The emotion regulation questionnaire: psychometric properties in general community samples

David A. Preece  
*Edith Cowan University*

Rodrigo Becerra

Ken Robinson  
*Edith Cowan University*

James J. Gross

Follow this and additional works at: https://ro.ecu.edu.au/ecuworkspost2013

This is an Accepted Manuscript of an article published by Taylor & Francis in the Journal of Personality Assessment on 4 February 2019, available online: http://www.tandfonline.com/10.1080/00223891.2018.1564319
This Journal Article is posted at Research Online. https://ro.ecu.edu.au/ecuworkspost2013/5915
The Emotion Regulation Questionnaire: Psychometric Properties in General Community Samples

David A. Preece¹, Rodrigo Becerra², Ken Robinson¹, James J. Gross³

¹Edith Cowan University, Perth, Australia.
²The University of Western Australia, Perth, Australia.
³Stanford University, Stanford, USA.
Abstract

The Emotion Regulation Questionnaire (ERQ) is a 10-item self-report measure of two emotion regulation strategies; cognitive reappraisal and expressive suppression. It is a widely used measure of emotion regulation, however, its factor structure has rarely been examined outside of university student samples, and some authors have recently questioned its factorial validity in general community samples. In this study, we examine the psychometric properties of the ERQ (original English version) in three Australian general community samples (Ns = 300, 400, 348). Confirmatory factor analyses in each sample demonstrated that the traditional 2-factor model (comprised of cognitive reappraisal and expressive suppression factors) was replicable and an excellent fit to the data. In all samples, ERQ cognitive reappraisal (α = .89-.90) and expressive suppression (α = .76-.80) scores had acceptable to excellent levels of internal consistency reliability. As expected, cognitive reappraisal scores were significantly negatively correlated with psychological distress and alexithymia, whereas expressive suppression scores were significantly positively correlated with psychological distress and alexithymia. We conclude that, similar to previous findings in student samples, the ERQ has strong psychometric properties in general community samples and can therefore be used confidently regardless of participants’ student status.
The Emotion Regulation Questionnaire: Psychometric Properties in General Community Samples

People use various strategies to attempt to modify the trajectory of their emotions and such attempts constitute emotion regulation (Gross, 2015a). Successful emotion regulation plays a crucial part in mental health (Aldao et al., 2010), and emotion regulation problems are a risk factor for a range of psychiatric disorders, including affective (Rottenberg, Gross, & Gotlib, 2005), anxiety (Mennin & Farach, 2007), substance use (Fox, Axelrod, Paliwal, Sleeper, & Sinha, 2007), eating (Bydlowski et al., 2005) and personality disorders (Linehan, 1993). Researchers and clinicians therefore need valid measures of emotion regulation.

Several self-report questionnaires have been developed for this purpose (for a review, see John & Eng, 2014), with one of the most commonly used being the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003). The ERQ was originally developed in English and has since been translated into 33 languages (see Stanford Psychophysiology Laboratory, 2018). To date though, few studies have formally examined the ERQ’s factor structure outside of university student populations, and Spaapen, Waters, Brummer, Stopa and Bucks (2014) have recently argued that it needs modifications to have adequate factorial validity in general community samples.

In the present study, we examine the psychometric properties of the ERQ (original English version) in general community samples. Prior to reporting the results of our study, we firstly describe the ERQ, and summarise the results of existing studies that have examined its concurrent/criterion validity, internal consistency reliability, and factor structure.

**Emotion Regulation Questionnaire**

The ERQ is a 10-item self-report questionnaire based on Gross’s (1998) process model of emotion regulation. This model categorises emotion regulation strategies based on how early they are activated in the emotion generation process, and hypothesises that
different regulation strategies might have different consequences. The ERQ is designed to measure people’s usage of two regulation strategies: an antecedent-focused strategy called *cognitive reappraisal* (6 items, e.g., “When I’m faced with a stressful situation, I make myself think about it in a way that helps me stay calm”) where a person attempts to change how he or she thinks about a situation in order to change its emotional impact, and a response-focused strategy called *expressive suppression* (4 items, e.g., “I keep my emotions to myself”) where a person attempts to inhibit the behavioural expression of his or her emotions (Gross & John, 2003). Separate scale scores are derived for these two regulation strategies. All items are answered on a 7-point Likert scale, ranging from 1 (strongly disagree) to 7 (strongly agree), with higher scores indicating higher usage of that strategy.

