavior; Schomerus, 2014), and the high prevalence of drinking behaviors in the US general population (Grant et al., 2017) may have contributed to its lesser discriminative ability across individuals endorsing higher vs. lower RSDB latent trait levels. Further, this single item, while a broad enough question for the purposes of PRBQ ("e.g., binge drank, which is having 4 or more drinks a day for women or 5 or more drinks a day for men; used alcohol in dangerous situations, such as driving"), could reference several facets of alcohol use for participants based on their experiences and perceptions (e.g., attempts to stop drinking, consequences/functions of alcohol use). Indeed, existing research on alcohol measures assessing these different facets has indicated differential functioning of items (Neal et al., 2006; Pilkonis et al., 2013). Referencing PRBQ Item 4, perhaps widespread use of certain technological devices (Pew Internet, 2009) for daily life purposes (e.g., work; Faulk et al., 2010) may have contributed to study findings. Also, prior factor-analytical research has indicated relatively lower factor loadings for items 1 and 4 on the RSDB latent trait (Contractor et al., 2020).

Notably, the PRBQ abides by recommended practices of having 6/7+ response categories for optimal reliability, validity, and discriminating power (Preston & Colman, 2000). This being said, study results suggest that some PRBQ items were classified as MHI vs. LI, primarily based

on if all response categories were being utilized (i.e., item threshold estimates). For the MTurk sample, items 3, 5, 8, 9, 10, 11, and 13 were classified as MHI; and for the IPV sample, items 9-11 were classified as MHI. As examples, participants chose only extreme response categories for Item 2 (problematic drug use); and extreme or middle response categories for items 1 (problematic alcohol use), 4 (problematic technology use), 6 (problematic eating behaviors), 8 (reckless spending), and 12 (reckless driving). Such results warrant an inquiry for research/clinical work into whether the currently utilized five response categories of all PRBQ items are necessary to measure extent of engagement in RSDBs; whether the response scale for certain PRBQ items may need to be revisited and modified/shortened; and whether response categories around the mid-point option are needed (Weijters et al., 2010). Notably, any reduction in response categories may facilitate responding and less participant burden/fatigue (thereby greater utility for intensive longitudinal studies); however, this will also reduce the breadth of experiences that can be reported by participants and statistically decreases variance in obtained scores (Preston & Colman, 2000). When comparing these costs of the potential loss of information vs. benefits of less participant fatigue, such response scale reduction may not be useful.

Some other points are noteworthy. While our results do not suggest deleting PRBQ items based on discrimination and threshold parameters, they imply that certain PRBQ items contributed less information for specific samples based on their characteristics. For instance, when using the PRBQ in a community sample of women currently experiencing IPV and substance use, PRBQ items 1, 2, 4, and 6 (two of these examine substance use) contributed little information to the entire instrument. This may be a result of the substance use inclusion criteria (overlapping with PRBQ items 1 and 2), and phone-based nature of the study (overlapping with

PRBQ Item 4). Further, PRBQ provides more information on participants endorsing greater extent of engagement in RDSBs. Thus, maximum information could be obtained from the PRBQ for participants with higher levels of the RSDB latent trait across both samples. Regarding person fit estimates, participants chose responses as expected for their RSDB latent trait levels and there was no evidence that the PRBQ kindled any adverse psychological reactions for participants. Lastly, in terms of DIF results, PRBQ items functioned well across gender-based groups as well as across different samples. Although several items were flagged for potential DIF, their effect sizes were low indicating that there was no difference in how two individuals from different groups (gender-based groups or samples) responded to the PRBQ items if they had the same RSDB latent trait level.

Current study results need to be interpreted in the light of limitations. First, study findings are limited by the demographic characteristics of the samples under consideration; future research needs to examine the study's research questions in other racially/ethnically and clinically diverse samples. Second, we used gender as a dichotomous variable primarily for sample size considerations; however, this approach did not capture functioning of the PRBQ across all gender-based groups. Third, future investigations may benefit from examining IRT-related psychometric properties in longitudinal investigations; such longitudinal item invariance to ensure that the RSDB construct can be compared across time is especially important when investigating intervention impacts. Fourth, GRM usually requires sample sizes of at least 200 (Kieftenbeld & Natesan, 2012). In this regard, the IPV sample size was lower than this recommended sample size; however, the MTurk sample met these guidelines. Fifth, the item discrimination parameter value cut-off estimate as greater than 0.75 is arbitrary. However, given

that all discrimination parameters were much higher than 0.75, we believe this limitation is of less consequence to the implications of the current study findings.

Sixth, the self-report LEC-5 used in the current study, especially in the context of the MTurk sample, is a screener for potentially traumatic events. For the MTurk sample, we did not use an interview format of a trauma measure; hence, cannot ascertain if individuals endorsed a Criterion A trauma as indicated in the *DSM-5*. Also, the LEC-5 has limited psychometric investigations and there are concerns about individuals not reporting potentially traumatic events on the LEC-5 reliably across time (Pugach et al., 2021). Lastly, recruitment via the MTurk platform has some concerns: the sample may include self-selected participants (Kraut et al., 2004); there is lack of control over the research environment (Thomas & Clifford, 2017); obtained data may have low quality (Chmielewski & Kucker, 2020); individuals may attempt the survey multiple times or fake study eligibility (Hauser et al., 2019); and computer programs may automatically complete research studies (Chmielewski & Kucker, 2020). While we followed recommended steps to enhance data quality such as use of validity checks to ensure attentive responding and comprehension, and excluded individuals missing too much data or attempting the survey multiple times (Aust et al., 2013; Buhrmester et al., 2011; Hauser et al., 2019; Oppenheimer et al., 2009), such steps also contributed to sample truncation (~52%). This potential selection bias in our study may limit generalizability of study results. Future research could use other quality checks such as restricting participation to MTurk workers with high reputation (>95% approval ratings; Peer et al., 2014).

Despite these limitations, current study results have important implications. Broadly, study results indicate all PRBQ items contributed information to the understanding and measurement of the RSDB construct and that the PRBQ items functioned well psychometrically

across the MTurk and the IPV samples, individually. Importantly, these findings indicate that the PRBQ can be administered meaningfully across populations with different socio-demographic characteristics and recruitment modality (online versus in-person). Further, the PRBQ items exhibited negligible DIF across gender-based groups or across the different samples. That is, given the same amount of the RSDB latent trait, the probability of endorsing any item on any response category was identical across men and women as well as across the MTurk and IPV samples. Thus, the results provide support for clinicians and researchers to compare PRBQ-assessed RSDBs across these groups. Lastly, all PRBQ items provided maximum information on participants with a greater extent of engagement (latent trait RSDB levels between 1 and 3). Therefore, it might be useful to include more items that can provide further information on participants reporting less engagement in RSDBs (i.e., lower values of RSDB latent trait). Broadly, results support strong psychometric properties of the PRBQ scores and PRBQ's use for clinicians and researchers.

References

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders* (5th ed.). American Psychiatric Association.

