Tables: 2

Addressing Psychometric Limitations of the Difficulties in Emotion Regulation Scale Through Item Modification

Joseph R. Bardeen[1,] Thomas A. Fergus[2], Susan M. Hannan[3], & Holly K. Orcutt[3]

[1]Department of Psychology, Auburn University, Auburn, AL

[2]Department of Psychology and Neuroscience, Baylor University, Waco, TX

[3]Department of Psychology, Northern Illinois University, DeKalb, IL

Author Note

Correspondence concerning this article should be addressed to Joseph R. Bardeen, Department of Psychology, Auburn University, 226 Thach, Auburn, AL 36849-5214. Phone: (334) 844-6647. E-mail: jbardeen@auburn.edu.

We would like to thank Dr. Kim Gratz for her valuable feedback on this manuscript.

Please contact the corresponding author to obtain a list of the modified DERS items used in the present study.

Abstract

Through its frequent use, a pattern has emerged showing psychometric limitations of the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004). The present three-part study sought to (a) determine whether these limitations are due to a method effect by rewording all reverse-coded items in a straightforward manner and submitting them to exploratory factor analysis (EFA), and (b) examine the tenability of an adaptation of the original measure. EFA results from Study 1 (N = 743) supported retention of 29 modified items across five factors. Consistent with the original theoretical underpinnings of the DERS, Awareness and Clarity items loaded on the same factor. In Study 2 (N = 738), confirmatory factor analysis (CFA) was used to examine the factor structure of the pool of items identified in Study 1. All of the modified subscales clustered strongly with one another and evidenced large loadings on a higher-order emotion regulation construct. These results were replicated in Study 3 (N = 918). Results from Study 3 also provided support for the reliability and validity of scores on the modified version of the DERS (i.e., internal consistency, convergent and criterion-related validity). The present findings provide psychometric support for a modified version of the DERS.

Keywords: Difficulties in Emotion Regulation Scale, emotion regulation, emotion dysregulation, self-report measure, psychometric properties

Emotion regulation difficulties – deficits in the ability to monitor, assess, and alter emotional experience in accordance with one's desired goals (Gratz & Roemer, 2004) – have been implicated in the development of a wide variety of maladaptive psychological outcomes, including Borderline Personality Disorder (Gratz, Rosenthal, Tull, Lejuez, & Gunderson, 2006), anxiety symptomatology (Bardeen & Fergus, 2014; Salters-Pedneault, Roemer, Tull, Rucker, & Mennin, 2006), depression (Tull, Stipleman, Salters-Pedneault, & Gratz, 2009), alcohol dependence (Fox, Hong, & Sinha, 2008), and a host of other maladaptive outcomes (see Aldao, Nolen-Hoeksema, & Schweizer, 2010 for a review). Moreover, emotion regulation difficulties prospectively predict depressive symptoms (Wenzlaff & Luxton, 2003) and posttraumatic stress symptoms following a potentially traumatic event (Bardeen, Kumpula, & Orcutt, 2013). Accordingly, a number of clinical interventions have been developed to alleviate psychological distress by increasing emotion regulation abilities (e.g., Gratz, Levy, & Tull, 2012; Hayes, Strosahl, & Wilson, 1999; Linehan, 1993). Given its clinical relevance and potential transdiagnostic status, the importance of having psychometrically sound self-report measures that assess the emotion regulation domain cannot be overstated.

The Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004) was developed as a multidimensional measure of emotion regulation. Gratz and Roemer (2004) developed the DERS based on a conceptual framework in which effective emotion regulation consists of the following four domains: identification and understanding of emotions, acceptance of emotions, the ability to continue to purse goal-directed behavior and inhibit impulsive behaviors when experiencing negative emotions, and perceived access to effective emotion regulation strategies. To test their hypothesized four-factor model of emotion regulation, Gratz and Roemer (2004) subjected a pool of 41 items to exploratory factor analysis (EFA) based on data collected from a sample of undergraduates students (*N* = 479). EFA results supported the retention of 36 items, and examination of the resultant scree plot, but not eigenvalue threshold, suggested that either a correlated six- or seven-factor solution fit the data (Gratz & Roemer, 2004). As noted by Gratz and Roemer (2004), the six-factor solution was retained because it was more easily interpreted. The six-factor solution included the hypothesized Nonacceptance (i.e., nonacceptance of emotional responses) and Strategies (i.e., perceived access to effective emotion regulation strategies) factors, but failed to support the two other hypothesized factors. Instead, the factor described as identification and understanding of emotions was divided into two distinct factors: Clarity (i.e., lack of emotional clarity) and Awareness (i.e., lack of awareness of emotions). The factor described as the ability to continue to purse goal-directed behavior and inhibit impulsive behaviors when experiencing negative emotions was also divided into two distinct factors: Goals (i.e., difficulty engaging in goal-directed behavior when experiencing negative emotions) and Impulse (i.e., impulse control difficulties when experiencing negative emotions).

Weinberg and Klonsky (2009) examined the factor structure of the DERS using EFA in a similarly sized sample (*N* = 428) of high school students. Consistent with Gratz and Roemer (2004), they found that both a six- or seven-factor solution fit the data. Weinberg and Klonsky (2009) also opted to retain the six-factor solution, but noted that, "our seventh factor appeared to represent an artifact of item format, rather than content (i.e., the factor consisted of reverse-coded items from four different DERS parent scales)" (p. 617). Importantly, an examination of the factor loadings of the six-factor model showed that all of the reverse-coded items loaded, or at least cross-loaded, on the same factor, thus suggesting the possibility of a method effect. In addition, only 30 of 36 items had loadings that would suggest item retention on their parent scale. Specifically, three items cross-loaded, one item loaded on a different scale, and two items did not load on any scale. Four of these six items were reverse-coded. Despite these findings, the authors concluded that their results largely supported the factor structure and reliability and validity of scores on the DERS.

The DERS factor structure has been examined in subsequent studies via CFA. Findings across these studies have been largely equivocal, with the interpretation of CFA results depending upon the specified fit statistic criteria. For example, the following criteria for determining adequate fit were suggested by Hu and Bentler (1999): comparative fit index (CFI) and Tucker-Lewis index (TLI)/ non-normed fit index (NNFI) close to .95 (ideally $\geq$ .95), RMSEA close to .06 (ideally $\leq$ .06), and SRMR close to .08 (ideally $\leq$ .08). However, more liberal criteria have also been suggested for evaluating model fit, with RMSEA values $\leq$ .10 (Browne & Cudeck, 1993; Meyers, Gamst, & Guarino, 2006) and CFI and TLI values $\geq$ .90 (MacCallum, Browne, & Sugawara, 1996; Meyers et al., 2006) being viewed as adequate. The use of different fit statistics, and fit criteria, for interpreting results related to the psychometric properties of the DERS can be seen across a number of studies. For example, Perez, Venta, Garnaat, and Sharp (2012) used CFA to examine the lower-order factor structure of the DERS in an inpatient psychiatric sample of adolescents (N = 218) and found that the CFI and RMSEA values indicated adequate fit based upon a-priori specified guidelines (i.e., CFI = .95, RMSEA = .08), while the weighted root mean square residual value did not (WRMR = 1.37). In a large sample of high-school students (N = 870), Neumann, van Lier, Gratz, and Koot (2010) found that the correlated factor structure of the DERS evidenced adequate fit when more liberal criteria for fit statistics were used (i.e., CFI = .92, TLI = .91, RMSEA = .045). In contrast, Giromini, Velotti, de Campora, Bonalume, and Zavattini (2012) found that the correlated factor structure of the DERS evidenced poor fit in a sample of Italian undergraduate students (*N* = 351; CFI = .86, NNFI = .84, RMSEA = .072). Kokonyei, Urban, Reinhardt, Jozan, and Demetrovics (2014) also found that the correlated factor structure of the DERS evidenced poor fit in a sample of chronic pain patients from Hungarian Hospitals (*N* = 207; CFI = .76, TLI = .74, RMSEA = .11).

