A Factor Analytic Evaluation of the Difficulties in Emotion Regulation Scale

Daniel J. Lee, Tracy K. Witte, Joseph R. Bardeen, Margaret T. Davis, and Frank W. Weathers

Auburn University

Objectives: The present study aimed to elucidate the factor structure of the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004)–a widely used measure of emotion dysregulation. **Method:** Participants were 3 undergraduate samples (N = 840, 78.33% female, mean age = 20.30). **Results:** We began by using confirmatory factor analysis (CFA) to examine 3 existing models, finding that none consistently demonstrated adequate fit across samples. Subsequently, we conducted an exploratory factor analysis, identifying a novel 5-factor model that consistently resulted in adequate fit across samples. We also ran several CFA models after removing the Awareness subscale items–which have performed inconsistently in prior research–finding that a reduced-measure variant of the model retained by Gratz and Roemer (2004) resulted in adequate fit across samples. No higher-order models consistently resulted in adequate fit across samples. **Conclusions:** Our findings are consistent with previous work in suggesting use of a DERS total score may not be appropriate. Additionally, further work is needed to examine the novel 5-factor model and the effect of reverse-scored items on the DERS factor structure. J. Clin. Psychol. 00:1–14, 2016.

Keywords: difficulties in emotion regulation; emotion dysregulation; emotion regulation; psychometrics; factor analysis

The Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004) is a multidimensional self-report measure that assesses several related domains of emotion dysregulation. This measure is largely based on Linehan's conceptual framework of emotion dysregulation (Linehan, 1993; Linehan, Bohus, & Lynch, 2007). Within this framework, emotion dysregulation involves difficulty modulating negative affect, difficulty modulating behavior while experiencing negative affect, limited emotional awareness and clarity, and nonacceptance of emotional responding.

Since its development, the DERS has been used in a variety of populations, and the DERS total score has been found to be associated with a range of mental disorders and related constructs. Specifically, emotion dysregulation, as operationalized by the DERS total score, has been found to be associated with borderline personality disorder (BPD; e.g., Bornovalova et al., 2008), substance abuse (e.g., Fox, Axelrod, Paliwal, Sleeper, & Sinha, 2007), anorexia nervosa (e.g., Harrison, Sullivan, Tchanturia, & Treasure, 2009), generalized anxiety disorder (e.g., Roemer, Lee, Salters-Pedneault, Erisman, Orsillo, & Menin, 2009), and posttraumatic stress disorder (PTSD; e.g., Bardeen, Kumpula, and Orcutt, 2013).

Additionally, the DERS total score has been found to be associated with a variety of related constructs including chronic worry (Salters-Pedneault, Roemer, Tull, Rucker, & Mennin, 2006), deliberate self-harm (Gratz & Roemer, 2008), marijuana use as a coping mechanism (Bonn-Miller, Vujanovic, & Zvolensky, 2008), and negative affect (e.g., Tull, Barrett, McMillan, & Roemer, 2007). Further, the DERS total score has been found to be sensitive to treatment effects. For example, changes in DERS scores over the course of treatment were found be to be associated with changes in substance abuse frequency in individuals with co-occurring substance dependence and BPD following a dialectical behavior therapy intervention (Axelrod, Perepletchikova, Holtzman, & Sinha, 2011). Likewise, another study found that successful inpatient treatment for cocaine dependence resulted in significant reductions in DERS total scores

Please address correspondence to: Daniel J. Lee, Department of Psychology, 226 Thach Hall, Auburn University, AL 36849-5214. E-mail: djlee@auburn.edu

Published online in Wiley Online Library (wileyonlinelibrary.com/journal/jclp). DOI: 10.1002/jclp.22297

(Fox et al., 2007). Collectively, these findings suggest that the DERS total score is associated with a host of mental disorders and related constructs and is also responsive to intervention.

Despite the widespread use of the DERS, existing research has raised a number of concerns regarding its factor structure. In their initial measure development article, Gratz and Roemer (2004) hypothesized that a four-factor lower-order model would provide the best fit to the data based on their theoretical framework of emotion dysregulation. This theoretical model specified that items would load onto four factors: nonacceptance of emotional responses while distressed (Nonacceptance); limited access to emotion regulation strategies perceived as effective while distressed (Strategies); limited awareness and understanding of emotions (Awareness/Clarity); and lack of ability to engage in goal-directed behavior and inhibit impulsive behavior while distressed (Goals/Impulse; see Figure 1).

However, when they conducted an exploratory factor analysis (EFA), they found that either a correlated six- or seven-factor solution best fit the data. Gratz and Roemer (2004) decided to retain the six-factor solution because it was more easily interpreted. Although this six-factor model included the predicted Nonacceptance and Strategies factors, it indicated that difficulties engaging in goal-directed behavior while distressed (Goals) and impulse control difficulties (Impulse) represented distinct factors, and that lack of emotional awareness (Awareness) and lack of emotional clarity (Clarity) represented distinct factors.

Several studies have used confirmatory factor analysis (CFA) to test the correlated six-factor lower-order model retained by Gratz and Roemer (2004; Bardeen, Fergus, & Orcutt, 2012; Fowler et al., 2014; Kökönyei, Urbán, Reinhardt, Józan, & Demetrovics, 2014; Snow, Ward, Becker & Raval, 2013). The fit of this model across studies can be evaluated according to standard guidelines in the CFA literature. Hu and Bentler (1999) and Kline (2011) proposed the following criteria for determining adequate fit: nonsignificant χ^2,root mean square error of approximation (RMSEA) $\leq .05$, comparative fit index (CFI), and Tucker-Lewis index (TLI) $\geq .95$.