Archer, J. (2004). Sex differences in aggression in real-world settings: A meta-analytic review. *Review of general psychology, 8*, 291-322. https://doi.org/10.1037/1089-2680.8.4.291

Armour, C., Greene, T., Contractor, A. A., Weiss, N. H., Dixon-Gordon, K., & Ross, J. (2020). Posttraumatic stress disorder symptoms and risky behaviors: A network analysis approach. *Journal of Traumatic Stress, 33*, 29-40. https://doi.org/10.1002/jts.22487

Aust, F., Diedenhofen, B., Ullrich, S., & Musch, J. (2013). Seriousness checks are useful to improve data validity in online research. *Behavior Research Methods, 45*, 527–535. https://doi.org/10.3758/s13428-012-0265-2

Baker, F. B. (2001). *The basics of item response theory.* (2nd ed.). ERIC Clearinghouse on Assessment and Evaluation.

Baker, F. B., & Kim, S.-H. (2004). *Item Response Theory: Parameter Estimation Techniques.* (2nd ed.). Marcel-Dekker.

Bakken, N. W., & Gunter, W. D. (2012). Self-cutting and suicidal ideation among adolescents: Gender differences in the causes and correlates of self-injury. *Deviant Behavior, 33*, 339-356. https://doi.org/10.1080/01639625.2011.584054

Becker, C. B., & Zayfert, C. (2001). Integrating DBT-based techniques and concepts to facilitate exposure treatment for PTSD. *Cognitive and Behavioral Practice, 8*, 107-122. https://doi.org/10.1016/S1077-7229(01)80017-1

Ben-Zur, H., & Zeidner, M. (2009). Threat to life and risk-taking behaviors: A review of empirical findings and explanatory models. *Personality and Social Psychology Review, 13*, 109-128. https://doi.org/10.1177/1088868308330104

Brady, K. T., & Randall, C. L. (1999). Gender differences in substance use disorders. *Psychiatric Clinics of North America, 22*, 241-252. https://doi.org/10.1016/S0193-953X(05)70074-5

Braquehais, M. D., Oquendo, M. A., Baca-García, E., & Sher, L. (2010). Is impulsivity a link between childhood abuse and suicide? *Comprehensive Psychiatry, 51*, 121-129. https://doi.org/10.1016/j.comppsych.2009.05.003

Brewerton, T. D. (2007). Eating disorders, trauma, and comorbidity: Focus on PTSD. *Eating Disorders, 15*, 285-304. https://doi.org/10.1080/10640260701454311

Brodey, B. B., First, M., Linthicum, J., Haman, K., Sasiela, J. W., & Ayer, D. (2016). Validation of the NetSCID: An automated web-based adaptive version of the SCID. *Comprehensive Psychiatry, 66*, 67-70. https://doi.org/10.1016/j.comppsych.2015.10.005

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3-5. https://doi.org/10.1177/1745691610393980

Camilli, G., & Shepard, L. A. (1994). *Methods for Identifying Biased Test Items.* Sage Publishers.

Casada, J. H., & Roache, J. D. (2005). Behavioral inhibition and activation in Posttraumatic Stress Disorder. *The Journal of Nervous and Mental Disease, 193*(2), 102-109. https://doi.org/10.1097/01.nmd.0000152809.20938.37

Chen, P., Coccaro, E. F., & Jacobson, K. C. (2012). Hostile attributional bias, negative emotional responding, and aggression in adults: Moderating effects of gender and impulsivity. *Aggressive Behavior, 38*, 47-63. https://doi.org/10.1002/ab.21407

Chmielewski, M., & Kucker, S. C. (2020). An MTurk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11*, 464-473. https://doi.org/10.1177/1948550619875149

Choi, S. W., Gibbons, L. E., & Crane, P. K. (2011). Lordif: An R package for detecting differential item functioning using iterative hybrid ordinal logistic regression/item response theory and Monte Carlo simulations. *Journal of Statistical Software, 39*, 1-30. https://doi.org/10.18637/jss.v039.i08

Clark, H. W., Masson, C. L., Delucchi, K. L., Hall, S. M., & Sees, K. L. (2001). Violent traumatic events and drug abuse severity. *Journal of Substance Abuse Treatment, 20*, 121-127. https://doi.org/10.1016/s0740-5472(00)00156-2

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* (2nd ed.). Erlbaum.

Cole, N. S. (1993). History and development of DIF. In P. Holland & H. Wainer (Eds.), *Differential item functioning* (pp. 25-29). Routledge.

Connor, D. F., Doerfler, L. A., Toscano, P. F., Volungis, A. M., & Steingard, R. J. (2004). Characteristics of children and adolescents admitted to a residential treatment center. *Journal of Child and Family Studies, 13*, 497-510.

Contractor, A. A., Armour, C., Forbes, D., & Elhai, J. D. (2016). PTSD's underlying dimensions and their relation with impulsivity facets. *The Journal of Nervous & Mental Disease, 204*, 20-25. https://doi.org/10.1097/NMD.0000000000000417

Contractor, A. A., Caldas, S., Fletcher, S., Shea, M. T., & Armour, C. (2018). Empirically-derived lifespan polytraumatization typologies: A systematic review. *Journal of Clinical Psychology, 74*, 1137-1159. https://doi.org/10.1002/jclp.22586

Contractor, A. A., Frankfurt, S., Weiss, N. H., & Elhai, J. D. (2017). Latent-level relations between DSM-5 PTSD symptom clusters and problematic smartphone use. *Computers in Human Behavior, 72*, 170-177. https://doi.org/10.1016/j.chb.2017.02.051

Contractor, A. A., Jin, L., Weiss, N. H., & O'Hara, S. (2021). A psychometric investigation on the diagnostic utility of the Posttrauma Risky Behaviors Questionnaire. *Psychiatry Research*. https://doi.org/10.1016/j.psychres.2020.113667

Contractor, A. A., & Weiss, N. H. (2019). Typologies of PTSD clusters and reckless/self-destructive behaviors: A latent profile analysis. *Psychiatry Research, 272*, 682-691. https://doi.org/10.1016/j.psychres.2018.12.124

Contractor, A. A., Weiss, N. H., Dolan, M., & Mota, N. (2019). Examination of the structural relations between PTSD symptoms and reckless/self-destructive behaviors. *International Journal of Stress Management*. https://doi.org/10.1037/str0000133

Contractor, A. A., Weiss, N. H., Dranger, P., Ruggero, C., & Armour, C. (2017). PTSD's risky behavior criterion: Relation with DSM-5 PTSD symptom clusters and psychopathology. *Psychiatry Research, 252*, 215-222. https://doi.org/10.1016/j.psychres.2017.03.008

Contractor, A. A., Weiss, N. H., Kearns, N. T., Caldas, S., & Dixon-Gordon, K. (2020). Assessment of posttraumatic stress disorder's E2 criterion: Development, pilot testing, and validation of the Posttrauma Risky Behaviors Questionnaire. *International Journal of Stress Management, 27*, 292-303. https://doi.org/10.1037/str0000145

Cooper, M. L., Agocha, V. B., & Sheldon, M. S. (2000). A motivational perspective on risky behaviors: The role of personality and affect regulatory processes. *Journal of Personality, 68*, 1059-1088.