**Concurrent/Criterion Validity and Reliability of the ERQ**

The process model of emotion regulation predicts that, when used habitually, cognitive reappraisal should generally have more beneficial outcomes than expressive suppression (John & Gross, 2004). Indeed, in many studies with participants from predominantly Western cultural backgrounds, high ERQ *cognitive reappraisal* scores have usually been associated with adaptive outcomes, such as lower levels of psychopathology symptoms (e.g., Badcock et al., 2011; Joormann & Gotlib, 2010), higher levels of positive affect (e.g., Balzarotti, Gross, & John, 2010; Preece, Becerra, & Campitelli, 2018) and better interpersonal functioning (e.g., Cabello, Salguero, Fernández-Berrocal, & Gross, 2013; Gross & John, 2003), whereas high *expressive suppression* scores have usually been associated with maladaptive outcomes in these same domains (for a review, see John & Eng, 2014). Such patterns have emerged across university student (e.g., Balzarotti et al., 2010; Gross & John, 2003), nonclinical community (e.g., Preece, Becerra, & Campitelli, 2018; Cabello et al., 2013) and clinical samples (e.g., Badcock et al., 2011; D’Avanzato, Joormann, Siemer, & Gotlib, 2013; Joormann & Gotlib, 2010; Svaldi, Griepenstroh, Tuschen-Caffier, & Ehring,
2012), though the extent to which these two strategies are “adaptive” or “maladaptive” does appear to vary based on some contextual factors, such as people’s cultural background or values. In samples with more Asian values, for example, expressive suppression appears to be less detrimental (Butler, Lee, & Gross, 2007; Soto, Perez, Kim, Lee, & Minnick, 2011; Su et al., 2015), and there is some evidence that cognitive reappraisal might be an ineffective strategy for some minority groups experiencing oppression (Perez & Soto, 2011; Soto et al., 2012). These results have been found regardless of whether these two strategies are operationalised using the ERQ or experimental paradigms (e.g., Butler et al., 2007; Gross, 2014), thus supporting the ERQ’s validity. Both ERQ scale scores have, moreover, frequently demonstrated acceptable levels of internal consistency reliability (Cronbach’s $\alpha \geq .70$) across a range of sample types and cultures (e.g., Balzarotti et al., 2010; Cabello et al., 2013; English & John, 2013; Gross & John, 2003; Spaapen et al., 2014; Wiltink et al., 2011). With respect to these particular psychometric markers, available evidence therefore suggests the ERQ performs well regardless of participants’ student status.

**Factor Structure of the ERQ**

The factorial validity of the ERQ has been less well studied. We identified twelve published studies (see Table 8.1) that have examined its factor structure, eight in university student samples and four in general community samples. The studies using university student samples have so far been conducted in US (Gross & John, 2003; Melka, Lancaster, Bryant, & Rodriguez, 2011; Moore, Zoellner, & Mollenholt, 2008), French (D’Argembeau & Van der Linden, 2006), German (Abler & Kessler, 2009; Sala et al., 2012), or Italian samples (Balzarotti et al., 2010; Sala et al., 2012), or have used combined data from 23 countries (Matsumoto et al., 2008). All these studies in university student samples have, to date, been broadly supportive of its intended 2-factor structure (i.e., items loading on separable “cognitive reappraisal” or “expressive suppression” factors), with exploratory factor analyses
(EFAs) or principal components analyses (PCAs) extracting these two factors (Abler & Kessler, 2009; Gross & John, 2003; D’Argembeau & Van der Linden, 2006; Matsumoto et al., 2008) and confirmatory factor analyses (CFAs; Balzarotti et al., 2010; Matsumoto et al., 2008; Melka et al., 2011; Moore et al., 2008; Sala et al., 2012) showing generally acceptable goodness-of-fit index values (based on commonly used fit index cut-off criteria like CFI [comparative fit index] ≥ .90 and RMSEA [root mean square error of approximation] ≤ .08; Bentler & Bonnet, 1980; Browne & Cudeck, 1992). The four available factor analytic studies in general community samples have, however, produced more mixed results in terms of fit index values (Cabello et al., 2013; John & Gross, 2004; Spaapen et al., 2014; Wiltink et al., 2011).

The first published factor analysis of the ERQ in a US community sample was an EFA involving older females (John & Gross, 2004), and the expected two factor solution was obtained. A later CFA in a German community sample using a German translation (Wiltink et al., 2011), however, found inadequate levels of fit for the traditional 2-factor model, because item 8 from the cognitive reappraisal scale cross-loaded (factor loadings > .40) on both factors. Though if this cross-loading was allowed in the model, adequate fit index values were achieved. Cabello et al. (2013) subsequently examined a Spanish translation (in a Spanish sample) using CFA and found adequate levels of model fit with no cross-loadings, thus suggesting that the cross-loading Wiltink et al. (2011) observed was likely unique to that German translation, rather than being present in all nonstudent samples. Indeed, the latest CFA study (Spaapen et al., 2014), which used the English ERQ in Australian and UK community samples, also reported no cross-loadings. Spaapen et al. (2014) did, however, find some fit index values below desired cut-offs. They reported that modification indices suggested that the error terms of item 1 (‘‘When I want to feel more positive emotion [such as joy or amusement], I change what I’m thinking about’’) and item 3 (‘‘When I want to feel less
negative emotion [such as sadness or anger], I change what I’m thinking about”) covaried in their data-sets. Spaapen et al. (2014) subsequently retested the 2-factor model with item 3 removed (i.e., to remove the error term covariance), which produced excellent fit index values. Based on their findings, Spaapen et al. (2014, p. 1, 7) argued that the “original ERQ factor structure was not supported”, that the ERQ is “not ideal in its original form”, and that an alternate 9-item version of the measure (with item 3 removed), which they call the ERQ-9, is “preferable [over the original ERQ] when measuring reappraisal and suppression in community samples”.