None of the above reviewed factor analytic studies reported the degree to which the six lower-order factors of the DERS loaded on a common higher-order factor. Because the six subscales of the DERS were intended to assess the same underlying construct, each subscale should evidence relatively robust loadings on a higher-order emotion regulation construct. Bardeen, Fergus, and Orcutt (2012) used a CFA to test the hierarchical factor structure of the DERS in a large university sample (*N* = 1,037). Bardeen et al. (2012) found that DERS-Awareness showed a much lower factor loading (.26) on a higher-order emotion regulation construct than the other five DERS subscales (loadings from .64 to .98). Further, and consistent with the pattern of associations reported in a number of other studies (reviewed below), Bardeen et al. (2012) found that DERS-Awareness shared much smaller correlations with the other DERS subscales (average *r* = .27) relative to the correlations among the other five DERS subscales (average *r* = .59).

Fowler et al. (2014) found that the higher-order model exhibited poor fit to the data in an inpatient psychiatric sample (*N* = 592). Fowler et al. (2014) described the results of the lower-order CFA as acceptable (i.e., CFI = .91, TLI = .91, RMSEA = .09). Importantly, Fowler et al.'s (2014) findings were consistent with Bardeen et al. (2012), suggesting that the interrelations among the six lower-order DERS constructs are not accounted for by a higher-order emotion regulation construct. Moreover, the DERS-Awareness subscale was the only subscale that did not significantly correlate with all of the other subscales, and the correlations between DERS-Awareness and the other subscales were significantly smaller (with the exception of the relation between DERS-Awareness and DERS-Clarity) than the correlations among the five remaining DERS subscales.

Because the six subscales of the DERS were intended to assess the same underlying construct, all of the subscales should share significant correlations and a consistent pattern of relations with constructs that are theoretically relevant to emotion regulation. The divergence of DERS-Awareness from the other DERS subscales has been consistently shown across a number of published studies. For example, in Gratz and Roemer's (2004) seminal article the DERS-Awareness subscale evidenced small to medium correlations with the other five DERS subscales (*r*s ranging from .08 to .46), whereas the five remaining subscales shared medium to large correlations (*r*s from .32 to .63). A similar pattern was observed in a sample of adolescents, with researchers finding small correlations between DERS-Awareness and the five other DERS subscales (*r*s from -.09 to .10), and medium to large correlations among the remaining subscales (*r*s from .34 to .54; Neumann et al., 2010). Similarly, in two adult samples, DERS-Awareness was not significantly associated with four of the DERS subscales (i.e., Impulse, Goals, Nonacceptance, Strategies), whereas the remaining subscales were all significantly associated (Tull, Barrett, McMillan, & Roemer, 2007; Tull, Gratz, Latzman, Kimbrel, & Lejuez, 2010).

In addition to evidencing only modest correlations with the other subscales of the DERS, DERS-Awareness has repeatedly shown a divergent pattern of relations with constructs theoretically related to emotion regulation. For example, unlike the other five DERS subscales, DERS-Awareness (a) does not predict Posttraumatic Stress Disorder status and is not significantly associated with posttraumatic stress symptom severity (McDermott, Tull, Gratz, Daughters, & Lejuez, 2009; Tull et al., 2007), (b) is not significantly associated with health anxiety (Bardeen & Fergus, 2014), defined as "the wide range of worry that people can have about their health" (p. 5; Asmundson & Taylor, 2005), (c) fails to predict Generalized Anxiety Disorder status (Salters-Pedneault et al., 2006), and (d) is not significantly associated with a history of childhood emotional abuse (Soenke, Hahn, Tull, & Gratz, 2010). Further, Tull et al. (2010) proposed that emotion regulation difficulties would be associated with behavioral inhibition system sensitivity. Results were only partially consistent with this proposition, as DERS-Awareness was the only DERS subscale that was not positively correlated with behavioral inhibition. Similarly, in a sample diagnosed with Binge Eating Disorder, DERS-Awareness was the only DERS subscale that was not positively correlated with food addiction (Gearhardt et al., 2012). Even when DERS-Awareness does have significant associations with theoretically related constructs, the magnitude of the relation is often times significantly smaller in comparison to the other DERS subscales. For example, Rusch, Westermann and Lincoln (2011) found that all of the DERS subscales were significantly positively correlated with social anxiety; however, DERS-Awareness shared a correlation with social anxiety that was much smaller in magnitude (r = .19) than were the correlations between the other five DERS subscales and social anxiety (rs from .38 to .56).

Taken together, the extant literature casts doubt on the construct validity of DERS-Awareness. Bardeen et al. (2012) suggested two primary hypotheses regarding why DERS-Awareness may not cluster with the other subscales of the DERS. First, the Awareness subscale may not belong to the same higher-order emotion regulation construct as the other factors of the DERS. Second, it may be that the manner in which the DERS-Awareness items are worded attenuates interrelations between this subscale and the other DERS subscales, as well as obscuring the relative importance of this construct to the higher-order emotion regulation construct. Specifically, DERS-Awareness is the only subscale for which all of the items are reverse-coded, with no other subscale having more than two reverse-coded items. Although reverse-coded items can serve the purpose of identifying random responding, they can also reduce scale validity by increasing the likelihood of systematic error. In fact, upon removal of reverse-coded items, the psychometric properties of self-report measures often improve (e.g., see Rodebaugh, Woods, & Heimberg, 2007; Weeks et al., 2005). Additionally, evidence from EFA, CFA, and multitrait-multimethod research has consistently demonstrated the tendency of reverse-coded items to load on their own separate factor, thus introducing a method effect that may result in the development of an arbitrary subscale or scale dimension (Dalal & Carter, 2015). Findings from Bardeen et al. (2012) and Fowler et al. (2014) indicate that, in its present form, DERS-Awareness does not belong to the same higher-order construct as the other DERS subscales. This finding is especially important because researchers often sum all of the DERS items to create a DERS total scale score when assessing emotion regulation (e.g., Fox et al., 2008; McDermott et al., 2009; Salters-Pedneault et al., 2006; Tull et al., 2007). As noted by Clark and Watson (1995), when deciding whether to use an overall score versus subscale scores it is important to “establish that all of the items – regardless of how they are placed in the various subscales – define a single general factor” (p. 318). Because DERS-Awareness does not appear to belong to the higher-order construct, including these items in the computation of the total score is not supported and may obscure potentially important findings.

Study 1

In Study 1 we sought to examine interrelations among the DERS items when the potential method effect (i.e., reverse-coded items) surrounding DERS-Awareness was addressed. To accomplish this goal, we reworded all reverse-coded items in a straightforward manner and subjected all 36 items to EFA. If the DERS-Awareness factor is not a byproduct of consisting solely of reverse-coded items, it should still be identified as a distinct factor in EFA even when its items are reworded in a straightforward manner. Interestingly, and discussed above, Gratz and Roemer (2004) initially hypothesized that Awareness and Clarity were best represented as belonging to the same latent construct. Thus, the forward coding of items may result in the items from these two scales loading on the same factor.

Method

Participants

The sample consisted of 743 community adults recruited via the Internet (see below). The sample was 54.4% female and had an average age of 34.7 years (*SD* = 11.7). In regard to racial/ethnic identification, 76.0% of the sample self-identified as White, 10.5% as Black/African-American, 10.2% as Asian, 1.3% as “other”, 1.2% as American Indian/Alaska Native, and 0.7% as Native Hawaiian or other Pacific Islander. Additionally, 7.4% of the sample identified their ethnicity as Hispanic/Latino.

