The Affect Labeling Questionnaire (ALQ): Decomposing affect labeling and implications for individual differences in socio-emotional well-being

Razia S. Sahi[1], João F. Guassi Moreira[1], Jared B. Torre[2], & Matthew D. Lieberman[1]*

[1]University of California Los Angeles, Department of Psychology, Los Angeles, CA, USA

[2]Adobe Inc., Seattle, Washington, USA

*Correspondence should be addressed to MDL (lieber@ucla.edu)
Department of Psychology, Pritzker Hall, University of California, Los Angeles, Los Angeles, California 90095 USA

Acknowledgments and Funding Information: We are grateful to Dr. Steven Reise and Dr. Wes Bonifay for their advice regarding our statistical analyses, particularly relating to factor analysis and item response theory. This material is based upon work supported by a NIMH grant to MDL (R01MH084116) and NSF-GRF (DGE-1144087) to JBT. Preparation of this manuscript was supported by a NSF-GRF (2016220797) and a NIH-T32 to JFGM. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or National Institutes of Health.

ABSTRACT

A number of studies have shown that putting feelings into words, a process known as affect labeling, can be an effective emotion regulation strategy for dampening negative affect. However, the components of this emotion regulation strategy and corresponding implications for social behavior, relationships, and other emotion regulation strategies remain poorly understood. We developed the Affect Labeling Questionnaire (ALQ) to measure individual differences in affect labeling and its relation to other psychological constructs. In study 1 (N_1 = 200), we used exploratory factor analysis to identify three factors underlying affect labeling: (a) affective awareness; (b) affect labeling tendency; and (c) affect labeling capacity. In study 2 (N_2 = 293), we used confirmatory factor analysis to validate our affect labeling measure with these three components. Item response theory was utilized in both studies for additional item validation and to ascertain psychometric properties of the scale. In study 3 (N_3 = 227), we correlated the ALQ with existing measures of social and emotional tendencies, and found that greater affect labeling use tracked with increased emotional expressivity, social support, and use of other emotion regulation strategies (i.e. reappraisal, coping). In addition to validating a measure of affect labeling for future use, these data shed light on the psychological structure underlying affect labeling and inform how it relates to other socio-emotional metrics of well-being.

Keywords: affect labeling, emotion regulation, measurement, factor analysis, item response theory

Emotion regulation – the process of managing our emotional experiences and expressions – is important for navigating daily challenges and maintaining health and well-being (Gross, 1998; Gross et al., 2006). Intuitively, one might try to distract themself from an undesirable emotional state (e.g. watching television) or try to change how they're thinking about an emotional event (e.g. "it's not that bad") to dampen an emotional response, but research suggests that simply putting feelings into words can also be a powerful tool for managing emotions (Torre & Lieberman, 2018). This emotion regulation strategy – known as affect labeling – can reduce self-reported distress, decrease fear-related neural responsivity, and lower skin conductance response to evocative stimuli (Burklund et al., 2014; Constantinou et al., 2014; Kircanski et al., 2012; Lieberman et al., 2011; Niles et al., 2015).

However, we still know relatively little about how various processes relating to emotion recognition and regulatory behavior – including awareness of one's emotional states, the tendency to use language to process emotional states, and the ability to find specific language to describe those states – contribute to affect labeling. Systematically decomposing the components of affect labeling and being able to assess their unique implications for social behavior, relationships, and emotion regulation more broadly is an important step in understanding how individuals use and benefit from affect labeling. One way towards this end is to develop and validate a measurement tool for individual differences in affect labeling. This approach facilitates a better mechanistic understanding of affect labeling, while also providing an inexpensive and easily implemented tool for future research that can aid in unpacking how affect labeling relates to different psychological behaviors and processes in the lab and in everyday life.

Decomposing Affect Labeling as an Emotion Regulation Strategy

The notion that affect labeling can help people regulate their emotions is not new. Indeed, therapeutic settings, in which people share their emotional experiences, and expressive writing, wherein people describe their thoughts and feelings, have long been thought to offer benefits in processing emotional states (Greenburg, 2001; Pennebaker, 1993). However, only in the last decade or so have researchers systematically examined affect labeling as a regulatory strategy. As this line of work evolved, affect labeling has been studied in terms of labeling one's own feelings (e.g. "I feel sad"), someone else's feelings (e.g. "That child looks sad"), or the emotionally evocative aspect of a stimulus (e.g. "That is a tragic situation"). Thus, affect labeling is broadly defined as putting feelings into words, and lab paradigms have studied affect labeling by asking participants to say, write, or choose between labels to describe emotional stimuli (Hariri et al., 2000; Torre & Lieberman, 2018).