---

20 In a factor model, an item error term contains the variance in the item score that is not accounted for by the specified substantive latent factors. A covariance between two item error terms therefore suggests that there is something similar about these items other than them both being markers of their specified latent factor (Gerbing & Anderson, 1984).
Table 8.1

A List of the Published Factor Analytic Studies of the Emotion Regulation Questionnaire.

<table>
<thead>
<tr>
<th>Study</th>
<th>Language version(s)</th>
<th>Analysis</th>
<th>Sample type(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matsumoto et al. (2008)</td>
<td>Various</td>
<td>CFA/EFA</td>
<td>University students from various countries (N=3018)</td>
</tr>
<tr>
<td>Moore et al. (2008)</td>
<td>English</td>
<td>CFA</td>
<td>US university students and trauma-exposed community (N=359)</td>
</tr>
<tr>
<td>Abler &amp; Kessler (2009)</td>
<td>German</td>
<td>EFA</td>
<td>German university students (Ns=113,167,174)</td>
</tr>
<tr>
<td>Balzarotti et al. (2010)</td>
<td>Italian</td>
<td>CFA</td>
<td>Italian university students (N=416)</td>
</tr>
<tr>
<td>Melka et al. (2011)</td>
<td>English</td>
<td>CFA</td>
<td>US university students (N=1188)</td>
</tr>
<tr>
<td>Wiltink et al. (2011)</td>
<td>German</td>
<td>CFA</td>
<td>German general community (N=2524)</td>
</tr>
<tr>
<td>Sala et al. (2012)</td>
<td>Italian or German</td>
<td>CFA</td>
<td>Italian or German university students (Ns= 127, 174)</td>
</tr>
<tr>
<td>Cabello et al. (2013)</td>
<td>Spanish</td>
<td>CFA</td>
<td>Spanish general community (N=866)</td>
</tr>
<tr>
<td>Spaapen et al. (2014)</td>
<td>English</td>
<td>CFA</td>
<td>Australian or UK general community (Ns=550, 483)</td>
</tr>
</tbody>
</table>

Note. PCA = principal components analysis, EFA = exploratory factor analysis, CFA = confirmatory factor analysis. Moore et al.’s (2008) sample was primarily comprised of university students (81.3%) so we categorise it as a university student sample when discussing Moore et al.’s results in text. In terms of the CFA estimation methods used, maximum likelihood (ML) was the most common method, either as normal theory ML (Melka et al., 2011; Wiltink et al., 2011) or Satorra and Bentler’s robust ML (Cabello et al., 2013; Balzarotti et al., 2010). Other CFA studies used asymptotic distribution free estimation (Spaapen et al., 2014), weighted least squares mean and variance adjusted estimation (Sala et al., 2012), or did not report the estimation method (Matsumoto et al., 2008).

As such, there is presently some contention about the suitability of using the ERQ outside of student samples. We, however, do not think that the existing body of evidence presently warrants the removal of item 3. We believe that further research is required, particularly since in three of the four factor analytic studies conducted in community samples, the removal of item 3 was not required to achieve good fit index values or an adequate 2-factor solution (Cabello et al. 2013; John & Gross, 2004; Wiltink et al., 2011) and, in the samples of Spaapen et al. (2014), removing item 3 reduced the internal consistency reliability of the cognitive reappraisal scale. Moreover, researchers generally agree that fit index cut-offs should not be used in isolation as “golden rules” for accepting or rejecting a model (e.g., Bagby et al., 2007; Marsh et al., 2004; Preece et al., 2017). This view extends to those
authors who initially introduced these fit index cut-offs (e.g., Bentler & Bonnet, 1980; Hu & Bentler, 1998). Hu and Bentler (1998, p. 449) note, for example, that “it is difficult to designate a specific cutoff value for each fit index because it does not work equally well with various types of fit indices, sample sizes, estimators, or distributions”, and Bagby et al. (2007, p. 260) similarly highlight that “absolute cut-off values for establishing fit are only rough markers for establishing the fit of a model. What is important in model testing is that different procedures to assess fit converge on more or less good fits across multiple samples”. Thus, whilst more factor analytic research in community samples is needed, we think that the ERQ’s traditional 2-factor model has, on balance, so far fulfilled this criterion.