Measures

A modified version of the DERS. The DERS (Gratz & Roemer, 2004) is a 36-item self-report measure used to assess six dimensions of emotion regulation: Nonacceptance, Goals, Impulse, Strategies, Clarity, and Awareness. Items are rated on a 5-point scale based on how often participants believe each item pertains to them (1 = *Almost never* to 5 = *Almost always*). As described, each original reverse-coded DERS item (11 total; six Awareness items, two Clarity items, one Goals item, one Impulse item, and one Strategies item) was reworded in a straightforward manner in the present study. We also took into account that the overwhelming majority of DERS items (75%) begin with the sentence stem, "When I'm upset." Specifically, five of five original DERS-Clarity items and three of the six original DERS-Awareness items do not begin with this sentence stem. To be consistent with the majority of DERS items, all of the modified items begin with, “When I’m upset.” In total, 14 of the original 36 DERS items were modified (11 reverse-coded items and 3 items that lacked the “When I’m upset” sentence stem).

Procedure

Participant recruitment took place using Amazon Mechanical Turk (MTurk), an online labor market where researchers can recruit general population adults to complete questionnaires in exchange for payment. MTurk samples tend to be more demographically diverse than American undergraduate samples (Buhrmester et al., 2011), and although a number of studies have demonstrated the reliability of data collected via MTurk (e.g., Behrend, Sharek, Meade, & Wiebe, 2011; Buhrmester, Kwang, & Gosling, 2011; Paolacci, Chandler, & Ipeirotis, 2010), one catch question was embedded in the online survey to ensure that participants were being attentive (i.e., “Please click the circle at the bottom of the screen. Do not click on the scale items that are labeled from 1 to 9”; Oppenheimer, Meyvis, & Davidenko, 2009). One catch question was used, rather than several, because of the survey’s relatively short time to complete (approximately 10 minutes). Nineteen participants failed to answer the catch question correctly, and thus, were removed from the final sample (*N* = 743). This study, as well as Study’s 2 and 3, was approved by the institutional review board at the first author’s academic institution. Recruitment was limited to MTurk users located within the United Stated and over the age of 19. Informed consent and questionnaires (i.e., demographics and the modified DERS items) were completed using a secure online survey program. Participants could complete the study from any computer with internet access. Participants were paid $0.40 upon study completion, an amount which is consistent with precedence for paying MTurk workers in similar questionnaire studies (Buhrmester et al., 2011).

Data Analytic Strategy

EFA (principal axis factoring) with oblique rotation (i.e., direct oblimin) was used to identify the initial factor structure of the modified pool of items. We initially considered factors with eigenvalues > 1.0 for retention (i.e., Kaiser-Guttman criterion; Kaiser, 1960). Because the Kaiser-Guttman criterion can result in overextraction, two alternate factor identification methods were used to facilitate the identification of an optimal factor structure of the modified pool of items. Specifically, as suggested by Fabrigar, Wegener, MacCallum, and Strahan (1999), both parallel analysis (O'Connor, 2000) and Velicer's (1976) minimum average partial (MAP) test were conducted following the initial EFA. O'Connor's (2000) SPSS syntax was used to conduct these analyses. Following identification of the factor structure of the modified pool of items, factor loadings were interpreted. We followed Gratz and Roemer's (2004) item retention criteria when they developed the original DERS, with item loadings $\geq$ .40 considered interpretable. Items with loadings $\geq$ .40 on more than one factor and items with loadings < .40 were excluded from further analysis. Because participants were prompted when questionnaire items were not fully completed, and participants only had to complete a demographics form and the modified version of the DERS, no data points were missing in Studies I and II.

Results

Evaluation of the assumptions of exploratory factor analysis were met (e.g., univariate and multivariate normality; see Kline, 2011). When considering the Kaiser-Guttman criterion, five factors from the modified 36-item pool were extracted with eigenvalues greater than 1.0 (18.67, 3.44, 2.13, 1.42, and 1.02). These five factors accounted for the following variance:

factor 1 = 51.85%, factor 2 = 3.13%, factor 3 = 1.86%, factor 4 = 1.19%, factor 5 = 0.75%. In contrast, parallel analysis indicated that a maximum of six factors should be extracted, as the seventh eigenvalue was larger than its randomly generated counterpart, but the sixth eigenvalue was not. This six factor solution was uninterpretable because items on the sixth factor did not have salient loadings based upon the above identified criteria for item retention. Consistent with the Kaiser-Guttman criterion, the MAP test indicated a five factor solution; the mean squared partial correlation decreased from the first to the fifth factor, but increased when six factors were extracted. As described by O'Connor (2000), when discrepancies are observed, the tendency of parallel analysis to err in the direction of overextraction should be considered. Given that both the Kaiser-Guttman criterion and MAP test identified an interpretable five factor solution and the six-factor solution was uninterpretable, the five factor solution was retained for further testing.

Factor loadings for the five factor solution are presented in Table 1. Item 17 and item 36 were excluded from further analysis due to low factor loadings (see Table 1). Item 30 was also discarded because it had salient cross-loading on two factors (factor 3 = .44, factor 6 = .43). Factor 1 consisted of the six items representing impulse control difficulties when experiencing negative emotions (i.e., Impulse scale). Factor 2 consisted of 10 items representing lack of awareness and clarity of emotion when experiencing negative affect (i.e., Clarity and Awareness scales). Factor 3 consisted of six items representing of the nonacceptance of emotional responses when experiencing negative emotions (i.e., Nonacceptance scale). Factor 4 consisted of five items representing difficulty engaging in goal-directed behavior when experiencing negative emotions (i.e., Goals scale). Factor 5 consisted of six items representing perceived access to effective emotion regulation strategies when experiencing negative emotions (i.e., Strategies scale). To facilitate consistency, the six highest loading items from the remaining 10 items of

Factor 2 were retained to represent the combined Awareness and Clarity items, henceforth referred to as the Identification factor; items 1, 9, 8, and 34 were excluded from further analysis, thus resulting in the retention of 29 items. Factor intercorrelations ranged from .38 to .71.

Study 1 Summary

In Study 1 we reworded all of the reverse-coded items of the DERS in a straightforward manner and subjected the 36 revised items to EFA. We also ensured that each DERS item began with the same sentence stem. Although we acknowledge that simultaneously addressing the wording of items and sentence stem precludes our ability to conclusively say that negatively-keyed items *per se* were responsible for the previously identified Awareness factor, Study 1 results are consistent with such a possibility. Specifically, Study 1 results are consistent with the possibility that DERS-Awareness may have been identified by Gratz and Roemer (2004) as being factorially distinct because it is the only scale for which all of the items are reverse-coded. It is possible that this method effect is responsible for the noted psychometric limitations related to the Awareness subscale (e.g., a significantly lower factor loading on a higher-order emotion regulation construct relative to the other five DERS subscales, much smaller correlations with the other DERS subscales, as well as with constructs that are theoretically relevant to emotion regulation, relative to the other DERS subscales). Importantly, Study 1 results are consistent with Gratz and Roemer's original conceptualization of emotion regulation in which the identification and understanding of emotions is encapsulated by a single factor. Study 2 sought to examine the tenability of the revised factor structure identified in Study 1 through use of CFA and examination of the interrelations among the latent factors.

Study 2

The purpose of Study 2 was to examine the factor structure of the five-factor model identified in Study 1, as well as the interrelations among the latent factors. Examination of latent factor interrelations, rather than intercorrelations among subscales, provides a clearer estimate of relations among constructs (Brown, 2006) because measurement error is taken into account (e.g., scale unreliability). We predicted that the factors would cluster strongly with one another and evidence loadings on a higher-order construct suggesting that the interrelations among the five lower-order constructs are accounted for by a higher-order emotion regulation construct. Such findings would support the use of a total scale score. Findings which support this factor structure, as well as the higher-order construct, will suggest that further psychometric evaluation of the five-factor model is warranted.