Danielson, C. K., Adams, Z., McCart, M. R., Chapman, J. E., Sheidow, A. J., Walker, J., Smalling, A., & de Arellano, M. A. (2020). Safety and efficacy of exposure-based risk reduction through family therapy for co-occurring substance use problems and posttraumatic stress disorder symptoms among adolescents: A randomized clinical trial. . *JAMA Psychiatry, 77*, 574-586. https://doi.org/10.1001/jamapsychiatry.2019.4803

Daw, J., Margolis, R., & Wright, L. (2017). Emerging adulthood, emergent health lifestyles: Sociodemographic determinants of trajectories of smoking, binge drinking, obesity, and sedentary behavior. . *Journal of Health and Social Behavior, 58*, 181–197. https://doi.org/10.1177/0022146517702421

Emons, W. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement, 32*, 224-247. https://doi.org/10.1177/0146621607302479

Engle, K., Talbot, M., & Samuelson, K. W. (2020). Is Amazon's Mechanical Turk (MTurk) a comparable recruitment source for trauma studies? *Psychological Trauma: Theory, Research, Practice, and Policy, 12*, 381=388. https://doi.org/10.1037/tra0000502

Faulk, D., Srikant, D., Lane, J., Guntupalli, A., Bott, J., & Holmes, M. (2010). *Media in the workplace: An analysis of the video Consumer Mapping Study Data.* Ball State University. http://cms.bsu.edu/Academics/CentersandInstitutes/BBR/CurrentStudiesandPublications.aspx

First, M. B., & Williams, J. B. W. (2016). *SCID-5-CV: Structured Clinical Interview for DSM-5 Disorders: Clinician Version*. American Psychiatric Association Publishing.

Furr, M. (2011). *Scale construction and psychometrics for social and personality psychology.* Sage Publications Ltd.

Germain, M. L. (2006). Stages of Psychometric Measure Development: The Example of the Generalized Expertise Measure (GEM). *Online Submission*.

Gidycz, C. A., McNamara, J. R., & Edwards, K. M. (2006). Women's risk perception and sexual victimization: A review of the literature. *Aggression and Violent Behavior, 11*, 441-456. https://doi.org/10.1016/j.avb.2006.01.004

Grant, B. F., Chou, S. P., Saha, T. D., Pickering, R. P., Kerridge, B. T., Ruan, W. J., Huang, B., Jung, J., Zhang, H., Fan, A., & Hasin, D. S. (2017). Prevalence of 12-month alcohol use, high-risk drinking, and DSM-IV alcohol use disorder in the United States, 2001-2002 to 2012-2013: Results From the National Epidemiologic Survey on Alcohol and Related Conditions. . *JAMA Psychiatry, 74*, 911–923. https://doi.org/10.1001/jamapsychiatry.2017.2161

Grant, B. F., Goldstein, R. B., Saha, T. D., Chou, S. P., Jung, J., Zhang, H., Pickering, R. P., Ruan, W. J., Smith, S. M., Huang, B., & Hasin, D. S. (2015). Epidemiology of DSM-5 alcohol use disorder: Results from the National Epidemiologic Survey on Alcohol and Related Conditions III. *JAMA Psychiatry, 72*, 757-766. https://doi.org/10.1001/jamapsychiatry.2015.0584

Hauser, D., Paolacci, G., & Chandler, J. J. (2019). *Common concerns with MTurk as a participant pool: Evidence and solutions.* Routledge.

Jacobson, C. M., Muehlenkamp, J. J., Miller, A. L., & Turner, J. B. (2008). Psychiatric

impairment among adolescents engaging in different types of deliberate self-harm.

Journal of Clinical Child & Adolescent Psychology, 37, 262-375.

https://doi.org/10.1080/15374410801955771

Kassambara, A., & Mundt, F. (2020). *Package 'factoextra'*. https://cran.r-

project.org/web/packages/factoextra/factoextra.pdf

Kerig, P. K. (2019). Linking childhood trauma exposure to adolescent justice involvement: The

concept of posttraumatic risk-seeking. *Clinical Psychology: Science and Practice,

28*(e12280). https://doi.org/10.1111/cpsp.12280

Kieftenbeld, V., & Natesan, P. (2012). Recovery of Graded Response Model parameters: A

comparison of Marginal Maximum Likelihood and Markov Chain Monte Carlo

Estimation. *Applied Psychological Measurement, 36*, 399-419.

https://doi.org/10.1177/0146621612446170

Klonsky, E. D. (2007). The functions of deliberate self-injury: A review of the evidence. *Clinical

Psychology Review, 27*, 226-239. https://doi.org/10.1016/j.cpr.2006.08.002

Kraut, R., Olson, J., Banaji, M., Bruckman, A., Cohen, J., & Couper, M. (2004). Psychological

research online: Report of Board of Scientific Affairs' Advisory Group on the Conduct of

Research on the Internet. *American Psychologist, 59*, 105-117.

https://doi.org/10.1037/0003-066X.59.2.105

Lang, P. J., & Cuthbert, B. N. (1984). Affective information processing and the assessment of

anxiety. *Journal of Behavioral Assessment, 6*, 369-395.

Lang, P. J., Levin, D. N., Miller, G. A., & Kozak, M. J. (1983). Fear behavior, fear imagery, and the psychophysiology of emotion: The problem of affective response integration. *Journal of Abnormal Psychology, 92*, 276-306. https://doi.org/10.1037/0021-843X.92.3.276

Lord, F. M. (1980). *Applications of Item Response Theory o Practical Testing Problems*. Lawrence Erlbaum Associates, Inc.

Lusk, J. D., Sadeh, N., Wolf, E. J., & Miller, M. W. (2017). Reckless self-destructive behavior and PTSD in veterans: The mediating role of new adverse events. *Journal of Traumatic Stress, 30*, 270-278. https://doi.org/10.1002/jts.22182

Marshall-Berenz, E. C., Vujanovic, A. A., & MacPherson, L. (2011). Impulsivity and alcohol use coping motives in a trauma-exposed sample: The mediating role of distress tolerance. *Personality and Individual Differences, 50*, 588-592. https://doi.org/10.1016/j.paid.2010.11.033

McCauley, J. L., Killeen, T., Gros, D. F., Brady, K. T., & Back, S. E. (2012). Posttraumatic stress disorder and co-occurring substance use disorders: Advances in assessment and treatment. *Clinical Psychology: Science and Practice, 19*, 283-304. https://doi.org/10.1111/cpsp.12006

McFall, M., Fontana, A., Raskind, M., & Rosenheck, R. (1999). Analysis of violent behavior in Vietnam combat veteran psychiatric inpatients with posttraumatic stress disorder. *Journal of Traumatic Stress, 12*, 501-517. https://doi.org/10.1023/A:1024771121189

Meade, A. W., & Craig, S. B. (2012). Identifying careless responses in survey data. *Psychological Methods, 17*, 437-455. https://doi.org/10.1037/a0028085

Meader, N., King, K., Moe-Byrne, T., Wright, K., Graham, H., Petticrew, M., Power, C., White, M., & Sowden, A. J. (2016). A systematic review on the clustering and co-occurrence of

multiple risk behaviours. *BMC public health, 16*, 1-9. https://doi.org/10.1186/s12889-016-3373-6

Mischra, S., & Carleton, N. (2017). Use of online crowdsourcing platforms for gambling research. *International Gambling Studies, 17*, 125-143. https://doi.org/10.1080/14459795.2017.1284250

Mokken, R. J. (1971). *A theory and procedure of scale analysis.* Walter de Gruyter. https://doi.org/10.1515/9783110813203

Moore, A. A., Gould, R., Reuben, D. B., Greendale, G. A., Carter, M. K., Zhou, K., & Karlamangla, A. (2005). Longitudinal patterns and predictors of alcohol consumption in the United States. *American Journal of Public Health, 95*, 458–465. https://doi.org/10.2105/AJPH.2003.019471

Morizot, J., Ainsworth, A. T., & Reise, S. (2007). Toward modern psychometrics: Application of item response theory models in personality research. In R. W. R. W. Robins, R. C. Fraley, & R. F. Kruger (Eds.), *Handbook of research methods in personality psychology.* (pp. 407–432). Guilford Press.