Factor structure is also only one of many markers of validity that must be considered when evaluating the ERQ’s utility, and from a content validity perspective, item 3 has an important role within the measure. This is because item 3 is the only ERQ item that defines the term negative emotion for respondents. Consequently, if item 3 is removed, this definition needs to be moved to another item, and it is presently unclear (i.e., untested) what psychometric impact this might have. Items 1 and 3 are also not directly interchangeable, because item 1 refers to up-regulating positive emotions, whereas item 3 refers to down-regulating negative emotions. People regulate both negative and positive emotions (e.g., Becerra et al., 2017; Preece, Becerra, Robinson, Dandy, & Allan, 2018b; Quoidbach, Berry, Hansenne, & Mikolajczak, 2010), so if item 3 is removed, this reduces the breadth of the construct assessed by the cognitive reappraisal scale. In sum then, if psychometric investigations can demonstrate that the 10-item ERQ functions adequately in nonstudent samples, we think it is preferable for item 3 to be retained.

The Present Study

The goal of the present study was to examine the psychometric properties of the 10-item ERQ. We examined its factor structure, concurrent validity, and internal consistency
reliability in three Australian general community samples ($N = 300, 400, 348$). The three samples were examined separately, rather than being combined, so that we could more fully examine the replicability of the ERQ’s psychometric performance. Concurrent validity was examined against some established measures of psychological distress (Depression Anxiety Stress Scales-21 [DASS-21]; Lovibond & Lovibond, 1995) or alexithymia (20-item Toronto Alexithymia Scale [TAS-20]; Bagby et al., 1994). Because depressive and anxiety disorders are characterised by emotion regulation problems (Rottenberg et al., 2005), we expected that, similar to other samples from countries with predominantly Western cultural backgrounds (e.g., Gross & John, 2003), people with a more maladaptive emotion regulation profile on the ERQ (i.e., low use of cognitive reappraisal, high use of expressive suppression) would report higher levels of psychological distress on the DASS-21. Similarly, because people with high levels of alexithymia have difficulty processing their emotions, alexithymia is a crucial rate-limiting factor for successful emotion regulation (Gross, 2014; Preece et al., 2017), so we expected that people with a more maladaptive emotion regulation profile on the ERQ would report higher levels of alexithymia on the TAS-20.

**Method**

**Participants and procedure.** All three samples (sample A, $N = 300$; sample B, $N = 400$; sample C, $N = 348$)\(^{21}\) completed the ERQ as part of a battery of psychological questionnaires administered in an online anonymous survey. All participants were recruited by an online survey recruiting company (Qualtrics panels) who attempted to recruit samples

---

\(^{21}\) In all three samples, some additional participants also completed the online survey. However, their data were excluded during quality screening because they failed an attention check question (which asked them to select a specific point on the Likert scale) or completed the survey impossibly quickly (at a rate of <2 seconds per question, suggesting inattentive responding). Across samples A, B, and C, data from 65 participants were excluded.
of adults representative of the general community in Australia.\textsuperscript{22} Demographic information for each sample is presented in Table 8.2.

\begin{table}[h]
\centering
\caption{Sample Demographics}
\begin{tabular}{lccc}
\hline
\textbf{Variable} & \textbf{Sample A (N = 300)} & \textbf{Sample B (N = 400)} & \textbf{Sample C (N = 348)} \\
\hline
\textbf{Age (years)} & \textit{M} = 51.59, \textit{SD} = 15.54, range = 18-82 & \textit{M} = 49.27, \textit{SD} = 16.55, range = 18-88 & \textit{M} = 45.62, \textit{SD} = 17.92, range = 18-83 \\
\hline
\textbf{Gender} & & & \\
\textbf{Male} & 139 (46.3) & 156 (39.0) & 124 (35.6) \\
\textbf{Female} & 161 (53.7) & 244 (61.0) & 224 (64.4) \\
\hline
\textbf{Highest level of education completed} & & & \\
\textbf{Primary school} & 1 (0.3) & 5 (1.3) & 6 (1.7) \\
\textbf{Year 10 high school} & 54 (18.0) & 62 (15.5) & 54 (15.5) \\
\textbf{Year 12 high school} & 52 (17.3) & 72 (18.0) & 89 (25.6) \\
\textbf{Technical diploma} & 127 (42.3) & 153 (38.3) & 106 (30.5) \\
\textbf{University Bachelor’s degree} & 49 (16.3) & 82 (20.5) & 73 (21) \\
\textbf{University postgraduate degree} & 17 (5.7) & 26 (6.5) & 20 (5.7) \\
\hline
\textbf{Current university student} & & & \\
\textbf{Yes} & 16 (5.3) & 30 (7.5) & 40 (11.5) \\
\textbf{No} & 284 (94.7) & 370 (92.5) & 308 (88.5) \\
\hline
\textbf{Place of birth} & & & \\
\textbf{Australia} & - & 306 (76.5) & 258 (74.1) \\
\textbf{United Kingdom} & - & 35 (8.8) & 31 (8.9) \\
\textbf{Asia} & - & 22 (5.5) & 22 (6.3) \\
\textbf{New Zealand} & - & 15 (3.8) & 9 (2.6) \\
\textbf{Continental Europe} & - & 13 (3.3) & 17 (4.9) \\
\textbf{Other} & - & 9 (2.3) & 11 (3.2) \\
\hline
\textbf{Primary language spoken at home} & & & \\
\textbf{English} & 280 (93.3) & - & - \\
\textbf{Other} & 20 (6.7) & - & - \\
\hline
\end{tabular}
\end{table}