Method

Participants

The sample consisted of 738 community adults recruited via the Internet (see below). The sample was 50.5% male and had an average age of 33.7 years (*SD* = 11.2). With regard to race, 82.0% of the sample self-identified as White, 8.3% as Black/African-American, 6.0% as Asian, 1.9% as “other”, 1.6% as American Indian/Alaska Native, and 0.3% as Native Hawaiian or other Pacific Islander. Additionally, 7.5% of the sample identified their ethnicity as Hispanic/Latino.

Measures

A modified version of the DERS. Study 1 supported a modified version of the DERS consisting of 29 items and assessing five dimensions of emotion regulation: Nonacceptance, Goals, Impulse, Strategies, and Identification. Participants used the same 5-point scale described in reference to the original DERS to indicate how often each item pertains to them.

Procedure

Participants were recruited via MTurk. Consistent with Study 1, one catch question was embedded in the online survey to ensure that participants were being attentive (Oppenheimer et al., 2009). Fifteen participants failed to answer the catch question correctly, and thus, were removed from the final sample ($N = 738$). Eligibility requirements were consistent with Study 1, with one caveat. Participants who completed Study 1 were unable to participate in the present study. Informed consent and questionnaires (i.e., demographics and the modified DERS items) were completed using a secure online survey program. Participants were paid $0.40 upon study completion.

Data Analytic Strategy

A higher-order CFA approach was used to examine the factor structure of the modified version of the DERS. Following Brown (2006), first-order CFA measurement models were first examined to test the adequacy of a parsimonious one-factor lower-order model and the five-factor lower-order model identified in Study 1, as well as to examine latent correlations among the five lower-order factors. Next, should the lower-order correlated five-factor model provide the best fit among the first-order measurement models, lower-order factor correlations were removed from the lower-order five-factor model and direct effects from the higher-order factor to each of the lower-order factors were added. This higher-order CFA was used to test whether a higher-order factor (i.e., emotion regulation) accounted for the latent correlations among the five lower-order constructs. The chi-square difference test was used to examine whether there was a significant difference in the fit of the lower-order and high-order models (Kline, 2011). However, because chi-square difference testing with large sample sizes may indicate significant differences when differences in parameter estimates are trivial in magnitude (Cheung & Rensvold, 2002), RMSEA 90% confidence intervals (CIs) were also compared. Differences in model fit are considered non-significant if models have overlapping 90% RMSEA CIs (Wang & Russell, 2005). In addition, the change in CFI (ΔCFI) was also examined to compare models (Kline, 2011). A ΔCFI value of less than or equal to .002 suggests trivial differences in model fit (Meade, Johnson, & Braddy, 2008).

Models were tested using LISREL 8.80 (Jöreskog & Sörbom, 2006). Robust maximum likelihood estimation (Satorra & Bentler, 1994), in which covariance and asymptotic covariance matrices are entered, was used to alleviate concerns regarding the influence of data non-normality on maximum likelihood estimation (Brown, 2006). Models were evaluated using standard goodness-of-fit statistics, including CFI, NNFI, RMSEA, and SRMR. As described earlier, Hu and Bentler (1999) provided the following guidelines for determining adequate fit: CFI and NNFI close to .95 (ideally ≥ .95), RMSEA close to .06 (ideally ≤ .06), and SRMR close to .08 (ideally ≤ .08).

Results

Lower-Order CFAs

The one-factor lower-order model did not provide an adequate fit to the data, χ^2 = 7110.69; Satorra-Bentler (SB) χ^2 = 7586.49 (377, p < .001), CFI = .906, NNFI = .899, RMSEA = .161 (90% CI = .158 - .164), SRMR = .107. All factor loadings were significant (ps < .001). The correlated five-factor model provided an adequate fit to the data, χ^2 = 1509.10; Satorra-Bentler (SB) χ^2 = 1071.62 (df = 367, p < .001), CFI = .991, NNFI = .990, RMSEA = .051 (90% CI = .048 - .055), SRMR = .056. All factor loadings were significant (ps < .001). In contrast to the six subscales of the original DERS, all of the subscales identified in Study 1 clustered well with one another, with latent correlations among all of the five subscales ranging from .39-83.

Higher-Order CFA

The five-factor higher-order CFA provided an adequate fit to the data, $\chi^2 = 1575.10$; Satorra-Bentler (SB) $\chi^2 = 1129.66$ (372, $p < .001$), CFI = .990, NNFI = .989, RMSEA = .053 (90% CI = .049 - .056), SRMR = .064. All lower-order factor loadings were significant (*p*s < .001). Additionally, factor loadings on the higher-order factor were all significant ($p < .001$) and generally large in magnitude: Identification = .64, Impulse = .89, Nonacceptance = .72, Goal = .79, and Strategies = .94.

The chi-square difference test, $\chi^2_D = 66.10$ ($df = 5$, $p < .001$), indicated that the five-factor lower-order model provided significantly better fit to the data than the higher-order model. However, the ΔCFI was < .002 and overlapping RMSEA CIs were observed. Moreover, all of the goodness-of-fit indices of the six-factor higher-order model met our a-priori benchmark indicating adequate fit. These results, in combination with our relatively large sample size (N = 738), which leaves the chi-square difference test prone to type I error, suggest a non-significant difference in fit between the two models.

Study 2 Summary

In Study 2 we sought to further examine the factor structure of the modified version of the DERS identified in Study 1. The lower-order five-factor model provided adequate fit to the data and the interrelations among the five lower-order factors were medium to large. In contrast to the original DERS, all of the subscales identified in Study 1 clustered relatively strongly with one another. Additionally, all of the lower-order factors evidenced loadings on the higher-order emotion regulation construct that were large in magnitude. Together, these findings suggest that the five lower-order factors represent distinct facets of the same higher-order emotion regulation construct. Importantly, results support the use of a total scale score to represent an overarching emotion regulation construct. Taken together, results from Study 2 support further evaluation of the psychometric properties of the modified version of the DERS.

Study 3

The purpose of the Study 3 was twofold. First, we sought replicate the factor structure identified in Study 1 and supported by the results of Study 2. Second, we sought to examine the reliability and validity of scores on the modified version of the DERS. Internal consistency of the total scale score and subscale scores was examined by calculating Cronbach's alpha (α). We sought to examine convergent validity by investigating associations between the modified version of the DERS and established measures of emotion regulation (i.e., Generalized Expectancy for Negative Mood Regulation Scale [NMR]; Catanzaro & Mearns, 1990; Trait Meta Mood Scale [TMMS]; Salovey, Mayer, Goldman, Turvey, & Palfai, 1995), as well as a measure of experiential avoidance which has been shown to be strongly related to measures of emotion regulation (i.e., Acceptance and Action Questionnaire-II [AAQ-II]; Bond et al., 2011).

Negative mood regulation expectancies, commonly considered a core component of contemporary conceptualizations of emotion regulation, are often assessed via the NMR. The Strategies subscale of the DERS parallels the construct of negative mood regulation expectancies, as both relate to one's perceived ability to cope with emotional distress. Consistent with findings from Gratz and Roemer (2004), we hypothesized that the Strategies subscale of the modified version of the DERS would share the strongest correlation with NMR in comparison to the four other subscales. Additionally, Gratz and Roemer also found that experiential avoidance was more strongly related to the Strategies subscale relative to the five other DERS emotion regulation dimensions. We therefore expected to observe a similar pattern of results in the present study. The TMMS assesses aspects of emotion regulation, such as the tendency to attend to and identify emotions, which the NMR does not. We hypothesized that these two dimensions of the TMMS would share the strongest correlations with the Identification subscale of the modified version of the DERS in comparison to the four other subscales. Next, we sought to examine the criterion validity of the modified version of the DERS via associations between the subscales and constructs theoretically relevant to emotion regulation (in this case, anxiety and depression: e.g., Bardeen & Fergus, 2014; Salters-Pedneault et al., 2006; Tull et al., 2009). We expected that all of the subscales would share moderate to strong correlations with these targeted criterion variables.

