Emotion expressivity in the Dutch

Kupper, Nina; Duijndam, Stefanie; Karreman, Annemiek

Published in:
Psychological Assessment

DOI:
10.1037/pas0000812

Publication date:
2020

Document Version
Peer reviewed version

Emotion expressivity in the Dutch: validation of the Dutch translation of the Berkeley Expressivity Questionnaire

Nina Kupper*, Stefanie Duijndam, and Annemiek Karreman

Center of Research on Psychological and Somatic diseases, Tilburg University, Tilburg, the Netherlands

* Corresponding author

Author note:

Nina Kupper, Center of Research on Psychological and Somatic diseases, Department of Medical & Clinical Psychology, Tilburg University, Tilburg, the Netherlands; Stefanie Duijndam, Center of Research on Psychological and Somatic diseases, Department of Medical & Clinical Psychology, Tilburg University, Tilburg, the Netherlands; Annemiek Karreman, Center of Research on Psychological and Somatic diseases, Department of Medical & Clinical Psychology, Tilburg University, Tilburg, the Netherlands

The datasets used for the current report are also used for other studies from the INHIBIT and PHEMORE research projects.

Funding sources: none, data repository: https://dataverse.nl/dataverse/depmedclin

Acknowledgements: We gratefully acknowledge the statistical advice and help from Paul Lodder (department of Methodology and Statistics, Tilburg University).

Contact information: Warandelaan 2, 5000LE Tilburg, the Netherlands, h.m.kupper@tilburguniversity.edu

ABSTRACT

When examining emotions and emotion regulation, we discriminate between emotion experience and emotion expressivity. Research shows that the two are modestly related. The Berkeley Expressivity Questionnaire (BEQ) was designed to assess positive and negative expressivity, as well as the intensity of the expressive behavior. The current paper reports on two studies that examined the reliability and validity of the Dutch translation of the BEQ. In study 1, we performed a confirmatory factor analysis of the BEQ items in two samples that differed in age (young adults and adults), correlated the facet and total scores with measures of behavioral control, affective response tendencies, and emotion regulation strategies, and examined sex differences. Results confirmed the three-factor structure and further showed that factors were correlated, and two items loaded on all factors. Internal consistency was good, and test-retest reliability was excellent. As expected, emotion expression was larger in women. Convergent and divergent validity were confirmed. Behavioral control measures were inversely related to emotion expression. While neuroticism and depression were associated with negative expressivity, extraversion, openness and agreeableness were associated with positive expressivity. Emotion regulation strategies generally showed association patterns in the expected directions. In study 2, we report on the predictive value of the BEQ facet scores for the emotional response to acute social evaluative stress. Results showed that only expression intensity was significantly associated with a larger emotional stress response. In conclusion, the Dutch version of Berkeley Expressivity Questionnaire is a reliable and valid instrument to be used in the Dutch setting.

Keywords: emotion expressivity; validation; Berkeley Expressivity Questionnaire; Dutch

Public significance: This study suggests that the Berkeley Expressivity Questionnaire is a valid instrument to assess emotion expressivity in the Netherlands. Results indicate that there are individual differences in emotion expressivity related to personality, mood, and emotion regulation tendencies.

Emotions play an important role in our adaptive response to internal and external environmental challenges or events (Frijda, 1988). Emotion expression is part of the behavioral component of emotion and may comprise facial expression, posture, and gesture, as well as vocal features (Russell, Bachorowski, & Fernandez-Dols, 2003). Emotion expression serves as a signal to the self (i.e., motivating behavior), and has a communicative (informative and evocative) function towards the social environment (Lewis, Haviland-Jones, & Barrett, 2008; Vingerhoets, Nyklíček, & Denollet, 2008).

The Berkeley Expressivity Questionnaire

The Berkeley Expressivity Questionnaire (BEQ) assesses emotion expression, i.e., positive and negative expressivity, and expression intensity (Gross & John, 1995). While positive and negative expressivity indicate the extent of expression of positively/negatively valenced emotions, expression intensity represents individual differences in the strength of emotional response tendencies (Gross & John, 1995). Only a few BEQ translations have been validated (Lin, Soi-Kawase, Narita-Ohtaki, Itoh, & Kim, 2016; Mohiyeddini, John, & Gross, 2008; Tunay Akan & Bariskin, 2017). These studies have shown some differences in factor structure (Tunay Akan & Bariskin, 2017), but generally confirmed converging and diverging associations with behavioral control, affective response tendencies, and emotion regulation strategies.

Individual differences – construct validity

Large individual differences exist in the level and manner of emotion expression. These differences relate to a wide range of intra- and interpersonal processes. Emotion expression may be affected at three levels: at behavioral control, at affective response tendencies (personality, mood episodes), and at emotion regulation.

Behavioral control - Research has shown that emotion expression subscales of the BEQ are negatively related to (emotional) self-control (Gross & John, 1997). Moreover, self-monitoring

has been related to increased positive expressivity and less expression intensity (Lin et al., 2016). Self-control is associated with mood, because exercising self-control can induce negative emotions (Tice & Bratslavsky, 2000). When attempts are made to regulate this negative emotional state this may come with the cost of failing in other areas of self-control (e.g., indulgence, relapse) (Tice & Bratslavsky, 2000)

Affective response tendencies - Studies have shown that extraversion, openness and agreeableness relate to larger positive expressivity. Conversely, neuroticism and conscientiousness relate to elevated negative expressivity (Gross & John, 1995; Lin et al., 2016; Mohiyeddini et al., 2008). Expression intensity was stronger for extraversion and openness in the Japanese validation study (Lin et al., 2016), while in other validation papers expression intensity was most strongly associated with neuroticism (Gross & John, 1995; Mohiyeddini et al., 2008), followed by extraversion, and openness (Gross & John, 1995). Previous studies report mixed findings for the association of emotion expression with depression (Lin et al., 2016; Mohiyeddini et al., 2008).

Emotion regulation – The expression of emotion may be reduced by emotion regulation, especially by expressive suppression, and to a lesser extent by reappraisal (Egloff, Schmukle, Burns, & Schwerdtfeger, 2006; Gross, 1998; Gross & Levenson, 1993). Conversely, problematic emotion regulation strategies have also been related to increased expression of emotion (Tunay Akan & Bariskin, 2017).

Criterion validity

Emotional experience may change as a function of the tendency to express emotion. A recent study found higher levels of self-reported emotional expressivity (assessed with the EEQ) was related to reduced cortisol reactivity, but unrelated to negative emotional reactivity (Wang & Lau, 2018).

Subgroups

Women in general show increased emotional expressivity for positive emotions, internalizing negative emotions, and impulse strength (Gross & John, 1995; Gross & John, 1997). Previous validation studies concur with these findings (Lin et al., 2016; Mohiyeddini et al., 2008; Tunay Akan & Bariskin, 2017). Sex is therefore thought to play a diverging role. While experimental research shows no apparent age-related differences in emotional responses in experiments where participants were asked to regulate their emotions, the strategies used to regulate emotions change with age, with older people more often using suppression in comparison to middle aged and young adults, which may affect emotion expressivity (Brummer, Stopa, & Bucks, 2014).

The current study: Research questions & hypotheses

We performed two studies to assess the validity of the Dutch translation of the BEQ scores. The first study aimed to investigate the validity (i.e., internal consistency, test-retest reliability, and construct validity) of the Dutch translation of the BEQ scores in a young adult and general population sample. Construct validity was examined by testing behavioral, affective, and cognitive correlates of emotion expression, while sex and age differences were additionally examined. The second study examined criterion validity by testing whether the BEQ total and facet scores were predictive of the emotional response to the stressful task.

We hypothesized that:

The factor structure of Dutch BEQ would concur with the originally reported factor structure: an hierarchical model with three latent factors and one overarching factor, with two general expressivity items loading on all factors (Gross & John, 1997) (study 1).

Emotion expressivity would be inversely associated with behavioral control variables, and with personality and mood characteristics, such that neuroticism, depressive affect and anxiety are associated with more negative expression, and extraversion and openness are associated with more

positive expression. We expect expression intensity to be associated with higher levels of neuroticism and depressed affect (Gross & John, 1995; Lin et al., 2016; Mohiyeddini et al., 2008) (study 1). Correlations between emotion expression and respectively emotional control, emotional experience and emotion regulation would be larger in women than in men (Gross & John, 1995; Gross & John, 1997; Vigil, 2009). With respect to age, we hypothesized older age to be increasingly characterized by a reduced affect intensity, possibly suggesting less intense emotion expression (study 1). Finally, we would expect higher intensity of expression to be related to increased negative emotional reactivity (study 2).

