Original Research Article

Assessment
1–18
© The Author(s) 2024
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/10731911241261168
journals.sagepub.com/home/asm
Sage

The Imperfect Yet Valuable Difficulties in Emotion Regulation Scale: Factor Structure, Dimensionality, and Possible Cutoff Score

Chen Erez[1] and Ilanit Gordon[1,2]

Abstract

The Difficulties in Emotion Regulation Scale (DERS) is frequently used to assess emotion regulation (ER) capabilities. Originally a multidimensional scale, many utilize its total score, without clear recommendations. We aimed to explore the DERS's structure, dimensionality, and utility and provide clinicians and researchers with clear guidelines. Self-report data on ER, personality, psychopathology, and life satisfaction were collected from 502 adults. Seventy also participated in a lab study evaluating group interactions, which included additional self-report and physiological monitoring. Findings suggested favoring the correlated-traits and bifactor models, the latter excelling in direct comparisons. The total score was found reliable and valid, explaining 53.3% of the variance, with a distinct emotional awareness subfactor, suggesting a non-pure unidimensional solution. A cutoff score of 95 identified significant ER difficulties, linked to psychopathology. We thus recommend using the DERS's total score and 95 as its cutoff, while calling for further validation in diverse and clinical samples.

Keywords

Difficulties in Emotion Regulation Scale, emotion regulation, scale psychometrics, cutoff scores, emotion dysregulation

Emotion regulation (ER) is the process by which we monitor, evaluate, and modify our emotional experiences and our expressions of them (Gross, 1998; Thompson, 1994). In recent decades, the ER literature has been extremely prolific, depicting ER's extensive role in adaptive functioning and mental health (Aldao et al., 2010; Dixon-Gordon et al., 2015; Sheppes et al., 2015). This literature is mostly based on self-report scales. One such popular ER scale is the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004), cited over 10,000 times to date (per Google Scholar, around 5,000 times per Scopus).

Partially based on earlier scales (Negative Mood Regulation Scale; Catanzaro & Mearns, 1990; Trait Meta-Mood Scale; Salovey et al., 2004), the DERS (Gratz & Roemer, 2004) conceptually emphasizes the functional nature of emotions. Its items represent a view that difficulties in ER encompass deficiencies in the capacity to experience, differentiate, and accept emotions, as well as inflexible use of ER strategies and inabilities to inhibit inappropriate behaviors deterring goal-attainment. The DERS was first presented as a multidimensional, 36-item scale assessing ER difficulties in six facets: nonacceptance of emotional responses (i.e., acceptance); difficulty engaging in goal-directed behavior (i.e., goals); impulse control difficulties (i.e., impulse); lack of emotional awareness (i.e., awareness); limited access to ER strategies (i.e., strategies); and lack of emotional clarity (i.e., clarity).

Despite its popularity, there are some remaining questions surrounding the structure and utilization of the DERS. Scholars have, at times, used it to derive six separate constructs, other times, to calculate a total score from all 36 items, and sometimes utilized both (e.g., Tull et al., 2012). While this has been common practice for decades in research utilizing self-report instruments, nowadays, researchers have been called to thoroughly check the structure, dimensionality, and utilization of total and subscores, especially with the 'rediscovery of bifactor measurement models' (Reise, 2012). As such, the issue demands further exploration and

[1]Bar-Ilan University, Ramat-Gan, Israel
[2]Yale University, New Haven, CT, USA

Corresponding Author:
Ilanit Gordon, Department of Psychology, Bar-Ilan University, Ramat-Gan 52900, Israel.
Email: ilanit.gordon@biu.ac.il

empirical evidence to provide the field with specific recommendations (and restrictions).

An additional issue concerns the convergence of the awareness subscale to the DERS, as it seems to be under-correlated with some of the questionnaire's other subscales. Hallion et al. (2018) found a very good fit for the DERS's structure by removing the items associated with the awareness subscale (and the latent factor). Osborne et al. (2017) found a good fit while retaining them, loaded only on their latent factor, and *not* loaded onto a general factor (total and/or second-order), and including (i.e., modeling) the correlation between the awareness and clarity factors. Scholars such as Bardeen (e.g., Bardeen et al., 2015) created modified versions of the DERS, suggesting that the distinctiveness of the awareness factor might result from the fact that all of its items are reveres-coded. These authors also found a good-fit solution by loading the items of both the awareness and clarity factors onto the same factor. At the same time, scholars from other fields suggest that emotional awareness should be regarded as a distinct theoretical construct. Lane and colleagues (Smith et al., 2018) have provided theoretical foundations depicting emotional awareness as a trait that varies in the population, with antecedents in variations in affective response generation and representation processes and in conscious access to such representations. These, in their view (Smith et al., 2020) are derived from the evolution of the human brain and its disproportionate cortical expansion, enabling the ability to simulate emotions, learn emotion concepts, and manipulate them in working memory when deciding how to act. In their empirical work, they found emotional awareness to be positively associated with reflective cognition, socio-emotional skills, and empathy, and negatively with psychopathy and a history of childhood trauma (Smith, Chuning, et al., 2022; Smith, Persich, et al., 2022).

Taking into account these bodies of work with respect to emotional awareness within and outside the confines of the DERS, we view the emotional awareness factor as a potentially reliable, valid, and important construct. Further, some of the findings presented above suggest that the awareness and clarity factors could converge onto a higher-ordered construct, representing an early sub-component of the ER process that could be of relevance in its own right.

Furthermore, despite all presented explorations, there is no definitive empirically rooted recommendation on whether one should use the DERS's total score or not. Since the debate surrounding the awareness subscale preoccupied most, and less attention was given to the use of the DERS's total and subscores, there is a lack of empirical assessment regarding the DERS's utility in classifying those with and without significant ER difficulties via specific cutoffs, which are of extreme importance to clinicians and researchers alike. Some found support for using the score of the strategies factor in predicting nonsuicidal self-injury among adolescents (Perez et al., 2012). Importantly, this finding was brought forth in a predetermined multidimensional conceptualization of the DERS's structure. Using a shorter version of the DERS, Lutz et al. (2018) found support for utilizing the total score for predicting disability and opioid misuse in a clinical sample. Nonetheless, to the best of our knowledge, currently, there is no agreed-upon cutoff score for clinicians and researchers to use when looking for indications of high probability for a person to be dealing with the most prevalent mental health concerns—anxiety and depression (World Health Organization, 2022).

To sum, the gaps in current literature regarding the DERS are whether its total score should be used, its subscores, or some combination of both. Meaning, which of its potential scores are reliable and appear to be an operationalization of a valid, robust construct. Specific concerns are attributed to the fit of the emotional awareness construct and its six associated items. Moreover, no cutoff scores for discerning those with significant ER difficulties are yet available.

The Current Study

In line with the above, the aim of the current study was to explore the psychometrics, factor structure, dimensionality (single, multi, etc.), and utility of the 36-item version of the DERS, including using the total score for distinction being those with and without significant difficulties in ER. To that aim, we collected self-reported data on ER, personality traits, psychopathology, and life-satisfaction. Some of the participants also attended another study featuring lab-set social interactions during which physiological activities were monitored. These participants also answered additional questionnaires regarding ER, attachment, and affective state. We first focused on the DERS's structure and aimed to test and compare several factorial structures by looking at the fit of newly collected data to each. We also looked into dimensionality indices and the reliability and validity parameters of the available scores. In case the data would support the DERS's total (i.e., general) score being a reliable and valid measure, we then aimed to test its utility as an indicator (i.e., via cutoff score) of maladaptive functioning, portrayed in poorer mental health. Receiver operating characteristic (ROC) analyses were performed to find appropriate cutoffs, which would later be used to test for inter-group (i.e., control vs. probable psychopathology) differences in related constructs (e.g., attachment, Big-5 personality) and in physiological indicators of adaptive regulatory processes.

Statistical Plan

One of our focal points in the process of assessing the psychometrics and utility of the DERS was to test its factor structure and dimensionality. This first entailed comparing various, competing factor structures by conducting confirmatory factor analyses (CFAs) via structural equation modeling (SEM). However, a preliminary issue was which estimation method should be used in this case.

Most work conducted in psychological research implemented the Maximum Likelihood (ML) estimator, probably due to the ability to obtain unbiased estimations when used in appropriate contexts (Bollen, 1989). However, there have been discussions on whether psychological questionnaire data can, or should be, treated as continuous data, a prerequisite for applying ML, rather than applying an estimator treating it as ordinal data such as diagonally weighted least squares; DWLS). That is, questions have been raised as to the accuracy and bias of both approaches.

While this issue is still under debate, simulation studies (e.g., Beauducel & Herzberg, 2006; Li, 2016; Rhemtulla et al., 2012) comparing robust ML (i.e., MLR) and robust categorical (i.e., cat-LS, WLSMV) methodologies for estimating CFA models, in conjunction with a recent summary of recommendations and pitfalls in conducting factor analyses in psychological assessment research (Sellbom & Tellegen, 2019), shed light on the matter. First and foremost, as stated by Rhemtulla et al. (2012), "it is a good practical strategy to apply robust corrections to normal theory ML when variables are categorical (in nature, as questionnaire responses are), because these variables are, by definition, nonnormal" (p. 357). Hence, we will not be discussing ML, but MLR as the estimator of choice in the continuous case. Furthermore, the aforementioned body of work suggests that in questionnaires scored using five categories (or more), MLR can produce similarly accurate estimations as those treating the data as ordinal. In addition, MLR could provide more accurate estimations of inter-factor correlations, a central issue in the model-comparison portion of this work. Thus, we use the MLR estimator in the current exploration. We also provide indicator and multivariate properties [analyzed using SPSS and psych package for R (Revelle, 2022)].