Others have suggested more liberal criteria for evaluating model fit, with RMSEA values $\leq .10$ (Browne, Cudeck, & Bollen, 1993; Meyers, Gamst, & Guarino, 2006) and CFI and TLI values $\geq .90$ (MacCallum, Browne, & Sugawara, 1996; Meyers et al., 2006) being viewed as adequate. Even when using more liberal criteria to evaluate model fit, the correlated six-factor lower-order model of the DERS has not consistently resulted in adequate fit (e.g., Kökönyei et al., 2014; RMSEA = .11, CFI = .76, TLI = .74). Further, several studies have found adequate fit only after conducting post hoc modifications of the models (e.g., allowing cross-loadings, removing items; Kökönyei et al., 2014; Neumann, van Lier, Gratz, & Koot; 2010; Perez, Venta, Garnaat, & Sharp, 2012). These findings raise concern about the acceptability of the six-factor lower-order model.

As noted earlier, researchers often sum all of the DERS items to create a DERS total scale score when assessing emotion dysregulation, which implies that a higher-order model would be appropriate. Accordingly, two studies have examined a CFA model with a higher-order Emotion Dysregulation factor and six lower-order factors corresponding to Gratz and Roemer's (2004) retained model. Although Bardeen and colleagues (2012) found that this model resulted in adequate fit (RMSEA = .07, CFI = .97, TLI = .96), Fowler and colleagues (2014) did not (RMSEA = .12, CFI = .83, TLI = .82). Thus, consistent evidence supporting this six-factor higher-order model is lacking.

To further explore the factor structure of the DERS, Weinberg and Klonsky (2009) used EFA in a large, community adolescent sample. Similar to Gratz and Roemer (2004), they found that both a six- and seven-factor solution fit the data. Weinberg and Klonsky (2009) 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). They too opted to retain the six-factor solution. Nevertheless, even in the model they retained, all of the reverse coded items loaded, or at least cross-loaded, on the same factor. This suggests that a method effect still may have been overlooked.

As noted by Clark and Watson (1995), when combining subscale scores to create a total score, it is essential to ensure that all factors represent a single general construct. This is particularly important in regard to the DERS because a number of studies have shown that the Awareness subscale does not consistently correlate with other DERS subscales and shows a differential

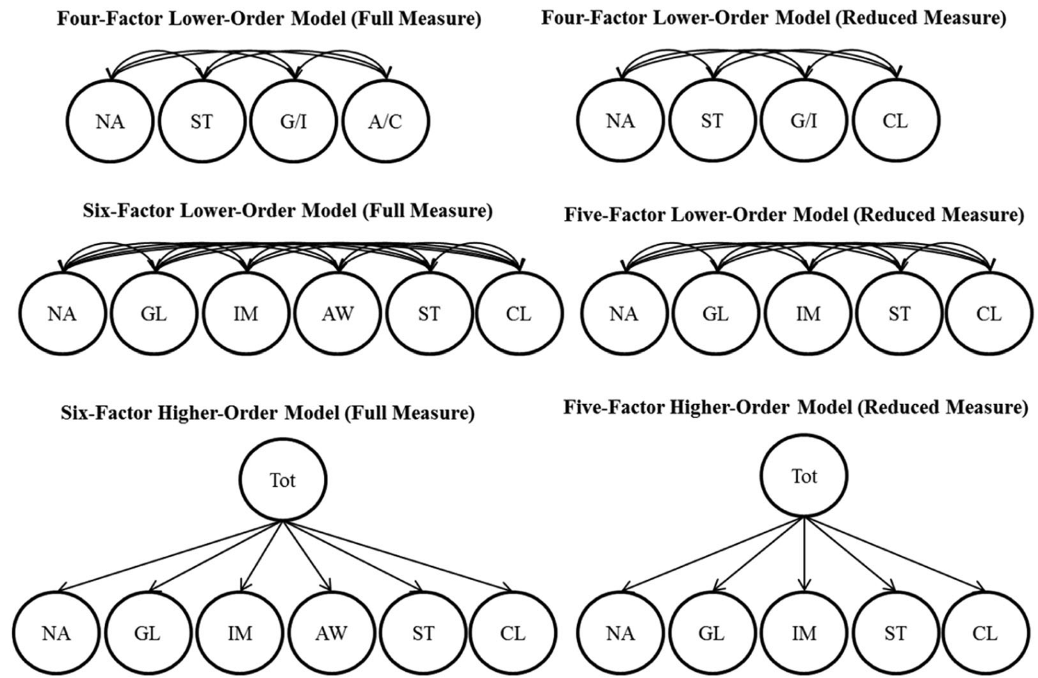

Figure 1. Examined measurement models
Note. A/C = limited awareness and understanding of emotions; AW = Lack of Emotional Awareness subscale; CL = Lack of Emotional Clarity subscale; G/I = lack of ability to engage in goal-directed behavior and inhibit impulsive behavior while distressed; GL = Difficulty Engaging in Goal-Directed Behavior While Distressed subscale; IM = Impulse Control Difficulties While Distressed subscale; NA = Nonacceptance of Emotional Responses While Distressed subscale; ST = Limited Access to Emotion Regulation Strategies Perceived as Effective While Distressed subscale; Tot = DERS total score; Reduced Measure = Awareness subscale items removed.

pattern of associations–compared to the five other subscales–with criteria of interest (see Bardeen et al., 2012, for a more detailed discussion of this issue). As such, the Awareness subscale of the DERS may not be assessing the same overarching construct as the other subscales of the DERS.