\textit{Note.} A small portion of people within each sample were university students; these students were retained within the samples because the proportion of students was quite small and, in a cross-section of a community, a portion of people are likely to be studying at university (Australian Bureau of Statistics, 2017a). If these university students were excluded from the samples, the pattern of results reported in the results section did not change.

\textsuperscript{22} Qualtrics panels recruit from multiple sources, primarily actively managed market research panels (see Qualtrics, 2014). Participants are sent an email inviting them to complete the online survey.
**Materials.** In sample A, the questionnaire battery included the ERQ, the DASS-21, and the TAS-20. In samples B and C, the questionnaire battery included the ERQ and DASS-21.

**Depression Anxiety Stress Scales-21.** The DASS-21 (Lovibond & Lovibond, 1995) is a 21-item self-report measure of psychopathology symptoms experienced in the past week. Three scale scores can be derived, corresponding to depression (7 items), anxiety (7 items) and stress (7 items) symptoms. All 21 items can also be summed into a total scale score as an overall marker of psychological distress. Items are answered on a 4-point Likert scale, with higher scores indicating more severe symptomatology. DASS-21 scores have demonstrated good validity and reliability (e.g., Kia-Keating et al., 2017).

**Toronto Alexithymia Scale-20.** The TAS-20 (Bagby et al., 1994) is a 20-item self-report measure of alexithymia. It is designed to assess the three components of alexithymia: difficulty identifying one’s own feelings (DIF; 7 items), difficulty describing feelings (DDF; 5 items), and an externally orientated thinking style (EOT; 8 items) whereby one rarely pays attention to their emotions. Subscale scores can be derived for each component, and all items can be summed into a total scale score as an overall marker of alexithymia. Items are answered on a 5-point Likert scale, with higher scores indicating higher levels of alexithymia (i.e., more emotion processing difficulties). Most aspects of the TAS-20 have demonstrated good validity and reliability, though the EOT subscale score usually has low reliability (Cronbach’s α < .70; e.g., Preece, Becerra, Robinson, & Dandy, 2018) and this was found in our data-set.

**Analytic strategy.** Analyses were conducted in each sample separately. LISREL 8.80 software was used to conduct the CFAs, and SPSS 24 was used for all other analyses. Scores for the ERQ items were reasonably normally distributed (sample A, max skew = -.58, max kurtosis = -.87; sample B, max skew = -.70, max kurtosis = -.92; sample C, max skew = -.78,
max kurtosis = -.99).

**Factor structure.** CFAs were conducted using maximum likelihood estimation (ML) based on a Pearson covariance matrix. As recommended by Curran et al. (1996), we calculated and reported results for both *normal theory ML* and Satorra and Bentler’s (1994) *robust ML* (RML). RML uses a scaled $\chi^2$ statistic ($SB\chi^2$) and performs better than ML when the data are not perfectly normally distributed (Curran et al., 1996; Han et al., 2009). Two models were tested for each sample (see Figure 8.1): (1) The traditional “2-factor model”, where items 1, 3, 5, 7, 8 and 10 were specified to load on a “cognitive reappraisal” factor, and items 2, 4, 6 and 9 were specified to load on an “expressive suppression” factor. These two factors were allowed to correlate. No item cross-loadings or error term covariances were specified. (2) A variant of the 2-factor model (“2-factor model+cov”), informed by the findings of Spaapen et al. (2014), whereby the error terms of item 1 and item 3 were allowed to covary.
Model goodness-of-fit was evaluated via the $\chi^2$ statistic (or SB$\chi^2$) and four fit indexes: CFI, NFI (normed fit index), RMSEA, and SRMR (standardised root mean square residual). A non-significant ($p>.05$) $\chi^2$ statistic suggests excellent fit, however in large sample sizes $\chi^2$ can over-reject models that are parsimonious and a reasonable approximation of the data (see Browne & Cudeck, 1992; Schumacker & Lomax, 2004). We therefore relied primarily on CFI, NFI, RMSEA and SRMR. CFI and NFI values around $\geq .90$ were judged to indicate acceptable fit and values around $\geq .95$ excellent fit. RMSEA and SRMR values around $\leq .08$ were judged to indicate acceptable fit and values around $\leq .06$ excellent fit (Bentler & Bonnet, 1980; Browne & Cudeck, 1992; Kline, 2005; Marsh et al., 2004). To directly compare the fit of the models the Akaike information criterion (AIC) was also used. AIC
penalises model complexity and lower AIC values indicate better fit (Byrne, 2013). Factor loadings ≥.40 were considered meaningful loadings (Stevens, 1992).