We explore the factor structure and dimensionality of the DERS by testing and comparing several factorial structures and looking into their fit to the data and into dimensionality indices. As is becoming common practice, our approach consisted of testing four models: A unidimensional model, in which all items load on a single latent factor (no other latent factors, nor correlated residuals); The original correlated-traits (i.e., multidimensional) model, consisting of six first-level factors, with each item loading onto one factor, and with all factors correlated; A second-ordered factor model, dissimilar from the correlated-traits model only by having all six first-ordered factors load on a single, higher-order factor; and a bifactor model, in which all items load both onto a single latent, general factor (like in the unidimensional model) and simultaneously onto a group (i.e., trait) factor (like in the correlated-traits model). In the classic bifactor model, all latent factors are orthogonal, and group factors are viewed as means to control residual variance. However, as discussed later, in some instances, data might be "unidimensional enough" (Rodriguez et al., 2016b) and group factors could also still be theoretically and statistically sound and be used alongside a general score. Importantly, seeing merit in past theory and research on the emotional awareness construct (and those including its items when scoring the DERS) yet anticipating some discrepancy between "its" items and some of the rest of the items, we had a priori chosen not to remove its items in any configuration. This approach was also encouraged by two additional considerations: The first being a view that post hoc removing items from a questionnaire that was, and probably still is, being administered in full via different methods (i.e., hardcopy, e-form by original order, e-form in randomized order) might confound the validity of our findings. The second is that throughout the DERS's history, researchers usually retained the awareness items when computing a total score. Hence, removing them here would hinder the ability to compare our findings with past evidence. We did however decide to test a fifth model—an "alternative bifactor model" that included the correlation between the awareness and clarity factors. Such a model however must be viewed with caution in interpreting dimensionality (Reise, 2012, pp. 691–692), especially considering the potential bias in modeling statistical "noise" instead of meaningful information. Also, as the debate is still ongoing, we provide additional results from models in which the awareness items were either removed or unloaded on the total factor but rather correlated with it (at the factor level) or with the Clarity factor (see Supplementary Tables S1 and S2).

The CFA models were tested via SEM. The analyses were conducted using the Lavaan package for R (version 0.6-12; Rosseel, 2012). Following the conventions set by Hoyle and Panter (1995) and Hu and Bentler (1999), the fit of the models to the data was evaluated using five goodness-of-fit indices. Three indices were absolute [χ^2 statistic, standardized root mean squared residual (SRMR) and root mean square error of approximation (RMSEA)]; and two were incremental [Comparative Fit Index (CFI) and Tucker–Lewis Index (TLI)]. A combination of a nonsignificant χ^2 (relevant in small samples), SRMR and RMSEA scores below .06, with CFI, and TLI above .95 indicates excellent fit, whereas a significant χ^2 and values below .08 and above .90, respectively, indicates adequate fit.

In addition to fit, there are various ways to compare competing models. The first, and perhaps most commonly used, is comparing the $\Delta\chi^2$ values per model as a function of Δdegrees of freedom. If the $\Delta\chi^2$ value is found significant, the less parsimonious model is deemed "better." Otherwise, the more parsimonious model is favored. In this case, using the MLR estimator required a "robustified" test. We used the formula suggested by Mplus for the hand-calculated Sattora-Bentler Scaled Chi-Square Difference test (TRd; Satorra & Bentler, 2010) (see online calculator at https://www.thestatisticalmind.com/calculators/SBChiSquareDifferenceTest.htm). We tested for significant differences at $< .005$ to adjust for multiple comparisons. A second approach is comparing differences in Akaike information criterion (AIC) or Bayesian information criterion (BIC). While there are no clear benchmarks, ΔBIC values of 2 to 5, 6 to 9, and 10 have been associated with positive, strong, and very strong evidence for significant differences, respectively (Raftery, 1995). Importantly, in whatever case, simulation studies have shown that the best-fitting model would usually be the one that matches the true structure of the data best (Morgan et al., 2015).

Alongside the fit of the models to the data, we explored the reliability and validity of latent factors and the DERS's dimensionality. That is, is it clearly unidimensional, multidimensional, or another "softer" distinction. Naturally, a combination in which (a) the bifactor model is found "the best" model, and (b) indices suggest that the general factor is a valid and reliable construct, and (c) that the data are mostly unidimensional, would suggest that only the general score is of merit and that the other group factors encapsulate less important portions of variance.

Factor reliability and validity and data dimensionality indices were calculated using the "Bifactor-IndicesCalculator" package for R (Dueber, 2021) based on Dueber's previously used (e.g., Constantinou & Fonagy, 2019) Excel calculator (Dueber, 2017). Like model fit, conclusions regarding these issues are deduced by converging evidence from multiple model-based indices, computed based on the standardized loadings in the unidimensional and bifactor models (augmenting the sometimes misleading coefficient alpha). First, to interpret factor *reliability*, one considers coefficient alpha but also coefficient omega and its derivatives. Coefficient omega is the model-based equivalent to coefficient alpha, suffering from similar limitations as the alpha. Coefficient omega hierarchical ("omegaH") however is affected more by item loadings than by parameters such as item redundancy and is considered one of the cornerstones for interpreting factor reliability. OmegaH represents the percentage of systematic variance in unit-weighted total scores that can be attributed to individual differences in a factor. That is, it is less affected by other latent factors, as opposed to Cronbach's alpha. Values of $.80 \leqslant$ suggest that the vast majority of reliable variance is attributable to a single common source, supporting a conclusion that a scale is mostly unidimensional when discussing the general factor. OmegaHs for group factors ("omegaH subscales") are computed as well, "reflecting the proportion of reliable systematic variance of a subscale score after partitioning out variability attributed to the general factor" (Reise et al., 2013, p. 225). Second, to interpret factor *validity*, factor determinacy (FD) and construct replicability ("H") were considered. Less of a concern for factors including many items is FD—'the correlation between factor score estimates and factors' (Rodriguez et al., 2016b, p. 146), but often relevant to group factors, especially ones composed of few items. It is recommended that factor score estimates should only be used when FD is $>.90$ (Gorsuch, 1983, p. 260). H represents "the correlation between a factor and an optimally-weighted item composite" (Rodriguez et al., 2016a, p. 230). That is, "how well the items reflect or account for the variance of the latent variable" (Rodriguez et al., 2016b, p. 146). Values $>.80$ suggest a well-defined latent variable. Finally, to explore *dimensionality*, three main parameters are considered: explained common variance (ECV), percent uncontaminated correlations (PUC), and absolute relative parameter bias (ARPB). Data unidimensionality in the context of a bifactor model is considered "unidimensionality enough" when the general factor explains a significant portion of the items' variance (high ECV), and when items load similarly on it in both the bifactor and the unidimensional models (i.e., little bias, or low ARPB). PUC is the percentage of "unique correlations in a correlation matrix that are influenced by a single factor divided by the total number of unique correlations. The higher the PUC, the more the matrix is saturated with information relevant to estimating the parameters of a single factor and the less likely the parameter estimates in a unidimensional model will be biased" (Rodriguez et al., 2016a, p. 146). Benchmarks for ECV and PUC have been reached mostly from studies testing how their interplay (along with omegaH) affects ARPB, with up to 15% bias considered acceptable (Reise et al., 2013; Rodriguez et al., 2016b). These studies suggested that ECVs $\geqslant .70$

indicate that data can be regarded unidimensional with little expected bias, especially combined with a PUC $\geqslant .70$. PUC also moderates the association between ECV and omegaH and ARPB, and findings suggest that when PUC is $\geqslant .80$, an ECV of .60 and an omegaH of .70 could be regarded as benchmarks for unidimensionality as well. In the case of the DERS, PUC is .85, thus, with other conditions met, an ECV lower than .60, would suggest that the data are not (purely) unidimensional, although ECVs even lower than .60 can sometimes be considered (e.g., Reise et al., 2010). ECVs are also calculated at the group factor level and the item level (i.e., I-ECV, variance explained by the general factor).

Should the DERS' general score be reliable and valid, and data would be "unidimensional enough," we will explore a possible cutoff score. It should be noted that whichever cutoff arises, it would be, at least to some degree, dependent on the specific characteristics of our general population, student-heavy sample. Hence, it might be relevant only to similar samples. We still see merit in testing for a cutoff score for several reasons. Generality speaking, at times researchers and clinicians wish to be able to categorize participants and patients according to their ER capabilities for research and treatment purposes (e.g., tailoring specific treatment plans according to ER capacity). More specifically, a cutoff score could be used by researchers looking into similar samples (past and present), and student samples are very commonly used. Moreover, we view this as a propositional cutoff score and wish to offer one so that future researchers can validate this score against scores of other samples. In sum, a crystallized understanding of cutoff scores between samples would be an advancement of the field and greatly useful for researchers and clinicians and we thus take advantage of the current opportunity to test for one. According to our review of studies using the DERS in general population and clinical samples (e.g., Bytamar et al., 2020; Fowler et al., 2014; Hannan & Orcutt, 2020; Lewczuk et al., 2021), its total score in population samples varied around 80, with *SD*s of ˜20, and a close-to-normal distribution. In studies comparing "healthy control" and "clinical" groups, group-level *SD* still fluctuated around 20, or slightly higher, with *M*s as low as 60, and as high as 120, respectively. Thus, we expected the computed sum of items to be normally distributed with an *M* of ˜80 and a *SD* of ˜20. Consequently, we expected 100 would be an adequate cutoff score indicating difficulties in ER.