To test this hypothesis, Bardeen and colleagues (2012) used CFA in a large university sample and found that the Awareness subscale showed (a) much smaller correlations with the other DERS subscales relative to the correlations among the other DERS subscales and (b) a significantly lower factor loading on a higher-order emotion dysregulation construct than the other five DERS subscales. Bardeen et al. (2012) proposed two possibilities regarding their findings. First, the Awareness subscale may not belong to the same higher-order emotion dysregulation construct as the other factors of the DERS. Second, because the Awareness scale is the only DERS subscale for which all items are reverse-keyed, the likelihood of systematic error may be increased, resulting in the idiosyncratic pattern of associations (e.g., see Rodebaugh, Woods, & Heimberg, 2007; Weeks et al., 2005).

Advancing this line of research, several studies have examined a reduced version of the DERS with the Awareness subscale items removed (e.g., Bardeen et al., 2012; Fowler et al., 2014). Bardeen and colleagues found that a both five-factor lower-order model (RMSEA = .06, CFI and TLI = .98) and a higher-order model (RMSEA = .06, CFI and TLI = .98) resulted in adequate fit. Likewise, Fowler and colleagues (2014) found evidence of adequate fit for the five-factor lower-order model (RMSEA = .10, CFI and TLI = .92). Additionally, although Fowler et al. (2014) attempted to examine the higher-order model, it failed to converge.

Despite replicated factor analytic examination across multiple populations, none of the measurement models described above have consistently resulted in adequate fit. Additionally, to our knowledge, no studies have examined Gratz and Roemer's (2004) four-factor model using CFA. Thus, the literature to date suggests that the latent factor structure of the DERS has not been firmly established and that further investigation is needed. Therefore, in the present study, we began by using CFA to examine the fit of the most prominent models of the DERS. Anticipating the possibility that none of these models would consistently result in adequate fit, we also planned to implement EFA to further explore the factor structure of the DERS. To evaluate the replicability of results, analyses were conducted in three independent samples.

Method

Participants and Procedure

Participants were undergraduate students enrolled in psychology courses at a large university in the southeastern United States who took part in one of two studies examining various aspects of posttraumatic stress disorder and emotion regulation. Results from one of these studies have been published elsewhere (Lee, Witte, Weathers, & Davis, 2015) and do not overlap with the current study. In both studies, participants were invited to participate in research examining individual differences in response to stressful life events and how individuals manage distress. Participants completed an online survey that included the DERS and other instruments not included in the present study, and they were compensated with course extra credit. Study procedures were approved by the university's institutional review board.

Sample 1 comprised 322 participants who completed at least the demographics questionnaire in the first study. Of these participants, 246 (76.40%) identified as female, 280 (86.96%) identified as Caucasian, 31 (9.63%) identified as African American, 5 (1.55%) identified as Asian American, 5 (1.55%) identified as another race, and 1 (0.31%) did not report race. The majority of participants reported being enrolled in school full-time (306; 95.03%) and identified as single (213; 66.15%). The mean age of the sample was 20.66 (standard deviation [*SD*] = 2.06, range = 18–38).

Five hundred and eighteen participants completed a second study. Of these participants, 412 (79.54%) identified as female, 456 (88.03%) identified as Caucasian, 36 (6.95%) identified as African American, and 26 (5.02%) identified as another race. The majority of participants reported being enrolled in school full-time (508; 98.07%) and identified as single (364; 70.28%). The mean age of these participants was 19.93 (*SD* = 2.23, range = 18–44). This larger sample was randomly halved to create Samples 2 and 3 ($N = 259$ in each). The purpose of splitting participants in the second study into two samples was to establish a third comparison group for analyses. Within CFA, multiple groups can be used to cross-validate measurement models. Each of these groups was adequately powered to conduct the planned analyses (see power analysis below). Accordingly, splitting the second study into two samples provides an additional sample to cross-check measurement models.

Measures

Participants completed a demographics questionnaire (e.g., sex, age, race) and the DERS (Gratz & Roemer, 2004). The DERS is a 36-item, self-report measure of several related aspects of emotion dysregulation. Participants designate what percent of the time each item applies to them on a 5-point scale ranging from 1(*almost never, 0–10%)* to 5 (*almost always, 91–100%)*.

Data Analytic Strategy

All analyses were conducted with Mplus version 6 (Muthén & Muthén, 1998–2011). In the first step of analyses, three competing measurement models of the full 36-item DERS were submitted to CFA in all three samples (see Figure 1). First, we examined Gratz and Roemer's (2004) hypothesized four-factor lower-order model described above because, to our knowledge, no studies have examined this model within a CFA framework. Second, we examined the six-factor lower-order model retained by Gratz and Roemer (2004). Third, we examined a hierarchical model with one higher-order emotion dysregulation factor and six lower-order factors corresponding to the aforementioned lower-order model.

In the event that no measurement models consistently resulted in adequate fit across samples, we planned to submit the full 36-item DERS to EFA in Sample 1 using an oblique rotation (geomin). Oblique rotation was selected because replicated findings have demonstrated correlation among DERS latent variables (see Fowler et al., 2014). As such, constraining DERS factors to be uncorrelated through the use of orthogonal rotation would not be appropriate (Fabrigar, Wegener, MacCallum, & Strahan, 1999). The Geomin form of oblique rotation was used because it tends to generalize to CFA well. However, it should be noted that most oblique rotation methods tend to produce similar findings (Fabrigar et al., 1999). Solutions including up to 10 factors (a conservative figure given that existing EFA research on the DERS has never supported measurement models comprising more than six factors; see Fowler et al., 2014, for a recent review) would be examined using eigenvalues, fit statistics, and patterns of factor loadings. An identified model using EFA procedures would then be submitted to CFA using Samples 2 and 3. In the event that salient cross-loadings are identified, these items would be permitted to cross-load in subsequent CFAs, rather than fixing meaningful cross-loadings to zero (see Asparouhov & Muthén, 2009 for a discussion).