Methods Study 1

Participants & Procedure***:*** The first sample comprised of 209 young adults (YA; 76.6% female; M_{age}, 20.32 ± 2.08; range 18-33 years) who participated for course credit. The second sample comprised a non-random selection of 586 adults (54.1% female; M_{age}, 46.63±15.94; range 18-81 years) from the general Dutch population (GP). Details on data collection and materials used in both samples can be found in the online Methods supplement.

Materials

Social Demographics - Demographic variables included age, sex, marital status (dichotomized: partner vs. no partner), and educational level (dichotomized at the level of high school or less vs. higher; general population only).

BEQ - With permission from the original authors (Gross & John, 1995), two independent translators translated the BEQ into Dutch (available from: https://spl.stanford.edu/resources). Two independent native English speakers then back-translated this version into English. Any discrepancies were discussed and resolved between translators and back-translators. The BEQ comprises 16 items and 3 subscales: impulse strength (six items), negative expressivity (six items),

and positive expressivity (four items). One general expressivity item (#10) has been placed in the positive subscale, while the second general expressivity item (#16) has been placed in the negative expressivity subscale. Responses are given on a seven-point Likert scale (1 = strongly disagree; 7 = strongly agree). Three items are reversely scored (items 3, 8, and 9), and thus recoded. The mean score of each subscale indicates each facet's salience, and general expressivity (BEQ full scale) results from averaging subscale scores. The Dutch questionnaire and the scoring syntax are available in Supplementary Table 1.

Validation questionnaires - For construct validation, we used Dutch versions of the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003)), the Cognitive Emotion Regulation Questionnaire Short-Form (CERQ-SF; Garnefski & Kraaij, 2006), the Brief Self-Control Scale (BSCS; Helmerhorst, De Vries Robbé, & De Vogel, 2011; Tangney, Baumeister, & Boone, 2004) the revised Self-Monitoring Scale (SMS-R; Jansen, Giebels, van Rompay, & Junger, 2018; Snyder & Gangestad, 1986), the Big Five Inventory (BFI; Denissen, Geenen, van Aken, Gosling, & Potter, 2008), the Patient Health Questionnaire (PHQ-9; Kroenke, Spitzer, & Williams, 2001), and the Generalized Anxiety Disorder scale (GAD-7; Spitzer, Kroenke, Williams, & Löwe, 2006). For details see the online Methods supplement.

Statistical approach

Statistical power considerations – For factor analysis, larger sample sizes will provide more precise and stable factor loadings (MacCallum, Widaman, Preacher, & Hong, 2001). No general sample to power ratio exists, as the size of communalities and over-determination have a strong influence on the required sample size. As our data were already collected for other purposes, we considered the dataset sizes as a given, with their limitations, and continued with factorability (see Methods Supplement). For the construct validity, we calculated correlations. Power analysis (α =

.05, β = .20, and r = .20) indicated that a sample of 200 would be needed to determine whether the expected correlations are different from 0.

Internal consistency - Cronbach's alphas and McDonald's $Omega_{total}$ values (R: psych package (R Core Team, 2016)) were calculated for the T0 and T1 assessments of the total and subscale scores of the BEQ and all validation questionnaires in in both samples. For $Omega_{total}$, we used polychoric correlations and the OLS analysis option (non-normality of the item data).

Test-retest reliability - Intra-class correlations were calculated to assess the agreement in BEQ total and subscale scores over a 6-month period in both the YA sample and the GP sample.

Construct validity – We conducted confirmatory factor analysis with the R Lavaan package (Rosseel, 2012). We used the WLSMV estimation method, because of the non-normality and ordered characteristics of the item data (Brown, 2006). Multiple fit indices were calculated, including χ^2/df ratio, TLI, and RMSEA. Latent factors were defined to have a mean of 0 and a variance of 1 (i.e., standardized), allowing free estimation of all factor loadings. Residuals were uncorrelated, and missing values were handled ('pairwise' option). We started with a 1-factor model that included all items, and then tested a correlated 2-factor model (i.e., including a general expressivity factor and impulse strength factor). Third, we fitted a correlated three-factor model, including BEQ's three facets, and then the original hierarchical model. We finally tested a correlated 3-factor model with the two general expressivity items loading on all factors. $\Delta\chi^2$ tests were used to compare models. Modification indices were requested for the best fitting model, for a post-hoc exploration of how model fit could be improved if necessary.

Cross-sectional (T0) correlations were calculated with affective response tendencies (personality, depression, anxiety), behavioral control (self-control, self-monitoring), and emotion regulation strategies to assess convergent and divergent validity. We examined the presence of sex

differences in the levels of emotion expression and intensity of expression, as well as in the magnitude of correlation with convergent and divergent constructs. Correlations with age were calculated.

Results

Descriptive statistics

Table S1 displays the demographic characteristics of both samples at inclusion (T0). The three facet means and the BEQ total mean item score (Table 1) were comparable between samples and measurement occasions. Correlation of the facets with the total score ranged between .81-.86 for the young adults (YA), and between .79-.84 for the general population (GP) sample. Inter-facet correlations were a bit lower, ranging between .47-.62 for YA, and between .27-.50 for the GP. Test-retest correlations across the 6-month follow-up period were excellent (YA: .84-.90; GP: .86-1.00). Cronbach's alphas showed good internal consistency (.86-.88 for the scale scores) and the McDonald's Omega total value was higher, ranging between .86 and .90, indicating reliability of the scale scores.

Confirmatory factor analysis

The 16 items of the BEQ were all positively correlated, with a mean inter-item correlation of .32 in both samples. In both samples, the model with a correlated three-factor model and overall loading for items 10 and 16 was the best fitting model (See Table S2, Figure S3), compared to all other models. The hierarchical model was only assessed in the general population sample and showed a worse fit. Factor covariances were .65 (YA) and .52 (GP; positive – negative), .72 (YA) and .56 (GP; strength – negative) and .43 (YA&GP; strength - positive). Table S3 shows the factor loadings for the best fitting model for both samples.

Post-hoc exploration - The modification indices from the hypothesized model suggest that both three-factor models might be improved by adding cross-loadings between several items and other

factors. A table with the five highest modification indices for both models is available in the online Supplement – Results Table S3. Exploratory results (Results supplement) showed that adding item 12 as an additional general expressivity item loading on all factors significantly improved model fit in YA and GP. Also, linking expression intensity with emotions that are inherently more intense (fear, anger) significantly improved the GP model.

Convergent and divergent validity

Behavioral correlates - We examined the correlations between BEQ and self-monitoring and self-control (Table 2, top panel). Self-control was unrelated to all facets of emotion expressivity in the young adult sample, while self-control was significantly related to valenced emotion expressivity and expression intensity in the general population, such that more self-control was associated with less expressivity and lower expression intensity.

Affective response tendencies - In young adults (second panel, Table 2), depression was unrelated to positive expressivity and expression intensity, but was related to increased negative expressivity. Importantly, this pattern of correlations was not observed in our general population sample, where depression was positively associated with expression intensity, but not with valenced expressivity. In both samples, anxiety was associated with increased intensity, while being unrelated to positive or negative expressivity. Neuroticism was associated with increased negative expressivity and especially increased intensity of expression in both samples. Extraversion was associated with increased negative and positive expressivity in both samples. Intensity of expression was positively related to extraversion (young adults). Openness to experience was associated with increased positive expressivity. Conscientiousness only showed significant relations in the young adult sample, being related to increased negative expressivity and increased expression intensity. Agreeableness was more strongly related to both positive and negative expressivity in the YA sample, and equally related across all subscales the GP sample,

with positive expressivity being more strongly related in the YA sample ($r = 30$ vs. .19*)*. Moreover, in the adult sample, expression intensity was also positively associated with agreeableness.

Emotion regulation - Suppression was strongly associated with reduced expression of positive and negative emotion, as well as expression intensity (bottom panel of Table 2.). Reappraisal showed a small significant correlation with increased positive expressivity. With respect to other cognitive regulatory styles, self-blame, blaming others, positive refocus, and perspective taking were largely unrelated to emotion expressivity. Acceptance was negatively associated with negative expressivity and expression intensity.

