Factor structure and clinical correlates of the original and 16-item version of the Difficulties In Emotion Regulation Scale in adolescent girls with eating disorders

Elin Monell1,2 | Andreas Birgegård1,2 | Line Nordgren3 | Hugo Hesser3,4 | Johan Bjureberg1,5

1Department of Clinical Neuroscience, Centre for Psychiatry Research, Karolinska Institutet and Stockholm Health Care Services, Region Stockholm, Stockholm, Sweden
2Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden
3Department of Behavioral Sciences and Learning, Linköping University, Linköping, Sweden
4School of Law, Psychology and Social Work, Center for Health and Medical Psychology, Örebro University, Örebro, Sweden
5Department of Psychology, Stanford University, Stanford, California, USA

Correspondence
Johan Bjureberg, Department of Psychology, Stanford University, Stanford, CA 94305, USA.
Email: Johan.Bjureberg@ki.se and Johan.Bjureberg@stanford.edu

Funding information
None

Abstract

Objectives: The Difficulties in Emotion Regulation Scale (DERS) is increasingly used in adolescents. This study is the first to examine the factor structure, measurement, and structural invariance across age, reliability, and validity of the original 36-item and 16-item version of the DERS in adolescents with eating disorders.

Methods: Several models were examined using confirmatory factor analysis. Measurement and structural invariance were studied across age groups, and Omega, Omega Hierarchical, and criterion validity were examined.

Results: A bifactor model, with five subscales, showed acceptable fit in both DERS versions. Measurement and structural invariance held across age. The general factor had high reliability and accounted for a large proportion of variance in eating pathology and emotional symptoms.

Conclusion: The Awareness subscale had a negative effect on fit in DERS, but both DERS versions were reliable and valid measures in both younger and older adolescents with eating disorders when using only five subscales.
Emotion dysregulation and its role in the development and maintenance of psychiatric disorders and pathological behaviors has been increasingly examined. Theoretical models of emotion dysregulation emphasize the importance of childhood socio-emotional development for future emotion regulation abilities (Thompson, 2019). However, although adolescence often is characterized by dysregulated affect and increased risk for developing psychopathology (e.g., McLaughlin et al., 2015), there are few validated measures of emotion dysregulation for adolescents. Existing measures mainly focus on a few specific regulation strategies (e.g., Cracco et al., 2015; Gullone & Taffe, 2012), rather than covering different aspects related to the multidimensional construct of emotion regulation. The popular questionnaire for adults, the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004), is increasingly used among adolescents (e.g., Bjureberg et al., 2017, 2018; Weinbach et al., 2018), but studies examining the validity and psychometric properties of the DERS in such populations are scarce. This is the first psychometric study of both the original and a 16-item version (DERS-16; Bjureberg et al., 2016; Gratz & Roemer, 2004) in adolescents with eating disorders.

1 | EMOTION DYSREGULATION: DEVELOPMENT AND VALIDATION OF THE DERS

Emotion dysregulation has been recognized as a transdiagnostic feature in psychiatry (Aldao et al., 2016), including eating disorders. Often developed during adolescence, eating disorders frequently entail substantial psychiatric, emotion-related comorbidity (e.g., depression, anxiety disorders; Peebles et al., 2006). Although the etiology of eating disorders is unclear, emotion related difficulties, including emotion dysregulation, has increasingly been conceptualized as crucial predisposing and/or maintaining factors in eating disorders (Haynos & Fruzzetti, 2011). For instance, the influential transdiagnostic theory of eating disorders by Fairburn et al. (2003) describes "mood intolerance" and emotional avoidance as common traits in eating disorders, levels of alexithymia (i.e., difficulties in emotional understanding and expression) are often elevated, and usage of dysfunctional emotion regulation strategies are common (Aldao et al., 2010; Nowakowski et al., 2013). Eating disorder symptoms such as binge-eating, restriction, and compulsive exercise are also increasingly conceptualized in terms of such dysfunctional emotion regulation strategies (Haynos & Fruzzetti, 2011). As a result, eating disorder interventions additionally targeting emotion dysregulation are increasingly being developed (e.g., Peterson et al., 2017).

Although an agreed upon definition of emotion dysregulation is still lacking, the multidimensional model outlined by Gratz and Roemer (2004) is extensively used in psychiatric research. Here, emotion dysregulation is defined as trait-level (a) difficulties in emotional awareness and understanding, (b) difficulties in emotional acceptance, (c) difficulties engaging in goal-directed behavior and refraining from impulsive actions when in distress, and (d) lack of effective and situationally functional strategies to modulate distress. In the original exploratory factor analysis (EFA; Gratz & Roemer, 2004) in an adult student sample, a correlated six-factor solution was preferred. These factors, and subsequent subscales, were called Non-Acceptance (non-accepting stance towards emotions, negative/judgmental secondary emotions in response to negative emotions), Goals (difficulties engaging in goal-focused activities when upset), Impulse (difficulties in impulse control when upset), Awareness (difficulties and/or unwillingness to attend to internal emotional signals), Strategies (perceived lack of effective strategies for emotion regulation when upset), and Clarity (low emotional understanding). Often, a total score is calculated by summing all items (Gratz & Roemer, 2004).
Supporting validity in both community and psychiatric adult and adolescent samples, the DERS Total and subscale scores have been associated with a wide range of psychopathology, including cognitive and behavioral eating disorder symptoms (Hansson et al., 2017; Monell et al., 2018; Racine & Wildes, 2013). The DERS has also been used in adolescents with eating disorders to examine treatment outcome (Murray et al., 2015), diagnostic differences (Weinbach et al., 2018), and to differentiate patients from controls (Pace et al., 2016). DERS scores have further been shown to change after successful treatments targeting emotion regulation in adolescents with self-harm and adults with eating disorders, respectively (Bjureberg et al., 2018; Peterson et al., 2017). However, despite being increasingly used within eating disorder research, and in adolescents, little psychometric research has been done in such populations.