Neal, D. J., Corbin, W. R., & Fromme, K. (2006). Measurement of alcohol-related consequences among high school and college students: application of item response models to the Rutgers Alcohol Problem Index. *Psychological Assessment, 18*, 402-414. https://doi.org/10.1037/1040-3590.18.4.402

Nolen-Hoeksema, S. (2004). Gender differences in risk factors and consequences for alcohol use and problems. *Clinical Psychological Review, 24*, 981-1010. https://doi.org/10.1016/j.cpr.2004.08.003

Noll, J. G., & Grych, J. H. (2011). Read-react-respond: An integrative model for understanding sexual revictimization. *Psychology of Violence, 1*, 202-215. https://doi.org/10.1037/a0023962

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*, 867-872. https://doi.org/10.1016/j.jesp.2009.03.009

Oshima, T., & Morris, S. (2008). Raju's Differential Functioning of Items and Tests (DFIT). *Educational Measurement: Issues and Practice, 27*, 43-50. https://doi.org/10.1111/j.1745-3992.2008.00127.x

Osório, F. L., Loureiro, S. R., Hallak, J. E. C., Machado-de-Sousa, J. P., Ushirohira, J. M., Baes, C. V., Apolinario, T. D., Donadon, M. F., Bolsoni, L. M., & Guimarães, T. (2019). Clinical validity and intrarater and test–retest reliability of the Structured Clinical Interview for DSM-5–Clinician Version (SCID-5-CV). *Psychiatry and clinical neurosciences, 73*(12), 754-760.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision Making, 5*, 411-419.

Pat-Horenczyk, R., Peled, O., Miron, T., Brom, D., Villa, Y., & Chemtob, C. M. (2007). Risk-taking behaviors among Israeli adolescents exposed to recurrent terrorism: Provoking danger under continuous threat? *American Journal of Psychiatry, 164*, 66-72. https://doi.org/10.1176/ajp.2007.164.1.66

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46*, 1023-1031. https://doi.org/10.3758/s13428-013-0434-y

Pew Internet. (2009). *Demographics of Internet Use.* Pew Internet and American Life Project, November 19-December 20 Tracking Survey. http://pewinternet.org/Data-Tools/Download

Pilkonis, P. A., Yu, L., Colditz, J., Dodds, N., Johnston, K. L., Maihoefer, C., Stover, A. M., Daley, D. C., & McCarty, D. (2013). Item banks for alcohol use from the Patient-Reported Outcomes Measurement Information System (PROMIS®): Use, consequences, and expectancies. *Drug and Alcohol Dependence, 130*, 167-177. https://doi.org/10.1016/j.drugalcdep.2012.11.002

Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica, 104*, 1-15. https://doi.org/10.1016/S0001-6918(99)00050-5

Prins, A., Bovin, M. J., Kimerling, R., Kaloupek, D. G., Marx, B. P., Pless Kaiser, A., & Schnurr, P. P. (2015). The Primary Care PTSD Screen for DSM-5 (PC-PTSD-5). *Instrument available from the National Center for PTSD at www.ptsd.va.gov.* http://www.ptsd.va.gov/professional/assessment/screens/pc-ptsd.asp

Pugach, C. P., Nomamiukor, F. O., Gay, N. G., & Wisco, B. E. (2021). Temporal stability of self-reported trauma exposure on the Life Events Checklist for DSM-5. *Journal of Traumatic Stress, 34*, 248-256. https://doi.org/10.1002/jts.22611

R Core Team. (2019). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Raju, N. S. (1988). The area between two item characteristic curves. *Psychometrika, 53*, 498-502. https://doi.org/10.1007/BF02294403

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement, 14*, 197-207. https://doi.org/10.1177/014662169001400208

Reckase, M. D. (1979). Unifactor latent trait models applied to multifactor test: Results and implications. *Journal of Educational Statistics, 4*, 207-230. https://doi.org/10.2307/1164671

Reeve, B. B., Hays, R. D., Bjorner, J. B., Cook, K. F., Crane, P. K., Teresi, J. A., Thissen, D., Revicki, D. A., Weiss, D. J., Hambleton, R. K., Liu, H., Gershon, R., Reise, S. P., Lai, J., Cella, D., & The PROMIS Cooperative Group. (2007). Psychometric evaluation and calibration of health-related quality of life item banks: Plans for the Patient-Reported Outcomes Measurement Information System (PROMIS). *Medical Care, 45*, S22–S31. https://doi.org/10.1097/01.mlr.0000250483.85507.04

Reise, S. P., & Waller, N. G. (1993). Traitedness and the assessment of response pattern scalability. *Journal of Personality and Social Psychology, 65*, 143-151. https://doi.org/10.1037/0022-3514.65.1.143

Reiss, S. P., Horan, W. P., & Blanchard, J. J. (2011). The challenges of fitting an Item Response Theory model to the Social Anhedonia Scale. *Journal of Personality Assessment, 3*, 213-224. https://doi.org/10.1080/00223891.2011.558868

Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software, 17*, 1-25.

Roberts, A., Sharman, S., Coid, J., Murphy, R., Bowden-Jones, H., Cowlishaw, S., & Landon, J. (2017). Gambling and negative life events in a nationally representative sample of UK men. *Addictive Behaviors, 75*, 95-102. https://doi.org/10.1016/j.addbeh.2017.07.002

Roley, M. E., Contractor, A. A., Weiss, N. H., Armour, C., & Elhai, J. D. (2017). Impulsivity facets' predictive relations with DSM-5 PTSD symptom clusters. *Psychological Trauma Theory Research Practice and Policy, 9*, 76-79. https://doi.org/10.1037/tra0000146

Sadeh, N., & Baskin-Sommers, A. (2017). Risky, impulsive, and self-destructive behavior questionnaire (RISQ): A validation study. *Assessment, 24*, 1080-1094. https://doi.org/10.1177/1073191116640356

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph Supplement, 34*(1). https://doi.org/10.1002/j.2333-8504.1968.tb00153.x

Schäfer, I., & Najavits, L. M. (2007). Clinical challenges in the treatment of patients with posttraumatic stress disorder and substance abuse. *Current Opinion in Psychiatry, 20*, 614-618. https://doi.org/10.1097/YCO.0b013e3282f0ffd9

Schomerus, G. (2014). The stigma of alcohol and other substance abuse. In P. W. Corrigan (Ed.), *The stigma of disease and disability: Understanding causes and overcoming injustices.* (pp. 57–72). American Psychological Association. https://doi.org/10.1037/14297-004

Selby, E. A., Bender, T. W., Gordon, K. H., Nock, M. K., & Joiner Jr, T. E. (2012). Non-suicidal self-injury (NSSI) disorder: A preliminary study. *Personality Disorders: Theory, Research, and Treatment, 3*, 167-175. https://doi.org/10.1037/a0024405