Method

Participants

The sample consisted of 918 community adults recruited via the Internet (see below). The sample was 68.2% female and had an average age of 36.2 years (*SD* = 11.6). In regard to racial/ethnic identification, 82.1% of the sample self-identified as White, 7.8% as Black/African-American, 5.4% as Asian, 3.5% as “other”, 1.0% as American Indian/Alaska Native, and 0.1% as Native Hawaiian or other Pacific Islander. Additionally, 6.4% of the sample identified their ethnicity as Hispanic/Latino.

Measures

A modified version of the DERS. Consistent with standard practice for scoring the DERS, the 29-items of the modified version were summed to create a total scale score with higher scores indicating greater emotion regulation difficulties. The items of each subscale were likewise summed. The mean total scale score was 58.53 (*SD* = 23.40, *range* = 29-145). Descriptive statistics for the subscales are as follows: Identification (*M* = 10.48, *SD* = 4.61, *range* = 6-30), Impulse (*M* = 10.75, *SD* = 5.51, *range* = 6-30), Nonacceptance (*M* = 12.72, *SD* = 6.25, *range* = 6-30), Goals (*M* = 12.74, *SD* = 5.61, *range* = 5-25), and Strategies (*M* = 11.84, *SD*

= 6.30, *range* = 6-30).

Depression, Anxiety, and Stress Scale-21 item Version (DASS – 21). The DASS-21 (Lovibond & Lovibond, 1995a) is a 21-item self-report measure that assesses depression, anxiety, and stress symptoms using three, 7-item scales. Each item is rated on a 4-point scale (0 = *Did not apply to me at all* to 3 = *Applied to me very much, or most of the time*). The DASS-21 has exhibited adequate psychometric properties in prior studies (e.g., see Henry & Crawford, 2005). In the present study, internal consistency for the DASS-21 Depression and Anxiety subscales was adequate (Depression α = .93; Anxiety α = .86). The mean DASS-21 Depression and Anxiety scores were 5.11 (*SD* = 5.48) and 3.49 (*SD* = 4.08), respectively.

Acceptance and Action Questionnaire-II (AAQ-II). The AAQ-II (Bond et al., 2011) is a 7-item measure that assesses experiential avoidance. Responses are provided on a 1 (*Never true*) to 7 (*Always true*) scale. Higher scores on the AAQ-II are indicative of higher levels of experiential avoidance (e.g., *I'm afraid of my feelings*). The AAQ-II has shown adequate psychometric properties in prior studies, including a near-perfect convergent correlation (*r* = .97) with the AAQ-I (Bond et al., 2011). In the present study, internal consistency was adequate for the AAQ-II (α = .94, *M* = 20.40, *SD* = 10.38).

Trait Meta Mood Scale (TMMS). The TMMS (Salovey et al., 1995) is a 30-item self-report measure that consists of three subscales designed to measure one's tendency to attend to (Attention: 13 items) and identify (Clarity: 11 items) affective states and emotions, as well as cope with negative emotions and affective states in such a way as to promote positive affect (Repair: 6 items). Items are rated on a 5-point scale based on how much participants agree with statements regarding attitudes toward their emotions and affective states (1 = S*trongly disagree* to 5 = *Strongly agree*). The TMMS has evidenced adequate psychometric properties (Salovey et al., 1995). In the present study, internal consistency was good for all three of the TMMS scales (α values ranging from .80 to .89). The mean TMMS-Awareness, -Clarity, and -Repair subscale scores were 46.19 (*SD* = 7.10), 40.64 (*SD* = 7.77) and 21.24 (*SD* = 5.00), respectively.

Generalized Expectancy for Negative Mood Regulation Scale (NMR). The NMR (Catanzaro & Mearns, 1990) is a 30-item self-report measure that assesses negative mood regulation expectancies (i.e., the degree to which respondents believe that they can make themselves feel better when experiencing a negative mood state). Items are rated on a 5-point scale (1 = *Strongly disagree* to 5 = *Strongly agree*). Higher scores indicate more positive emotion regulation expectancies. The NMR has evidenced adequate psychometric properties in previous research (Catanzaro, 1994; Catanzaro & Mearns, 1990). In the present study, internal consistency was adequate for the NMR (α = .93, *M* = 99.46, *SD* = 20.63).

Procedure

Participants were recruited via MTurk. As noted, participants took approximately 10 minutes to complete Study's 1 and 2. In contrast, Study 3 took approximately 40 minutes to complete, thus providing greater potential for fluctuations in the degree to which participants were appropriately attending to survey questions. With this in mind, and consistent with Bardeen, Fergus, and Orcutt (2013) in their MTurk study of similar length, we embedded three catch questions in this online survey, with each question having only one correct, obvious answer (e.g., "While watching television, have you ever had a fatal heart attack? Please click the circle at the bottom of the screen. Do not click on the scale items that are labeled from 1 to 9"; Oppenheimer et al., 2009; Paolacci et al., 2010). Participants needed to correctly answer two of three of these catch questions to be included in the final sample (Bardeen et al., 2013). Of the 949 participants who completed the modified DERS items in the present study, 31 (3.3%)

answered 1 or less catch questions correctly, and were thus excluded from the final sample (N = 918). Eligibility requirements were consistent with Study's 1 and 2, except that participants who completed either Study 1 or 2 were unable to participate in the present study. Informed consent and questionnaires were completed using a secure online survey program. Participants were paid $1.50 upon study completion.

Data Analytic Strategy

First, the data analytic strategy described in Study 2 was repeated to examine the replicability of the latent structure of the emotion regulation construct across independent samples. Specifically, lower- and higher-order models were tested via CFA. Fit statistics and chi-square difference results were evaluated using the guidelines described above. Second, bivariate correlations were calculated to examine the construct validity of the total scale score and subscale scores. To determine whether the hypothesized differences in the strength of correlations was statistically significant, Meng, Rosenthal, and Rubin's (1992) test for dependent correlations was used. Based on recommendations from Allison (2002) and Enders (2010), a multiple imputation procedure was used to estimate all missing values (1.22% of data points). Specifically, the Missing Values add-on available in IBM SPSS software (Version 19) was used to impute 5 datasets, with each dataset containing a different estimate for each missing value. The parameter estimates and standard errors from each dataset were then combined to create a single set of estimated values.

Results

Lower-Order CFAs

The one-factor lower-order model did not provide an adequate fit to the data, χ^2 = 9661.80; Satorra-Bentler (SB) χ^2 = 9911.26 (377, p < .001), CFI = .905, NNFI = .901, RMSEA =

.166 (90% CI = .163 - .169), SRMR = .103. The correlated five-factor model provided an adequate fit to the data, χ^2 = 2904.27; Satorra-Bentler (SB) χ^2 = 1447.50 (367, $p < .001$), CFI = .989, NNFI = .986, RMSEA = .057 (90% CI = .054 - .060), SRMR = .054. All factor loadings were significant (*p*s < .001). Consistent with the results of Study 2, all of the subscales clustered well with one another, with latent correlations among all of the five subscales ranging from .42 to .78.

Higher-Order CFA

The five-factor higher-order CFA provided an adequate fit to the data, χ^2 = 2154.36; Satorra-Bentler (SB) χ^2 = 1501.074 (372, $p < .001$), CFI = .989, NNFI = .988, RMSEA = .058 (90% CI = .055 - .061), SRMR = .059. All lower-order factor loadings were significant (*p*s < .001). Additionally, factor loadings on the higher-order factor were all significant ($p < .001$) and generally large in magnitude: Identification = .64, Impulse = .88, Nonacceptance = .78, Goal = .77, and Strategies = .90.

The chi-square difference test, χ^2_D = 60.15 (df = 5, $p < .001$), indicated that the lower-order model provided significantly better fit to the data than the higher-order model. However, the ΔCFI was < .002 and overlapping RMSEA CIs were observed. Moreover, goodness-of-fit indices of the six-factor higher-order model met our a-priori benchmark indicating adequate fit. These results, in combination with our relatively large sample size (N = 918), which leaves the chi-square difference test prone to type I error, suggest a non-significant difference in fit between the two models.