2 | PSYCHOMETRIC PROPERTIES AND FACTOR STRUCTURE OF THE DERS

DERS psychometrics have been examined in several studies, although primarily in adult community samples, with far fewer studies using psychiatric samples. Using confirmatory factor analysis (CFA), the original first-order six-factor correlated traits model has showed good fit in community and student adults (e.g., Bardeen et al., 2012; Miguel et al., 2017), acceptable fit in community adolescents (Neumann et al., 2010), and acceptable fit in adult both in- and outpatients with various psychiatric disorders (Fowler et al., 2014; Hallion et al., 2018; Osborne et al., 2017). However, poor fit has also been reported (adolescent and adult students, Cooper et al., 2014; community adults, Medrano & Trogolo, 2016). In the only two previous studies in psychiatric adolescent samples (Charak et al., 2019; n = 636 inpatients aged 12–17 years with severe mental illness; Perez et al., 2012; n = 218 inpatients aged 12–17 years with various psychiatric disorders), as well as in a combined sample of adult patients with eating disorders (n = 134) and healthy controls (n = 74; Wolz et al., 2015), the first-order six-factor correlated traits model showed acceptable fit. However, in the only study conducted in a solely eating disorder sample (Nordgren et al., 2020; n = 857 adult patients, the Swedish DERS version), it showed poor fit according to some fit indices.

In the aforementioned studies, the Awareness subscale has tended to have lower internal consistency, weaker intercorrelations with the other subscales (with the exception of Clarity), as well as lower or no correlation with clinical measures in both community and psychiatric samples. Subsequently, in some studies using adult samples Awareness has been excluded, resulting in somewhat better fit in both first- and second-order, and bifactor models (e.g., Bardeen et al., 2012; Fowler et al., 2014; Hallion et al., 2018; Lee et al., 2016; Sörman et al., 2021). Awareness is the only DERS scale only including reverse-scored items, which has been hypothesized to contribute to the divergence. In a modified DERS (all reverse-scored items reworded, seven items removed), Clarity and Awareness items instead formed the factor “Identification”, showing superior fit compared to the original DERS in both first- and second-order, and bifactor models (Benfer et al., 2019). This DERS has not been examined in psychiatric samples, but other modifications have been made. For instance, Osborne and colleagues (2017) found good fit for a bifactor model where Awareness was allowed to correlate with Clarity (but Awareness items were not allowed to load on the general factor). Similarly, in the eating disorder specific study by Nordgren and colleagues (2020), a modified bifactor model allowing Awareness to correlate with Clarity, and its items to load onto the general factor, was preferred (although not meeting all indices of good fit) when contrasted with a range of other models (first- and second-order six- and five-factors, unmodified six-factor bifactor model). The general factor accounted for most of the variance; little unique information was added by Strategies, Goals, and Impulse, whereas Awareness, Clarity, and Non-Acceptance provided unique information.

Thus, in adult psychiatric samples, good fit has been indicated for modified bifactor models, although these results still need replication in additional samples. Further, additional studies in psychiatric adolescent samples are needed, particularly studies contrasting different structures. Additionally, factor structures among adolescents that differ from adults have been found (e.g., in eating disorder symptoms; Forsén Mantilla at al., 2017).
Further emphasizing the need to evaluate this issue in adolescents with eating disorders specifically, a distinguishing feature of eating disorders is that patients may "deny" illness, or minimize symptom load. For example, this may be due to deficits in the ability to gauge the emotional impact of the disorder, or unwillingness to undergo treatment, factors that may be more common in younger patients (e.g., Vandereycken, 2006), or due to a phenotype that may not match traditional measures of symptomatology (Watson et al., 2019). Further, emotion dysregulation is a prominent characteristic of eating disorder and therefore, investigating how self-reported both psychological and symptom domains function in eating disorders, particularly among younger patients, is needed.

3 | PSYCHOMETRIC PROPERTIES, LATENT STRUCTURE, AND VALIDITY OF A DERS SHORT-FORM

In the interest of ease of use and avoiding repetition, the DERS-16 was developed (Bjureberg et al., 2016), comprising 16 original DERS items based on item-total correlations and item content aiming to preserve internal consistency, decrease error variance, and maintain multidimensionality. The DERS-16 includes no items from Awareness (due to low item-total correlations), but all other original subscales are represented with two or more items. Findings in both community and psychiatric samples have shown excellent internal consistency, strong convergence with the (original) DERS-36, and both DERS versions were significantly and similarly correlated with psychiatric measures (e.g., anxiety, depression, self-harm frequency), indicating construct validity and equivalence (Bjureberg et al., 2016; Shahabi et al., 2018). So far, the DERS-16 has shown good bifactor fit in one additional adult psychiatric sample and one adult eating disorder sample (Hallion et al., 2018; Nordgren et al., 2020). In an adolescent psychiatric in-patient sample, a first-order five-factor model has showed good fit (Charak et al., 2019). However, no study has examined the factor structure of the DERS-16 in an adolescent sample with eating disorders.

Finally, few studies have examined measurement and structural invariance between age groups for any DERS version. One recent study found that the factor structure and factor loadings of the DERS-16 were similar across adolescence and adulthood (Charak et al., 2019) and another found evidence for both measurement and structural invariance across ages among adults with psychosis (Lawlor et al., 2021). However, it is to our knowledge not known if the DERS shows measurement and structural invariance in younger and older adolescents. The significant emotional and cognitive development during adolescence may differentially impact DERS ratings depending on age. For instance, while older adolescents generally show improved metacognition, more emotion regulation strategies, and decreased emotional instability when compared to younger adolescents (Bailen et al., 2019; Weil et al., 2013; Zimmermann & Iwanski, 2014), they also seem to experience more negative emotional states and encounter more potential emotional stressors (McLaughlin et al., 2015). This may indicate that older adolescents would rate less difficulties in several DERS factors, although their increased negative emotionality may bias ratings or rating patterns, thus contributing to potential variance between age groups. Relatedly, lower metacognitive skills in younger adolescents could possibly also contribute to lower reliability and/or more inconsistencies in ratings in this age group. The establishment of age-related invariance would allow adolescents to be compared across age groups for both scientific and therapeutic reasons.