Sex and age differences

Student *t*-tests showed significant sex differences in the mean scores of all facets (both samples; Results Supplement - Figure S1). There were also sex differences for the association of expressivity with emotion regulatory styles (e.g., positive reappraisal, refocused planning), personality traits (e.g., neuroticism), and affect (Supplemental Results: Table S5). Age was unrelated to the BEQ total score (r=-.05, p=.24), expression intensity (r=.01, p=.73) and negative expressivity (r=-.04, p=.34), but was negatively associated with positive expressivity (r=-.11, p=.007), suggesting reduced expressivity with increasing age.

Conclusion

Study 1 examined construct validity of the Dutch translation of the BEQ. We found that validity was adequate, as we found a correlated three factors model with two general expressivity items that loaded on all factors, in two independent samples that differed in age and background. Though this is different from the original hierarchical model, the fit of the hierarchical model was comparable to the correlated three-factor model without cross-loadings but worse than our final model, in which the two general expressivity items loaded on all factors (based on Gross and John (1997)). Exploratory analyses based on the modification indices showed the plausibility of a third

general expressivity item (item 12). Caution is needed in interpreting and using modification indices, as they often do not replicate between samples (MacCallum, Roznowski, & Necowitz, 1992). Convergent and divergent construct validity was established by showing negative correlations with behavioral control, negative affective response tendencies, and suppressive emotion regulation. Positive correlations were found for positive affective response tendencies and reappraisal, and adaptive emotion regulation styles. Sex differences are omnipresent for the BEQ, and for the correlations with convergent and divergent constructs. Positive expressivity reduced with age.

Methods Study 2

Participants & procedure study 2. In total, 123 undergraduate students (77% women, age = 20.1± 2.7) participated in the study in exchange for course credits. (See Methods supplement).

Materials – The BEQ (see study 1 for details) was administered before the stress protocol started. After rest, and stress, participants rated items reflecting their level of affective arousal on a 7-point Likert scale. (See Methods supplement).

Statistical approach – Emotional responses were ln transformed before analysis. Pearson correlations assessed the correlation of the BEQ total and facet scores with baseline emotions. To assess criterion validity, we used repeated measures ANOVAs with two measurement occasions (rest – stress) to test whether the BEQ total score (dependent variable) was predictive of the emotional stress response. Post-hoc analyses assessed which emotion expressivity facet was driving this result and whether there were differences between individual emotions.

Results

Criterion validity - The stress test induced a significant negative emotional response ($F(1, 122)=311.405$; $p < .0001$; partial $\eta^2 = .719$). RM-ANOVA of the total negative emotion score at rest and stress found a trend relation with total emotion expression ($F(1,122)=2.777$; $p = .098$; partial $\eta^2 = .022$). Subscale analysis revealed that only expression intensity was associated with increased negative responding ($F(1, 119)=5.775$; $p = .018$; partial $\eta^2 = .046$). A significant between-subjects effect was found for expression intensity ($F(1, 119)=7.925$; $p = .006$; partial $\eta^2 = .062$) only. This means that individuals with higher expression intensity scores had higher emotional responding offsets and a steeper response slope. The baseline level of negative emotions was unrelated to the BEQ total score and its facets. For the individual emotions similar patterns emerged (Online Results Supplement Table S6, Figure S2).

Conclusion

Study 2 examined criterion validity of the Dutch BEQ, and revealed that only the intensity of expression, not valenced expressivity, was associated with increased negative emotional responses to acute mental stress. Expression intensity was most strongly related to the emotions anger and being annoyed.

General discussion

The current paper addressed the reproducibility of the BEQ's original factor structure and reliability, the assessment of construct validity, and the assessment of criterion validity.

Reliability - As hypothesized, our data fit a correlated three-factor model with two general expressivity items well, which differs from the originally proposed model, but is consistent with Gross and John (1997)'s finding that two items reflect general expressivity. Importantly, because the study by Gross and John (1997) did not test a correlated three-factor model, we do not know

whether their data would fit that model as well. A more recent CFA of the BEQ failed to find an adequate model fit (Dobbs, Sloan, & Karpinski, 2007), which might have to do with not modeling the two general expressivity items.

Exploratory analysis based on the modification indices of the final model showed the plausibility of a third general expressivity item, i.e., item 12 '*Can't sometimes hide my feelings, even if I wanted to*'. The presence of this overall cross-loading is concurrent with previous research (Dobbs et al., 2007), but inconsistent with the hypothesized model (Gross & John, 1997). Looking at the content of the item though, it may relate more to generally expressivity (like item 10: *'I am an emotionally expressive person*'). Future research should investigate this further.

Internal consistency of the Dutch BEQ was good. Like in other studies, the values for the total scale score were larger than for the subscale scores (Dobbs et al., 2007). Test-retest correlations were substantially higher than in previous validation studies (Gross & John, 1995; Lin et al., 2016; Tunay Akan & Bariskin, 2017), and inter-correlations between BEQ facets concurred with prior studies (Gross & John, 1995; Gross & John, 1997).

Construct validity - Construct (convergent, divergent, criterion) validity was confirmed, and our results revealed important differences between the young adult sample and the sample from the general population. The convergent and divergent construct validity of the Dutch translation of the BEQ was good. We examined the associations of the Dutch BEQ scores with variables representing behavioral control, affective response tendencies, and cognitive control, i.e., emotion regulation strategies. Positive expressivity was more strongly related to positive personality dimensions as well as adaptive emotion regulation strategies, while negative expressivity was more associated with their negative counterparts. The same was observed in the original BEQ validation paper (Gross & John, 1997).

Criterion validity – The second study revealed that intensity of expression, but not valenced expressivity, was associated with increased negative emotional responses to acute mental stress. Our results closely follow and extend findings from a recent study in which higher levels of self-reported positive and negative emotional expressivity were unrelated to negative emotional reactivity (Wang & Lau, 2018). Our results extend the current knowledge by showing that especially the intensity of negative emotion expressivity was associated with the size of the experienced emotional stress response.

Sex differences - Sex differences were omnipresent for the BEQ total and facet scores, as well as for the correlations with convergent and divergent constructs. Our findings correspond with previous studies showing enhanced emotional expressivity in women. Our results add to the literature by showing that there are sex differences in the strengths of the correlations between emotional expressivity and behavioral control variables, affective response tendencies, and emotion regulation strategies. Future research may examine measurement equivalence in the factor structure.

Limitations & Strengths - The current paper comes with limitations and strengths. We did not use Monte Carlo simulation to establish an a priori required sample size. Since we used existing databases, we used rules of thumb to decide whether to include the sample in the CFA. Future research could do a simulation study first. Several sample characteristics may explain the difference in results between the young adult and general population sample. The majority (77%) of the young adult sample was female, and the sample was quite homogeneous, as only first-year psychology students participated. The general population sample was quota-sampled with respect to sex and age, and contained 54% women. Moreover, this latter sample was more diverse in terms of socio-economic status, educational/professional background, and age (lifespan range from 18-

85). Based on the current results, one needs to be aware that emotion expression questions may be interpreted differently by young adults, as compared to (older) people from the general population. Another limitation is that we did not examine measurement equivalence in factor structure, means, variances, and co-variances for sex and age cohorts. While we did analyze sex differences in correlations revealing several subtle but important sex differences in emotion expression and its correlates, we could not do the same for age. A larger sample would be needed for that. Similarly, a deeper investigation into the role of socio-economic disparities in these relations is warranted. We also did not include a clinical sample, so representativeness may be limited to the general population. While inclusiveness is of huge importance for the generalizability of study findings, we did not measure ethnicity, which is a limitation. Only a small minority (3-5%) of first year university students has an immigration background, so in the future, additional recruitment efforts should be made to get a sufficiently large subsample. We did include both a young adults sample and a general population sample, which is a strength because of the broader generalizability to the indigenous Dutch population. Other strengths of the paper are the inclusion of criterion validity, which often is omitted from validation papers, and the large sample size of our general population sample.

Conclusion - The Dutch translation of the BEQ rendered data with satisfactory psychometric properties in young adults and the general population, suggesting the usability of the Dutch BEQ in experimental and community settings. This measure will facilitate further research on the interplay between emotion experience and emotional expression. Future research is encouraged to extend validity testing of the BEQ, testing measurement equivalence between gender and age groups, and assess individual differences related to sociodemographic and ethnicity factors. Moreover, studies to somatic and psychological clinical samples will enable individual differences

in emotional expressivity to be be taken into account in designing interventions to improve mental and physical health.