The next set of analyses focused on replicating previous analyses in which the six items corresponding to the Awareness subscale were removed. We compared three competing measurement models of the 30-item DERS (see Figure 1). The first model was a four-factor lower-order model based on the Gratz and Roemer's (2004) a priori hypothesized model. Second, we examined a five-factor lower-order model retained by Gratz and Roemer (2004) with the Awareness subscale removed. Third, we examined a hierarchical five-factor model with one higher-order emotion dysregulation factor and five lower-order factors corresponding to the aforementioned five-factor model.

Individual items were used as indicators in all models. For all analyses, DERS items were treated as ordinal, rather than continuous, because of the small number of response options and because individual items were not normally distributed (Flora & Curran, 2004; Wirth & Edwards, 2007). Accordingly, parameters were estimated with the mean- and variance-adjusted weighted least squares (WLSMV) estimator, which provides a robust χ^2. Model fit was assessed using multiple indices (i.e., χ^2, TLI, CFI, and RMSEA) and examination of all fit statistics for each model (Brown, 2006). Nested models were compared using the DIFFTEST command

Table 1
Fit Indices for DERS Measurement Models Examined using CFA

Model/sample	χ^2	*df*	*p*	CFI	TLI	RMSEA (90% CI)
			Full measure			
Correlated four-factor lower-order model						
Sample 1	2343.03	588	< .001	.91	.91	.10 [.09, .10]
Sample 2	2434.63	588	< .001	.85	.84	.12 [.11, .12]
Sample 3	2054.38	588	< .001	.88	.87	.10 [.10, .11]
Correlated six-factor lower-order model						
Sample 1	1517.54	579	< .001	.95	.95	.07 [.07, .08]
Sample 2	1857.04	579	< .001	.90	.89	.10 [.09, .10]
Sample 3	1674.82	579	< .001	.91	.90	.09 [.08, .09]
Six-factor higher-order model						
Sample 1	2258.71	588	< .001	.92	.91	.10 [.09, .10]
Sample 2	2630.22	588	< .001	.83	.82	.12 [.12, .13]
Sample 3	2103.93	588	< .001	.87	.86	.10 [.10, .11]
EFA-derived five-factor model						
Sample 2	1371.14	578	< .001	.94	.93	.08 [.07, .08]
Sample 3	1409.01	578	< .001	.93	.92	.08 [.07, .08]
			Reduced measure (Awareness subscale items removed)			
Correlated five-factor lower-order model						
Sample 1	959.06	395	< .001	.97	.97	.07 [.06, .07]
Sample 2	1110.04	395	< .001	.94	.93	.09 [.08, .09]
Sample 3	1121.48	395	< .001	.94	.93	.09 [.08, .09]
Correlated four-factor lower-order model						
Sample 1	1477.45	399	< .001	.94	.94	.09 [.09, .10]
Sample 2	1418.59	399	< .001	.91	.90	.10 [.10, .11]
Sample 3	1454.15	399	< .001	.91	.90	.10 [.10, .11]
Five-factor higher-order model						
Sample 1	1577.26	401	< .001	.94	.93	.10 [.09, .10]
Sample 2	1389.33	401	< .001	.91	.90	.10 [.10, .11]
Sample 3	1611.78	401	< .001	.90	.89	.11 [.11, .12]

Note. df = degree of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation.

in Mplus, which allows for comparison of nested models while using the WLSMV estimator (Brown, 2006).

Missing data were handled using pairwise deletion because of the small portion of the data missing (i.e., covariance coverage ranged from .98 to 1.00 for all item pairs), absence of techniques for pooling most fit indices across estimates in multiple imputation (Enders, 2010), and inability to use full information maximum likelihood procedures with the WLSMV estimator. An a priori power analysis was conducted according to criteria established by MacCallum and colleagues (1996). Specifically, for all models tested, all three samples exceeded the identified minimum sample sizes to achieve .80 power using the test of not-close fit.

Results

Full Measure CFA

None of the measurement models examined using all 36 items of the original DERS consistently resulted in adequate fit across samples (see Table 1). The six-factor lower-order model appeared to provide the best fit to the data. Using relatively liberal criteria for model fit (RMSEA values $\leq .10$; CFI and TLI values $\geq .90$), this model demonstrated adequate fit in Samples 1 and 3; however, it did not demonstrate adequate fit in Sample 2. By comparison, neither the six-factor

Table 2
Zero-Order Correlations Among DERS Latent Variables in the Six-Factor Lower-Order CFA Model in Each Sample

	Sample	Nonacceptance	Goals	Impulse	Awareness	Strategies
Goals	1	.61*	–	–	–	–
	2	.59*	–	–	–	–
	3	.50*	–	–	–	–
Impulse	1	.67*	.39*	–	–	–
	2	.70*	.65*	–	–	–
	3	.62*	.66*	–	–	–
Awareness	1	.30*	.05	.22*	–	–
	2	.07	.25*	.15*	–	–
	3	.22*	.13*	.21*	–	–
Strategies	1	.82*	.56*	.59*	.24*	–
	2	.80*	.65*	.84*	.08	–
	3	.74*	.70*	.87*	.17*	–
Clarity	1	.57*	.32*	.32*	.52*	.39*
	2	.36*	.24*	.47*	.69*	.44*
	3	.48*	.39*	.56*	.67*	.55*

Note. DERS = Difficulties in Emotion Regulation Scale; CFA = confirmatory factor analysis; Awareness = Lack of Emotional Awareness subscale; Clarity = Lack of Emotional Clarity subscale; Goals = Difficulty Engaging in Goal-Directed Behavior While Distressed subscale; Impulse = Impulse Control Difficulties While Distressed subscale; Nonacceptance = Nonacceptance of Emotional Responses While Distressed subscale; Strategies = Limited Access to Emotion Regulation Strategies Perceived as Effective While Distressed subscale.
$^{*}p < .05$.

higher-order model nor the four-factor lower-order model resulted in adequate fit in any sample. Results of chi-square difference testing demonstrated that the six-factor lower-order model fit significantly better than the six-factor higher-order model in Samples 1 ($\chi^2 = 256.14$, degree of freedom [*df]* = 9, $p < .001$), 2 ($\chi^2 = 294.87$, $df = 9$, $p < .001$), and 3 ($\chi^2 = 183.58$, $df = 9$, $p < .001$).