4 | STUDY AIMS AND HYPOTHESES

The overall aim was to examine the psychometric properties, factor structure, and validity of the DERS-36 and DERS-16 in younger and older adolescents with eating disorders. This has not been done before for either DERS version. For the DERS-36, we examined the first-order six- and five-factor correlated traits models, as well as three bifactor models (including Awareness with or without adjustments and excluding Awareness). Briefly, in a bifactor
model, items are assumed to first measure a general factor that accounts for covariation among items, and thereafter they measure uncorrelated other factors (e.g., subscales), that is, these factors account for covariation not measured by the general factor. Bifactor models are therefore particularly useful for examining the appropriateness of both total and subscales (Rodriguez et al., 2016). It is important to note that bifactor models are generally inappropriately favored when model comparisons are solely based on fit indices (Bonifay et al., 2017; Bornovalova et al., 2020). Thus, bifactor models should only be used if there is reason to believe that the true structure is bifactor, since unidimensional models only differ from these by the extent to which subfactors are given emphasis (Reise et al., 2007). Gratz and Roemer (2004) explicitly state that the goal of the DERS was to assess emotion dysregulation, and chose items that reflected different dimensions of emotion dysregulation. As noted above, ample studies have since demonstrated the utility of the total score as well as separate facets; the general factor accounts for substantial common variance but a unidimensional model has shown poor model fit. Thus, we expected that the overall DERS score would comprise a strong general factor and that non-acceptance, goals, impulse, strategies, and clarity are relevant and distinct secondary dimensions and could thus be appropriately studied in a bifactor model.

We also extended previous findings by examining bifactor models in the DERS-16, and testing for measurement and structural invariance for the DERS-36 across younger and older adolescents. Last, we tested the criterion validity of the best fitting bifactor model by examining associations between the latent factors and eating disorder psychopathology and emotional symptoms.

Based on prior findings, we hypothesized that (1) the bifactor models would produce good fit for both DERS versions, (2) there would be measurement and structural invariance across younger and older adolescents, and (3) the latent factors of the DERS would be positively associated with eating disorder psychopathology and emotional symptoms.

5 | METHODS

5.1 | Participants and procedure

Participants (N = 581) were adolescent girls aged 13–17 drawn from a nationwide clinical register for specialized eating disorder treatment units in Sweden (the Stepwise database; Birgegård et al., 2010). Stepwise inclusion criteria are self- or medical referral to a treatment unit, a DSM-IV eating disorder diagnosis (American Psychiatric Association, 2000), and intent to treat from the unit. Stepwise includes patients in all ages entering treatment since 2005. Initial assessment and Stepwise registration are performed by trained eating disorder professionals by the patient’s 3rd visit to the treatment unit. This assessment includes interviews, clinical ratings, and self-ratings (both obligatory and optional ratings) and takes around 45 min. A semi-structured interview with good validity (Structured Eating Disorder Interview; de Man Lapidoth & Birgegard, 2010) is used as a basis for diagnoses.

Data was extracted on April 17th, 2018. Registrations before April 7th, 2014 were excluded (date of DERS inclusion into Stepwise), as were registrations for patients outside the age span of interest, leaving 1755 registrations. Thereafter, the following exclusions were made: no research consent (n = 152), types of Eating Disorders Not Otherwise Specified (EDNOS) irrelevant for this study (“other” n = 49, “chewing and spitting” n = 34), and double/multiple registrations (n = 33; last registration was kept), leaving 1487 patients with no missing data at item-level (technically prohibited in the Stepwise system). Lastly, patients without DERS ratings (optional in Stepwise on individual patient basis; n = 884) and boys (n = 22) were excluded, leaving 581 patients in the final sample (41% of potential patients). Criteria for individual clinicians’ decisions to include DERS in the assessment were not recorded. The present sample (descriptives in Table 1) did not significantly differ from patients without DERS in age, BMI, diagnostic distribution, self-rated emotional symptoms, or eating disorder symptoms (see measures; examined by χ² and independent samples t-tests). Analyses at diagnostic subgroup level are presented in Tables S1 and S2.
5.2 | Instruments

DERS (Gratz & Roemer, 2004) consists of 36 items assessing emotion dysregulation (e.g., “When I am upset, my emotions feel overwhelming”) scored from 1 (almost never [0%–10%]) to 5 (almost always [91%–100%]), where higher scores indicate greater difficulties, and provides a Total Score and the six subscales Non-Acceptance, Goals, Impulse, Awareness, Strategies, and Clarity. Subscale scores are sums of constituent items, and the Total score the sum of all items. See Table S3 for a complete list of items.

DERS-16 (Bjureberg et al., 2016) was extracted from DERS-36, retaining 3 items from the Non-Acceptance subscale, 3 items from the Goals subscale, 3 items from the Impulse subscale, 5 items from the Strategies subscale, 2 items from the Goals subscale, and no items from the Awareness subscale (Table S3).

Eating Disorder Examination Questionnaire (EDE-Q: Fairburn & Beglin, 1994) adolescent version (Carter et al., 2001) consists of 36 items assessing cognitive and behavioral eating disorder symptoms over the past 14 days. It is scored 0–6 where higher scores indicate more frequent symptoms. In this study, we computed the mean Global Score measuring overall eating disorder psychopathology, and used it to examine convergent validity since the DERS has shown moderate associations with EDEQ Global score in adults with eating disorders (Monell et al., 2020). The EDE-Q adolescent version has adequate psychometric properties (Forsén Mantilla et al., 2017). Internal consistency was good in the present sample (Cronbach’s α = 0.96).

Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997) self-report version (Goodman et al., 1998) consists of 25 items screening for psychosocial difficulties the last 6 months. It is scored 0–2 (Not true; Somewhat true; Certainly true) and provides one Total Difficulties Score and five subscales. The subscale Emotional symptoms (psychosomatic symptoms, worry, fear; α = 0.72 in the present sample) consists of five items, where higher scores indicate more difficulties, and was expected to be associated with emotion dysregulation since it can be said to target the consequences of poor emotion regulation. The original and Swedish SDQ self-report versions have shown good psychometric properties (Goodman, 2001; Hagquist, 2007).

5.3 | Statistical analyses

5.3.1 | Factor structure of DERS-36 and DERS-16

CFA was used to investigate the factor structure of DERS-36 and DERS-16. A series of CFA models were tested based on prior examinations in clinical samples (first-order correlated traits and bifactor models). Models were fitted using Mplus 8.0 and the non-normality maximum likelihood estimator, MLR (Muthén & Muthén, 2017).

Factor Structure of DERS-36. We examined the original first-order six-factor correlated traits model (model 1.1) and a bifactor model with six specific factors (model 2.1). Based on previous findings, a priori modifications were made regarding the Awareness subscale, either excluding these items (five factors; correlated traits model 1.2 and bifactor model 2.3) or allowing correlations between the Awareness and Clarity subscales (six specific factors; bifactor model 2.2). See Table 2 for all models.

Factor Structure of DERS-16. The first-order five-factor correlated traits model (Model 3) and the bifactor model with five specific factors (Model 4) were examined. Given that the Clarity subscale was only represented by two items, loadings were constrained to 1 for these items to make the model globally identified.

Following recommendations, several goodness-of-fit indices were examined: the chi square test, the comparative fit index (CFI), the standardized root mean square residual (SRMR), and the root mean square error of approximation (RMSEA) with 90% confidence intervals (CIs). The following guidelines were used to determine good fit (Hu & Bentler, 1999; Wang & Wang, 2020): a combination of CFI >0.95, SRMR 0.90, and RMSEA <0.08. We also compared nested models using a scaled χ² difference test (Δχ²; Satorra & Bentler, 2001), where a significant difference between two models indicates worse fit of the more restricted model.
5.3.2 | Measurement and structural invariance

We examined measurement and structural invariance in younger (13–15 years; n = 338) and older adolescents (16–17 years; n = 243). First, the best-fitting bifactor model identified in the total sample was fitted in younger and older adolescents to confirm that the overall factor structure did not statistically significantly differ across age groups (Δχ²). We then tested configural invariance and strong measurement invariance (i.e., scalar invariance; invariant intercepts and loadings; Meredith, 1993) in multigroup CFA. We also fitted the configural model as a baseline model that we compared more restricted subsequent models with. The baseline model had the same factor structure in each group, but intercepts, loadings, and residuals and structural parameters (factor variance) were allowed to vary across groups. We then fitted a model in which we constrained both item intercepts and factor loadings to be invariant across age groups. This model was compared to the baseline model, and if it was not statistically significantly worse (Δχ²), that was taken as evidence for strong measurement invariance.

We then continued to test the structural parameters of the best-fitted bifactor CFA model by first testing the factor variance and then testing the factor means in multigroup CFA (using models in which intercepts and loadings were constrained to be equal). We examined factor variance invariance by testing if a model in which factor variance was constrained to be invariant across age groups was statistically significantly different than a model in which factor variance were free parameters. Factor mean invariance was examined by testing if a model that constrained all factor means to be equal across groups with a model in which factor means differed as a function of age group were statistically significantly different from each other. If the models constraining factor variance and means to be equal across age groups did not statistically significantly degrade model fit, structural invariance was presumed to be established.

5.3.3 | Scale reliability and explained common variance (ECV) of DERS-36

We examined the reliability of the Total and subscales for the best bifactor model by calculating Omega (Raykov, 1997) and Omega Hierarchical (OmegaH; Rodriguez et al., 2016), which range from 0 (no reliability) to 1 (perfect reliability). These are calculated instead of coefficient alphas because OmegaH seems particularly useful for multidimensional models including both total and subscale factors (Rodriguez et al., 2016). Further, unlike alpha they do not assume equal measurement error and provide a more accurate reliability estimate (Raykov, 1997). Omega uses all sources of reliable variance and can be thought of as a weighted coefficient alpha, while OmegaH uses reliable variance that is specifically accounted for by a particular scale. Thus, for the Total scale, OmegaH indicates the reliable variance explained by all items loading on the general factor excluding reliable variance explained by the subscales, while subscale OmegaH indicates reliable variance explained by specific subscale items excluding variance explained by these items’ loading on the general factor. Both global- and subscales are appropriate to use if items load substantially on both the general and the specific factors. There are no specific criteria for loadings on specific factors offered in the literature, but >0.50 might serve as such a value (although values closer to 0.75 may be more optimal; Reise et al., 2013).

According to recommendations for evaluating bifactor models, we also calculated ECV for the final bifactor model (Rodriguez et al., 2016). ECV can be considered as an indicator of the degree to which a measure is unidimensional. High ECV (greater than 0.70–0.80) indicates that most of the common variance is explained by the general factor. Low ECV indicates that most of the variance is explained by the subscales (Rodriguez et al., 2016).

5.3.4 | Criterion validity of the DERS-36

In the best-fitting bifactor original DERS-36 CFA model, we examined how much variability each latent factor accounted for variance in eating pathology (EDE-Q) and emotional symptoms (SDQ Emotional symptoms). This was
done by separately including EDE-Q and SDQ Emotional symptoms as dependent variables in the CFA model. We used the latent factors instead of observed scores of the scales to be able to test whether each latent factor accounted for unique variance in the criterion variables. This method also accounts for measurement error and thus eliminates regression bias (Raykov et al., 2018).