References

Brown. (2006). *Confirmatory factor analysis for applied research.* . New York: Guildford.

Brummer, Stopa, & Bucks. (2014). The influence of age on emotion regulation strategies and psychological distress. *Behavioural and Cognitive Psychotherapy, 42*(6), 668-681. doi:10.1017/S1352465813000453

Denissen, Geenen, van Aken, Gosling, & Potter. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of personality assessment, 90*(2), 152-157. doi:10.1080/00223890701845229

Dobbs, Sloan, & Karpinski. (2007). A psychometric investigation of two self-report measures of emotional expressivity. *Personality and Individual Differences, 43*(4), 693-702. doi:https://doi.org/10.1016/j.paid.2007.01.010

Egloff, Schmukle, Burns, & Schwerdtfeger. (2006). Spontaneous emotion regulation during evaluated speaking tasks: associations with negative affect, anxiety expression, memory, and physiological responding. *Emotion, 6*(3), 356-366. doi:10.1037/1528-3542.6.3.356

Frijda. (1988). The laws of emotion. *American Psychologist, 43*, 349-358.

Garnefski, & Kraaij. (2006). Cognitive emotion regulation questionnaire - development of a short 18-item version (CERQ-short). *Personality and Individual Differences, 41*(6), 1045-1053. doi:10.1016/j.paid.2006.04.010

Gross. (1998). Antecedent- and response-focused emotion regulation: divergent consequences for experience, expression, and physiology. *Journal of Personality and Social Psychology, 74*(1), 224-237. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9457784

Gross, & John. (1995). Facets of emotional expressivity: three self-report factors and their correlates. *Personality and Individual Differences, 19*(4), 555-568. doi:https://doi.org/10.1016/0191-8869(95)00055-B

Gross, & John. (1997). Revealing feelings: facets of emotional expressivity in self-reports, peer ratings, and behavior. *Journal of Personality and Social Psychology, 72*(2), 435-448. Retrieved from https://www.ncbi.nlm.nih.gov/pubmed/9107009

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

Gross, & Levenson. (1993). Emotional suppression: physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology, 64*(6), 970-986.

Helmerhorst, De Vries Robbé, & De Vogel. (2011). *Zelf-controle schaal en zelf-controle schaal observatie*: Van der Hoeven Stichting.

Jansen, Giebels, van Rompay, & Junger. (2018). The Influence of the presentation of camera surveillance on cheating and pro-social behavior. *Frontiers in psychology, 9*, 1937. doi:doi:10.3389/fpsyg.2018.01937

Kroenke, Spitzer, & Williams. (2001). The PHQ-9. *Journal of General Internal Medicine, 16*(9), 606-613. doi:10.1046/j.1525-1497.2001.016009606.x

Lewis, Haviland-Jones, & Barrett. (2008). *Handbook of emotions* (3rd ed. ed.). New York: Guilford Press.

Lin, Soi-Kawase, Narita-Ohtaki, Itoh, & Kim. (2016). Reliability and validity of a self-report emotional expressivity measure: The Japanese version of the Berkeley Expressivity Questionnaire. *Japan Journal of Nursing Science, 13*(1), 196-201. doi:10.1111/jjns.12094

MacCallum, Roznowski, & Necowitz. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504. doi:10.1037/0033-2909.111.3.490

MacCallum, Widaman, Preacher, & Hong. (2001). Sample size in factor analysis: the Role of model error. *Multivariate Behavioral Research, 36*(4), 611-637. doi:10.1207/S15327906MBR3604_06

Mohiyeddini, John, & Gross. (2008). Der "Berkeley Expressivity Questionnaire": Deutsche adaption und erste validierungsbefunde. [German version of the Berkeley Expressivity Questionnaire.]. *Diagnostica, 54*(3), 117-128. doi:10.1026/0012-1924.54.3.117

R Core Team. (2016). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rosseel. (2012). Lavaan: an R package for structural equation modeling. *Journal of Statistical Software, 48*(2), 1-36. Retrieved from http://www.jstatsoft.org/v48/i02/

Russell, Bachorowski, & Fernandez-Dols. (2003). Facial and vocal expressions of emotion. *Annual Review of Psychology, 54*, 329-349. doi:10.1146/annurev.psych.54.101601.145102

Snyder, & Gangestad. (1986). On the nature of self-monitoring: matters of assessment, matters of validity. *Journal of Personality and Social Psychology, 51*(1), 125-139. doi:10.1037/0022-3514.51.1.125

Spitzer, Kroenke, Williams, & Löwe. (2006). A Brief Measure for Assessing Generalized Anxiety Disorder: The GAD-7. *JAMA Internal Medicine, 166*(10), 1092-1097. doi:10.1001/archinte.166.10.1092

Tangney, Baumeister, & Boone. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of personality, 72*(2), 271-324. doi:10.1111/j.0022-3506.2004.00263.x

Tice, & Bratslavsky. (2000). Giving in to Feel Good: The Place of Emotion Regulation in the Context of General Self-Control. *Psychological Inquiry, 11*(3), 149-159. doi:10.1207/S15327965PLI1103_03

Tunay Akan, & Bariskin. (2017). [Reliability and Validity Indicators of Berkeley Expressivity Questionnaire in the Context of Culture and Gender]. *Turk psikiyatri dergisi = Turkish journal of psychiatry, 28*(1), 43-50.

Vigil. (2009). A socio-relational framework of sex differences in the expression of emotion. *The Behavioral and Brain Sciences, 32*(5), 375-390. doi:10.1017/s0140525x09991075

Vingerhoets, Nyklíček, & Denollet. (2008). *Emotion regulation : conceptual and clinical issues*. New York, NY :: Springer.

Wang, & Lau. (2018). Ethnicity moderates the benefits of perceived support and emotional expressivity on stress reactivity for Asian Americans and Euro Americans. *Cultural Diversity & Ethnic Minority Psychology, 24*(3), 363-373. doi:10.1037/cdp0000197

Table 1. Descriptive and reliability statistics for the Dutch BEQ

	Mean (SD) T0	*Mean (SD) T1*	*Cronbach's alpha T0*	*Omega total T0*	*Average ICC*
Young adults (N=209 (T0); N=166 (T1))					
BEQ - negative expression	3.93 (1.03)	4.05 (.99)	.75	.79	.86 (.81-.90)
BEQ - positive expression	5.38 (.86)	5.37 (.88)	.71	.80	.84 (.79-.88)
BEQ - impulse strength	4.68 (1.17)	4.75 (1.17)	.80	.66	.89 (.85-.92)
BEQ total	4.66 (.85)	4.72 (.84)	.87	.89	.90 (.87-.93)
General Population (N=586 (T0); N=113 (T1))					
BEQ - negative expression	3.97 (.96)	4.10 (.94)	.70	.69	.86 (.80-.90)
BEQ - positive expression	5.26 (.89)	5,42 (.83)	.70	.75	.89 (.85-.93)
BEQ - impulse strength	4.22 (1.21)	4.40 (1.11)	.83	.88	.90 (.85-.93)
BEQ total	4.49 (.83)	4.48 (.83)	.86	.90	1.00 (1.00-1.00)

Note: SD = standard deviation; ICC= intra-class correlation

Table 2. Construct validity – Behavioral, affective and cognitive correlates of emotion expression

		BEQ total		**Positive expressivity**		**Negative expressivity**		**Expression intensity**	
		YA	GP	YA	GP	YA	GP	YA	GP
Behavioral control	Self-monitoring	.022	.101	.087	**.192**	-.027	.001	-.087	.066
	Self-control[1]	.064	**-.217**	.095	**-.154**	.064	**-.130**	.016	**-.233**
Emotional response tendency	Depression	.037	.099	-.034	-.062	**-.129**	-.054	-.046	**.289**
	Anxiety	.111	**.213**	-.027	.040	-.046	.018	**.303**	**.392**
	BFI – Neuroticism	**.270**	**.322**	-.018	.085	**.182**	**.172**	**.441**	**.467**
	BFI – Extraversion	**.382**	**.238**	**.544**	**.441**	**.366**	**.239**	**.111**	-.019
	BFI – Openness	.054	.069	**.185**	**.228**	-.042	.025	.019	-.042
	BFI - Conscientiousness	**.148**	-.022	.089	.087	**.127**	-.046	**.147**	-.072
	BFI - Agreeableness	**.176**	**.198**	**.302**	**.185**	**.133**	**.135**	.045	**.167**
Emotion regulation	ERQ - suppression	**-.590**	**-.551**	**-.545**	**-.502**	**-.604**	**-.581**	**-.354**	**-.300**
	ERQ - reappraisal	-.021	.045	**.116**	**.137**	-.069	-.042	-.070	.025
	CERQ - self blame	-.023	.011	-.066	.007	-.009	-.053	.006	.059
	CERQ - Acceptance	**-.118**	.034	.010	.013	**-.143**	-.028	**-.138**	.083
	CERQ - Rumination	**.211**	**.247**	.060	**.135**	**.100**	.063	**.326**	**.355**
	CERQ - Positive refocus	.023	.094	.069	**.136**	.038	.044	-.035	.057
	CERQ - refocus planning	.065	.013	**.197**	**.119**	.054	-.003	-.050	-.059
	CERQ - Positive reappraisal	-.019	**.118**	**.164**	**.185**	-.049	.029	**-.118**	.083
	CERQ - Perspective	-.020	-.056	.045	-.026	.077	-.068	**-.144**	-.042
	CERQ - Catastrophizing	.099	**.163**	-.064	.038	.042	.039	**.227**	**.273**
	CERQ - Blaming others	.031	.10	.009	**.104**	.028	.027	.037	**.118**