Additionally, the six-factor lower-order model fit significantly better than the four-factor lower-order model in Samples 1 ($\chi^2 = 264.11$, $df = 9$, $p < .001$), 2 ($\chi^2 = 253.38$, $df = 9$, $p < .001$), and 3 ($\chi^2 = 201.43$, $df = 9$, $p < .001$). Zero-order correlations among latent variables in the six-factor, lower-order model ranged from .05 to .82 (see Table 2). With the exception of the Awareness subscale, significant associations were observed between all subscales across all samples.

Full Measure EFA

Because none of the above-described measurement models consistently resulted in adequate fit across samples, DERS items were submitted to EFA using Sample 1. Eigenvalues were above 1.0 for solutions with one to five factors (eigenvalues for one- through six-factor models were 16.04, 4.60, 2.12, 1.95, 1.43, and 0.98, respectively), indicating a maximum of five factors should be extracted. Of these models, the five-factor solution offered the best fit ($\chi^2 = 950.50$, $df = 460$, $p < .001$; RMSEA = .06 90% confidence interval [.05, .06]; CFI = .98; TLI = .97) and was interpretable. In this model, all items had salient loadings (parameter estimates > .3; Kline, 2011) on at least one factor. Additionally, six items (items 1r, 7r, 17r, 20r, 22r, 30) had salient cross-loadings on a second factor (see Table 3). Rather than being removed, items with salient cross-loadings on a second factor were allowed to cross-load in subsequent analyses.

This revised five-factor lower-order model was similar to the original six-factor measurement model retained by Gratz and Roemer (2004) in a number of ways. First, Factors 2 and 5 were identical to the Clarity and Goals subscales in the original six-factor model, respectively.

Table 3
DERS Items GEOMIN Rotated Factor Loadings for Sample 1

Item	F1	F2	F3	F4	F5
1r. I am clear about my feelings[a]	**.608**	**.334**	.026	−.027	.138
2r. I pay attention to how I feel	**.878**	.146	−.056	.031	.023
6r. I am attentive to my feelings	**.783**	.152	−.041	.045	.017
7r. I know exactly how I am feeling[a]	**.497**	**.486**	−.156	.016	.200
8r. I care about what I am feeling	**.712**	.056	.022	.092	−.171
10r. When I'm upset, I acknowledge my emotions	**.579**	.044	.050	.119	−.153
17r. When I'm upset, I believe that my feelings are . . . [a]	**.560**	−.046	.021	**.397**	−.270
20r. When I'm upset, I can still get things done[a]	**.323**	−.029	.071	−.117	**.542**
22r. When I'm upset, I know that I can find a way . . . [a]	**.366**	−.024	**.515**	−.146	.074
34r. When I'm upset, I take time to figure what I'm . . .	**.557**	.035	.222	−.118	−.147
4. I have no idea how I am feeling	.111	**.807**	.121	.006	−.051
5. I have difficulty making sense out of my feelings	.043	**.817**	.083	.044	.030
9. I am confused about how I feel	.035	**.735**	.032	.022	.186
3. I experience my emotions as overwhelming . . .	−.240	.288	**.658**	.015	−.007
14. When I'm upset, I become out of control	.137	−.074	**.851**	.025	−.011
15. When I'm upset, I believe that I will remain . . .	−.009	−.089	**.627**	.238	.206
16. When I'm upset, I believe that I'll end up feeling . . .	.043	−.048	**.519**	.273	.236
19. When I'm upset, I feel out of control	.046	.040	**.790**	.088	−.017
24r. When I'm upset, I feel like I can remain in control . . .	.258	−.116	**.635**	−.156	.022
27. When I'm upset, I have difficulty controlling . . .	−.055	.123	**.958**	−.022	−.146
28. When I'm upset, I believe there is nothing I can do . . .	.144	−.040	**.633**	.234	.120
30. When I'm upset, I start to feel very bad about myself	.126	.017	**.302**	**.525**	.154
31. When I'm upset, I believe that wallowing in it is . . .	−.038	.070	**.555**	.228	.125
32. When I'm upset, I lose control over my behaviors	.031	.121	**.975**	−.007	−.198
35. When I'm upset, it takes me a long time to feel better	−.106	.074	**.539**	.190	.238
36. When I'm upset, my emotions feel overwhelming	−.091	.106	**.616**	.090	.284
11. When I'm upset, I become angry with myself . . .	.020	−.061	.037	**.785**	.081
12. When I'm upset, I become embarrassed . . .	.013	.037	−.074	**.898**	.043
21. When I'm upset, I feel ashamed with myself . . .	−.057	.129	.045	**.871**	−.088
23. When I'm upset, I feel like I am weak	−.003	.062	.138	**.573**	.111
25. When I'm upset, I feel guilty for feeling that way	−.001	.069	−.010	**.940**	−.103
29. When I'm upset, I become irritated with myself . . .	.032	−.011	.133	**.784**	.081
13. When I'm upset, I have difficulty getting work done	.067	−.027	.055	.120	**.782**
18. When I'm upset, I have difficulty focusing on . . .	−.021	.162	−.002	−.023	**.855**
26. When I'm upset, I have difficulty concentrating	−.007	.149	.034	.061	**.798**
33. When I'm upset, I have difficulty thinking about . . .	−.046	.053	.217	.072	**.707**

Note. DERS = Difficulties in Emotion Regulation Scale. Salient (parameter estimates >.3 and significant) loadings are bold.
[a]Cross-loaded item.