In this vein, we would perform three Receiver Operating Characteristic (ROC) analyses, testing the accuracy as well as the sensitivity (true positive) and specificity (true negative) of the DERS in discerning between control and probable clinical psychopathology subgroups. This, as indicated by three self-report instruments with acceptable cutoffs for assessing social anxiety, generalized anxiety, and major depression. A scale is considered of low diagnostic accuracy when the area under the curve (AUC) is $<.70$, acceptable between .70 and .80, excellent between .80 and .90, and outstanding $>.90$. The AUC is augmented by sensitivity and specificity of a scale, with 80% in both considered the minimal optimum. If the two do not converge, we shall put more emphasis on specificity, fostering what we see as more relevant to our field—forgoing some "true positives" to avoid "false positives."

For further exploration of the suggested cutoff score, additional t-tests will be performed on a sub-sample of participants who have filled out the DERS, along with additional questionnaires, as a part of participation in an ongoing larger, community, group psychophysiological study. Within this study, people who scored within a range of 65-85 or 95 or above on the DERS were invited to partake. They had filled out additional self-report instruments assessing personality, social inclinations, and ER, and formed groups who had undergone a negative affect induction manipulation and were asked to perform a joint task while their physiological activity was monitored (followed by additional, mostly group perception scales). Here, we focus on self-reported differences and similarities between below and above-cutoff groups, as well as differences in a physiological indicator of ER—root mean square of successive differences between normal heartbeats (RMSSD; see below and Method). These analyses, and the ROCs, will be performed using SPSS. Be advised that additional self-reported and physiological data are available.

Our focus on physiological activities alongside self-reported measures stems from a long and well-established field of inquiry looking into the link between physiological processes and ER. In short, two major conceptual models have provided frameworks to study neurophysiological processes that are central in all that relates to experiencing and regulating emotions—Thayer and Lane's (2000, 2009) "neurovisceral integration model" and Porges's polyvagal Theory (PVT; see Porges, 2007 for detailed description). Both focused on neural regulation (see Central Autonomic Network in Thayer and Lane's work) of psychophysiological processes by focusing on the central connection to the Autonomic Nervous System (ANS). The ANS is comprised of the Sympathetic and Parasympathetic Nervous Systems (SNS; PNS). The SNS is considered to take over during threat and allow for "fight or flight" responses. As such, the SNS is key in the quick allocation of resources in response to demanding events, while the PNS is dominant during calming down and routine function ("rest and digest") (Dienstbier, 1989; Meijen et al., 2020; Palumbo et al., 2017). The dynamics between the SNS and PNS, and their impact on attention, vigilance, and ER, are the central functions of the neural regulatory network. This function is posited to be predominantly mediated by the input from the vagus nerve to peripheral organs. Most specific to the current study, the PNS inputs can be assessed by measuring changes in high-frequency heart rate variability, usually via Respiratory Sinus Arrhythmia, but also time measures such as RMSSD, which thus provides proxies for systemic flexibility. Indeed, empirical work supports the notion of an association between higher PNS activity and better ER within community and clinical populations (see for example Balzarotti et al., 2017; Bylsma et al., 2014; Campbell & Wisco, 2021; Gordon et al., 2021). Hence, higher RMSSD should also correspond with better ER abilities.

Method

We report how we determined sample size, all data exclusions, manipulations, and all measures in our large sample. The project aimed at addressing a cutoff score for the DERS was not pre-registered. However, this project is a part of a larger project which was pre-registered here: https://osf.io/3q7j4/?view_only=70a0e79492ec453e91ed240088ee8d24

Data is available in OSF (https://osf.io/x29sp/?view_only=90d486100671429abfd4841495251c7f), as are codes and scripts, and the questionnaires.

Required Sample Size

There are various views regarding sample size and power considerations in SEM. A general rule of thumb suggests a minimum of 200 participants. Some argue for a ratio of five participants per parameter-to-be-estimated (in our case 72–109, suggesting a required N of roughly 547), but estimation method and data characteristics also come into play. In all, a sample of 500 participants is considered sufficient for most cases. For the ROCs, we used MedCalc for Windows version 20.014 (MedCalc Software, Ostend, Belgium) to calculate the required sample size. Not knowing the DERS's projected AUC, or the exact ratio between "control" and "clinical" cases, we ran two power analyses. In both, we set alpha at .01, to adjust for three ROCs performed on the same data set, power at .8 and ratio at 11.5. In one, AUC was .725, and in the other .75. Both AUC values were thus within the lower range of acceptable accuracy, to avoid a type 2 error. Results suggested required sample sizes of 213 and 263, respectively, much less than that required for the CFAs. Hence, we decided on a sample of 500 participants to suffice for the primary statistical procedures performed in this study.

Participants and Procedure

Between November 2021 and August 2022, native Hebrew speaking adults not diagnosed with dyslexia or intellectual disability were invited via social media and university platforms to fill out an online survey for a chance to win 150$ in a raffle, as well as for course credit (for psychology undergraduates). These inclusion criteria were imperative as this study was a part of a larger study on group processes that includes word-assembly tasks. In efforts to gather a diverse sample, we routinely posted ads in general social media forums unrelated to academia. After informed consent, participants answered several self-report scales. Five hundred and two participants answered the majority of the survey and were thus included in the sample. Most (74.3%) identified as women (the rest as men, besides two—one a-binary and one transexual man), childless (84.5%), aged between 18 and 57 (M = 24.59, SD = 4.96), studying for a bachelor's degree (78.3%). Socio-economic status (SES) was normally distributed and 51% indicated their SES was mediocre (or average).

Of these 502, 70 participants later arrived at a different lab-based study exploring behavioral and physiological processes during social group interactions, including two trials of a joint word-assembly game separated by viewing a negative emotional movie. In the game, each participant received three letters that in and of themselves could not form a word, and participants were instructed to assemble as many words as possible from their nine shared letters within 3 minutes (longer words granted more points; scoring and word assembly rules were borrowed from the game "Boggle"). Between games, participants watched an upsetting scene from Disney's The Lion King, instructed to be attentive and try to "get into it." As a part of this latter study, participants filled out additional self-report scales and their physiological activity was monitored for approximately 40 minutes, separated into a baseline measurement stage, Game Trial 1, emotional viewing, and Game Trial 2. All interactions and studies were approved by the Institutional Review Board of Bar-Ilan University.

Materials

Questionnaires Answered by Entire Sample. Several questionnaires were filled out in addition to the DERS (being the third scale in the battery, the rest are presented in the order answered by participants).

Demographics. We asked participants for their age, gender, education, SES and relationship and parenthood statuses. We intentionally did not ask for ethnicity or origins. This decision was preceded by participants in previous studies expressing discontent and even anger when asked such questions, probably due to the significant blurring of former traditional characterizations in Israeli society these past couple of decades.

The Big-5 Inventory (BFI). The BFI (John et al., 1991) is a 44-item scale used to assess the Big-5 personality traits of extroversion, agreeableness, openness to experiences, conscientiousness, and neuroticism. Participants rate the degree to which an adjective (e.g., "is talkative") describes them on a 5-point scale ranging from 1 (*disagree strongly*) to 5 (*agree strongly*). Scores on sets of items are averaged, as higher scores indicate higher agreement. Cronbach's αs ranged from .74 to .88.

The Social Phobia Inventory (SPIN). The SPIN (Connor et al., 2000) is a 17-item scale commonly used to assess social anxiety (Cronbach's α = .92). Participants are asked to indicate how much each statement (e.g., 'I avoid talking to people I don't know') applied to them over the past couple of weeks on a 5-point scale from 0 (*not at all*) to 4 (*extremely*). Scores are summed, with higher scores indicating higher social anxiety, and 31 or higher considered a cutoff score for probable clinical social anxiety.

The Generalized Anxiety Disorder-7 (GAD-7). The GAD-7 (Spitzer et al., 2006) is a 7-item scale frequently used to measure anxiety (Cronbach's α = .88). This instrument assesses experiences over the past 2 weeks (e.g., "Trouble relaxing") ranging from 0 (*not at all*) to 3 (*all day*). Scores are summed, with higher scores indicating higher anxiety, and 10 or higher considered a cutoff score for probable clinical anxiety.

The Patient Health Questionnaire-2 (PHQ-2). The PHQ-2 (Arroll et al., 2010) is a two-item scale frequently used for measuring depression (r_p = .66). Both items ("Little interest or pleasure in doing things" and "Feeling down, depressed or hopeless") are rated in reference to experiences occurring over the past 2 weeks, rated on a 4-point Likert-type scale ranging from 0 (*not at all*) to 3 (*all day*). Scores are summed, with higher scores indicating higher depression and 3 or higher considered a cutoff score for probable clinical depression.

The Single-Item Life Satisfaction Scale (SILS). The SILS (Jovanović & Lazić, 2020) asks participants to indicate how satisfied they are with their lives ranging from 1 (*not satisfied at all*) to 10 (*completely satisfied*). The SILS was shown to be highly correlated with longer, commonly used scales, such as the Satisfaction with Life Scale (Diener et al., 1985).

Questionnaires and Parameters for Sub-Sample

Emotion Regulation Questionnaire (ERQ). The 10-item ERQ (Gross & John, 2003) measures the tendency to regulate emotions in two ways: Cognitive reappraisal (e.g., "I control my emotions by changing the way I think about the situation") (Cronbach's α = .86) and Expressive suppression (e.g., ". . . careful not to express them") (Cronbach's α = .77), regarded adaptive and maladaptive, respectively. Participants answer each item on a scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*), and scores are averaged.