Note: [1]Measured at T1. YA = young adults, GP = general population. Bold faced indicates $p < .05$

SUPPLEMENTARY MATERIALS

Online Supplement – Methods

Study 1: Participant data collection procedure

Sample 1: Participants filled out (most) questionnaires twice via an emailed link after giving informed consent, at inclusion (T0) and 6 months later (T1). Approval for the INHIBIT study was obtained from the institutional ethics review board (ERB; protocol number: EC-2015 64a).

Sample 2: comprised a non-random selection of 586 adults (54.1% female; M_{age}, 46.63±15.94; range 18-81 years) from the general Dutch population. Quota sampling was successfully applied to ensure that different age and sex groups were equally represented. Approval for this study was obtained from the institutional ERB (protocol number: EC-2015 64a). Research assistants were responsible for distributing the questionnaires (online or on paper) and for the quota sampling. Research assistants were relatively free in choosing who to approach, how (personally or by phone), and where to approach them, as long as participants were not employees of the university or friends. After explaining the purpose of the study, participants received an informed consent form and a questionnaire either in person or by mail, which were sent back to the research assistants in closed envelopes. For both samples, online questionnaires were of a forced entry format, so that no questions could be missed. We ran a data validation procedure, checking for duplicates, missingness, and poor responding. Questionnaires on paper were collected predominantly for the highest age cohort of 75-85, due to lack of digital skills. Paper questionnaires were entered into the database by others, guaranteeing anonymity. Returned questionnaires did not contain any explicit identifiers (i.e., names) but rather were coded by number for purposes of data collection

tracking. Part of the sample indicated that they would like to participate in follow-up research, and left their email address for this purpose. The datasets used for this publication have been stored at the Tilburg University Dataverse repository.

Materials

Social demographics

Lower education was defined as completing primary school, prevocational education or high school, and higher education as completing vocational education, college, or university. A positive marital status included being married, living together, or being in a longstanding relationship.

Questionnaires study 1

ERQ - The Emotion Regulation Questionnaire (ERQ; (Gross & John, 2003)) was used to assess the emotion regulation strategies suppression and reappraisal. Emotion regulation was rated on a seven-point Likert scale ranging from 1 (strongly disagree) to 7 (strongly agree). Suppression was assessed with four items, and reappraisal with six items. Cronbach's alphas for the current study ranged between .74-.88 ($Omega_{total}$=.82 (YA), .83 (GP)).

CERQ - To measure cognitive emotion regulation, the Cognitive Emotion Regulation Questionnaire Short-Form (CERQ-SF; Garnefski & Kraaij, 2006) was used. This 18-item questionnaire is a short form of the original 36-item questionnaire. The questionnaire was rated on a five-point Likert scale, ranging from 1 ((almost) never) to 5 ((almost) always). Nine subscales were distinguished (Self-blame, Other-blame, Rumination, Catastrophizing, Positive refocusing, Planning, Positive reappraisal, Putting into perspective and Acceptance) and every subscale consists of two items. Because of the small number of items (i.e., two) per subscale, inter-item

correlations were calculated instead of Cronbach's alphas, which ranged between .47 and .70 for the respective subscales ($Omega_{total}$= .89 (YA), .93(GP)), suggesting good consistency.

Self-control - The Brief Self-Control Scale (BSCS) was used to measure self-control. This scale is assessed to scale individual differences in the trait self-control in the four major domains of self-control; controlling emotions, thoughts, behaviors and impulses. The BSCS consists out of 13 items on a five-point Likert scale from 1 (strongly disagree) to 5 (strongly agree) (Tangney et al., 2004). The BSCS has shown to be a valid instrument (Maloney, Grawitch, & Barber, 2012). In the current dataset, we used the total score, which had sufficient internal consistence (Cronbach's alpha = .77-.78; Omega= .86).

Self monitoring - The Self-Monitoring Scale (SMS-R; Gangestad & Snyder, 1985) was used to measure self-expression. Participants gave true/false answers to indicate whether each of the 18 items described them. The scale consists of two subscales (Public Performance and Other Directedness), however, we decided to only use the total score because of its better (though still moderate) psychometric properties (Cronbach's alpha = .61-.67; $Omega_{total}$= .82 (YA), .86 (GP)). Cronbach's alphas were comparable to those in existing literature (Gangestad & Snyder, 1985).

Big-5 - The BFI was used to measure the Big Five components extraversion, agreeableness, conscientiousness, neuroticism, and openness. The questionnaire consists of 44 items to measure the Big Five components. The items are answered on a five-point Likert scale (1 = strongly disagree, 5 = strongly agree). The scales of the BFI have shown good internal consistency and convergent validity with corresponding scales of Goldberg's adjectives and Costa and McCrae's NEO-FFI (Denissen, Geenen, van Aken, Gosling, & Potter, 2008). Cronbach's alphas ranged between .75 and .85 ($Omega_{total}$= .84 (YA), .92 (GP)).

Depressive symptoms - The 9-item Patient Health Questionnaire (PHQ-9) was used to assess depressive symptoms. This questionnaire scores each of the 9 DSM-IV criteria on a four-point Likert scale from 0 (not at all) to 3 (nearly every day) (Kroenke, Spitzer, Williams, 2001). This reliable and valid measure of depression severity had a Cronbach's alpha of .82 in the current study (Omega=.90 (YA), .93(GP)).

Anxiety symptoms – Anxiety symptoms were measured using the seven-item Generalized Anxiety Disorder scale (GAD-7; Spitzer et al., 2006). Items on this scale are rated on a four-point Likert scale from 0 (not at all) to 3 (almost daily), with total scores ranging from 0 to 21. The GAD-7 is a reliable scale, with good criterion-, construct-, factorial-, and procedural validity (Spitzer et al., 2006). Cronbach's alpha for the GAD-7 was .87 (Omega=.90 (YA), .93(GP)).

Factorability

In SPSS, the Kaiser-Mayer-Olkin test indicates suitability of data for factor analysis. In our general population sample, the KMO statistic = .61, while in the young adult sample, it was .84, suggesting they were respectively fair and excellent samples for factor analysis.

Study 2:

Participants & procedure

Data are part of a larger stress study examining individual differences in physiological and emotional stress reactivity (PHEMORE). Participants underwent the Trier Social Stress task. Data were collected during the winter of 2015. None of the participants reported being in poor health at the time of the experiment.

The study protocol and its amendments were approved by the Institutional Ethics Review Board (EC-2011.01a). All participants gave informed consent before participating and were debriefed afterwards.

Materials

Acute emotional responses - Participants rated items reflecting their level of affective arousal (tension, anxiety, irritation, fatigue, annoyance, sad, angry, stressed, at ease), task engagement (engaged/stimulated, interested), and task difficulty (effort, burden, difficulty) on a Likert-type scale of 1 (not at all) to 7 (very much). Fusing the affective arousal items, we calculated a negative emotion total score for the resting period and for stress (adding up average scores for the math and speech test). These negative emotion scores were internally consistent scores (Cronbach's alpha .74 (rest), and .84 (stress)). In addition, we computed individual emotion scores for the rest and stress period, to calculate individual reactivity scores. We added this post-hoc analysis because of a meta-analysis showing stimulus-associated emotion-specific responses (Siegel et al., 2018).