Internal consistency reliability. Cronbach’s alpha internal reliability coefficients were calculated for the ERQ cognitive reappraisal and expressive suppression scale scores. Reliability coefficients ≥.70 were judged to indicate acceptable reliability, ≥.80 good reliability, and ≥.90 excellent reliability (Groth-Marnat, 2009).

Concurrent validity. In sample A, Pearson correlations were calculated between ERQ scores and DASS-21/TAS-20 scores. In the other two samples, Pearson correlations were calculated between ERQ scores and DASS-21 scores.
Results

Descriptive statistics for the administered measures are displayed in Table 8.3.

Table 8.3

Descriptive Statistics and Cronbach’s Alpha (α) Reliability Coefficients for the Administered Measures.

<table>
<thead>
<tr>
<th>Measure/subscale</th>
<th>Sample A (N=300)</th>
<th></th>
<th></th>
<th>Sample B (N=400)</th>
<th></th>
<th></th>
<th>Sample C (N=348)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>α</td>
<td>M</td>
<td>SD</td>
<td>α</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>ERQ</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cognitive reappraisal</td>
<td>29.00</td>
<td>6.68</td>
<td>.89</td>
<td>28.97</td>
<td>7.09</td>
<td>.89</td>
<td>28.61</td>
<td>7.32</td>
</tr>
<tr>
<td>Expressive suppression</td>
<td>15.97</td>
<td>5.16</td>
<td>.79</td>
<td>15.69</td>
<td>5.41</td>
<td>.80</td>
<td>15.88</td>
<td>5.14</td>
</tr>
<tr>
<td>DASS-21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total scale</td>
<td>13.44</td>
<td>13.70</td>
<td>.96</td>
<td>16.13</td>
<td>14.97</td>
<td>.96</td>
<td>16.79</td>
<td>14.76</td>
</tr>
<tr>
<td>Depression</td>
<td>4.73</td>
<td>5.52</td>
<td>.94</td>
<td>5.63</td>
<td>5.83</td>
<td>.94</td>
<td>6.14</td>
<td>5.96</td>
</tr>
<tr>
<td>Anxiety</td>
<td>3.63</td>
<td>4.31</td>
<td>.87</td>
<td>4.43</td>
<td>4.94</td>
<td>.90</td>
<td>4.36</td>
<td>4.74</td>
</tr>
<tr>
<td>Stress</td>
<td>5.09</td>
<td>4.94</td>
<td>.94</td>
<td>6.07</td>
<td>5.41</td>
<td>.92</td>
<td>6.28</td>
<td>5.36</td>
</tr>
<tr>
<td>TAS-20</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total scale</td>
<td>48.96</td>
<td>12.24</td>
<td>.87</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DIF</td>
<td>15.10</td>
<td>6.13</td>
<td>.88</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DDF</td>
<td>13.16</td>
<td>4.75</td>
<td>.83</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>EOT</td>
<td>20.70</td>
<td>4.25</td>
<td>.54</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note. ERQ = Emotion Regulation Questionnaire, DASS-21 = Depression Anxiety Stress Scales-21, TAS-20 = 20-item Toronto Alexithymia Scale, DIF = Difficulty identifying feelings, DDF = Difficulty describing feelings, EOT = Externally orientated thinking. The TAS-20 was not administered in samples B or C.

Factor structure. CFAs of the ERQ items produced a similar pattern of results in all three samples (for CFA fit index values see Table 8.4, for CFA factor loadings and factor intercorrelations see Table 8.5). The pattern of results was similar across ML and RML analyses, so we summarise only the RML results here. The traditional 2-factor model was an excellent fit according to all examined fit indexes (i.e., SBχ², CFI, NFI, RMSEA, and SRMR), suggesting that this model was a good representation of the data in all three datasets. All items loaded well on their intended factor. Allowing the error terms of item 1 and item 3 to covary (i.e., 2-factor model+cov) slightly improved fit index values in all three
samples. However, because model fit was already excellent without this modification, the more parsimonious traditional 2-factor model was our preferred solution.