Negative Mood Regulation (NMR). The 30-item NMR (Catanzaro & Mearns, 1990) measures expectancies to alleviate negative mood. Participants are asked to indicate the degree to which they believe their use of various coping strategies can counteract a negative mood state. Items are scored on a 5-point Likert-type scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*) with a statement completing the stem, 'When I'm upset I believe that . . .' (e.g., "I can do something to feel better"). Scores are summed, and higher scores indicate a strong belief that one can alleviate one's own negative moods (Cronbach's α = .87).

The Experiences in Close Relationships Revised (ECR-R). The ECR-R (Mikulincer & Florian, 2000) is a 36-item scale that assesses attachment styles on two dimensions (18 items each): avoidance (e.g., "I find it difficult to allow myself to depend on romantic partners") and anxiety (e.g., "My desire to be very close sometimes scares people away"). Cronbach's αs were .91 and .89, respectively. Scores are summed, with higher scores indicating greater attachment insecurity.

The Positive Affect Negative Affect Schedule (PANAS). The PANAS (Watson et al., 1988) is a 20-item scale assessing current affect. Items converge into two composites: positive (e.g., "enthusiastic") and negative (e.g., "upset") affect. Due to their relevance to the current study, we chose to add "happy" and "sad," for a total of 22 items. Cronbach's αs were .86 and .84, respectively. Participants filled out the PANAS at the beginning of the lab session.

RMSSD. During the lab session participants were continuously monitored for their physiological activity using MindWare Mobile Impedance Cardiograph recording devices (MindWare Technology, Gahanna, OH) transmitting wirelessly to a computer in a control

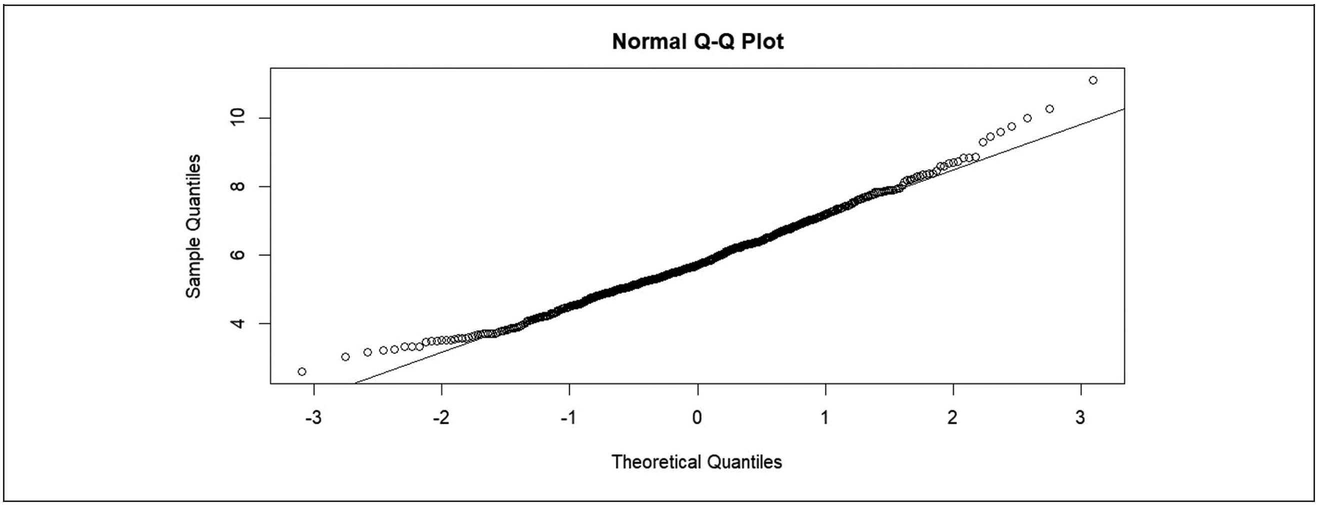

Figure 1. Q-Q Plot for Multivariate Distribution of the DERS's Total Score.

Table 1. CFA Model Statistics and Fit Indices.

Models/ parameters	Robust χ^2 (*df*)	Standard χ^2	Scaling factor	SRMR	RMSEA [LL, UL]	CFI	TLI	AIC	BIC
Unidimensional	4,560.10 (594)	5,212.544	1.143	.121	.115 [.112,.118]	.573	.547	47,828.21	48,131.95
Correlated traits	1,723.91 (579)	1,944.37	1.128	.076	.067 [.063,.070]	.878	.868	44,590.03	44,957.05
Second-ordered	1,932.50 (588)	2,182.011	1.129	.093	.072 [.068,.075]	.857	.847	44,809.67	45,138.72
Bifactor (Classical)	1,540.87 (558)	1,726.322	1.120	.084	.063 [.059,.066]	.896	.883	44,413.98	44,869.59
Bifactor (Alternative)	1,313.22 (557)	1,469.666	1.119	.061	.055 [.051,.059]	.920	.910	44,159.33	44,619.16

Note. df = degrees of freedom. LL and UL = Lower Limit and Upper Limit, respectively.

room at a sampling rate of 500Hz. Since respiratory data was corrupted in some instances (hindering the usage of Respiratory Sinus Arrhythmia), we focused on intervals between heartbeats, derived from one's electrocardiogram, collected via three electrodes placed on the torso according to the modified lead II configuration. The signal is filtered for muscle and electrical noise, with visual inspection and manual editing ensuring proper data collection, all performed and analyzed within Mindware's designated application (as done, for example, in Gordon et al., 2020; Milstein & Gordon, 2020). RMSSD is a time-domain indicator of parasympathetic ("rest and digest") regulation (Shaffer & Ginsberg, 2017). It is obtained by first calculating each successive time difference between heartbeats in milliseconds. Then, each value is squared, and the result is averaged before the square root of the total is obtained. Higher scores suggest higher parasympathetic activity.

Results

Data Characteristics

At the indicator and raw general score levels, the psychometrical properties of the DERS are roughly as expected. Responses on all 36 items but one can be considered normality distributed ($\pm$ 1.96 skewness and kurtosis). For item 32, kurtosis was 2.82. Items'*M*s ranged from 1.55 to 3.11, with *Mo*s usually at 1 or 2, indicating some category threshold asymmetry. The summed ("total") score can also be considered normality distributed with a slight skew (M = 81.84, Md = 78, and SD = 22.28). Results from Mardia's test for multivariate normality however suggest significant nonnormality ($MV_{Skewness}$ = 176.08, *Mardia's Skewness Statistic* = 14731.7, $p < .001$; $MV_{Kurtosis}$ = 1569.62, *Mardia's Kurtosis Statistic* = 43.18, $p < .001$). Visual inspection of the Q-Q plot suggests the same (see Figure 1).

Data Structure, Factor Indices, and Dimensionality

Results of all CFA analyses are summarized in Table 1. Results first indicated that the unidimensional model fitted the data poorly and had the "poorest" χ^2 and BIC by a considerable margin. The other models could all be argued to fit the data to some degree. Also of note, all pairwise TRds detailed below were significant at $<.001$, and ΔBICs suggested similar meaningful differences.

Table 2. Indices for the Classical Bifactor Model—Reliability, Validity, and Dimensionality.

						ECV		
Factor/index	Alpha	Omega	OmegaH	FD	H	GS	SG	SS
General factor	.94	.96	.83	.96	.95	53.32%	53.32%	53.32%
Nonacceptance	.90	.91	.43	.91	.76	50.76%	09.61%	49.24%
Goals	.88	.88	.40	.86	.66	54.99%	06.83%	45.01%
Impulse	.86	.88	.27	.84	.62	66.60%	05.73%	33.40%
Awareness	.81	.82	.79	.91	.81	04.50%	12.32%	95.50%
Strategies	.90	.92	.10	.79	.51	82.25%	04.25%	17.75%
Clarity	.78	.80	.57	.86	.72	29.64%	07.93%	70.36%

Note. GS = Proportion of common variance of the items *in each specific factor* which is due to the *general factor*; SG = Proportion of common variance of *all items* which is due to the *specific factor*; SS = Proportion of common variance of the *items in each factor* which is due to *that factor.*

For the correlated-traits model, the absolute fit indices, SRMR and RMSEA, indicated adequate fit, but the incremental indices CFI and TLI did not (although close). For the second-ordered factor model, fit was significantly poorer. The RMSEA did surpass the threshold for adequate fit, yet the rest were descriptively worse compared with that of the correlated-traits model. Its χ^2(TRd(9) = 226.46) and BIC (ΔBIC = 181.22) were also significantly (or meaningfully) poorer compared with the correlated-traits model. For the classical bifactor model, RMSEA values indicated adequate fit. However, its fit as indicated by SRMR and CFI values was just borderline adequate, while in terms of model comparison, it was better than the correlated-traits model (TRd(21) = 193.77; ΔBIC = 87.46). Finally, the alternative bifactor model had excellent fit as indicated by the RMSEA, and adequate fit per SRMR, CFI and TLI. Its absolute fit was also better than that of the classical bifactor model (TRd(1) = 172.29; ΔBIC = 250.43), probably due to the high correlation between the awareness and clarity factors (r_p = .82, $p < .001$).

In sum, results indicated that the better fitting models were the correlated-traits, classical bifactor, and alternative bifactor models, with significantly better fit in this order. The alternative bifactor model is however (potentially) more challenging to intercept in terms of unidimensionality. Regardless, findings did indicate that unidimensionality should be explored. Hence, we proceeded to explore the factors' reliability and validity and the data's dimensionality, first as indicated by the classical bifactor model, and then by the alternative.