Internal Consistency and Convergent and Criterion Validity

Internal consistency was adequate for scores on both the total scale (α = .97) and the five subscales (α values ranging from .88 to .95). Bivariate correlations are present in Table 2. As predicted, the Strategies subscale shared the strongest correlation with the NMR (r = -.74) in comparison to relations among NMR and (a) the Identification subscale (r = -.33; z = 14.24, p < .001), (b) the Impulse subscale (r = -.54; z = 10.81, p < .001), (c) the Nonacceptance subscale (r = -.51; z = 10.75, p < .001), and (d) the Goals subscale (r = -.59; z = 8.57, p < .001). Consistent with Gratz and Roemer (2004), the Strategies subscale of the modified version of the DERS was more strongly related to the AAQ-II (r = .71) in comparison to relations among AAQ-II and (a) the Identification subscale (r = .39; z = 11.13, p < .001), (b) the Impulse subscale (r = .61; z = 5.51, p < .001), (c) the Nonacceptance subscale (r = .65; z = 4.57, p < .001), and (d) the Goals subscale (r = .63; z = 4.74, p < .001). Also, as predicted, both TMMS-Attention and TMMS-Clarity were more strongly related to Identification (r = -.42, r = -.56, respectively) in comparison to relations among TMMS-Attention and TMMS-Clarity and (a) the Strategies subscale (r = -.23, z = .547, p < .001, and r = -.39, z = 5.39, p < .001, respectively), (b) the Impulse subscale (r = -.21, z = 6.46, p < .001, and r = -.47, z = 3.50, p < .001, respectively), (c) the Nonacceptance subscale (r = -.22, z = 5.85, p < .001, and r = -.44, z = 3.95, p < .001, respectively) and (d) the Goals subscale (r = -.11, z = 7.97, p < .001, and r = -.34, z = 6.28, p < .001, respectively). Finally, all of the subscales evidenced moderate to strong associations with DASS-21 Anxiety and DASS-21 Depression (rs ranging from .34 to .68, ps < .001).

Study 3 Summary

Results from Study 2 were replicated in Study 3. Specifically, the lower-order five-factor model provided adequate fit to the data, interrelations among the five lower-order factors were medium to large, all of the subscales clustered well with one another, and all of the lower-order factors evidenced loadings on the higher-order emotion regulation construct that were large in magnitude. Additionally, results from Study 3 provide support for the modified version of the

DERS as a reliable and valid measure of emotion regulation difficulties. The subscales showed good internal consistency, converged with established measures of emotion regulation and experiential avoidance, and shared moderate to strong correlations with constructs theoretically relevant to emotion regulation (i.e., anxiety, depression). As hypothesized, of the subscales, Strategies shared the strongest correlations with the NMR and the AAQ-II. Additionally, TMMS-Attention and TMMS-Clarity shared the strongest correlations with the Identification subscale. Moreover, all of the subscales evidenced significant relations with DASS-21 Anxiety and Depression

General Discussion

Given the clinical relevance and potential transdiagnostic status of emotion regulation, it is not surprising that the DERS, a multidimensional self-report measure of emotion regulation, has been widely used. Through its frequent use, a pattern has emerged showing psychometric limitations related to the Awareness subscale of the DERS, as well as evidence that an overarching emotion regulation construct fails to account for the interrelations among the six DERS subscales, thus precluding use of a DERS total score based on summing all of the items from these lower-order factors. As noted, it is essential to establish that all scale items represent the same overarching factor when using a total score. Thus, given evidence to the contrary, the purpose of the present study was to (a) determine whether the described psychometric limitations of the DERS are due to a method effect (i.e., reverse-coded items) by rewording all of the reverse-coded items in a straightforward manner and submitting them to EFA, and (b) examine the tenability of an adaptation of the original measure.

Results from Study 1 support Bardeen et al.’s (2012) hypothesis that the reverse-coded items of the DERS contribute to the described psychometric limitations. Additionally, the results of Study 1 confirm Gratz and Roemer’s (2004) original conceptualization of emotion regulation in which one construct accounted for the identification and understanding of emotions (i.e., Identification). That is, the only distinguishing feature between the original DERS-Awareness and Clarity subscales appears to be that the Awareness subscale consists solely of reverse-coded items. Importantly, results from the CFA in Study's 2 and 3 suggest that rewording all of the reverse-coded items in a straightforward manner, and combining the Clarity and Awareness subscales as described above, result in a multidimensional factor structure in which all of the lower-order factors belong to the same higher-order emotion regulation construct, thus supporting use of a summed total score. Whereas Bardeen et al. (2012) found that the original DERS-Awareness subscale had a factor loading on the higher-order construct that was small in magnitude (i.e., .26), in the present study, the Identification subscale evidenced a factor loading on the higher-order construct that was large in magnitude (i.e., .64) and was consistent with the other four subscales. The pattern of strong scale intercorrelations among the five subscales is noteworthy, as prior studies have found that the original DERS-Awareness scale did not cluster strongly with the other five DERS subscales. Importantly, the adequacy of the modified factor structure was supported across independent samples, thus increasing confidence in the replicability of the factor structure of the modified version of the DERS.

The results of Study 3 provide additional psychometric support for the modified version of the DERS in terms of scores indicating reliability (i.e., internal consistency) and construct validity (i.e., convergent, discriminant, and criterion validity). In particular, the association between the Identification subscale and the TMMS-Attention and TMMS-Clarity subscales, as well as the association between the Identification subscale and the NMR, provides additional psychometric evidence in favor of combining the noted Clarity and Awareness items to form a single subscale. Specifically, as hypothesized by Gratz and Roemer (2004), one latent factor (i.e., Identification) assesses the identification and understanding of emotions rather than mood regulation expectancies. Conversely, the Strategies subscale appears to be assessing negative mood regulation expectancies (i.e., NMR) rather than the identification and understanding of emotions.

The associations among the modified subscales and criterion variables (i.e., depression and anxiety) provide additional psychometric evidence in favor of the modified version of the DERS, as dimensions of the same underlying construct should share a consistent pattern of relations with constructs that are theoretically relevant to the underlying construct. Whereas Neumann et al. (2010) found that the DERS-Awareness subscale was the only DERS subscale that was not significantly associated with measures of anxiety and depression, in Study 3, the Identification subscale evidenced a similar pattern of significant positive associations as the other subscales with depression and anxiety. These results provide evidence of criterion-related validity and highlight the importance of using the modified version of the DERS. Given the frequency with which DERS-Awareness exhibits a divergent pattern of correlations with criterion variables (e.g., Bardeen & Fergus, 2014; Gratz and Roemer, 2004; McDermott et al., 2009; Neumann et al., 2010; Salters-Pedneault et al., 2006; Soenke et al., 2010; Tull et al., 2007, 2010), use of the original DERS may lead to spurious conclusions about the relation between the identification of emotions and psychopathology. In other words, from studies in which the DERS has been used, one may draw the conclusion that the identification of negative emotions may be unnecessary for adaptive regulation of emotion. However, such a conclusion would beg the question, how does one strategically alter emotions without first identifying the emotional experience? It would seem that those who have difficulty identifying emotional experiences would be less likely to purposely up- or down-regulate emotional states in order to effectively direct behavior. As a consequence of failing to identify and strategically regulate emotions, such individuals would be more likely to experience prolonged negative affect and subsequent psychopathology. Having psychometrically sound measures of emotion identification and understanding may be particularly useful when assessing outcomes of psychological interventions for which the central tenet is that mindful awareness, including awareness of emotions, is essential for maintaining psychological well-being (e.g., dialectical behavior therapy: Linehan, 1993; acceptance and commitment therapy: Hayes et al., 1999).