However, no work to our knowledge has assessed how people naturally implement affect labeling to regulate their emotions. There are likely several complementary processes that fold into effectively using this strategy that can be identified by developing a measure for affect labeling. First, individuals must have some *awareness* of their emotional states and how they are affected by them. Some people might go through the day feeling moody or upset without recognizing how they are feeling or what may have triggered their emotions, but others may be able identify the events or interactions that affected them (e.g. "When I am upset, I can pinpoint what is bothering me"). Such awareness is likely to play an integral role in the use of affect labeling in daily life since it draws attention to specific emotional states that one can process and potentially regulate.

Second, individuals may set an intention around regulating their emotions that allows them to process their feelings through language. Existing measures for emotion regulation strategies often assess this *tendency*, or habitual use, of emotion regulation strategies (Gross & John, 2003; Williams et al., 2018). Deliberately trying to change one's emotional states is not necessary for implicit regulatory strategies like affect labeling to alter emotional experiences (Koole & Rothermund, 2011). However, the tendency to use language to regulate emotions (e.g. "When I want to feel less upset, I will describe how I am feeling.") is likely to play an important role in downstream well-being outcomes, since it reflects a degree of effort that a person is investing in using affect labeling to maintain their emotional health.

Third, it is important to consider whether an individual can effectively translate their emotions into words, a process that is typically characterized in terms of the *capacity* to use a particular emotion regulation strategy. In other words, one might frequently attempt to regulate their emotions by describing them, but struggle to find the appropriate words to label their feelings (e.g. "It takes a lot of effort to convert my feelings into words."). Indeed, difficulty describing one's emotions (i.e. alexithymia) is associated with worse mental health and poorer therapeutic outcomes (Leweke et al., 2011; Ogrodniczuk et al., 2011), suggesting that this aspect of affect labeling plays a pivotal role in emotion regulation efficacy. Given the importance of this distinction between tendency and capacity in the use of emotion regulation strategies, researchers are increasingly disentangling them in how they measure individual differences in emotion regulation and related well-being outcomes (Mcrae, 2013; Silvers & Guassi Moreira, 2019; Troy et al., 2017). Thus, it will be important to assess both tendency and capacity in a useful measure of affect labeling.

Affect Labeling and Socio-Emotional Tendencies

Affect labeling can benefit individuals by dampening unwanted or disproportionate emotional responses, but it remains unknown how this regulatory strategy relates to other social and emotional behaviors and abilities that can facilitate overall health and well-being. First, it is useful to unpack how affect labeling relates to social dispositions. Since affect labeling often unfolds in social interactions such as therapy or conversations with others (Fan et al., 2019), it is possible that sociable personality traits such as extraversion and lower social anxiety track with putting feelings into words (Bainbridge et al., 2022). Relatedly, empathic behavior may track with affect labeling since identifying emotional states plays a role in understanding and sharing in others' emotional experiences (Morelli et al., 2017; Zaki, 2020). Thus, a measure of affect labeling would allow us to examine the association between affect labeling and possibly related measures of personality, giving us insight into who uses and potentially benefits the most from this emotion regulation strategy.

Second, it is useful to characterize how affect labeling relates to affect, mental health, and emotion regulation more broadly. Since affect labeling has been demonstrated to be an effective strategy for dampening negative affect, it should correspond to less daily negative affect and potentially less depression, and anxiety. Examining this association is important for establishing the cumulative benefits of affect labeling in everyday life. Additionally, greater affect labeling should correspond to greater use of other adaptive emotion regulation strategies such as cognitive reappraisal. Past research assessing the relationship between affect labeling and cognitive reappraisal during lab tasks suggests that there is a relationship between the efficacy of these strategies within individuals (Burklund et al., 2014; Lieberman et al., 2011), so it is informative from a validation perspective to test whether this positive relationship is also captured by questionnaires measuring these emotion regulation strategies.