The DERS's general factor was found highly reliable by all parameters; Cronbach's α = .94, omega = .96, and omegaH = .83. It also demonstrated high validity; FD = .96, and H = .95. It explained 53.3% of the variance, a below-threshold value for "pure" unidimensionality. ARPB was at 18.14%, indicating some bias. This somewhat high average difference in loadings between the unidimensional and bifactor models mainly stemmed from two low-loading (0.04 and 0.03) outlier (Z-scores 3.21 and 3.47) awareness items. PUC is, as mentioned, at .85. Thus, results indicated some support for unidimensionality, but also some multidimensionality (see indices in Table 2, and item loadings, I-ECV, and bias in Table 3).

Looking into this "non-pure" unidimensionality, findings showed that the general factor explained only 4.5% of the variance of the items affiliated with the awareness factor and 29.64% of those affiliated with the clarity factor. Moreover, the awareness factor was found psychometrically sound, with acceptable/good reliability and validity. These suggest that it holds as a construct in and of itself. The clarity factor is less robust—with omegaH at .57, and validity indices below thresholds. Still, 70.36% of the variance of "its" items is explained by it, a sizable portion. On an item level, resonating with other findings, three of the six awareness items loaded insignificantly on the general factor. The omegaHs, validity, and ECVs for the other four group factors were low, suggesting they do not represent separate, distinct constructs.

Taking a similar approach in inspecting the results of the alternative bifactor model, indices regarding the general factor remained unchanged beside the ECV decreasing to 52.81% (the 0.5% difference now explained mostly by the clarity factor), and importantly, ARPB increased to 29.61%, challenging interpretation. We also assessed whether the fit of a CFA model with awareness and clarity loading onto an additional second-order factor was better than the one in which they merely correlated, and results indicated that it did. These combined entice future research to explore the two as preliminary components in the chain of processes that materialize into individuals' ER capacity, but this direction will not be further discussed herein, especially considering the high ARPB.

Thus, results suggested the DERS could be "unidimensional enough," while the variance of the awareness

Table 3. Bifactor Model—Item Loadings, I-ECV, and Relative Bias.

I#	Uni.	Gen.	Accept.	Goals	Imp.	Aware.	Strat.	Clarity	I-ECV	Bias
1	.31	.26						.49	.22	.20
2	.26	.18				.65			.07	.46
3	.67	.70			.14				.96	.04
4	.38	.31						.65	.19	.21
5	.54	.50						.43	.58	.07
6	.29	.23				.73			.09	.25
7	.31	.25						.69	.11	.26
8	.19	.13				.70			.03	.53
9	.47	.44						.52	.42	.07
10	.14	*.07*				.63			.01	1.18
11	.60	.56	.52						.54	.08
12	.56	.50	.65						.37	.13
13	.60	.58		.56					.52	.04
14	.63	.59			.50				.58	.08
15	.76	.74					.46		.72	.03
16	.76	.73					.47		.71	.05
17	.12	*.06*				.58			.01	1.10
18	.65	.64		.59					.54	.01
19	.74	.72			.38				.78	.03
20	.42	.38		.43					.44	.10
21	.59	.53	.68						.37	.12
22	.52	.48					.39		.60	.09
23	.64	.65	.15						.95	.01
24	.47	.45			.21				.82	.04
25	.64	.61	.61						.50	.06
26	.58	.56		.59					.47	.05
27	.68	.63			.58				.54	.08
28	.67	.64					.39		.73	.05
29	.64	.60	.61						.49	.07
30	.80	.81					*.03*		1.00	.01
31	.71	.71					.23		.91	.00
32	.66	.59			.60				.49	.10
33	.69	.68		.41					.73	.01
34	.15	*.08*				.55			.02	.80
35	.69	.71					.18		.94	.02
36	.69	.76					−.17		.95	.10

Note. Italicized are nonsignificant loadings. I = Item; Uni. = Unidimensional; Gen. = General factor (bifactor model); Accept. = Nonacceptance; Imp. = Impulse; Aware. = Awareness; Strat. = Strategies.

factor is unexplained by the general composite. Still, we proceeded to check for a possible cutoff score demarcating significant difficulties in ER, by performing three 3 ROCs testing the DERS's accuracy, sensitivity, and specificity in predicting social anxiety, generalized anxiety, and major depression.

Cutoff Score

In the current sample, out of valid responses in each domain, 27.31% (136, four missing), 15.45% (76, 10 missing), and 21.59% (106, 11 missing) met the criteria for probable social anxiety, generalized anxiety, and depression, respectively. 38.8% of participants met the criteria for at least one of the three, suggesting some comorbidity. The DERS's accuracy was acceptable in the classification of "control" and "clinical" subgroups in social anxiety (AUC = .79, 95% CI [.75, .83], $p <$ .001), and excellent with regards to generalized anxiety (AUC = .84, 95% CI [.80, .89], $p < .001$), and major depression (AUC = .82, 95% CI [.77, .87], $p < .001$). Sensitivity and specificity did not however converge to 80% in either of the three analyses. Hence, we focused on 80% specificity as a focal point, looking for optimum sensitivity within that context (see Table 4).

For social anxiety, a score of 91.5 (the analyses average two consecutive ordered observed values, i.e., 91 and 92) held 80.7% specificity and 58.8% sensitivity. For generalized anxiety, 94.5 was found to correspond with 80.3% specificity and 71.1% sensitivity. For depression, 90.5 was found to correspond with 80.0% specificity and 72.6% sensitivity. Results then first suggested that the predicted cutoff score of 100 would be exceedingly high. Also, they suggested that cutoffs in

Table 4. ROCs—Sensitivity and Specificity for Relevant Range.

DERS score/ scale	Social anxiety		Generalized anxiety		Depression	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
77.50	.816	.608	.908	.563	.858	.584
78.50	.787	.624	.908	.587	.858	.610
79.50	.765	.649	.895	.613	.840	.636
80.50	.728	.666	.868	.635	.821	.660
81.50	.706	.685	.868	.659	.811	.683
82.50	.699	.702	.842	.671	.802	.699
83.50	.676	.721	.842	.695	.783	.719
84.50	.676	.746	.842	.716	.783	.743
85.50	.676	.757	.842	.726	.783	.753
86.50	.676	.765	.842	.733	.783	.761
87.50	.669	.771	.829	.738	.783	.769
88.50	.647	.782	.803	.750	.764	.782
89.50	.632	.787	.803	.760	.755	.790
90.50	.610	.798	.763	.769	**.726**	**.800**
91.50	**.588**	**.807**	.763	.784	.708	.810
92.50	.574	.815	.750	.793	.708	.823
93.50	.574	.818	.750	.796	.698	.823
94.50	.566	.831	**.711**	**.803**	.679	.834
95.50	.537	.837	.697	.815	.670	.847
96.50	.493	.845	.671	.832	.632	.860
97.50	.485	.854	.645	.837	.613	.865
98.50	.485	.862	.632	.841	.594	.868
99.50	.463	.873	.605	.853	.557	.875
100.50	.449	.881	.592	.863	.509	.875

Note. Boldfaced are possible, optimal cutoff scores considering sensitivity and specificity.

each domain do not perfectly match, with a possible range between 90 and 95 (roughly 0.25 in SDs). Since each could be of value within specific contexts, we decided to look further into comparing the two.

Considering a score of 90 as a cutoff for significant ER difficulties would suggest classifying 32.73% of participants as such, roughly a third. A score of 95 would suggest classifying 27.64%. We performed independent samples T-tests comparing scores on psychopathology, neuroticism, and life satisfaction among participants in the 90 to 94 range (N = 25–26) and those scoring between 95 and 99 (N = 29), as well as when using 90 or 95 as cutoffs across the entire range of values.

Results indicated no significant differences between the 90 to 94 and 95 to 99 subgroups. However, effect sizes mostly pointed to an intermediately small effect (i.e., Cohen's *d*s of 0.3–0.4, see Tables 5 and 6), with 95 to 99 reporting higher levels, indicating potential differences in the grander population. Surprisingly though, the 95 to 99 subgroup reported descriptively higher life satisfaction. Findings from T-tests comparing groups below and above the cutoff performed using either 90 or 95 as general cutoffs are more coherent. All analyses indicated significant differences, in the expected directions. Apart from the life satisfaction domain, all effect sizes were descriptively higher when using the 95 cutoff. Moreover, subgroup descriptives (M and Md in particular) suggest using the 95 cutoff provides further separation between "control" and "clinical" subgroups. These, along with the overall distribution and density, suggest favoring the 95 cutoff score.

Although not our main objective, we also performed one-way analyses of variance (ANOVAs) comparing levels in each domain among three subgroups: <65, 65 to 94 and >95, looking for potential high and average ER capacities cutoffs. All main effects and pairwise comparisons (Bonferroni correction) were found significant in to-be-expected directions (see Tables S3–S6 in Supplemental materials). This issue however requires further testing.

Finally, to further test the validity and utility of the 95 cutoff score, we performed additional independent samples t-tests on a sub-sample of 70 participants who had also taken part in a separate study (see Table 7). As mentioned, as a part of this study, they had filled out additional self-report instruments and formed groups who jointly watched an upsetting scene and performed a task while their physiological activity was monitored.

The below cutoff group was comprised of 46 participants ($DERS_{range}$ = 65–85), and the >95 group included 24 ($DERS_{range}$ = 95–143). The 95 or above group reported significantly lower ER capacity on the cognitive

Table 5. T-tests—Comparing 90 and 95 Cutoffs.