Factor 1 included all of the items corresponding to the original Awareness subscale, but it also included two items from the original Clarity subscale (the only two items from this subscale that are reverse-scored)–one item from the original Goals subscale and one item from the original Strategies subscale. Of note, Factor 1 exclusively comprised reverse-scored items and included all but one of the reverse-scored items on the full measure. Factor 3 comprised all of the items corresponding to the original Strategies and Impulse subscales. Accordingly, this factor appeared to index the perceived inability to successfully regulate both emotion and behavior while distressed. Factor 4 comprised all of the items corresponding to the original Nonacceptance subscale–one item from the original Awareness subscale and one item from the original Strategies subscale.

This five-factor lower-order model was then submitted to CFA using Samples 2 and 3 and resulted in adequate fit in both samples (see Table 1). Because this five-factor model is not

Table 4
Correlations Among the Latent Variables in the EFA-Derived Five-Factor Model as Examined Using CFA in Sample 2 (and Sample 3)

	F1	F2	F3	F4
F2	.53* (.62*)	–		
F3	.45* (.55*)	.76* (.69*)	–	
F4	.32* (.46*)	.66* (.69*)	.60* (.51*)	–
F5	.21* (.45*)	.41* (.47*)	.47* (.52*)	.49* (.43*)

Note. EFA = exploratory factor analysis; CFA = confirmatory factor analysis.
*$p < .05$.

Table 5
Zero-Order Correlations Among the Reduced 30-Item DERS Latent Variables in the Five-Factor Lower-Order CFA Model in Each Sample

	Sample	Nonacceptance	Goals	Impulse	Strategies
Goals	1	.61*	–	–	–
	2	.59*	–	–	–
	3	.50*	–	–	–
Impulse	1	.67*	.58*	–	–
	2	.70*	.64*	–	–
	3	.62*	.66*	–	–
Strategies	1	.82*	.75*	.87*	–
	2	.80*	.65*	.84*	–
	3	.74*	.70*	.87*	–
Clarity	1	.57*	.47*	.54*	.58*
	2	.39*	.26*	.50*	.46*
	3	.50*	.40*	.58*	.57*

Note. DERS = Difficulties in Emotion Regulation Scale; CFA = confirmatory factor analysis; Clarity = Lack of Emotional Clarity subscale; Goals = Difficulty Engaging in Goal-Directed Behavior While Distressed subscale; Impulse = Impulse Control Difficulties While Distressed subscale; Nonacceptance = Nonacceptance of Emotional Responses While Distressed subscale; Strategies = Limited Access to Emotion Regulation Strategies Perceived as Effective While Distressed subscale.
*$p < .05$.

nested within the other examined models, chi-square difference testing could not be conducted to compare fit. Zero-order correlations among latent variables in this model ranged from .21 to .76 (see Table 4). In contrast to the original six factor lower-order model, all of the latent variables were significantly correlated with one another.

Reduced Measure CFA

The next series of analyses focused on examining the latent factor structure of the reduced DERS with the six Awareness items removed. Using relatively liberal criteria to determine fit, the five-factor lower-order model resulted in adequate fit in all three samples. By comparison, the five-factor higher-order model did not consistently demonstrate adequate fit across samples, and the four-factor lower-order model had adequate fit only in Sample 1. Results of chi-square difference testing demonstrated that the five-factor lower-order model fit significantly better than the five-factor higher-order model in Samples 1 ($\chi^2 = 201.50$, $df = 6$, $p < .001$), 2 ($\chi^2 = 125.02$, $df = 6$, $p < .001$), and 3 ($\chi^2 = 192.73$, $df = 6$, $p < .001$).

Additionally, the five-factor lower-order model fit significantly better than the four-factor lower-order model in Samples 1 ($\chi^2 = 1441.98$, $df = 4$, $p < .001$), 2 ($\chi^2 = 122.46$, $df = 4$, $p < .001$), and 3 ($\chi^2 = 127.26$, $df = 4$, $p < .001$). Zero-order correlations among latent variables in the five-factor lower-order model ranged from .26 to .87 (see Table 5). Significant associations were observed between all five factors across all three samples.

Discussion

This study examined the latent factor structure of the DERS, the most widely used self-report measure of emotion dysregulation, in three independent undergraduate samples. Of the three measurement models of the full 36-item DERS compared using CFA, none consistently resulted in adequate fit across samples. Of these models, the original six-factor lower-order model retained by Gratz and Roemer (2004) appeared to result in the best fit to the data. However, because none of the examined measurement models consistently resulted in adequate fit across samples, DERS items were then submitted to EFA. The resulting measurement model, a five-factor solution, was then submitted to CFA using Samples 2 and 3, both of which resulted in adequate fit. In an effort to replicate previous results, the factor structure of a reduced 30-item DERS (with the six items of the original Awareness subscale removed) was examined. Of the three measurement models compared using CFA, the five-factor lower-order model retained by Gratz and Roemer (2004) with the Awareness subscale removed appeared to best fit the data and resulted in adequate fit across all three samples.