6 | RESULTS

6.1 | Sample descriptives

The sample consisted of 581 girls with eating disorders (demographic and clinical variables in Table 1). EDNOS was the most prevalent ED diagnosis followed by anorexia nervosa (AN)-R. Within EDNOS, the most prevalent type was Type 2 (atypical AN) followed by Type 4 (inappropriate compensatory behaviors without binge-eating).

6.2 | Factor structure of the DERS-36

In total, five models were examined using CFA. See Table 2 for fit indices for each model.

First-Order Correlated Traits Models. Neither the original six-factor correlated traits model (Model 1.1) nor the five-factor correlated traits model (Model 1.2, excluding Awareness items) fully reached criteria for acceptable fit (CFI was <0.90).

Bifactor models. The first bifactor model with six specific factors (Model 2.1) showed poor model fit (CFI was <0.90). Allowing Awareness to correlate with Clarity in the second bifactor model with six specific factors (Model 2.2) provided good RMSEA and SRMR, and acceptable CFI; thus the model fit was considered acceptable. This adjustment improved fit significantly ($\Delta \chi^2(1) = 250.433, p < 0.001$). Excluding the Awareness items in the third bifactor model with five specific factors (Model 2.3; see Table S3 for factor loadings) also resulted in acceptable fit, and was statistically significantly better than model 2.2 ($\Delta \chi^2(182) = 482.148, p < 0.001$).

As model 2.3 provided statistically better fit compared to model 2.2, it was subsequently analyzed for reliability and criterion validity. A bifactor model was chosen for these subsequent analyses because it was deemed theoretically sensible and offers the opportunity to evaluate OmegaH, ECV, and can be used for determining to what extent latent factor scores are reliable after accounting for the general factor.

6.3 | Factor structure of the DERS-16

Two models were examined, see Table 2 for model fit indices for the respective models. The correlated five-factor model (Model 3) provided good fit according to all fit indices, as did the bifactor model (Model 4; see Table S3 for factor loadings).

6.4 | Reliability and ECV of the DERS-36

The reliability of the DERS-36 bifactor Model 2.3 (Awareness excluded) were examined by obtaining Omega- and OmegaH-values, presented in Table 3. Omegas were high for all scales. OmegaH for the Total scale was high, indicating that most reliable variance was accounted for by the assumed general emotion dysregulation factor
As for subscales, only the Clarity factor accounted for substantial unique variance independent of the general factor (i.e., OmegaH values >0.50 leading to higher proportion of reliable variance). In contrast, Strategies accounted for little unique variance independent of the general factor, while Non-Acceptance, Goals, and Impulse accounted for some unique variance.

The estimates of ECV statistic was 0.898 for the DERS-36 bifactor Model 2.3. This suggested that most of the common variance were accounted for by the general factor and thus only a small proportion of the variance (1–0.898 = 0.102) seems to be explained by subscale factors beyond the general factor.

### TABLE 1  
Descriptive and clinical variables; number and percentage of eating disorder diagnoses, means or sums (DERS scales), SDs and minimum to maximum values of age, BMI, DERS-36 and -16 scales, and measures of psychopathology

<table>
<thead>
<tr>
<th></th>
<th>$M$/sum</th>
<th>$SD$</th>
<th>Min–Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td>15.16</td>
<td>1.32</td>
<td>13–17</td>
</tr>
<tr>
<td>BMI</td>
<td>18.86</td>
<td>3.31</td>
<td>11.70–42.60</td>
</tr>
<tr>
<td>AN-R$^a$</td>
<td>165</td>
<td>27.4%</td>
<td></td>
</tr>
<tr>
<td>AN-BP$^a$</td>
<td>20</td>
<td>3.3%</td>
<td></td>
</tr>
<tr>
<td>BN$^a$</td>
<td>50</td>
<td>8.3%</td>
<td></td>
</tr>
<tr>
<td>EDNOS$^{a,b}$</td>
<td>368</td>
<td>61%</td>
<td></td>
</tr>
<tr>
<td>DERS-36 Non-Acceptance</td>
<td>15.39</td>
<td>6.29</td>
<td>6–30</td>
</tr>
<tr>
<td>DERS-36 Goals</td>
<td>17.50</td>
<td>5.41</td>
<td>5–25</td>
</tr>
<tr>
<td>DERS-36 Impulse</td>
<td>15.74</td>
<td>6.71</td>
<td>6–30</td>
</tr>
<tr>
<td>DERS-36 Awareness</td>
<td>19.00</td>
<td>5.28</td>
<td>6–30</td>
</tr>
<tr>
<td>DERS-36 Strategies</td>
<td>22.90</td>
<td>8.21</td>
<td>8–40</td>
</tr>
<tr>
<td>DERS-36 Clarity</td>
<td>14.55</td>
<td>4.83</td>
<td>5–25</td>
</tr>
<tr>
<td>DERS-36 Total Score</td>
<td>105.08</td>
<td>28.04</td>
<td>41–175</td>
</tr>
<tr>
<td>DERS-16 Non-Acceptance</td>
<td>8.22</td>
<td>3.36</td>
<td>3–15</td>
</tr>
<tr>
<td>DERS-16 Goals</td>
<td>10.52</td>
<td>3.38</td>
<td>3–15</td>
</tr>
<tr>
<td>DERS-16 Impulse</td>
<td>7.55</td>
<td>3.75</td>
<td>3–15</td>
</tr>
<tr>
<td>DERS-16 Strategies</td>
<td>14.11</td>
<td>5.39</td>
<td>5–25</td>
</tr>
<tr>
<td>DERS-16 Clarity</td>
<td>5.68</td>
<td>2.28</td>
<td>2–10</td>
</tr>
<tr>
<td>DERS-16 Total Score</td>
<td>46.08</td>
<td>14.85</td>
<td>16–80</td>
</tr>
<tr>
<td>EDE-Q Global</td>
<td>3.51</td>
<td>1.55</td>
<td>0–6</td>
</tr>
<tr>
<td>SDQ Emotional symptoms</td>
<td>6.03</td>
<td>2.43</td>
<td>0–10</td>
</tr>
</tbody>
</table>

Note: $N = 581$.