Senn, T. E., Braksmajer, A., Urban, M. A., Coury-Doniger, P., & Carey, M. P. (2017). Pilot test of an integrated sexual risk reduction intervention for women with a history of childhood sexual abuse. *AIDS and Behavior, 21*, 3247–3259. https://doi.org/10.1007/s10461-017-1854-y

Shankman, S. A., Funkhouser, C. J., Klein, D. N., Davila, J., Lerner, D., & Hee, D. (2018). Reliability and validity of severity dimensions of psychopathology assessed using the Structured Clinical Interview for DSM-5 (SCID). *International journal of methods in psychiatric research, 27*. https://doi.org/10.1002/mpr.1590

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science, 1*, 213-220. https://doi.org/10.1177/2167702612469015

Sommer, J. L., El-Gabalawy, R., Contractor, A. A., Weiss, N. H., & Mota, N. (2020). PTSD's risky behavior criterion: Associated risky and unhealthy behaviors and psychiatric correlates in a nationally representative sample. *Journal of Anxiety Disorders, 73*. https://doi.org/10.1016/j.janxdis.2020.102247

Spitzer, E. G., Benfer, N., Zuromski, K. L., Marx, B. P., & Witte, T. K. (2020). Examining unique and prospective relationships among self-injurious thoughts and behaviors and posttraumatic stress disorder: a network analysis in two trauma-exposed samples. *Psychological Medicine*. https://doi.org/10.1017/S0033291720002263

Stein, D. J., Chiu, W. T., Hwang, I., Kessler, R. C., Sampson, N., Alonso, J., Borges, G., Bromet, E., Bruffaerts, R., de Girolamo, G., Florescu, S., Gureje, O., He, Y., Kovess-Masfety, V., Levinson, D., Matschinger, H., Mneimneh, Z., Nakamura, Y., Ormel, J., Posada-Villa, J., Sagar, R., Scott, K. M., Tomov, T., Viana, M. C., Williams, D. R., & Nock, M. K. (2010). Cross-national analysis of the associations between traumatic events and suicidal behavior: Findings from the WHO World Mental Health Surveys. *PLOS ONE, 5*. https://doi.org/10.1371/journal.pone.0010574

Tendeiro, J. N., Meijer, R. R., & Niessen, A. S. M. N. (2016). PerFit: An R package for person-fit analysis in IRT. *Journal of Statistical Software, 74*(5). https://doi.org/10.18637/jss.v074.i05

Thomas, K. A., & Clifford, S. (2017). Validity and mechanical turk: An assessment of exclusion methods and interactive experiments. *Computers in Human Behavior, 77*, 184-197. https://doi.org/10.1016/j.chb.2017.08.038

[Record #2244 is using a reference type undefined in this output style.]

[Record #2245 is using a reference type undefined in this output style.]

van der Ark, L. A. (2007). Mokken scale analysis in R. *Journal of Statistical Software, 20*, 1-19. http://www.jstatsoft.org/v20/i11/paper

van der Flier, H. (1980). *Vergelijkbaarheid van individuele testprestaties [Comparability of individual test performance]*. Swets and Zeitlinger.

van der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology, 13*, 267-298. https://doi.org/10.1177/0022002182013003001

Van der Kolk, B., Greenberg, M., Boyd, H., & Krystal, J. (1985). Inescapable shock, neurotransmitters, and addiction to trauma: Toward a psychobiology of posttraumatic stress. *Biological Psychiatry, 20*, 314-325. https://doi.org/10.1016/0006-3223(85)90061-7

van Stolk-Cooke, K., Brown, A., Maheux, A., Parent, J., Forehand, R., & Price, M. (2018). Crowdsourcing trauma: Psychopathology in a trauma-exposed sample recruited via Mechanical Turk. *Journal of Traumatic Stress, 31*, 549-557. https://doi.org/10.1002/jts.22303

Walker, C. (2011). What's the DIF? Why differential item functioning analyses are an important

    part of instrument development and validation. *Journal of Psychoeducational*

    Assessment, 29, 364-376. https://doi.org/10.1177/0734282911406666

Weathers, F. W., Blake, D. D., Schnurr, P. P., Kaloupek, D. G., Marx, B. P., & Keane, T. M.

    (2013). The Life Events Checklist for DSM-5 (LEC-5). *Instrument available from the*

    National Center for PTSD at www.ptsd.va.gov.

Weijters, B., Cabooter, E., & Schillewaert, N. (2010). The effect of rating scale format on

    response styles: The number of response categories and response category labels.

    International Journal of Research in Marketing, 27, 236-247.

    https://doi.org/10.1016/j.ijresmar.2010.02.004

Weiss, N. H., Tull, M. T., Dixon-Gordon, K., & Gratz, K. L. (2016). Assessing the negative and

    positive emotion-dependent nature of risky behaviors among substance dependent

    patients. *Assessment*, 1-14. https://doi.org/10.1177/1073191116665906

Table 1

Demographic Information for the Amazon Mechanical Turk Sample (n = 464)

	M (SD)	n (%)*
Age	35.7 (11.12)	
Years of Education	15.23 (2.34)	
PRBQ Total Score	21.04 (9.78)	
Gender		
Female		258 (55.6%)
Male		199 (42.89%)
Male to Female Transgender		1 (0.21%)
Female to Male Transgender		4 (0.86%)
Other		2 (0.43%)
Ethnicity		
Hispanic		60 (12.93%)
Non-Hispanic		396 (85.34%)
Unknown		8 (1.7%)
Race		
White		355 (76.5%)
African American		43 (9.2%)
Asian		52 (11.2%)
American Indian or Alaska Native		21 (4.5%)
Native Hawaiian or Other Pacific Islander		3 (0.6%)
Other, Unknown, or Multiple Races		8 (1.7%)
Employment		
Full-time		325(70%)
Part-time		73 (15.73%)
Retired		14 (3.02%)
Unemployed		41 (8.83%)
Unemployed Student		11 (2.37%)
Income		
< $15,000		48 (10.3%)
$15,000 to $24,999		59 (12.71%)
$25,000 to $34,999		71 (15.3%)
$35,000 to $49,999		63 (13.58%)
$50,000 to $64,999		87 (18.75%)
$65,000 to $79,999		43 (9.27%)
≥ $80,000		93 (20%)
Relationship Status		
Not Dating		77 (16.59%)
Casually Dating (Non-Monogamous)		34 (7.33%)
Seriously Dating (Monogamous)		114 (24.57%)
Married		201 (43.32%)
Divorced		19 (4.09%)
Separated		11 (2.37%)
Widowed		8 (1.72%)
Potentially Traumatic Event (Index)		

Natural Disaster	67 (14.66%)
Fire or Explosion	24 (5.25%)
Transportation Accident	74 (16.20%)
Serious Accident	16 (3.5%)
Exposure to Toxic Substance	2 (0.43%)
Physical Assault	46 (10%)
Assault with Weapon	15 (3.28%)
Sexual Assault	58 (12.7%)
Other Unwanted/Uncomfortable Sexual Experience	10 (2.19%)
Combat or Exposure to War-Zone	5 (1.09%)
Captivity	2 (0.44%)
Life-Threatening Illness/Injury	30 (6.56%)
Severe Human Suffering	8 (1.75%)
Sudden Violent Death	31 (6.78%)
Sudden Accidental Death	36 (7.88%)
Serious Injury, Harm, or Death you Caused to Someone Else	8 (1.75%)
Other Very Stressful Event/Experience	25 (5.47%)