Study procedure study 2.

Procedure. Upon arrival, all participants were welcomed, placed in a quiet, dimly lit waiting room, and asked to sign for informed consent. Then, a psychological survey was administered, including dedicated questions on demographics (age, sex, partner status), health behaviors (exercise, smoking, weekly alcohol consumption, daily coffee consumption), body composition (length, weight), medication use, mood disorders (anxiety, depression: Has a medical doctor or registered psychologist told you that you have depression/anxiety, or are you being treated for anxiety/depression?), and a series of standardized psychological questionnaires. Then, participants

were fitted with the cardiovascular measurement equipment. Participants were examined in a sitting position. After a 10-min resting period, during which a physiological baseline was recorded, participants took part in a stress test battery.

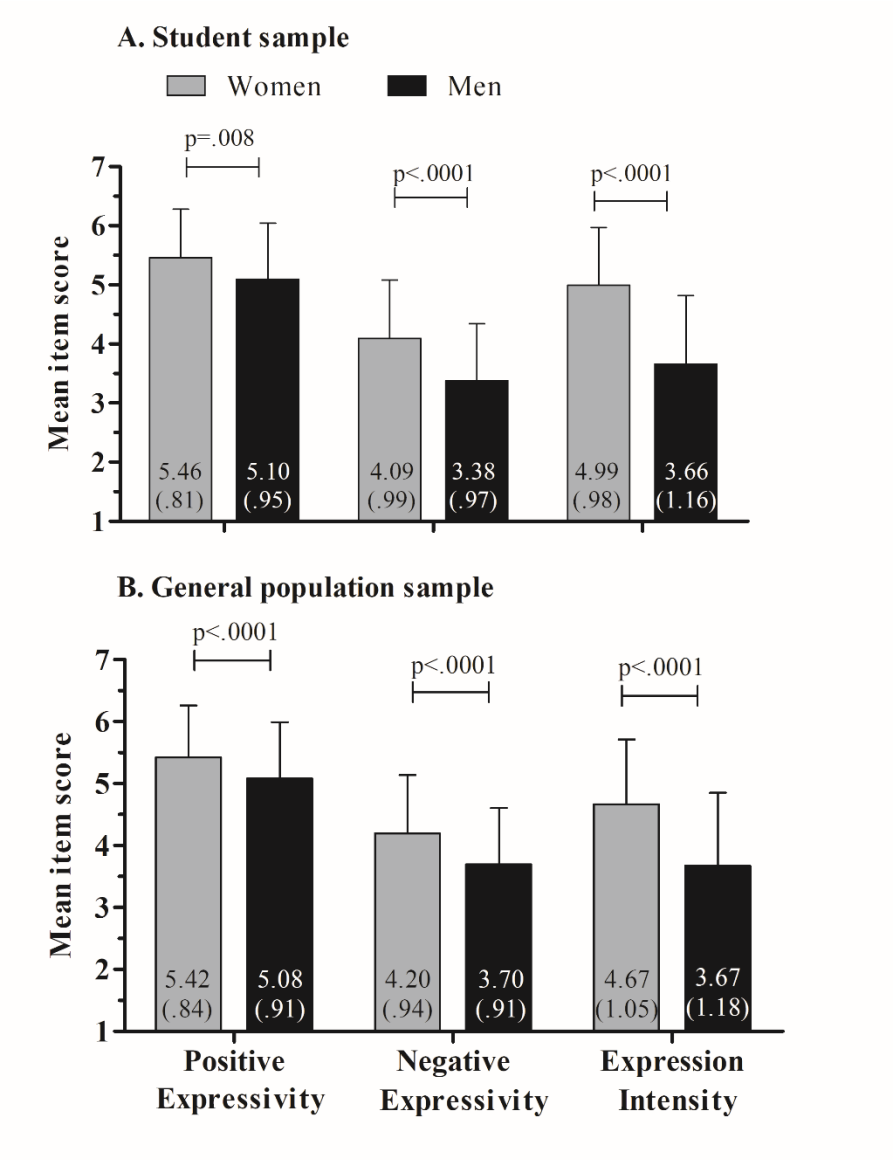
***Adapted Trier social stress test*.** Briefly, the Trier Social Stress test is a social stressor during which a participant is asked to perform a math task and to give a prepared speech (Kirschbaum, Pirke, & Hellhammer, 1993). We adapted the original protocol of the Trier Social Stress test in two ways. First, we asked participants to remain seated throughout the entire procedure, as this is vital for the hemodynamic measures (not used in the current analysis). Second, instead of a job interview, we asked participants to prepare (3-minute preparation period) and give a speech on their own positive and negative social skills (5 minutes), in front of a 2-person audience. Previous research has shown that the current procedure induces a significant cardiovascular stress response (Kupper, Pelle, & Denollet, 2013). Further, we randomized task order, so that about half of the participants first did the speech task, while the other half started with the 5-minute math task.

Study 2 measures

Cronbach's alpha for the BEQ total score was .89 (Omega=.89). For negative expressivity, positive expressivity, and impulse strength, Cronbach's alphas were .75 (Omega=.83), .74 (Omega=.74), and .84 (Omega=.67) respectively, indicating good internal consistency in this sample.