	90-94 vs. 95-99			
Parameter/comparison	*t*	*df*	*p*	*Cohen's D [LL, UL]*
Life satisfaction	−1.14	52	.26	−0.31 [−0.85, 0.23]
Neuroticism	−1.16	53	.25	−0.31 [−0.85, 0.22]
Social anxiety	−1.14	52	.26	−0.31 [−0.85, 0.23]
Generalized anxiety	−0.06	52	.95	−0.01 [−0.55, 0.52]
Depression	−1.49	52	.14	−0.41 [−0.95, 0.14]
	Below 90 vs. 90 and above			
Parameter/comparison	*t*	*df*	*p*	*Cohen's D [LL, UL]*
Life satisfaction	11.5	489	.00	1.10 [0.90, 1.31]
Neuroticism	−14.8	500	.00	−1.4 [−1.61, −1.19]
Social anxiety	−11	496	.00	−1.05 [−1.25, −0.85]
Generalized anxiety	−13.2	490	.00	−1.27 [−1.48, −1.06]
Depression	−13.5	489	.00	−1.30 [−1.51, −1.09]
	Below 95 vs. 95 and above			
Parameter/comparison	*t*	*df*	*p*	*Cohen's D [LL, UL]*
Life satisfaction	10	489	.00	1.01 [0.80, 1.22]
Neuroticism	−14.8	500	.00	−1.48 [−1.70, −1.25]
Social anxiety	−10.9	496	.00	−1.09 [−1.30, −0.88]
Generalized anxiety	−12.7	490	.00	−1.28 [−1.49, −1.06]
Depression	−13.3	489	.00	−1.34 [−1.56, −1.12]

Note. df = degrees of freedom. LL and UL = Lower Limit and Upper Limit, respectively.

reappraisal subscale of the ERQ and on the NMR—two frequently used ER scales—with effect sizes of .72 and 1.17, respectively. They also scored descriptively higher on the (maladaptive) tendency to suppress one's emotions, with an effect size of .39. Moreover, they reported significantly higher attachment insecurity and current negative (but not positive) affect. RMSSD comparisons were found not significant, with small effect sizes in the expected direction (i.e., lower among the above cutoff). Also, significant, somewhat surprising differences were found in extraversion and agreeableness, with the 95 or above group reporting lower levels with moderate effect sizes, while nonsignificant differences were found for conscientiousness and openness to experience.

Thus, it appears that for the most part findings converged to depict the 95 cutoff as corresponding with other ER assessments and related constructs. That is, apart from positive affect. They also suggest a differentiation between two sets of personality traits, which could perhaps be classified as interpersonal-related, and more intra-personal related.

Discussion

In this study, we aimed to test the DERS's psychometrics, structure, and dimensionality, as well as proper utilization (i.e., total vs. subscores) and potential cutoffs for demarcating significant difficulties in ER. To that end, we compared the fit of newly gathered data to four factor structures via CFAs: a unidimensional model, a correlated-traits model, a second-ordered model, and a bifactor model. We also tested an alternative bifactor model in which two group factors (awareness and clarity) were correlated.

We found that the correlated-traits and bifactor models showed the best fit, which ranged from good, to marginal, and insufficient, depending on the specific parameter. Overall, fit as assessed by absolute indices was better than fit assessed by incremental ones. Importantly, in comparing models by TRds and looking into ΔBICs, the bifactor model rose as the best model. The data fitted the alternative bifactor model best, but it was biased to a degree of irrelevance for interpreting dimensionality. Further examination of the bifactor model and the dimensionality indices suggested that the DERS's total score is reliable and valid, yet it is not purely unidimensional, as indicated by the lower-than-threshold ECV due to the awareness factor, which was not explained by the total score. Notwithstanding, results suggested that it is "unidimensional enough" to warrant testing the utility of the total score in predicting maladaptive functioning.

Table 6. Group Descriptives—Comparing 90 and 95 Cutoffs.

	90-94 vs. 95-99					
Parameter	Group	*N*	*M*	*MD*	*SD*	*SE*
Life Satisfaction	90-94	25	6.08	6	1.91	0.38
	95-99	29	6.62	7	1.57	0.29
Neuroticism	90-94	26	25.04	24.5	6.19	1.21
	95-99	29	26.97	28	6.09	1.13
Social anxiety	90-94	25	25.48	25	12.1	2.42
	95-99	29	29.14	30	11.37	2.11
Generalized anxiety	90-94	25	7.04	6	4.56	0.91
	95-99	29	7.1	7	3.59	0.67
Depression	90-94	25	1.88	1	1.92	0.38
	95-99	29	2.52	2	1.18	0.22

	Below 90 vs. 90 and above					
Parameter	Group	*N*	*M*	*MD*	*SD*	*SE*
Life satisfaction	< 90	330	7.73	8	1.42	0.08
	90 ⩾	161	6.01	6	1.81	0.14
Neuroticism	< 90	337	20.59	21	6.05	0.33
	90 ⩾	165	29.09	30	6.08	0.47
Social anxiety	< 90	335	18.34	17	11.77	0.64
	90 ⩾	163	31.09	31	12.87	1.01
Generalized anxiety	< 90	331	3.96	3	3.08	0.17
	90 ⩾	161	8.65	8	4.73	0.37
Depression	< 90	330	0.92	1	1.08	0.06

	Below 95 vs. 95 and above					
Parameter	Group	*N*	*M*	*MD*	*SD*	*SE*
	90 ⩾	161	2.55	2	1.54	0.12
Life satisfaction	< 95	355	7.61	8	1.52	0.08
	95 ⩾	136	6	6	1.79	0.15
Neuroticism	< 95	363	20.91	21	6.16	0.32
	95 ⩾	139	29.85	31	5.77	0.49
Social anxiety	< 95	360	18.84	17	11.92	0.63
	95 ⩾	138	32.11	32	12.79	1.09
Generalized anxiety	< 95	356	4.18	4	3.29	0.17
	95 ⩾	136	8.95	8	4.72	0.40
Depression	< 95	355	0.99	1	1.18	0.06
	95 ⩾	136	2.67	3	1.44	0.12

The DERS' total score was found to have acceptable diagnostic accuracy in predicting social anxiety, and excellent in predicting generalized anxiety and depression. However, its sensitivity and specificity did not converge to the 80% benchmark, mostly in predicting probable social anxiety. We chose to contain more false negatives to avoid false positives, a decision we believe would resonate with clinicians and researchers. After comparing usages of 90 and 95 as cutoff scores, the latter was found more appropriate in terms of expected prevalence in the population and associations with related constructs. In a smaller sub-sample, comparisons between below vs. equal or above 95 in related domains were as expected—the group above the cutoff reported lower ER and satisfaction with life, higher attachment insecurity and neuroticism, and had descriptively lower parasympathetic regulation during social interaction. Also, upon exploration, we found that those below the cutoff reported significantly higher levels of extraversion and agreeableness.

These findings add to the lively discussion regarding the validity and utility of the DERS. Since it is extensively utilized, and since there is great value in assessing individuals' emotional awareness, we did not modify or exclude any of the DERS's 36 original items in our analyses. However, like in previous studies, it seems apparent that the items affiliated with the awareness factor are virtually unaccounted for by the total factor. Also, the awareness factor is found the only subfactor that could

Table 7. Group Comparison—Independent Samples T-tests Below and Above 95 Cutoff (Subsample).

Variable	Group	*N*	*M*	*MD*	*SD*	*t*	*p*	*Cohen's D [LL, UL]*
BFI—Extraversion	< 95	46	26.87	25.5	5.34	2.08	.04	0.52 [0.02, 1.03]
	95 ⩾	24	23.92	23	6.18			
BFI—Agreeableness	< 95	46	35.72	36.5	4.43	2.66	.01	0.67 [0.15, 1.18]
	95 ⩾	24	33	33	3.23			
BFI—Consc.	< 95	45	33.18	34	5.34	1.18	.24	0.30 [−0.20, 0.79]
	95 ⩾	24	31.67	32.5	4.56			
BFI—Openness to Experience	< 95	46	37.65	38.5	5.55	0.43	.67	0.11 [−0.39, 0.60]
	95 ⩾	24	37	37.5	6.89			
ERQ—Cognitive Reappraisal	< 95	45	5.14	5.17	0.97	2.82	.01	0.72 [0.19, 1.24]
	95 ⩾	23	4.41	4.5	1.10			
ERQ—Emotion Suppression	< 95	45	2.94	3	1.24	−1.53	.13	−0.39 [−0.90, 0.12]
	95 ⩾	23	3.42	3.5	1.22			
NMR—Total score	< 95	45	106.76	107	12.10	4.51	.00	1.17 [0.60, 1.73]
	95 ⩾	22	92.64	91.5	11.85			
ECR-R—Attachment Avoidance	< 95	45	52.09	50	16.97	−2.47	.02	−0.64 [−1.17, −0.11]
	95 ⩾	22	63.14	62.5	17.74			
ECR-R—Attachment Anxiety	< 95	45	64.98	63	14.89	−3.65	.00	−0.95 [−1.49, −0.40]
	95 ⩾	22	80.82	81.5	19.96			
PANAS—Positive Affect	< 95	46	2.67	2.64	0.69	0.05	.96	0.01 [−0.48, 0.51]
	95 ⩾	24	2.66	2.59	0.67			
PANAS—Negative Affect	< 95	46	1.42	1.27	0.39	−2.34	.02	−0.59 [−1.09, −0.08]
	95 ⩾	24	1.68	1.64	0.53			
RMSSD—Baseline	< 95	46	35.26	29.27	19.76	0.91	.37	0.23 [−0.27, 0.72]
	95 ⩾	24	31.01	28.66	16.17			
RMSSD—Game Trial 1	< 95	46	33.66	30.72	17.90	0.64	.52	0.16 [−0.33, 0.65]
	95 ⩾	24	30.96	28.99	13.85			
RMSSD—Negative Affect Man.	< 95	46	41.48	28.53	41.17	1.08	.29	0.27 [−0.23, 0.76]
	95 ⩾	24	32.08	28.68	15.91			
RMSSD—Game Trial 2	< 95	46	36.40	33.42	21.85	1.26	.15	0.32 [−0.18, 0.81]
	95 ⩾	24	30.22	29.85	13.58			

Note. Results are from two-sided tests, uncorrected for multiple comparisons. Consc. = Conscientiousness; Man. = Manipulation.

be considered valid, while the others converge into a general ER composite. The clarity factor is also under-explained by the total factor yet was not found to be reliable and valid. Though not a sound factor, it did significantly correlate with the awareness factor, like in Bardeen et al.'s (2015) findings where the two converged. In all, our findings do not offer a clear-cut solution for these issues. We do, however, urge researchers to retain the items associated with the awareness factor since (a) it is a growing field aiming to fill crucial gaps in our knowledge; (b) simply using the DERS without them (i.e., 30 items) could significantly impact the operationalization, and thus its validity; and (c) for further exploration into the connection and distinctions between emotional awareness and clarity.