Study limitations must be acknowledged. In the present study we were fundamentally interested in whether the original DERS could be strengthened by addressing concerns of its more basic properties (i.e., item wording, reverse-keyed items) that have been raised in the existing literature. However, the use of alternative techniques for refining the original DERS may be warranted in future research. For example, even though estimates from item response theory and classical test theory consistently correlate at .90 or higher (Reise & Waller, 2009), there may be interest in future research that examines possible differential item functioning among the original DERS items using item-response theory. In addition, in Study 1 we used the same factor loading criterion for item retention as Gratz and Roemer (2004). Although the .40 criterion has precedence in the extant literature, more conservative criteria for item retention exist (see Matsunaga, 2010), and thus, may be used in future research to examine the factor structure of the DERS.

The DERS has been widely used in both nonclinical and clinical samples. Although the use of three general population samples should be considered a relative strength of the present set of studies, future research should replicate study findings in clinical samples to ensure generalizability. Additionally, evidence of criterion-related validity was demonstrated with two closely related constructs (i.e., anxiety, depression). A wide variety of criterion variables should be used in future research to demonstrate a consistent pattern of associations to those observed in the present set of studies. Given that the original DERS has been used in both cross-sectional and longitudinal research (e.g., emotion regulation-related treatment-outcome research), it will be important to use the 29-item modified version of the DERS in similar longitudinal studies to determine its sensitivity to change and retest reliability. It will be especially important to examine the Identification subscale in relation to constructs which have been shown to be unrelated to the original DERS-Awareness subscale. In regard to the use of reverse-coded items to reduce the likelihood of response biases (e.g., yea- and nay-saying), other methods of data quality control have been identified (e.g., embedding catch questions in survey batteries) which do not result in the psychometric limitations produced by the use of reverse-coded items (e.g., Oppenheimer et al., 2009; Paolacci et al., 2010).

Gratz and Roemer (2004) provided a critical first step in improving the measurement of emotion regulation by developing a much needed multidimensional self-report questionnaire of this construct. It is only through the widespread use of the DERS that psychometric limitations of the DERS were identified. The present findings provide psychometric support for a 29-item modified version of the DERS in which all of the reverse-coded items of the original DERS are reworded in a straight forward manner. The use of the modified version of the DERS in future research will likely help researchers further explicate correlates of emotion regulation difficulties and determine the importance of directly targeting emotion regulation in treatment. Towards this aim, use of the modified version of the DERS may assist in providing researchers and clinicians alike with a more accurate assessment of emotion regulation.

References

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review, 30*, 217-237.

Allison, P. D. *(*2002*). Missing data. Thousand Oaks. CA: Sage.*

Asmundson, G. J. G., & Taylor, S. (2005). *It's not all in your head: How worrying about your health could be making you sick – and what you can do about it*. New York, NY: Guilford.

Bardeen, J. R., & Fergus, T. A. (2014). An examination of the incremental contribution of emotion regulation difficulties to health anxiety beyond specific emotion regulation strategies. *Journal of Anxiety Disorders, 28,* 394-401.

Bardeen, J. R., Fergus, T. A., & Orcutt, H. K. (2013). Experiential avoidance as a moderator of the relationship between anxiety sensitivity and perceived stress. *Behavior Therapy, 44,* 459-469.

Bardeen, J. R., Fergus, T. A., & Orcutt, H. K. (2012). An examination of the latent structure of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment, 34,* 382-392.

Bardeen, J. R., Kumpula, M. J., & Orcutt, H. K. (2013). Emotion regulation difficulties as a prospective predictor of posttraumatic stress symptoms following a mass shooting. *Journal of Anxiety Disorders, 27,* 188-196.

Behrend, T. S., Sharek, D. J., Meade, A. W. & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43,* 1–14.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K,…& Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action

Questionnaire-II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42,* 676–688.

Brown, T.A. (2006). *Confirmatory factor analysis for applied research*. New York: Guilford Press.

Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. *Sage Focus Editions*, *154*, 136-136.

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.

Catanzaro, S. J. (1994). Discrimination of mood regulation expectancies from dysphoria: Confirmatory factor analytic findings. *Assessment, 1,* 53-58.

Catanzaro, S. J., & Mearns, J. (1990). Measuring generalized expectancies for negative mood regulation: Initial scale development and implications. *Journal of Personality Assessment, 54,* 546–563.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233-255.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7,* 309-319.

Dalal D. K. & Carter, N. T. (2015). Negatively worded items negatively impact survey research. In C. E. Lance, & R. J. Vandenberg (Eds.), *More Statistical and Methodological Myths and Urban Legends* (pp. 112-132). New York: Taylor & Francis.

Enders, C. K. (2010). *Applied missing data analysis.* New York: Guilford Press.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Fox, H. C., Hong, K. A., & Sinha, R. (2008). Difficulties in emotion regulation and impulse control in recently abstinent alcoholics compared with social drinkers. *Addictive Behaviors, 33,* 388-394.

Fowler, J. C., Charak, R., Elhai, J. D., Allen, J. G., Frueh, B. C., & Oldham, J. M. (2014). Construct validity and factor structure of the Difficulties in Emotion Regulation Scale among adults with severe mental illness. *Journal of Psychiatric Research, 58,* 175-180.

Gearhardt, A. N., White, M. A., Masheb, R. B., Morgan, P. T., Crosby, R. D., & Grilo, C. M. (2012). An examination of the food addiction construct in obese patients with Binge Eating Disorder. *International Journal of Eating Disorders, 45,* 657-663.

Giromini, L., Velotti, P., de Campora, G., Bonalume, L., & Zavattini, G. C. (2012). Cultural adaptation of the difficulties in emotion regulation scale: Reliability and validity of an Italian version. *Journal of Clinical Psychology, 68,* 989-1007.

Gratz, K. L., Levy, R., & Tull, M. T. (2012). Emotion regulation as a mechanism of change in an acceptance-based emotion regulation group therapy for deliberate self-harm among women with borderline personality pathology. *Journal of Cognitive Psychotherapy, 26,* 365-380.

Gratz, K. L. & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment, 26,* 41-54.

Gratz, K. L., Rosenthal, M., Tull, M. T., Lejuez, C. W., & Gunderson, J. G. (2006). An experimental investigation of emotion dysregulation in borderline personality disorder. *Personality Disorders: Theory, Research, and Treatment, 1,* 18-26.

Hayes, S. C., Strosahl, K., & Wilson, K. G. (1999). *Acceptance and commitment therapy: An experiential approach to behavior change.* New York, NY: Guilford Press.

Henry, J. D., & Crawford, J. R. (2005). The short-form version of the Depression Anxiety Stress Scales (DASS-21): Construct validity and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 44,* 227-239.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Jöreskog, K. G., & Sörbom, D. (2006). LISREL (Version 8.80) [Computer software]. Chicago: Scientific Software Inc.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement, 20,* 141-151.

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed.). New York: Guilford Press.

Kokonyei, G., Urban, R., Reinhardt, M., Jozan, A., & Demetrovics, Z. (2014). The Difficulties in Emotion Regulation Scale: Factor structure in chronic pain patients. Journal of Clinical Psychology, 70, 589-600.

Linehan, M. (1993). *Cognitive–behavioral treatment of borderline personality disorder.* New York: Guilford Press.

Lovibond, S. H., & Lovibond, P. F. (1995a). *Manual for the Depression Anxiety Stress Scales.* (2nd. Ed.) Sydney: Psychology Foundation.

Matsunaga, M. (2010). How to factor-analyze your data right: Do's and don'ts, and how-to's. *International Journal of Psychological Research, 3*, 97-110.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*, 130-149.

McDermott, M. J., Tull, M. T., Gratz, K. L., Daughters, S. B., & Lejuez C. W. (2009). The role of anxiety sensitivity and difficulties in emotion regulation in posttraumatic stress disorder among crack/cocaine dependent patients in residential substance abuse treatment. *Journal of Anxiety Disorders*, *23*, 591-599.