Abbreviations: AN-BP, anorexia nervosa binge/purge subtype; AN-R, anorexia nervosa restricting subtype; BMI, body mass index; BN, bulimia nervosa; DERS, Difficulties in Emotion Regulation Scale; EDE-Q, Eating Disorder Examination Questionnaire; EDNOS, eating disorders not otherwise specified; SDQ, Strengths and Difficulties Questionnaire.

$^a$= Presented as $n$ and %.

$^b$= EDNOS distribution: Type 1 (AN without amenorrhea), $n = 44$; Type 2 (atypical AN), $n = 159$; Type 3 (low-frequency BN), $n = 21$; Type 4 (inappropriate compensatory behaviors without binge-eating), $n = 139$; Type 6 (binge eating disorder), $n = 5$.
6.5.1 Measurement invariance

The DERS-36 bifactor solution omitting the Awareness factor (Model 2.3) showed acceptable fit in both the younger and older subsamples (see Table 2). The configural multigroup CFA model without equality restrictions on factor loadings and intercepts across the two groups also demonstrated acceptable fit (see Table 2), warranting further tests of invariance. Similarly, multigroup CFA analyses revealed that a scalar model in which factor loadings...
and intercepts were set to be equal across age groups (see Table 2) did not have significantly worse model fit compared to the configural model, $\Delta \chi^2(78) = 52.271, p = 0.989$. This indicated strong measurement invariance between age groups and thus further tests of potential differences in structural parameters across age groups were deemed justified.

### 6.5.2 | Structural invariance

Structural invariance of the bifactor model excluding awareness was first tested in a multigroup CFA by comparing a model in which factor variance was held equal across age groups ($\chi^2(834) = 1644.851, p < 0.001; \text{RMSEA} = 0.058 [95\% \text{ CI} : 0.054, 0.062]; \text{CFI} = 0.923; \text{SRMR} = 0.059$) to an identical model in which factor variance could vary in each group ($\chi^2(828) = 1637.020, p < 0.001; \text{RMSEA} = 0.058 [95\% \text{ CI} : 0.054, 0.062]; \text{CFI} = 0.923; \text{SRMR} = 0.055$). The constrained model did not significantly degrade model fit ($\Delta \chi^2(6) = 7.493, p = 0.27$). Next, a model which constrained factor loadings, intercepts, factor means, and factor variance to be equal across groups ($\chi^2(840) = 1666.589, p < 0.001; \text{RMSEA} = 0.058 [95\% \text{ CI} : 0.054, 0.062]; \text{CFI} = 0.921; \text{SRMR} = 0.057$) was compared to a model in which factor means were free parameters ($\chi^2(834) = 1658.550, p < 0.001; \text{RMSEA} = 0.058 [95\% \text{ CI} : 0.054, 0.062]; \text{CFI} = 0.921; \text{SRMR} = 0.057$). Results suggested that the models did not differ significantly from each other, $\Delta \chi^2(6) = 7.769, p = 0.256$. Thus, there were no statistically significant differences in factor dispersion and means of factor scores between the two age groups.

### 6.6 | Criterion validity of the DERS-36

The general factor was statistically significantly associated with total scores on the EDE-Q ($z = 12.037, p < 0.001$) and accounted for 36.8% of the variance ($R^2 = 0.368$). The model in which EDE-Q was regressed on all the subfactors and the general factor was statistically significantly better than the model in which EDE-Q was only regressed on the general factor according to the scaled $\chi^2$ difference test ($\Delta \chi^2(5) = 20.033, p = 0.001$). The other subscales accounted for an additional 1.1% beyond the general factor. Clarity was the only other latent factor that was statistically significantly associated with the total score of the EDE-Q ($z = 3.932, p < 0.001$).

The general factor was also associated with SDQ emotional symptoms ($z = 11.743; p < 0.001$) and accounted for 44.5% of the variance ($R^2 = 0.445$). Regressing emotional symptoms on all the subfactors and the general factor statistically significantly improved the model fit ($\Delta \chi^2(5) = 23.379, p < 0.001$). The other subscales accounted for an additional 1.0% beyond the general factor. Goals ($z = 1.999, p = 0.046$), Strategies ($z = 2.136, p = 0.033$), and Clarity ($z = 3.014, p = 0.003$) were statistically significantly associated with the SDQ emotional symptoms.

### 7 | DISCUSSION

In this study, the psychometric properties of the original DERS and the short-form DERS-16 were examined in adolescent girls with eating disorders. Psychometric examinations of both versions in clinical samples are sparse, particularly in clinical adolescents. We concluded that a bifactor model with five subscales showed acceptable fit in both DERS versions. Measurement and structural invariance held across younger and older adolescents. The general factor (in the model excluding Awareness) had high reliability and accounted for large proportion of variance in eating pathology and emotional symptoms.
7.1 | Factor structure of the DERS-36 and -16

As hypothesized, bifactor models showed acceptable fit for the DERS-36, but a model excluding Awareness was the best fitting one. As such, our results add to evidence of the suitability of a bifactor model of the DERS in clinical samples (Hallion et al., 2018; Nordgren et al., 2020; Osborne et al., 2017). The only prior studies examining the DERS-36 factor structure in clinical adolescents (Charak et al., 2019; Perez et al., 2012) only examined the first-order six-factor correlated traits model. This structure showed acceptable fit, similar to our results. The adjusted bifactor model examined in this study (Awareness allowed to correlate with Clarity) was first examined by Nordgren et al. (2020); currently, the only study focusing solely on adult patients with eating disorders. Here, the adjusted bifactor model showed the best fit relative to a series of other models. However, a bifactor model excluding Awareness altogether was not examined. Otherwise in eating disorders, only the six-factor correlated traits model has been examined, showing acceptable fit in a combined adult patient and control sample (Wolz et al., 2015).