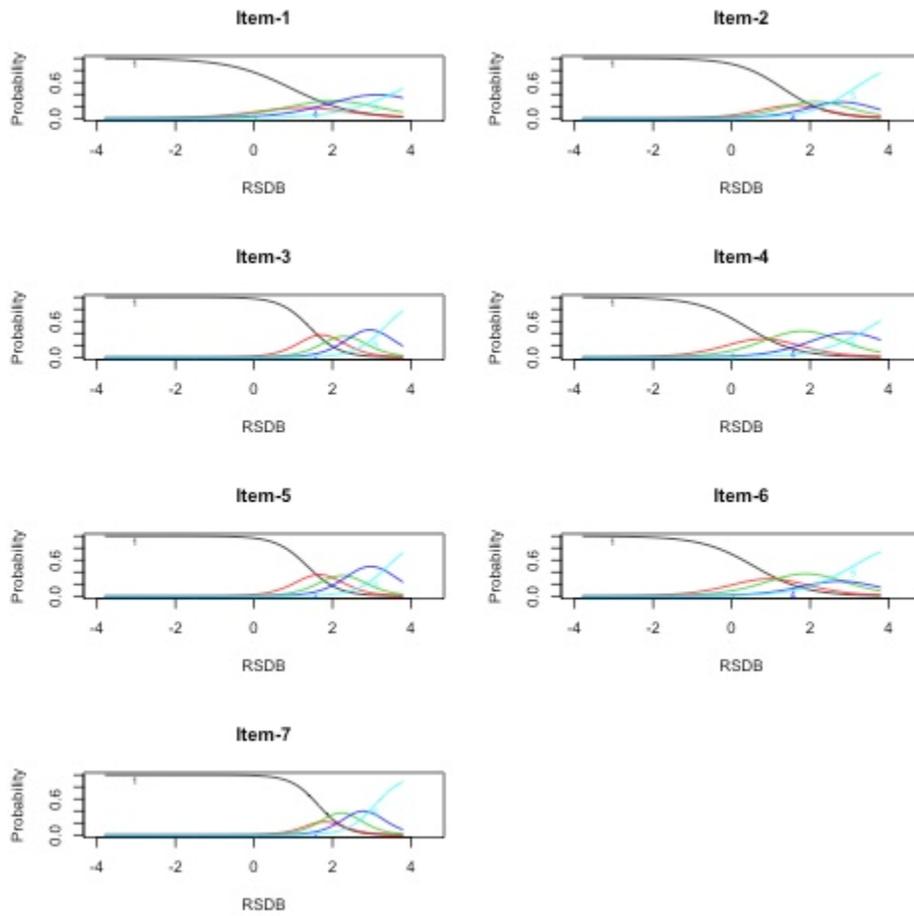*Note.* PRBQ is Posttrauma Risky Behaviors Questionnaire.

Table 2

Demographic Information for the Community Sample (n = 171)

	M (SD)	n (%)
Age	46.03 (74.42)	
Years of Education	12.45 (1.98)	
PRBQ Total Score	31.27 (47.53)	
Gender		
Female		166 (97.1%)
Female to Male Transgender		1 (.6%)
Male to Female Transgender		2 (1.2%)
Gender Queer		1 (0.6%)
Other		1 (0.6%)
Ethnicity		
Hispanic/Latinx		30 (17.5%)
Not Hispanic/Latinx		129 (75.4)
Race		
White		70 (40.9%)
African American/Black		53 (31.0%)
American Indian/Alaskan Native		13 (7.6%)
Hispanic/Latinx		19 (11.1%)
Other		13 (7.6%)
Employment Status		
Full-time		9 (5.3%)
Part-time		20 (11.7%)
Unemployed		107 (62.6%)
Not in labor force (student, homemaker)		23 (13.5%)
Annual Household Income		
< $15,000		67 (39.8%)
$15,000 to $24,999		19 (11.2%)
$25,000 to $34,999		5 (3%)
$35,000 to $49,999		6 (3.5%)
$50,000 to $64,999		3 (1.8%)
$65,000 to $79,999		1 (0.6%)
$\geq$ $80,000		4 (2.4%)
Relationship Status		
Married		14 (8.2%)
Unmarried		127 (74.3%)
Separated or Divorced		13 (7.6%)
Potentially Traumatic Event (Index)		
Natural Disaster		5 (3.9%)
Fire or Explosion		8 (6.25%)
Transportation Accident		8 (6.25%)
Serious Accident		3 (2.34%)
Exposure to Toxic Substance		0 (0%)

Physical Assault	33 (25.78%)
Assault with Weapon	5 (3.9%)
Sexual Assault	37 (28.9%)
Other Unwanted/Uncomfortable Sexual Experience	4 (3.12%)
Combat or Exposure to War-Zone	1 (0.78%)
Captivity	1 (0.78%)
Life-Threatening Illness/Injury	6 (4.69%)
Severe Human Suffering	2 (1.56%)
Sudden Violent Death	7 (5.47%)
Sudden Accidental Death	6 (4.69%)
Serious Injury, Harm, or Death you Caused to Someone Else	0 (0%)
Other Very Stressful Event/Experience	2 (1.56%)

Note. PRBQ is Posttrauma Risky Behaviors Questionnaire.

Table 3

Posttrauma Risky Behaviors Questionnaire Item Parameters for the MTurk Sample and Intimate Partner Violence Sample from the Unconstrained Graded Response Model.

PRBQ Items	MTurk Sample					Intimate Partner Violence Sample				
	b_1	b_2	b_3	b_4	a	b_1	b_2	b_3	b_4	a
PRBQ 1	0.95	1.52	2.48	3.80	1.26	-0.44	0.76	2.07	3.41	0.81
PRBQ 2	1.31	1.81	2.48	3.13	1.73	-0.82	0.26	1.03	2.05	1.00
PRBQ 3	1.42	2.02	2.57	3.33	2.62	1.17	1.79	2.48	4.76	1.27
PRBQ 4	0.39	1.21	2.40	3.51	1.56	0.63	1.19	2.54	3.28	0.91
PRBQ 5	1.36	1.96	2.54	3.41	2.49	0.36	0.97	1.92	2.59	1.65
PRBQ 6	0.57	1.36	2.41	3.09	1.49	0.28	0.78	1.62	2.89	1.03
PRBQ 7	1.64	1.95	2.49	3.08	2.86	0.32	1.20	2.12	2.64	1.60
PRBQ 8	0.70	1.56	2.52	3.52	1.65	-0.07	0.59	1.43	2.31	1.48
PRBQ 9	1.45	1.95	2.53	3.94	2.78	0.39	1.04	1.67	2.41	2.40
PRBQ 10	1.01	1.67	2.37	3.09	2.31	-0.54	0.33	1.24	2.07	1.76
PRBQ 11	1.62	2.11	2.61	3.43	3.20	0.67	1.34	1.95	2.58	3.08
PRBQ 12	1.32	1.97	2.54	2.96	2.70	1.13	1.31	1.96	2.58	2.96
PRBQ 13	1.55	1.92	2.30	2.84	3.44	1.06	1.50	2.13	2.74	2.09
PRBQ_14	1.73	2.07	2.74	3.65	2.57	1.48	1.76	2.38	3.64	1.56