Meade, A. W., Johnson, E. C., & Braddy, P. W. (2008). Power and sensitivity of alternative fit indices in tests of measurement invariance. *Journal of Applied Psychology*, *93*, 568-592.

Meng, X. L., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111,* 172-175.

Meyers, L. S., Gamst, G., & Guarino, A. J. (2006). *Applied multivariate research: Design and interpretation*. Sage.

Neumann, A., van Lier, P. A. C., Gratz, K. L., & Koot, H. M. (2010). Multidimensional assessment of emotion regulation difficulties in adolescents using the Difficulties in Emotion Regulation Scale. *Assessment, 17*, 138-149.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behaviour Research Methods and Instrumentation, 32,* 396–402.

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.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk. *Judgment and Decision Making, 5,* 411–419.

Perez, J., Venta, A., Garnaat, S., & Sharp, C. (2012). The difficulties in Emotion Regulation Scale: Factor structure and association with nonsuicidal self-injury in adolescent inpatients. *Journal Of Psychopathology And Behavioral Assessment*, *34*(3), 393-404.

Rodebaugh, T. L., Woods, C. M., & Heimberg, R. G. (2007). The reverse of social anxiety is not always the opposite: The reverse-scored items of the Social Interaction Anxiety Scale do not belong. *Behavior Therapy, 38*, 192-206.

Rusch, S., Westermann, S., and Lincoln, T. M. (2011). Specificity of emotion regulation deficits in social anxiety: An internet study. *Psychology and Psychotherapy: Theory, Research and Practice*, *85*, 268-277.

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (1995). Emotional attention, clarity and repair: Exploring emotional intelligence using the trait meta-mood scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, and health* (pp. 125–154).Washington, DC: APA Books.

Salters-Pedneault, K., Roemer, L., Tull, M. T., Rucker, L., & Mennin, D. S. (2006). Evidence of broad deficits in emotion regulation associated with chronic worry and generalized anxiety disorder. *Cognitive Therapy and Research, 30,* 469-480.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variable* *analysis: Applications for developmental research* (pp. 399-419). Thousand Oaks, CA: Sage.

Soenke, M., Hahn, K. S., Tull, M. T., & Gratz, K. L. (2010). Exploring the relationship between childhood abuse and analogue generalized anxiety disorder: The mediating role of emotion dysregulation. *Cognitive Therapy and Research, 34,* 401-412.

Tull, M. T., Barrett, H. M., McMillan, E. S., & Roemer, L. (2007). A preliminary investigation between emotion regulation difficulties and posttraumatic stress symptoms. *Behavior Therapy, 38,* 303-313.

Tull, M. T., Gratz, K. L., Latzman, R. D., Kimbrel, N. A., & Lejuez, C. W. (2010). Reinforcement Sensitivity Theory and emotion regulation difficulties: A multimodal investigation. *Personality and Individual Differences, 49,* 989-994.

Tull, M. T., Stipleman, B. A., Salters-Pedneault, K., & Gratz, K. L. (2009). An examination of recent non-clinical panic attacks, panic disorder, anxiety sensitivity, and emotion regulation difficulties in the prediction of generalized anxiety disorder in an analogue sample. *Journal of Anxiety Disorders, 23,* 275-282.

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41,* 321–327.

Weinberg, A., & Klonsky, E.D. (2009). Measurement of emotion dysregulation in adolescents. *Psychological Assessment, 21,* 616-621.

Wang, M., & Russell, S. S. (2005). Measurement equivalence of the job descriptive index across Chinese and American workers: Results from confirmatory factor analysis and item response theory. *Educational and Psychological Measurement, 65*, 709-732.

Weeks, J. W., Heimberg, R. G., Fresco, D. M., Hart, T. A., Turk, C. L., Schneier, F. R., et al. (2005). Empirical validation and psychometric evaluation of the Brief Fear of Negative Evaluation Scale in patients with social anxiety disorder. *Psychological Assessment, 17*, 179-190.

Wenzlaff, R. M., & Luxton, D. D. (2003). The role of thought suppression in depressive rumination. *Cognitive Therapy and Research, 27*, 293-308.

Table 1

Factor Loadings for the pool of modified DERS items.

	Factor				
	1	2	3	4	5
1C	.03	**.62**	.04	.16	-.01
4C	.11	**.77**	-.06	-.01	.08
5C	-.03	**.72**	.06	.17	.06
7C	.01	**.85**	-.04	.03	.05
9C	-.01	**.67**	.12	.15	.01
6A	.01	**.87**	-.02	-.06	-.01
8A	.10	**.69**	-.01	-.02	-.02
2A	-.00	**.87**	-.01	-.08	-.06
10A	-.09	**.75**	.09	-.03	.02
17A	.09	.28	.35	-.07	.21
34A	.12	**.57**	.04	.06	.15
3I	**.50**	.05	.10	.18	.10
14I	**.88**	.03	-.02	-.02	.02
19I	**.81**	-.02	.05	.08	.01
24I	**.63**	.06	.14	.07	.05
27I	**.85**	.03	.03	.06	-.02
32I	**.86**	.04	-.03	-.05	.11
11N	.14	.13	**.68**	.04	-.11
12N	.06	.01	**.85**	-.02	-.06
21N	.03	-.02	**.88**	-.00	.03
23N	-.07	.04	**.50**	.10	.29
25N	-.01	.05	**.86**	-.04	.03
29N	.01	-.01	**.77**	.08	.07
13G	.02	.06	-.01	**.88**	-.05
33G	.06	.01	-.03	**.70**	.20
18G	.08	-.04	.03	**.82**	.02
20G	.03	.04	.00	**.86**	-.00

26G	-.01	.03	.03	**.89**	.00
15S	.19	-.07	.04	.09	**.60**
16S	.07	-.04	.13	.02	**.68**
22S	.06	.07	.03	-.03	**.77**
28S	.12	.08	.02	-.02	**.75**
30S	-.08	-.02	**.44**	.16	**.43**
31S	.08	.12	.01	.04	**.66**
35S	.00	.10	-.05	.16	**.71**
36S	.26	-.07	.11	.32	.34

Note: $N = 743$ (Sample 1). Items loading on each factor are bolded. C = DERS Clarity item; A = DERS Awareness item; I = DERS Impulse item; N = DERS Nonacceptance item; G = DERS Goals item; S = DERS Strategies item.

Table 2

Bivariate Correlations

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Total	--												
2. M-Identification	.68	--											
3. M-Impulse	.88	.55	--										
4. M-Nonacceptance	.84	.50	.66	--									
5. M-Goals	.81	.37	.66	.58	--								
6. M-Strategies	.89	.48	.76	.68	.71	--							
7. AAQ-II	.74	.39	.61	.65	.63	.71	--						
8. DASS-Anxiety	.55	.34	.51	.49	.41	.50	.57	--					
9. DASS-Depression	.66	.35	.52	.58	.54	.68	.72	.66	--				
10. NMR	-.67	-.33	-.54	-.51	-.59	-.74	-.68	-.40	-.63	--			
11. TMMS-Attention	-.28	-.42	-.21	-.22	-.11	-.23	-.21	-.16	-.22	.30	--		
12. TMMS-Clarity	-.53	-.56	-.47	-.44	-.34	-.39	-.48	-.36	-.38	.44	.43	--	
13. TMMS-Repair	-.55	-.28	-.46	-.42	-.45	-.63	-.60	-.33	-.62	.74	.38	.43	--

Note. *N* = 918. All correlations are significant at *p* < .001. M = scale items have been modified; AAQ-II = Acceptance and Action Questionnaire-II; DASS = Depression, Anxiety, and Stress Scale-21 item Version; NMR = Generalized Expectancy for Negative mood Regulation Scale; TMMS = Trait Meta Mood Scale.