Table 8.4

**Goodness-of-fit Index Values from Confirmatory Factor Analyses (RML and ML) of the Emotion Regulation Questionnaire Items**

<table>
<thead>
<tr>
<th>Model</th>
<th>$\chi^2$ or $SB\chi^2$ (df)</th>
<th>$p$</th>
<th>CFI</th>
<th>NFI</th>
<th>RMSEA (90% CI)</th>
<th>SRMR</th>
<th>AIC</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sample A (N = 300)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor model</td>
<td>14.558(34)</td>
<td>.999</td>
<td>1.00</td>
<td>.993</td>
<td>.000(.000-.000)</td>
<td>.0492</td>
<td>56.558</td>
</tr>
<tr>
<td>2-factor model+cov</td>
<td>11.396(33)</td>
<td>1.00</td>
<td>1.00</td>
<td>.994</td>
<td>.000(.000-.000)</td>
<td>.0453</td>
<td>55.396</td>
</tr>
<tr>
<td>ML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor model</td>
<td>89.607(34)</td>
<td>&lt;.001</td>
<td>.972</td>
<td>.956</td>
<td>.0731(.0546-.0920)</td>
<td>.0492</td>
<td>130.275</td>
</tr>
<tr>
<td>2-factor model+cov</td>
<td>66.132(33)</td>
<td>&lt;.001</td>
<td>.983</td>
<td>.967</td>
<td>.0595(.0391-.0796)</td>
<td>.0453</td>
<td>111.902</td>
</tr>
<tr>
<td><strong>Sample B (N = 400)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor model</td>
<td>9.869(34)</td>
<td>1.00</td>
<td>1.00</td>
<td>.996</td>
<td>.000(.000-.000)</td>
<td>.0670</td>
<td>51.869</td>
</tr>
<tr>
<td>2-factor model+cov</td>
<td>5.809(33)</td>
<td>1.00</td>
<td>1.00</td>
<td>.998</td>
<td>.000(.000-.000)</td>
<td>.0587</td>
<td>49.809</td>
</tr>
<tr>
<td>ML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor model</td>
<td>165.878(34)</td>
<td>&lt;.001</td>
<td>.952</td>
<td>.941</td>
<td>.094(.0795-.110)</td>
<td>.0670</td>
<td>196.684</td>
</tr>
<tr>
<td>2-factor model+cov</td>
<td>93.069(33)</td>
<td>&lt;.001</td>
<td>.978</td>
<td>.967</td>
<td>.0653(.0492-.0818)</td>
<td>.0587</td>
<td>133.161</td>
</tr>
<tr>
<td><strong>Sample C (N = 348)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor model</td>
<td>8.499(34)</td>
<td>1.00</td>
<td>1.00</td>
<td>.996</td>
<td>.000(.000-.000)</td>
<td>.0550</td>
<td>50.499</td>
</tr>
<tr>
<td>2-factor model+cov</td>
<td>6.798(33)</td>
<td>1.00</td>
<td>1.00</td>
<td>.997</td>
<td>.000(.000-.000)</td>
<td>.0513</td>
<td>50.798</td>
</tr>
<tr>
<td>ML</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-factor model</td>
<td>107.857(34)</td>
<td>&lt;.001</td>
<td>.969</td>
<td>.955</td>
<td>.0796(.0631-.0966)</td>
<td>.0550</td>
<td>150.692</td>
</tr>
<tr>
<td>2-factor model+cov</td>
<td>80.559(33)</td>
<td>&lt;.001</td>
<td>.980</td>
<td>.966</td>
<td>.0659(.0483-.0838)</td>
<td>.0513</td>
<td>126.668</td>
</tr>
</tbody>
</table>

*Note.* RML = Satorra and Bentler’s (1994) robust maximum likelihood estimation, ML = maximum likelihood estimation, CFI = comparative fit index, NFI = normed fit index, RMSEA = root mean square error of approximation, SRMR = standardised root mean residual, AIC = Akaike information criterion, CI = confidence interval.
Internal consistency reliability and concurrent validity. In all three samples, the ERQ cognitive reappraisal ($\alpha_{range} = .89-.90$) and expressive suppression ($\alpha_{range} = .76-.80$) scale scores had acceptable to excellent levels of internal consistency reliability (see Table 8.3). The ERQ scores also correlated with scores from the DASS-21 or TAS-20 as expected. Expressive suppression scores were significantly ($p < .05$) positively correlated with psychological distress (sample A, $r = .20$; sample B, $r = .18$; sample C, $r = .13$) and alexithymia (sample A, $r = .41$). Cognitive reappraisal scores were significantly ($p < .01$)
negatively correlated with psychological distress (sample A, $r = -0.19$; sample B, $r = -0.30$; sample C, $r = -0.27$) and alexithymia (sample A, $r = -0.19$).