Note. PRBQ is Posttrauma Risky Behaviors Questionnaire; PRBQ Item 1 references problematic alcohol use; PRBQ Item 2 references problematic drug use; PRBQ Item 3 references problematic gambling; PRBQ Item 4 references problematic technology use; PRBQ Item 5 references impulsive/risky sexual behaviors; PRBQ Item 6 references problematic eating behaviors; PRBQ Item 7 references illegal behaviors; PRBQ Item 8 references reckless spending; PRBQ Item 9 references physically aggressive behaviors; PRBQ Item 10 references verbally aggressive behaviors; PRBQ Item 11 references property destruction; PRBQ Item 12 references reckless driving; PRBQ Item 13 references deliberately injuring self without intent to kill one self; PRBQ Item 14 references suicidal behaviors.
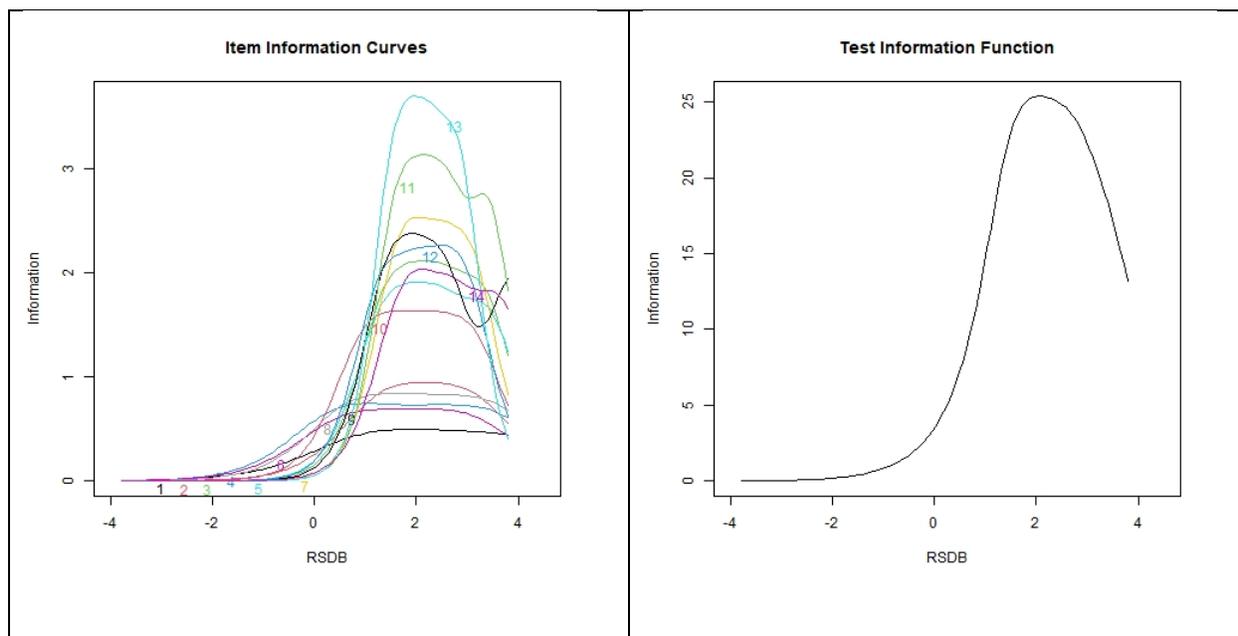
Figure 1

Category Response Curves of Posttrauma Risky Behaviors Questionnaire (PRBQ) Items (MTurk

Sample).

Item-8

Item-9

Item-10

Item-11

Item-12

Item-13

Item-14
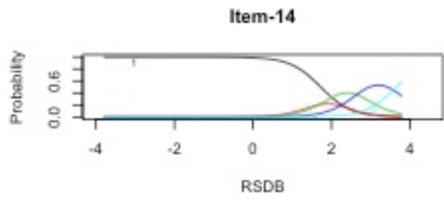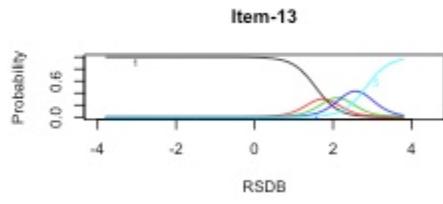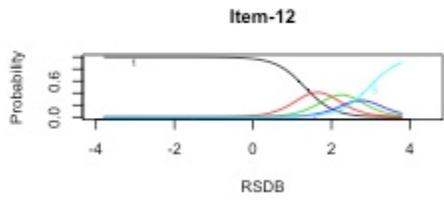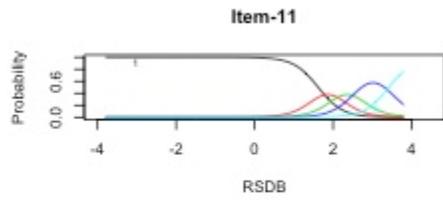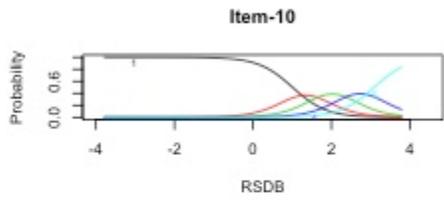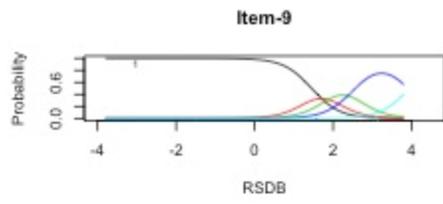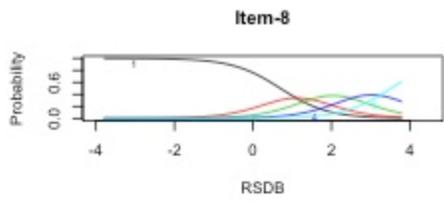
Figure 2