Because the bifactor models examined have varied across studies in adults, and no other examination in adolescents exists, comparison is difficult. Although ideally in bifactor models, the specific factors should not be allowed to correlate, the adjusted versions allowing a correlation between Awareness and Clarity resulted in slightly better fit compared to other models in Nordgren et al. (2020) and Osborne et al. (2017). In our sample, this modified bifactor model showed slightly but significantly worse fit than a bifactor model excluding Awareness altogether. Together with previous studies showing problems with the Awareness factor (e.g., Bardeen et al., 2012; Fowler et al., 2014; Osborne et al., 2017), this model excluding Awareness was chosen for subsequent reliability and validity analyses. For the DERS-16, both the bifactor model and the first-order five-factor correlated traits model showed good fit. Thus, the bifactor model for the DERS-16 has now shown good fit in three clinical samples including the present one (Hallion et al., 2018; Nordgren et al., 2020). The first-order correlated traits model has shown good fit in psychiatric adolescents (Charak et al., 2019). It has also showed good fit for two other DERS short-forms in both psychiatric and community adolescents (Charak et al., 2019; Kaufman et al., 2015), as has two bifactor models in community adolescents (including and excluding Awareness items; Moreira et al., 2020). It is important to note that both the best-fitting bifactor model and the DERS-16 do not include any items from the Awareness factor. Hence, the total scores derived from either factor solution do not directly compare to the original DERS-36 total score. To date, three short versions of the DERS exists but have only been compared in undergraduates (Skutch et al., 2019). Whether one short form has better validity and reliability among clinical adolescents has yet to be examined.

7.2 | Measurement and structural invariance of the DERS-36

The present study is the first to test for measurement and structural invariance of the DERS-36 for younger and older adolescents with eating disorders. Findings showed that the assumption of strong measurement invariance was met. Also, variation in factor scores and mean factor scores were similar across age groups. Hence, the DERS-36 (omitting the Awareness subscale) assesses the same latent construct in both younger and older adolescents and show similar emotion regulation difficulties. This means younger and older adolescents with eating disorders are comparable in terms of emotion dysregulation, which is important for both cross-sectional and longitudinal studies of emotion dysregulation in general and the DERS in particular. These findings are in line with research showing that the DERS-16 is measurement invariant between adolescent and adult samples (Charak et al., 2019). Although other measures of emotion regulation have been adapted from adult language to young adolescent language (e.g., Gullone & Taffe, 2012), our results indicate that the same DERS version may be used for both younger and older adolescents.
7.3 | Reliability and ECV of the DERS-36

Reliability for the Total score was good for both examined bifactor models. High OmegaH for Total scores and high ECV indicate that the majority of the ECV can be attributed to the Total score. Regarding subscales, OmegaH showed that only Clarity accounted for substantial reliable variance on its own. Variance in Strategies was almost entirely attributable to the general factor. Our OmegaH-values and ECV were similar to those obtained in adult patients with eating disorders (Nordgren et al., 2020). Osborne et al. (2017), which to our knowledge is the only other study examining OmegaH-values for the DERS in a clinical sample (although excluding Awareness), also found that Clarity as well as Non-Acceptance accounted for substantial reliable variance on their own. It is noteworthy that the relatively high subscale Omegas of the DERS suggests that the subscales are reliable indicators of their unique constructs, although the total score may explain more variability than the individual subscales.

Thus, in our eating disorder sample and that of Nordgren et al. (2020), DERS Total score seems reliable and appropriate to use. As for the subscales, only Clarity items loaded substantially on both the general and the specific factors. That is, only this subscale contributed substantial information over and above the general emotion dysregulation factor (Reise et al., 2013), and the relevance of the other scores over and above the general factor should be investigated further.

7.4 | Validity of the DERS-36

The results from the best fitting bifactor model showed that the general factor in DERS-36 was strongly associated with eating pathology (36.8%). Overall, although contributions from other subscales were statistically significant, they accounted for little variance (1.1%) beyond the general factor, with Clarity alone significantly contributing in accounting for eating pathology. Our results corroborate previous findings in psychiatric adolescents (Perez et al., 2012), suggesting that the DERS is a valid measure of global emotion dysregulation among adolescents relevant to eating disorders. As expected, the general factor was also strongly linked to emotional symptoms (44.5%). The subfactors accounted for an additional 1% beyond the general factor where Clarity, Goals, and Strategies were statistically significant contributors on their own.

8 | IMPLICATIONS

Overall, our results indicate that the DERS-36 Total score is a reliable and valid measure in adolescents with eating disorders, and although the Total score may explain more variability than the individual subscales, the subscales are still reliable indicators of their unique constructs. This adds to prior findings of the usefulness of the DERS among adolescents with psychiatric disorders (Charak et al., 2019; Perez et al., 2012). Within eating disorders, the DERS-36 seems to function quite similarly in both younger and older adolescent groups. In comparison to adults, our adolescents descriptively rated slightly higher emotion dysregulation than the adults in Nordgren et al. (2020), which may reflect that dysregulated affect is generally accentuated in adolescence (McLaughlin et al., 2015). Otherwise, factor structure and reliability measures were similar in both age groups. The DERS-16 also seemed to function adequately for these adolescents, so the DERS-16 is more time efficient for this population.

We found similar levels of emotion dysregulation as the inpatient adolescents in the study of Perez et al. (2012). Due to the small number of boys with DERS ratings in the Stepwise registry, we could not examine the psychometric properties of the DERS in boys. Therefore, whether or not the DERS works similarly among girls and boys with eating disorders remains to be examined.