In sum, the ERQ performed similarly across these three data-sets, and performed well on every marker of validity and reliability that we tested.

**Discussion**

Our purpose in this study was to examine the psychometric properties of the 10-item ERQ in general community samples. Overall, it performed well on every marker of validity and reliability that we tested.

The ERQ’s factor structure was replicable and consistent with its theoretical basis in our CFAs. The traditional 2-factor model displayed excellent goodness-of-fit across three different Australian community samples. Our findings across these three data-sets are therefore consistent with the results of most existing factor analytic studies, which have supported the traditional 2-factor model in mostly university student samples (e.g., Balzarotti, Gross, & John, 2010; Gross & John, 2003; Matsumoto et al., 2008). Like Spaapen et al.’s (2014) results in their Australian or UK community samples, there was a covariance between the error terms of ERQ items 1 and 3 in our data-sets, but unlike Spaapen et al.’s results, it did not need to be added into the model for fit index cut-off values to be reached. Even if this error term covariance is added into the model though, many researchers agree that the presence of a few error term covariances in a model, practically speaking, often has little impact on a measure’s utility, and such error term covariances are frequently added into models when these covariances are theoretically justified (e.g., Aldao & Nolen-Hoeksema, 2010; Bagby et al., 2007; Gerbing & Anderson, 1984; Gullone & Taffe, 2011; Matsumoto et al., 2008; Podsakoff et al., 2003; Preece et al., 2017). Error term covariances, for example, may reflect a method effect that is caused by similarities in item wording or sentence structure (Podsakoff et al., 2003). We think this is likely the case here, as ERQ items 1 and 3
are unique from the rest of the items, in that both these items include the phrase “…I change what I’m thinking about” and both include a bracketed section that defines the terms negative or positive emotion. Whilst the error terms of these two ERQ items are therefore likely to covary, this does not exclude these items from contributing unique and useful variance to the “cognitive reappraisal” factor score (Podsakoff et al., 2003); indeed, in our data-sets, and all other published factor analyses (e.g., John & Gross, 2004; Spaapen et al., 2014), ERQ items 1 and 3 both load strongly (factor loadings > .40) on the “cognitive reappraisal” factor. Available community data therefore appear to support the factorial validity of the ERQ.

The ERQ cognitive reappraisal and expressive suppression scale scores also correlated in expected ways with scores from established measures of psychological distress (DASS-21; Lovibond & Lovibond, 1995) or alexithymia (TAS-20; Bagby et al., 1994) in our data-sets, and had acceptable to excellent levels of internal consistency reliability. Taken alongside the large body of similar findings from mostly university student samples (e.g., Gross & John, 2003; Matsumoto et al., 2008), our findings therefore suggest that the ERQ may be used confidently regardless of participants’ student status. Because ERQ item 3 is important from a content validity perspective, we think on balance it is hence preferable for researchers and practitioners to continue using the ERQ in its traditional 10-item form (Gross & John, 2003), rather than Spaapen et al.’s (2014) modified and less well tested 9-item version.

One limitation of our study, however, is that our results only apply to the English version of the ERQ, and to samples from a country with a predominantly Western cultural background (Australian Bureau of Statistics, 2017b).23 In university student samples, the ERQ’s factor structure has generally performed similarly across various cultural groups and

---

23 We did not collect data on the ethnicity or cultural values of our participants, however, our collected birthplace data was similar to recent Australian census data (Australian Bureau of Statistics, 2017b), whereby most participants reported being born in Australia or the UK.
language versions (see Matsumoto et al., 2008), but further research is needed to examine whether this is also the case for community samples. More factor analytic work in Asian community samples would be useful, for example, given that the consequences of expressive suppression appear to be different in this group (Soto et al., 2011; Su et al., 2015). Our results also only apply to nonclinical samples. Most forms of validity and reliability have already been established in clinical samples (e.g., Badcock et al., 2011; Joormann & Gotlib, 2010), but the ERQ’s factor structure has not yet been examined. Factorial invariance testing across clinical, community, and university student samples would hence be beneficial.

Whilst further work is therefore needed to confirm our findings across different population types, available evidence presently suggests that the ERQ has strong validity and reliability regardless of participants’ student status.
References


doi:10.1207/s15328007sem1203_7


psychopathology: A transdiagnostic approach to etiology and treatment (pp. 13–37).
New York, NY: Guilford.