Results from this study are consistent with the existing literature in that the six-factor lower-order model of the full measure retained by Gratz and Roemer (2004) failed to consistently result in adequate fit across samples. Given the inconsistency with which this model has demonstrated adequate fit across studies, caution is warranted in using this scoring approach. Likewise, the six-factor higher-order model did not fit the data well in any of three samples, even after removing the Awareness subscale items. These results replicate findings of Fowler and colleagues (2014) and suggest that the DERS items do not represent a single-factor emotion dysregulation construct. Scale items should only be combined to create a total score once it has been demonstrated that all factors represent a single general factor (Clark & Watson, 1995). Given the now replicated findings that the six-factor higher-order model does not fit the data well, these results collectively suggest that this commonly used single-factor scoring method may not be appropriate.

The absence of an adequate higher-order model raises questions about how to interpret the observed associations between the DERS total score and mental disorders. For example, Bornovalova and colleagues (2008) used the DERS total score as a predictor of BPD diagnostic status among inner-city substance users in residential treatment. Although the DERS total score emerged as a significant predictor of diagnostic status, interpretation of this finding is challenging because factor analytic evidence of the DERS to date collectively indicates that emotion dysregulation as measured by the DERS is not a unitary construct. Accordingly, future work examining the association between the DERS and external correlates would benefit from use of DERS subscales, rather than a total score, to gain more meaningful insight into the function of specific dimensions of emotion dysregulation. Our understanding of relations between emotion dysregulation, as measured via the 36-item DERS total score, and variables of interest may be obscured because all lower-order DERS factors do not appear to represent the same overarching construct.

Findings from the present study are consistent with a wealth of previous research showing that, in comparison to the other five DERS subscales, the Awareness subscale has shown a divergent pattern of relations with criterion relevant to the emotion regulation domain and evidence that the Awareness subscale does not belong to the same higher-order emotion regulation construct (see Bardeen et al., 2012).

Additionally, as noted, the five-factor lower-order model with the Awareness subscale removed provided the best fit to the data across all three samples in the present study. As such, continued use of the Awareness subscale may lead one to draw erroneous conclusions about the awareness of negative emotional states in relation to maladaptive outcomes. For example, Tull et al. (2007) found that the Awareness subscale was the only DERS subscale not associated with PTSD and hypothesized that perhaps emotional awareness is not necessary for adaptive emotion regulation. However, it would seem necessary to first identify an emotional experience in order to successfully up- or down-regulate one's affective state. Thus, awareness of negative emotions may be particularly relevant in the context of the development of PTSD, and the results of Tull et al. (2007) may be better explained by the noted psychometric limitations of this subscale. This is only one example among many in which hypotheses were made regarding unexpected null findings as they relate to the Awareness subscale (see Bardeen et al. 2012).

To our knowledge, this is the first study to examine the fit of Gratz and Roemer's (2004) hypothesized four-factor lower-order model using CFA. Given the results using the six-factor model, it was not surprising that the more parsimonious four-factor lower-order model generally did not result in adequate fit. As such, this measurement model is not recommended for use.

Two measurement models that consistently demonstrated adequate fit across samples were the five-factor lower-order model of the reduced measure (i.e., Awareness subscale items removed) and the EFA-derived five-factor model of the full measure. Consistent adequate fit of the five-factor lower-order model of the reduced measure across samples is in keeping with previous research (e.g., Bardeen et al., 2012), including work demonstrating that the construct validity of this model remained strong in an inpatient sample (Fowler et al., 2014). These results collectively support this measurement model as having the strongest factor analytic support to date.

The EFA-derived five-factor model, which resulted in adequate fit across two samples, was similar to the original six-factor model. Two factors were identical to the Clarity and Goals subscales in the original six-factor model. However, this model integrated the Strategies and Goals subscales into a single factor indexing inability to successfully regulate both emotion and behavior while distressed and pulled together what appeared to be a method factor of reverse-scored items. This measurement model was distinct from the other solutions examined in that several items were allowed to cross-load. The fit of this model while allowing cross-loadings suggests several DERS items are not unique indicators for one latent variable. Although this EFA-derived model was validated in two independent samples in the current study, more work is needed to identify the effect of cross-loading DERS items in other samples, particularly clinical samples in which emotion dysregulation is more prevalent. Given that this model was the only examined measurement model of the full 36-item measure to consistently result in adequate fit across samples, this novel five-factor model warrants further examination.

Results from the current study provide some support for Bardeen and colleagues' (2012) suggestion that the psychometric limitations of the DERS may be due, in part, to a method effect. Specifically, results from the first EFA, using all 36 of the original DERS items, showed that 10 of the 11 reverse-keyed items from four of six of the subscales loaded on the same factor. This finding is consistent with evidence that reverse-keyed items reduce scale validity by increasing the likelihood of systematic error (Hinkin, 1995). The psychometric properties of self-report measures have been shown to improve when reverse-keyed items are removed, leading some researchers to advocate for not using reverse-keyed items (e.g., Rodebaugh et al., 2007; Weeks et al., 2005).

Of note, Bardeen, Fergus, Hannan, & Orcutt (in press) recently conducted a study in which they reworded all reverse-coded items of the DERS in a straightforward manner and subjected all 36 items to EFA. As predicted, the Awareness subscale was no longer identified as a distinct factor once reverse-coding was removed. Interestingly, and in line with Gratz and Roemer's initial hypothesis, all Awareness and Clarity items loaded on the same factor, thus resulting in a five-factor model. Bardeen et al.'s (in press) five-factor model evidenced adequate fit of a lower-order model and a higher-order model across multiple samples. Additionally, all factors clustered strongly within one another and showed a consistent pattern of significant relations with constructs theoretically relevant to emotion regulation. These findings provide evidence in support of the hypothesis that the psychometric limitations of the DERS are due, at least in part, to a method effect (i.e., use of reverse-coded items).