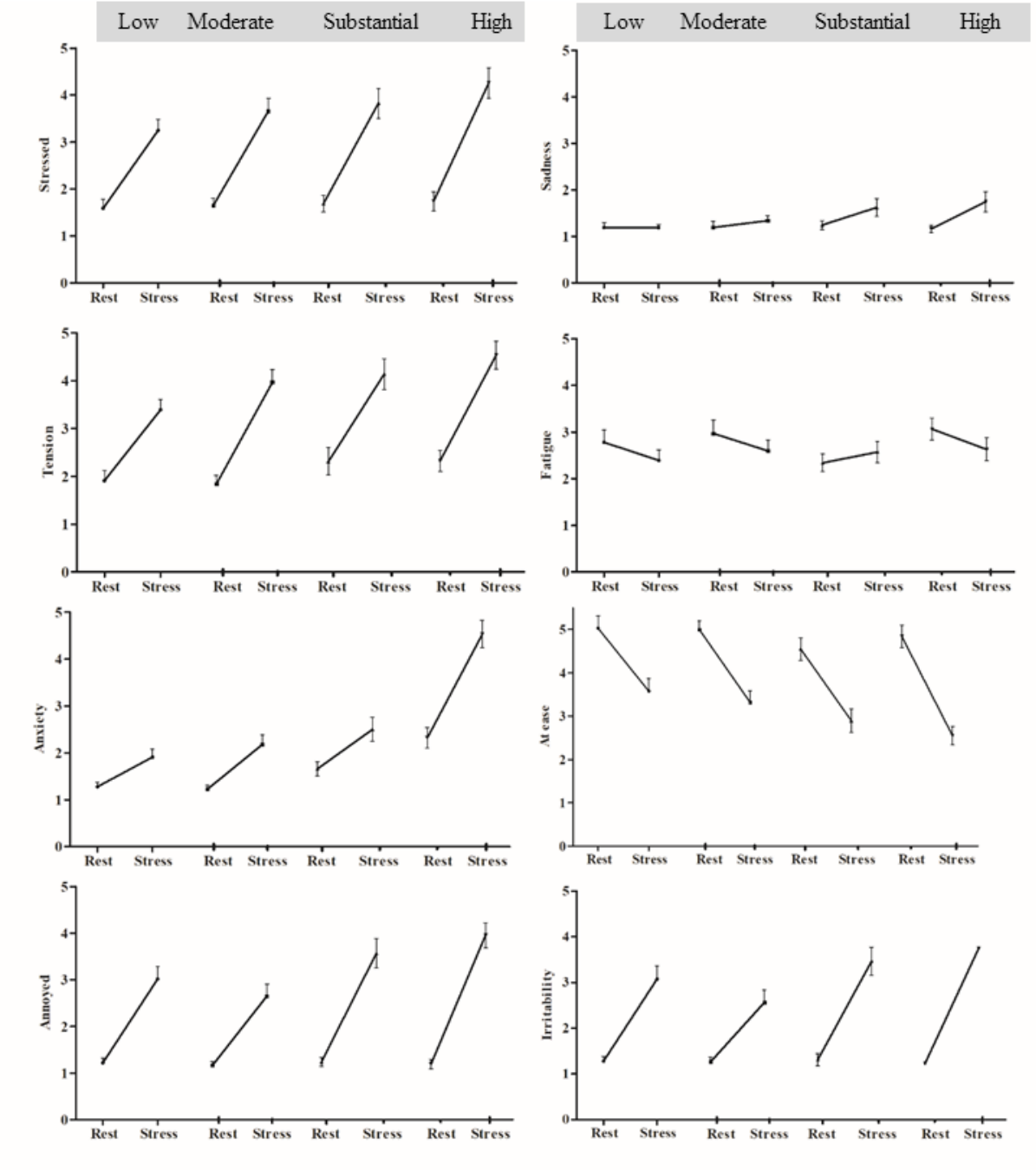
Supplemental results

Figure S1. Sex differences in expressivity scores



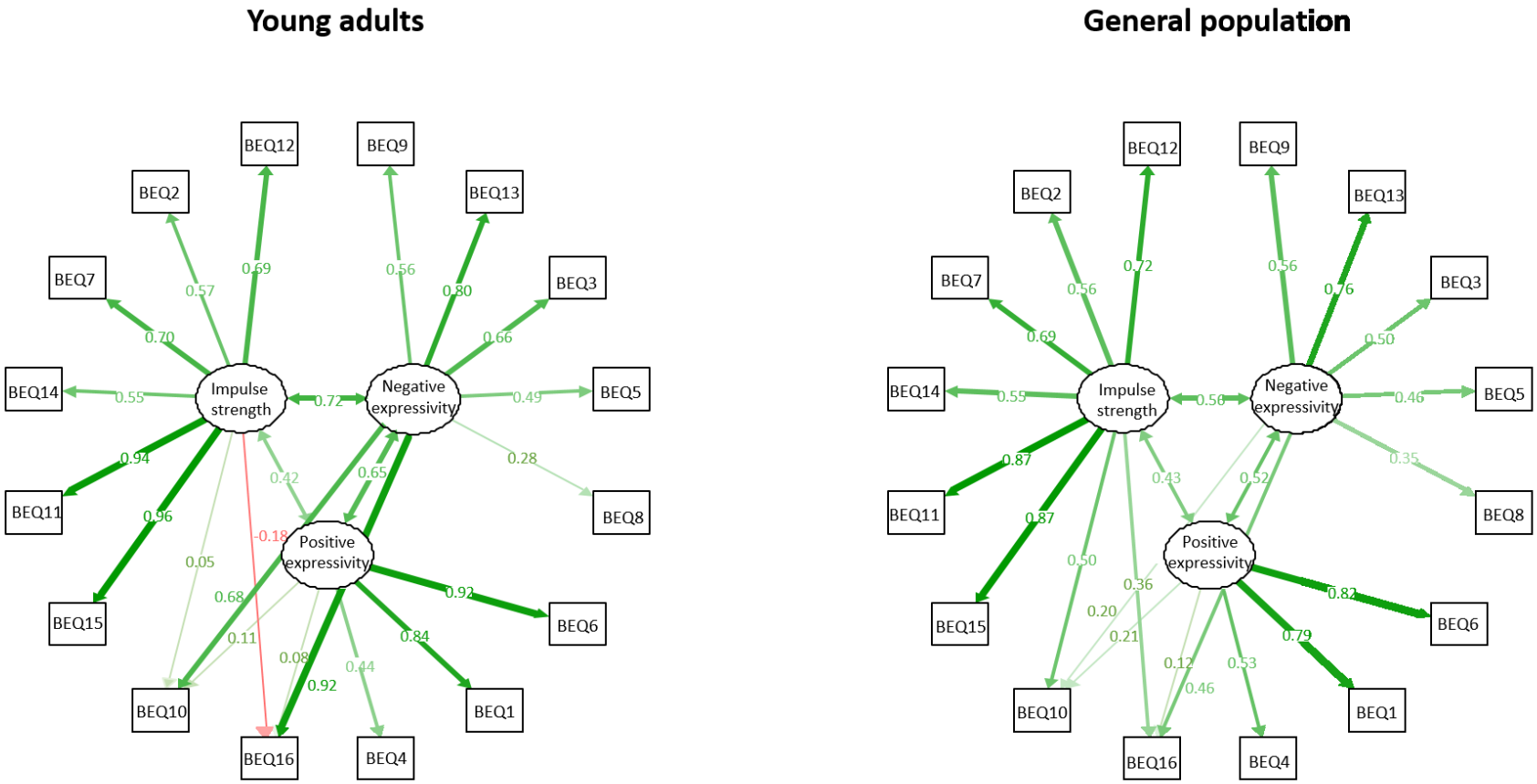
Note: Means ± SD for each subscale and each gender are presented inside the bars.

Figure S2. Stress responses of individual emotions, stratified by four categories of emotion expression intensity.



Note: Lines represent the means (error bars = SEM) of the reported emotions during rest and stress, stratified by four categories of emotion expression (i.e., low, moderate, substantial and high).

Figure S3. Best fitting models with factor loadings for Young adult sample and General population sample



Note: Green lines reflect positive factor loadings, while red lines reflect negative factor loadings. Darker colors and broader arrows reflect stronger loadings

Table S1. Baseline characteristics of the samples

	Study 1		*Study 2*
	Young adult Survey sample (N=209)	**General population Survey sample (N=586)**	**Stress test Young adult sample (N=123)**
Female sex (%, (N))	77% (160)	54% (315)	77% (95)
Age in years *(Mean ±SD)*	20.3 ± 2.1	46.6 ± 15.9	20.1 ± 2.7
With partner (%, (N))	42% (88)	79% (463)	48% (59)
Higher education (%, (N))	100% (209)	50% (290)	100% (123)
Current psychological or psychiatric treatment (%, (N))	7% (14)	4% (25)	12% (15)

Table S2A&B. Model fitting results CFA Young adults (N=209) and the General population (N=570)

Model	df	Chi square	Δ Chi square	Δ df	Significance[5]	Robust fit indices		
						χ^2 / df	TLI	RMSEA
3-factor+cross[1]	97	246.62				2.54	.97	.079
3-factor[2]	101	315.81	69.19	4	<.0001	3.13	.96	.088
2-factor[3]	103	340.66	94,04	2	<.0001	3.24	.96	.090
1-factor	104	954.68	708,06	1	<.0001	9.01	.80	.20
Model	**df**	**Chi square**	**Δ Chi square**	**Δ df**	**Significance[5]**	**Robust fit indices**		
						χ^2 / df	TLI	RMSEA
3-factor+cross[1]	97	425.27				4.38	.95	.079
Higher order - 3F[4]	101	636.12	210.85	4	<.0001	6.30	.92	.096
3-factor[2]	101	671.96	125.434	4	<.0001	6.65	.92	.10
2-factor[3]	103	780.61	53.155	2	<.0001	7.58	.91	.10
1-factor	104	1014.47	101.770	1	<.0001	9.75	.89	.12

Note: [1] Original = correlated three-factor model with two general expressivity items loading on all factors. [2] Correlated three-factor model without cross-loadings of the general items. General items were loading on the subscale where they were placed at in the questionnaire scoring instructions. [3] Correlated two-factor model, one for valenced expressivity, one for expressive intensity. [4] hierarchical model, only tested in General population due to small sample size of Young adult sample. [5] Significance indicates the significant of differences in fit in comparison with the original model.