Item Information Curves and Test Information Curve for the Posttrauma Risky Behaviors
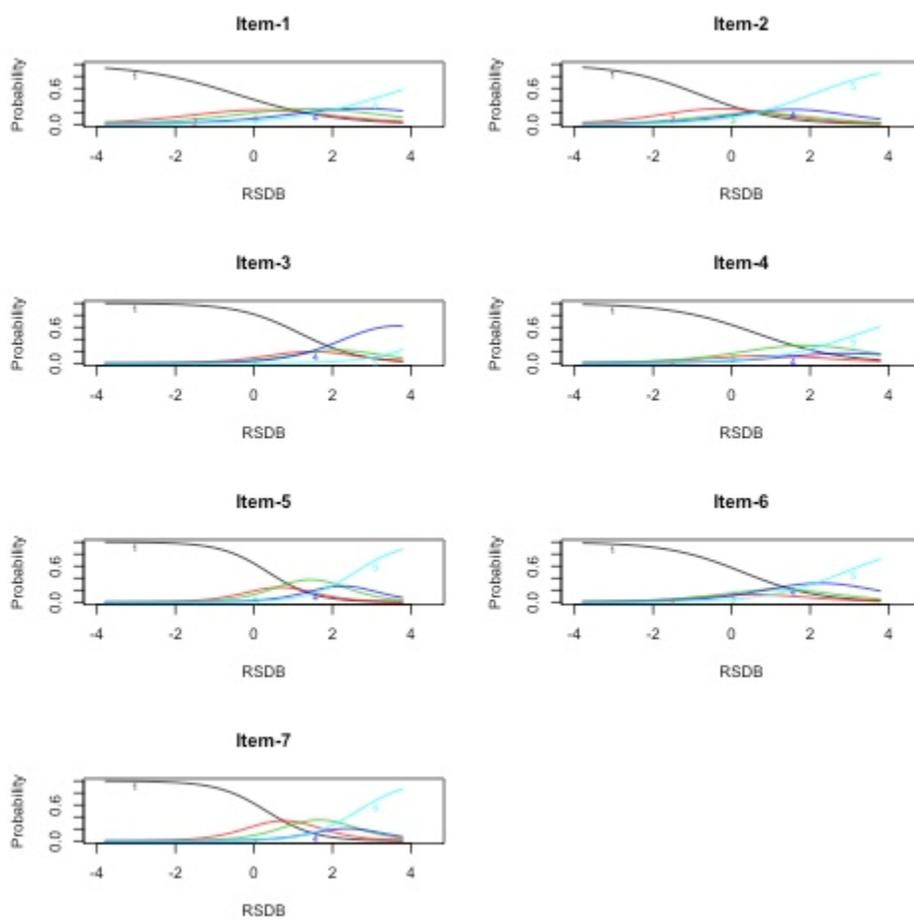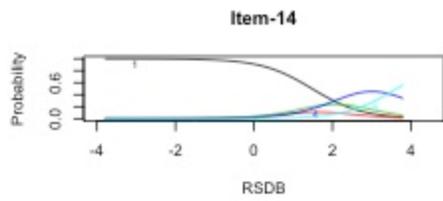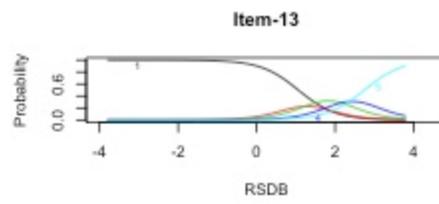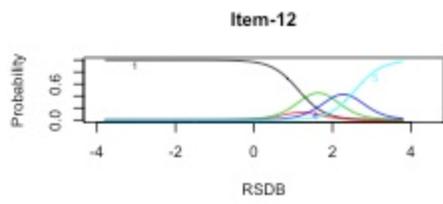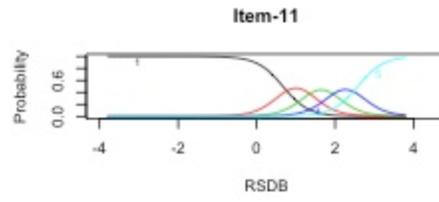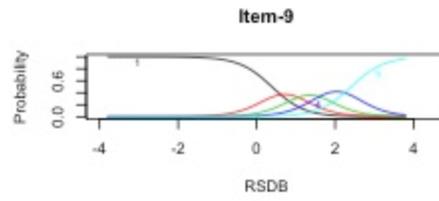
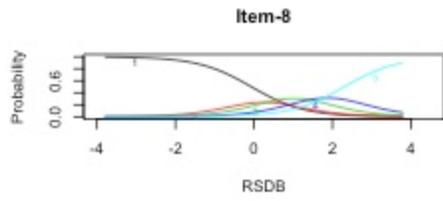*Questionnaire (PRBQ) Items (MTurk Sample)*

Note. RSDB is reckless and self-destructive behaviors.

Figure 3

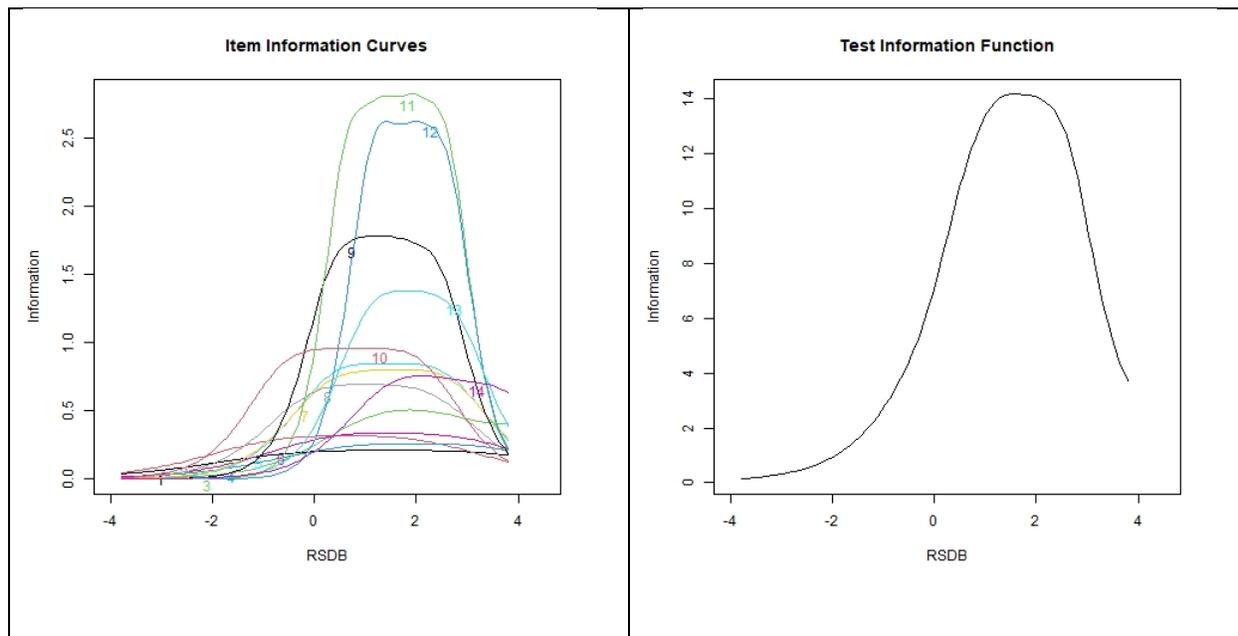*Category Response Curves of Posttrauma Risky Behaviors Questionnaire (PRBQ) Items*

(Intimate Partner Violence Sample)

Item-8

Probability

Item-9

Probability

RSDB

RSDB

Item-10

Probability

Item-11

Probability

RSDB

RSDB

Item-12

Probability

Item-13

Probability

RSDB

RSDB

Item-14

Probability
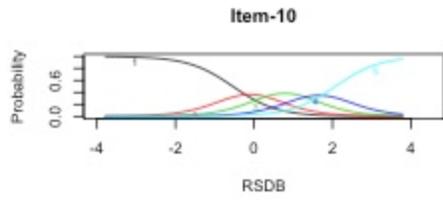
RSDB

Figure 4

Item Information Curves and Test Information Curve for the Posttrauma Risky Behaviors

Questionnaire (PRBQ) Items (Intimate Partner Violence Sample)

Note. RSDB is reckless and self-destructive behaviors.