Results from the current study contribute to the existing literature examining the latent factor structure of the DERS in a number of ways. First, the commonly used six-factor lower-order and related higher-order measurement models of the full 36-item DERS failed to consistently result in adequate fit across samples. This is consistent with previous work described earlier and indicates a stable factor structure for the full 36-item measure is yet to be established. Second, the EFA-derived five-factor model demonstrated adequate fit across multiple samples and warrants further examination. In particular, examination of this model among clinical populations represents an important direction of future research.

Third, a higher-order DERS measurement model consistently failed to result in adequate fit. This is consistent with Fowler and colleagues' (2014) findings and suggests that use of a

DERS total score may not be appropriate. Finally, after removing the Awareness subscale items, a reduced five-factor lower-order model resulted in adequate fit across all three samples. This is the third study to demonstrate this finding (Bardeen et al., 2012; Fowler et al., 2014). As such, this measurement model appears to have the best empirical support of any existing models.

Limitations

The present study has a number of limitations. First, all three samples consisted of undergraduate students who were predominantly female and White. Although the majority of participants in Gratz and Roemer's (2004) examination of the DERS also identified as female and White, caution is warranted in generalizing study findings to more gender and racially diverse samples. As noted by Fowler and colleagues (2014), many of the published examinations of the DERS factor structure have been conducted using undergraduate samples that experience relatively limited emotion dysregulation compared to clinical samples. Accordingly, examination of the factor structure of the DERS in a clinical sample represents an important future direction in this line of research.

Conclusion

Despite the absence of a close-fitting measurement model, the DERS has been shown to predict the development and maintenance of a wide array of psychopathology and maladaptive outcomes, as well as response to intervention. Thus, future psychometric examinations of the DERS should focus on examining the role of reverse-scored items in factor structure of the measure as well as both convergent and predictive validity. Given that a measurement model fit the data well across multiple samples for the modified version of the DERS (Bardeen et al., in press), this version of the measure warrants further examination. Finally, future studies examining associations between the DERS and external correlates should focus on established subscales rather than a DERS total score.

References

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. Structural Equation Modeling: A Multidisciplinary Journal, 16, 397–438.

Axelrod, S. R., Perepletchikova, F., Holtzman, K., & Sinha, R. (2011). Emotion regulation and substance use frequency in women with substance dependence and borderline personality disorder receiving dialectical behavior therapy. The American Journal of Drug and Alcohol Abuse, 37, 37–42.

Bardeen, J. R., Fergus, T. A., Hannan, S. M., & Orcutt, H. K. (in press). Addressing psychometric limitations of the Difficulties in Emotion Regulation Scale through item modification. Journal of Personality Assessment.

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., Kumpala, 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.

Bonn-Miller, M. O., Vujanovic, A. A., & Zvolensky, M. J. (2008). Emotional dysregulation: Association with coping-oriented marijuana use motives among current marijuana users. Substance Use and Misuse, 43, 1656–1668.

Bornovalova, M. A., Gratz, K. L., Daughters, S. B., Nick, B., Delany-Brumsey, A. D., Lynch, T. R., . . . & Lejuez, C. W. (2008). A multimodal assessment of the relationship between emotion dysregulation and borderline personality disorder among inner-city substance users in residential treatment. Journal of Psychiatric Research, 42, 717–726.

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.

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

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.

Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. Psychological Methods, 9, 466.

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.

Fox, H. C., Axelrod, S. R., Paliwal, P., Sleeper, J., & Sinha, R. (2007). Difficulties in emotion regulation and impulse control during cocaine abstinence. Drug and Alcohol Dependence, 89, 298–301.

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., & Roemer, L. (2008). The relationship between emotion dysregulation and deliberate self-harm among female undergraduate students at an urban commuter university. Cognitive Behaviour Therapy, 37, 14–24.

Harrison, A., Sullivan, S., Tchanturia, K., & Treasure, J. (2009). Emotion recognition and regulation in anorexia nervosa. Clinical Psychology and Psychotherapy, 16, 348–356.

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. Journal of Management, 21, 967–988.

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.

Kline, R. B. (2011). Principles and practice of structural equation modeling (2nd ed.). New York: Guilford.

Kökönyei, G., Urbán, R., Reinhardt, M., Józan, A., & Demetrovics, Z. (2014). The Difficulties in Emotion Regulation Scale: Factor Structure in Chronic Pain Patients. Journal of Clinical Psychology, 70, 589–600.

Lee, D. J., Witte, T. K., Weathers, F. W., & Davis, M. T. (2015). Emotion regulation strategy use and posttraumatic stress disorder: Associations between multiple strategies and specific symptom clusters. Journal of Psychopathology and Behavioral Assessment, 37, 533–544.

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

Linehan, M. M., Bohus, M., & Lynch, T. R. (2007). Dialectical behavior therapy for pervasive emotion dysregulation. In Gross, J. (Ed.), Handbook of emotion regulation (581–605). New York: Guilford Press.

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.

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

Muthén, L. K., & Muthén, B. O. (1998–2011). Mplus user's guide (6th ed.). Los Angeles, CA: Muthén & Muthén.

Neumann, A., van Lier, P. 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(1), 138–149.

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, 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.

Roemer, L., Lee, J. K., Salters-Pedneault, K., Erisman, S. M., Orsillo, S. M., & Mennin, D. S. (2009). Mindfulness and emotion regulation difficulties in generalized anxiety disorder: Preliminary evidence for independent and overlapping effects. Behavior Therapy, 40, 142–154.

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.

Snow, N. L., Ward, R. M., Becker, S. P., & Raval, V. V. (2013). Measurement invariance of the Difficulties in Emotion Regulation Scale in India and the United States. Journal of Educational and Developmental Psychology, 3, 147–157.

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

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

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

Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: Current approaches and future directions. Psychological Methods, 12, 58–79.