As for the underlying structure of the DERS-36, a bifactor model excluding Awareness best fitted our data. Which model should be preferred is still unclear pending more studies in similar populations. It may also depend on
the estimated clinical utility of the Awareness factor, that is, if emotional awareness is of particular interest. Our results further indicated that the Total score is both valid and reliable in adolescents with eating disorders, and that Clarity seems to measure unique aspects of emotion dysregulation. As for the remaining subscales, and Strategies in particular, our and Nordgren et al.’s (2020) results indicate that within eating disorder populations, there are reasons to be cautious when interpreting results from these subscales as they mimic general emotion dysregulation. As this study is the first to examine underlying structures other than the first-order five-factor model in the DERS-16 (or any other DERS short-form) in clinical adolescents, these findings need replication.

8.1 | Strengths and limitations

A major strength is our relatively large adolescent eating disorder sample, which is lacking in previous research. Our sample was drawn from a nationwide clinical registry, strengthening the ecological validity of our findings. Further, we examined the psychometric properties of two DERS versions, and replicated prior findings on the suitability of a bifactor model, as previously shown in adult eating disorders (Nordgren et al., 2020). The bifactor structure is also particularly useful for examination of multidimensional constructs (Rodriguez et al., 2016), and it allowed us to examine reliability by calculating Omega and OmegaH. Lastly, we extended the findings by Nordgren et al. (2020) by also examining the convergent and divergent validity of both DERS versions in an eating disorder sample.

There are also several limitations. First, as the DERS is an optional measure in the Stepwise database administered to patients on an individual basis, and the sample might be biased in unknown ways. However, a prior study of 999 adult patients with DERS data from the same database did not find any meaningful differences between patients with or without DERS ratings; instead, clinician/clinic variables (e.g., interest, time constraints) were likely more influential on DERS administration than patient characteristics (Monell et al., 2018). In our study, 41% of patients were administered the DERS, and patients with or without DERS ratings did not differ in any meaningful way on a variety of clinically relevant variables, suggesting representativeness. Second, the DERS-16 was extracted from the DERS-36, and replication of our findings using separate ratings are needed. Third, when examining the validity of the DERS-36, we had no other measures of emotion dysregulation. Also, instead of using the rather brief and nonspecific Emotional symptoms subscale of the SDQ, it likely would have been informative to include more specific measures of emotion related pathology (e.g., anxiety, depression). Fourth, as noted, our sample only included no male patients, and we were therefore unable to examine potential gender differences in terms of reliability and validity. Also, our findings may not generalize to boys, or to girls with or without psychiatric disorders other than eating disorders. Last, both conceptual and statistical concerns have been raised regarding bifactor models. The former relates to difficulties in how to interpret the meaning of a bifactor structure in psychological constructs; as for the latter, in terms of fit indices, bifactor structures tend to outperform other models, possibly due to statistical bias (Bonifay et al., 2017; Murray & Johnson, 2013). Even though the bifactor models showed the best fit in our sample and was theoretically justified, fit indices for first-order models were not that different. Our results replicated findings in adults with eating disorders (Nordgren et al., 2020), but we cannot conclude that the bifactor structure is a valid representation of the true latent structure in the population. Even so, bifactor models are useful for evaluation of multidimensional constructs by allowing calculation of indices such as Omega and ECV that can help determine the appropriateness of total and subscales (Rodriguez et al., 2016).

9 | CONCLUSIONS

Our results indicate that the DERS-36 overall is a reliable measure in both younger and older adolescent girls with eating disorders, also showing satisfactory convergent validity. The DERS-16 also seemed to function adequately for these adolescents, showing high concordance with the full DERS version. A bifactor structure fit data for both
versions in our sample. The Total score seemed reliable, but most subscales suffered from low unique reliable variance, suggesting caution when interpreting results.

CONFLICT OF INTERESTS
The authors declare that there are no conflict of interests.

ETHICS AND PATIENT CONSENT STATEMENT
All procedures involving human participants were in accordance with the ethical standards of the APA, the institutional and/or national research committee, and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. The study was approved by the Stockholm regional ethics committee (reference numbers 2015/928-31/4; 2018/2039-32). Data for this study were extracted from the Stepwise database, covering the majority of eating disorder specialist units in Sweden since 2005. Thus, participants were not informed about this specific study. However, participants and their legal guardians are asked for general research consent that their data anonymously may be used in future research and publication in international journals when Stepwise assessment is performed. All participants in this study have given such general research consent.

DATA AVAILABILITY STATEMENT
Data belongs to the Swedish Stepwise clinical database and are not available for sharing. Code for confirmatory factor analyses and reliability analyses are available by demand.

ORCID
Elin Monell  http://orcid.org/0000-0002-3305-4844
Andreas Birgegård  http://orcid.org/0000-0003-1220-9680
Line Nordgren  http://orcid.org/0000-0002-9960-4431
Hugo Hesser  http://orcid.org/0000-0002-9736-8228
Johan Bjureberg  http://orcid.org/0000-0003-2747-9990

PEER REVIEW
The peer review history for this article is available at https://publons.com/publon/10.1002/jclp.23286

ENDNOTE
1Estimating sample size for SEM is difficult, as there is no universal criterion for a suitable sample size that applies to all circumstances and models. However, when it comes to multigroup modeling, the rule of thumb that is often stated is a minimum of 100 cases per group. (Wang & Wang, 2020).

REFERENCES


**SUPPORTING INFORMATION**

Additional supporting information may be found in the online version of the article at the publisher’s website.

---

**How to cite this article:** Monell, E., Birgegård, A., Nordgren, L., Hesser, H., & Bjureberg, J. (2021). Factor structure and clinical correlates of the original and 16-item version of the Difficulties In Emotion Regulation Scale in adolescent girls with eating disorders. *Journal of Clinical Psychology, 1–19.*

https://doi.org/10.1002/jclp.23286