Table S3. Unstandardized (SE) and standardized factor loadings from the three-factor models on both samples

Latent factor	*Indicator*	*Young adults*		*General population*	
		B (SE)	**Beta**	**B (SE)**	**Beta**
Negative	(9r)	.56 (.07)	.56	.56 (.03)	.51
Negative	(13)	.80 (.04)	.81	.76 (.03)	.76
Negative	(16*)	.92 (.22)	.92	.46 (.06)	.46
Negative	(3r)	.66 (.04)	.66	.50 (.05)	.50
Negative	(5)	.49 (.05)	.49	.46 (.06)	.46
Negative	(8r)	.28 (.07)	.28	.35 (.04)	.35
Negative	(10*)	.68 (.12)	.68	.20 (.06)	.20
Positive	(6)	.92 (.05)	.92	.82 (.02)	.82
Positive	(1)	.84 (.04)	.84	.80 (.03)	.80
Positive	(4)	.44 (.11)	.44	.53 (.04)	.53
Positive	(10)*	.11 (.13)	.11	.21 (.05)	.21
Positive	(16)*	.08 (.17)	.08	.12 (.05)	.12
Impulse strength	(15)	.96 (.02)	.96	.87 (.02)	.87
Impulse strength	(11)	.94 (.02)	.94	.87 (.02)	.87
Impulse strength	(14)	.55 (.06)	.55	.55 (.03)	.55
Impulse strength	(7)	.70 (.04)	.70	.69 (.02)	.69
Impulse strength	(2)	.57 (.05)	.57	.56 (.03)	.56
Impulse strength	(12)	.69 (.04)	.69	.72 (.02)	.72
Impulse strength	(10)*	-.18 (.14)	-.18	.36 (.05)	.36
Impulse strength	(16)*	.05	.08	.50 (.04)	.50

Note: * = general expressivity item loading on all factors

Table S4. Modification indices for the three-factor model of the Dutch BEQ in two samples

Latent Factor	Operator	Item	Modification Index	Expected parameter change
Young adults				
Negative	=~	BEQ12 *'can't hide feelings'*	82.24	.78
Positive	=~	BEQ12 *'can't hide feelings'*	49.07	.55
Negative	=~	BEQ15 *'strong experience of emotions'*	25.31	.58
Strength	=~	BEQ3r *'people unaware of what I feel'*	12.80	-.42
Negative	=~	BEQ11 *'strong emotions'*	12.66	-.40
General population				
Strength	=~	BEQ5 *'difficult to hide fear'*	75.18	.48
Strength	=~	BEQ8r *'better to suppress anger'*	57.29	-.37
Positive	=~	BEQ14 *'moments that I can't stop crying'*	30.27	-.24
Positive	=~	BEQ3r *'people unaware of what I feel'*	25.49	.32
Negative	=~	BEQ12 *'can't hide feelings'*	16.81	.20

Note: All modification indices (MI) suggested adding cross-loadings. For young adults, these were the highest five MI values corresponding to three items of the BEQ. In the general population, 5 items were suggested for adding cross-loadings.

Exploratory results

As Table S3 shows, both samples show a range of substantial modification indices, of which we show the five largest values per sample. Adhering to the guideline that one should only make changes that are theoretically sensible, there are several options that can be explored.

Items 11, 12, 14, and 15 are all expression intensity items. In both samples, these items also may relate to positively valenced expression and negatively valenced expression. Theoretically, the expression of certain negative emotions (anger, fear) may naturally be expressed more strongly than low arousal emotions, and therefore may be correlated. Not being able to hide any feelings, suggests the increased expression of both negative and positive emotions. Exploring the addition of item 12 as a third general expressivity item loading on all factors, improved the young adult model significantly ($\Delta\chi^2 = 77.6$, df = 2, $p < .0001$) and improved its fit (TLI = .98, RMSEA = .061). In the general population sample, adding a loading of item 5 ('*difficult to hide fear*') on strength (largest MI) led to a significant improvement in model fit ($\Delta\chi^2 = 74{,}29$, df = 1, $p < .0001$, TFI=.96, RMSEA = .06). Also, adding a cross-loading between item 8r and the strength factor (highest MI in latest model),

further improved the model ($\Delta\chi^2 = 74{,}29$, df = 1, $p < .0001$). These two additions to the model signify that some negative emotions such as anger and fear are by definition of higher intensity.

Table S5: Sex differences in correlations between BEQ facets and constructs of behavioral control, emotional response tendencies, and emotion regulation.

			BEQ total		**Positive expressivity**		**Negative expressivity**		**Expression intensity**	
			YA	GP	YA	GP	YA	GP	YA	GP
Behavioral control	Self-monitoring	Women	.11	.19	.15	**.30**	.08	**.**08	.05	.08
		Men	.04	.16	.12	.19	-.04	.01	.02	.19
	Self-control	Women	.09	**-.26**	.12	-.15	.10	-.14	.02	**-.29**
		Men	-.08	-.13	-.002	-.10	.14	-.05	-.07	*-.15*
Emotional response tendency	Depression	Women	.03	.08	-.01	-.08	-.18	-.13	**.25**	**.35**
		Men	-.05	.01	-.19	*-.12*	-.07	-.04	.10	.15
	Anxiety	Women	.05	.19	-.05	-.02	-.13	-.03	**.29**	**.44**
		Men	.11	.16	-.06	.04	.07	-.01	**.23**	**.30**
	BFI Neuroticism	Women	.16	.18	-.06	-.11	.05	.01	**.36**	**.46**
		Men	**.25**	**.40**	-.10	**.24**	**.30**	**.29**	**.39**	**.43**
	BFI Extraversion	Women	**.45**	**.24**	**.57**	**.52**	**.41**	**.26**	**.15**	-.18
		Men	**.40**	**.26**	**.54**	**.38**	**.36**	**.21**	**.14**	**.11**
	BFI Openness	Women	.12	**.23**	**.22**	**.40**	.00	.14	.10	.03
		Men	**.20**	.04	**.25**	.13	.05	-.03	.19	-.002
	BFI Conscientiousness	Women	.10	-.09	.05	.003	.10	-.07	.08	-.13
		Men	.19	.04	.12	.18	.11	-.02	**.22**	-.03
	BFI Agreeableness	Women	.13	.02	.28	.12	.12	-.001	-.05	-.06
		Men	**.22**	**.24**	**.34**	.19	.09	.18	.13	**.22**
Emotion regulation	ERQ-suppression	Women	**-.60**	**-.53**	**-.56**	**-.52**	**-.62**	**-.56**	**-.29**	-.19
		Men	**-.50**	**-.50**	**-.44**	**-.55**	**-.48**	**-.44**	**-.34**	**-.28**
	ERQ reappraisal	Women	-.08	-.03	.11	.13	-.09	-.06	-.18	-.12
		Men	-.03	.09	.09	.13	-.15	-.07	-.02	.12
	CERQ self blame	Women	-.01	.08	-.06	.03	.04	.02	.02	.11
		Men	.03	.04	-.05	.03	-.09	-.10	.18	.13
	CERQ acceptance	Women	-.05	.05	.09	.08	-.07	-.04	-.11	.07
		Men	-.11	-.01	-.09	-.08	-.21	-.04	-.003	.08
	CERQ Rumination	Women	.16	**.29**	.03	.19	.04	.08	**.30**	**.40**
		Men	**.28**	.19	.10	.06	.18	.02	**.39**	**.33**

	CERQ Positive refocus	Women	.10	.08	.06	.17	.01	.07	-.07	-.03
		Men	.11	.08	.09	.09	.15	-.01	.04	.11
	CERQ Refocus planning	Women	.10	.09	**.22**	.19	.05	.09	.01	-.04
		Men	.19	-.03	**.24**	.06	.26	-.09	.01	-.03
	CERQ Positive reappraisal	Women	-.05	.16	.17	.24	-.08	.03	-.17	.11
		Men	**.32**	.09	**.26**	.13	**.26**	.03	**.28**	.07
	CERQ Perspective	Women	.14	-.03	.17	-.02	**.22**	.002	-.05	-.04
		Men	**-.33**	-.11	-.23	-.05	**-.29**	-.17	**-.30**	-.05
	CERQ Catastrophizing	Women	.02	.12	-.09	-.003	-.07	-.02	.19	**.27**
		Men	**.21**	.19	-.06	.06	**.29**	.07	**.28**	**.29**
	CERQ Blaming others	Women	.09	.14	.08	.10	.03	.05	.10	.16
		Men	-.10	.14	-.19	.14	.03	.04	-.09	.16

Note: Correlations significant at $p<.01$ are highlighted in bold faced typing. YA = young adult; GP = General population. The intensity of emotion expression was most strongly associated with the emotional response on being annoyed, and feeling angry, and unrelated to fatigue, irritability, and being interested in the task. In correlational analysis, it was further shown that negative expressivity was positively associated with baseline levels of tension ($r=.24$, $p=.007$), and that positive expressivity was inversely correlated with the level of baseline sadness ($r=-.19$, $p=.036$).

Table S6. Unadjusted eta squared values from repeated measures analysis of reactivity of individual emotions

N=123	Negative expressivity	Positive expressivity	Impulse strength
Stressed	.01	.001	**.04**
Tense	*.03*	.002	**.04**
Anxiety	.001	.02	**.03**
Irritability	.00	.00	*.03*
Fatigue	.001	.01	.002
Annoyed	.01	.000	**.06**
Sad	.000	.001	**.04**
Anger	.02	.00	**.09**
At ease	.008	.001	**.03***
Interest	.004	.004	.001

* At ease demonstrated a reverse response, bold-faced indicates a significant effect ($p<.05$).