One central issue is the meaning and caution needed when using the 95-cutoff score that we point to here. The 95 score fosters a more conservative approach to characterizing an individual as having lower ER resources. The benefit of this approach is that such a score can give confidence to a clinician or researcher in identifying ER difficulties, especially when interested in the ramifications of generalized anxiety and depression. The drawback is that the 95 score may "miss" several people who should be identified as having low ER. We thus recommend further assessments of people with scores between 85 and 95 (half a *SD* toward the mean), for example, when used for clinical evaluations. Also, as mentioned, the DERS is better at predicting generalized anxiety and depression than it is at predicting social anxiety. This raises a need for further exploration of the 95-cutoff score with regard to other disorders and conditions. This is of course challenging since many require in-depth interviews and assessments by mental health professionals.

In designing such studies, or when thinking of incorporating the DERS into an entry self-report battery in the clinical intake process, one should keep in mind our surprising finding that those who scored above the cut-off also reported lower extraversion and agreeableness, the more socially oriented traits of the Big-five. In our opinion, there are several main explanations for this finding. First, while ER has been shown to be crucial for both social and objective (i.e., task) functioning, it might be more central to social dynamics and processes than to, for example, task performance. This means some people with low ER could be experiencing propound challenges in their relationships while managing their workload. Second, we focused on anxiety and depression considering their prevalence. It could be that people who answer anxiety and depression questionnaires and score above the cutoffs have a distinctive profile, for example, high reflectiveness (at times along with self-criticism, or obsessiveness) and willingness to communicate their hardships. Other people could also be suffering yet could be less aware, or willing to acknowledge or share their struggles. Third, as mentioned, although we considered the prevalence in the general population, our main empirical "anchors" were self-reported anxiety and depression. Further information is needed regarding the robustness of this cutoff to other conditions and phenomena.

Having said that, the discussion revolving around the cutoff is still preliminary and should be perceived as such. In this work, we present one sample of its own specific characteristics. It is drawn from the general adult population, and the majority of participants were students. As such, the 95 cutoff might be valid only for this particular (though heavily studied) population. We recommend that every attempt to use a cutoff, having not yet been crystallized, be tested further with similar methodologies used here (e.g., ROCs). With the appropriate caution, results do indicate that 95 could prove a valuable cutoff. The descriptive statistics in the current sample were roughly as expected. We were in fact surprised that a 100 cutoff score was shown too restrictive in the ROCs, being roughly one *SD* above the mean and hence presumably a solid anchor. Moreover, while this sample was derived from the general population, a substantial number of participants met the criteria for social anxiety, generalized anxiety, and depression. This both surprised us but provided some reassurance that testing for a cutoff is indeed worthwhile in the current work with the hope that it could be fruitful not just in student samples.

Some additional limitations should be noted. First, some bias, above the recommended 15% threshold was found that might deter from using the DERS's total score. However, this was mostly due to the exceptionally low loadings of some of the awareness items on the general factor, along with extreme relative differences between the unidimensional and bifactor solutions. These acted as outliers, "inflating" the bias calculations. Thus, aside from the previous discussion on the complexities associated with computing the awareness factor into the total score, we see little cause for concern. Second, our findings are based almost exclusively on self-report measures, which could be biased. Future studies could potentially compare groups below and above the 95-cutoff in ER tasks, cognitive functions, and objective measures to test its predictive validity. Third, we did not control for other, well-established and highly related constructs (e.g., neuroticism) while testing the DERS' structure and utility. Hence, it could be argued that other crucial factors should have been addressed as they could affect the results. Our decision to not control for additional variables stemmed from a view of ER as overlapping with many constructs as opposed to it being contaminated by and contaminating them, and that controlling for them would result in removing a good deal of valid variance from a measure of ER. Nonetheless, we cannot refute the possibility of contamination by other factors. Finally, the current sample was comprised of mostly young women pursuing a bachelor's degree. Hence, as suggested, our findings, perhaps mostly the specifics of the accuracy of using different cutoffs for different conditions, merit further examination within larger, more diverse samples as well as in clinical samples to be solidified.

To conclude, the DERS is mostly unidimensional, and its total score is a reliable measure of a valid construct. The emotional awareness factor converges less with the other ER domains and is the only factor representing a valid and distinct construct. Hence, as per our findings, it is the only usable DERS subscore. The interplay between general ER abilities and emotional awareness requires further examination, as both fields evolve. In addition, scoring 95 or higher in the DERS' is a sound indicator of experiencing significant difficulties in ER. The robustness of this cutoff needs to be examined further, specifically among people experiencing hardships dissimilar to anxiety and depression. Nonetheless, if we utilize this cutoff based "solely" on psychometrics and prediction of probable depression and anxiety, the most prevalent mental health disorders, its impact on clinical and scholarly work could be profound. Hence, we encourage people to use it when trying to understand people better—whether in conceptualizing one's psyche and devising appropriate interventions or studying how diminished ER impacts people's well-being.

Acknowledgment

The free version of "Grammarly" was used to check for typos.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: Ilanit Gordon received funding from the ISF #434/21.

ORCID iD

Ilanit Gordon https://orcid.org/0000-0002-4888-1308

Supplemental Material

Supplemental material for this article is available online.

References

Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review*, *30*(2), 217–237. https://doi.org/10.1016/j.cpr.2009.11.004

Arroll, B., Goodyear-Smith, F., Crengle, S., Gunn, J., Kerse, N., Fishman, T., Falloon, K., & Hatcher, S. (2010). Validation of PHQ-2 and PHQ-9 to screen for major depression in the primary care population. *Annals of Family Medicine*, *8*(4), 348–353. https://doi.org/10.1370/afm.1139

Balzarotti, S., Biassoni, F., Colombo, B., & Ciceri, M. R. (2017). Cardiac vagal control as a marker of emotion regulation in healthy adults: A review. *Biological Psychology*, *130*, 54–66. https://doi.org/10.1016/j.biopsycho.2017.10.008

Bardeen, J. R., Fergus, T. A., Hannan, S. M., & Orcutt, H. K. (2015). Addressing psychometric limitations of the Difficulties in Emotion Regulation Scale through item modification. *Journal of Personality Assessment*, *98*(3), 298–309. https://doi.org/10.1080/00223891.2015.1091774

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Bollen, K. A. (1989). *Structural equations with latent variables*. John Wiley & Sons. https://doi.org/10.1002/9781118619179

Bylsma, L. M., Salomon, K., Taylor-Clift, A., Morris, B. H., & Rottenberg, J. (2014). Respiratory sinus arrhythmia reactivity in current and remitted major depressive disorder. *Psychosomatic Medicine*, *76*(1), 66–73. https://doi.org/10.1097/PSY.0000000000000019

Bytamar, J. M., Saed, O., & Khakpoor, S. (2020). Emotion regulation difficulties and academic procrastination. *Frontiers in Psychology*, *11*, Article 524588. https://doi.org/10.3389/FPSYG.2020.524588/BIBTEX

Campbell, A. A., & Wisco, B. E. (2021). Respiratory sinus arrhythmia reactivity in anxiety and posttraumatic stress disorder: A review of literature. *Clinical Psychology Review*, *87*, 102034. https://doi.org/10.1016/j.cpr.2021.102034

Catanzaro, S. J., & Mearns, J. (1990). Measuring generalized expectancies for negative mood regulation: Initial scale development and implications. *Journal of Personality Assessment*, *54*(3–4), 546–563. https://doi.org/10.1080/00223891.1990.9674019

Connor, K. M., Davidson, J. R. T., Churchill, L. E., Sherwood, A., Weisler, R. H., & Foa, E. (2000). Psychometric properties of the Social Phobia Inventory (SPIN). *British Journal of Psychiatry*, *176*(4), 379–386. https://doi.org/10.1192/bjp.176.4.379

Constantinou, M., & Fonagy, P. (2019). Evaluating bifactor models of psychopathology using model-based reliability indices. *PsyArXiv*. https://doi.org/10.31234/osf.io/6tf7j

Diener, E. D., Emmons, R. A., Sem, R. J. L., & Griffin, S. (1985). The satisfaction with life scale. *Journal of Personality Assessment*, *49*(1), 71–75. https://doi.org/10.1207/s15327752jpa4901_13

Dienstbier, R. A. (1989). Arousal and physiological toughness: Implications for mental and physical health. *Psychological Review*, *96*(1), 84–100.

Dixon-Gordon, K. L., Weiss, N. H., Tull, M. T., DiLillo, D., Messman-Moore, T., & Gratz, K. L. (2015). Characterizing emotional dysfunction in borderline personality, major depression, and their co-occurrence. *Comprehensive Psychiatry*, *62*, 187–203. https://doi.org/10.1016/j.comppsych.2015.07.014

Dueber, D. M. (2017). *Bifactor indices calculator: A Microsoft Excel-based tool to calculate various indices relevant to bifactor CFA models*. https://doi.org/https://doi.org/10.13023/edp.tool.01

Dueber, D. M. (2021). *BifactorIndicesCalculator: Bifactor indices calculator*. https://cran.r-project.org/package=BifactorIndicesCalculator

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. https://doi.org/10.1016/J.JPSYCHIRES.2014.07.029

Gordon, I., Gilboa, A., Cohen, S., Milstein, N., Haimovich, N., Pinhasi, S., & Siegman, S. (2020). Physiological and behavioral synchrony predict group cohesion and performance. *Scientific Reports*, *10*(1), 8484. https://doi.org/10.1038/s41598-020-65670-1

Gordon, I., Horesh, D., Milstein, N., Tomashin, A., Mayo, O., & Korisky, A. (2021). Pre-pandemic autonomic nervous system activity predicts mood regulation expectancies during COVID-19 in Israel. *Psychophysiology*, *58*, Article e13910. https://doi.org/10.1111/PSYP.13910

Gorsuch, R. L. (1983). *Factor analysis* (2nd ed.). Lawrence Erlbaum Associates.

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*(1), 41–54. https://doi.org/https://doi.org/10.1023/B:JOBA.0000007455.08539.94

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, *85*(2), 348–362. https://doi.org/10.1037/0022-3514.85.2.348

Hallion, L. S., Steinman, S. A., Tolin, D. F., & Diefenbach, G. J. (2018). Psychometric properties of the Difficulties in

Emotion Regulation Scale (DERS) and its short forms in adults with emotional disorders. *Frontiers in Psychology*, *9*, Article 539. https://doi.org/10.3389/fpsyg.2018.00539

Hannan, S. M., & Orcutt, H. K. (2020). Emotion regulation in undergraduate students with posttraumatic stress symptoms: A multimethod study. *Psychological Trauma: Theory, Research, Practice and Policy*, *12*(6), 643–650. https://doi.org/10.1037/TRA0000577

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Sage.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory—Versions 4a and 54*. University of California, Berkeley, Institue of Personality and Social Research.

Jovanović, V., & Lazić, M. (2020). Is longer always better? A comparison of the validity of single-item versus multiple-item measures of life satisfaction. *Applied Research in Quality of Life*, *15*(3), 675–692. https://doi.org/10.1007/s11482-018-9680-6

Lewczuk, K., Kobylińska, D., Marchlewska, M., Krysztofiak, M., Glica, A., & Moiseeva, V. (2021). Adult attachment and health symptoms: The mediating role of emotion regulation difficulties. *Current Psychology*, *40*(4), 1720–1733. https://doi.org/10.1007/S12144-018-0097-Z/FIGURES/2

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, *48*(3), 936–949. https://doi.org/10.3758/s13428-015-0619-7

Lutz, J., Gross, R. T., & Vargovich, A. M. (2018). Difficulties in emotion regulation and chronic pain-related disability and opioid misuse. *Addictive Behaviors*, *87*, 200–205. https://doi.org/10.1016/J.ADDBEH.2018.07.018

Meijen, C., Turner, M., Jones, M. V., Sheffield, D., & McCarthy, P. (2020). A theory of challenge and threat states in athletes: A revised conceptualization. *Frontiers in Psychology*, *11*, Article 126. https://doi.org/10.3389/fpsyg.2020.00126

Mikulincer, M., & Florian, V. (2000). Exploring individual differences in reactions to mortality salience: Does attachment style regulate terror management mechanisms? *Journal of Personality and Social Psychology*, *79*(2), 260–273. https://doi.org/10.1037//0022-3514.79.2.260

Milstein, N., & Gordon, I. (2020). Validating measures of electrodermal activity and heart rate variability derived from the Empatica E4 utilized in research settings that involve interactive dyadic states. *Frontiers in Behavioral Neuroscience*, *14*, Article 566230. https://doi.org/10.3389/FNBEH.2020.00148/BIBTEX

Morgan, G. B., Hodge, K. J., Wells, K. E., & Watkins, M. W. (2015). Are fit indices biased in favor of bi-factor models in cognitive ability research? A comparison of fit in correlated factors, higher-order, and bi-factor models via Monte Carlo simulations. *Journal of Intelligence*, *3*(1), 2–20. https://doi.org/10.3390/jintelligence3010002

Osborne, T. L., Michonski, J., Sayrs, J., Welch, S. S., & Anderson, L. K. (2017). Factor structure of the Difficulties in Emotion Regulation Scale (DERS) in adult outpatients receiving Dialectical Behavior Therapy (DBT). *Journal of Psychopathology and Behavioral Assessment*, *39*(2), 355–371. https://doi.org/10.1007/s10862-017-9586-x

Palumbo, R. V., Marraccini, M. E., Weyandt, L. L., Wilder-Smith, O., McGee, H. A., Liu, S., & Goodwin, M. S. (2017). Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review*, *21*(2), 99–141. https://doi.org/10.1177/1088868316628405

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. https://doi.org/10.1007/S10862-012-9292-7/FIGURES/2

Porges, S. W. (2007). The polyvagal perspective. *Biological Psychology*, *74*(2), 116–143. https://doi.org/10.1016/j.biopsycho.2006.06.009

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, *25*, 111–164. https://doi.org/10.2307/271063

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Measurement*, *73*(1), 5–26. https://doi.org/10.1177/0013164412449831

Revelle, W. (2022). *psych: Procedures for psychological, psychometric, and personality research*. https://cran.r-project.org/package=psych

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods*, *17*(3), 354–373. https://doi.org/10.1037/A0029315

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016a). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment*, *98*(3), 223–237. https://doi.org/10.1080/00223891.2015.1089249

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016b). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Salovey, P., Mayer, J. D., Goldman, S. L., Turvey, C., & Palfai, T. P. (2004). Emotional attention, clarity, and repair:

Exploring emotional intelligence using the Trait Meta-Mood Scale. In J. W. Pennebaker (Ed.), *Emotion, disclosure, & health* (pp. 125–154). American Psychological Association. https://doi.org/10.1037/10182-006

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/S11336-009-9135-Y

Sellbom, M., & Tellegen, A. (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, *31*(12), 1428–1441. https://doi.org/10.1037/pas0000623

Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in Public Health*, *5*, Article 258. https://doi.org/10.3389/FPUBH.2017.00258/ABSTRACT

Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion regulation and psychopathology. *Annual Review of Clinical Psychology*, *11*, 379–405. https://doi.org/10.1146/annurev-clinpsy-032814-112739

Smith, R., Chuning, A. E., Tidwell, C. A., Allen, J. J. B., & Lane, R. D. (2022). Psychopathic tendencies are selectively associated with reduced emotional awareness in the context of early adversity. *PLoS One*, *17*(12), e0277475. https://doi.org/10.1371/journal.pone.0277475

Smith, R., Killgore, W. D. S., & Lane, R. D. (2018). The structure of emotional experience and its relation to trait emotional awareness: A theoretical review. *Emotion*, *18*(5), 670–692. https://doi.org/10.1037/EMO0000376

Smith, R., Persich, M., Lane, R. D., & Killgore, W. D. S. (2022). Higher emotional awareness is associated with greater domain-general reflective tendencies. *Scientific Reports*, *12*(1), 1–10. https://doi.org/10.1038/s41598-022-07141-3

Smith, R., Steklis, H. D., Steklis, N. G., Weihs, K. L., & Lane, R. D. (2020). The evolution and development of the uniquely human capacity for emotional awareness: A synthesis of comparative anatomical, cognitive, neurocomputational, and evolutionary psychological perspectives. *Biological Psychology*, *154*, 107925. https://doi.org/10.1016/J.BIOPSYCHO.2020.107925

Spitzer, R. L., Kroenke, K., Williams, J. B. W., & Löwe, B. (2006). A brief measure for assessing generalized anxiety disorder: The GAD-7. *Archives of Internal Medicine*, *166*(10), 1092–1097. https://doi.org/10.1001/archinte.166.10.1092

Thayer, J. F., & Lane, R. D. (2000). A model of neurovisceral integration in emotion regulation and dysregulation. *Journal of Affective Disorders*, *61*(3), 201–216. https://doi.org/10.1016/S0165-0327(00)00338-4

Thayer, J. F., & Lane, R. D. (2009). Claude Bernard and the heart-brain connection: Further elaboration of a model of neurovisceral integration. *Neuroscience and Biobehavioral Reviews*, *33*(2), 81–88. https://doi.org/10.1016/j.neubiorev.2008.08.004

Thompson, R. A. (1994). Emotion regulation: A theme in search of definition. *Monographs of the Society for Research in Child Development*, *59*(2), 25–52. https://doi.org/https://doi.org/10.1111/j.1540-5834.1994.tb01276.x

Tull, M. T., Weiss, N. H., Adams, C. E., & Gratz, K. L. (2012). The contribution of emotion regulation difficulties to risky sexual behavior within a sample of patients in residential substance abuse treatment. *Addictive Behaviors*, *37*(10), 1084–1092. https://doi.org/10.1016/j.addbeh.2012.05.001

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*(6), 1063–1070. https://doi.org/10.1037/0022-3514.54.6.1063

World Health Organization. (2022). *Mental disorders*. https://www.who.int/news-room/fact-sheets/detail/mental-disorders