Finally, there are several existing scales for measuring emotional expressivity more generally that should relate to affect labeling. While some of these measures, such as measures of alexithymia (Bagby et al., 1994) should be strongly inversely related to affect labeling to establish convergent validity, it would be important to observe more moderate associations with measures of general expressivity (Gross & John, 1997) for discriminant validity. Together, examining how a measure of affect labeling relates to other measures of socio-emotional tendencies and well-being can shed light on what it captures, how it differs from other constructs, and how it benefits individuals in everyday life.

Research Overview

Across three experiments, we examined how individuals use affect labeling to regulate their emotions. To assess individual differences in the different components of affect labeling, we developed a questionnaire including items corresponding to (a) affective awareness (i.e. awareness of one's emotional states); (b) affect labeling tendency (i.e. the tendency to use language to process emotional states); and (c) affect labeling capacity (i.e. the ability to find language to describe those states). We then proceeded to validate our measure by using a combination of exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and item response theory (IRT).

In study 1 (*N* = 200), we used EFA to determine the general factor structure of our measure (i.e. how many subscales to retain), and then IRT to determine which items to retain in each subscale. In study 2 (*N* = 293), we validated our pared down measure in a new sample using CFA. We tested two different measurement models to determine the factor structure of the model. These models, in conjunction with IRT, clarified the factor structure of affect labeling and confirmed the final items for each scale. In study 3 (*N* = 227), we correlated our final measure with other validated measures of social and emotional behaviors. We found that greater affect labeling use tracked with increased social support, emotional expressivity, and use of adaptive emotion regulation strategies such as reappraisal and coping.

Methods

General Procedure. We applied the same overall procedure and exclusion criteria across all three studies. Participants completed a set of questionnaires that took approximately 1 hour to complete, including a measure of affect labeling: the Affect Labeling Questionnaire (ALQ) (see Measures section, below, for complete list of questionnaires). For the ALQ, they read each of the test items and two attention check questions in a pseudo-randomized order and indicated the extent to which they agreed with each item on a 5-point Likert scale ranging from strongly disagree (1) to strongly agree (5). All analyses were conducted using the statistical package R (Version 1.2.1335). De-identified data, analysis scripts, and study materials are hosted on Open Science Framework (OSF; osf.io/fj3cz; Sahi et al., 2022). All procedures were approved by the local IRB committee and informed consent was obtained from all participants.

Exclusion Criteria & Sample Size. Data for all three studies was cleaned simultaneously using the same exclusionary criteria.[1] To ensure validity and replicability, we considered criteria relating to "careless responding" (Jaso et al., 2021): participants were excluded if they responded to greater than 50% of questions with a midline response (i.e. choosing 3 on a 5-point Likert scale) ($N = 36$), failed the attention check questions (e.g. "You must pick the leftmost option on this item.") ($N = 41$), finished the survey in under 20 minutes or over 3 hours ($N = 72$) (i.e., participants were excluded for completing the survey in less than 1/3 or more than 3 times the anticipated survey time of 1 hour), or finished less than 90% of the survey ($N = 5$). We additionally excluded participants if they were not proficient in English ($N = 75$) since reading and understanding each item was critical to providing informative responses. To attain sufficient statistical power and representation across a range of demographic variables (Comrey, 1988; Tinsley & Tinsley, 1987), taking into consideration anticipated data loss due to our exclusionary criteria, we recruited upwards of 300 participants each through Amazon Mechanical Turk (mTurk) and our undergraduate subject pool with the goal of retaining at least 200 participants for each study.

Measures. We included questionnaires related to personality, affect/mood, emotion regulation, empathy, and expressivity. For personality, we measured the big five dimensions (Big Five Inventory (BFI); John & Srivastava, 1999), social desirability (Social Desirability Scale (MC-SDS-10); Crowne & Marlowe, 1960), self-esteem (Rosenberg Self Esteem Scale (RSE); Rosenberg, 1979), and social anxiety (Social Interaction Anxiety Scale (SIAS); Mattick & Clarke, 1998). For empathy, we measured

[1] Since our samples for study 1 and study 2 were pooled for study 3, data was aggregated across our two samples and cleaned together before being subset into a study 1 dataset (mTurk), study 2 dataset. (undergraduates), and study 3 dataset (pooled data) for subsequent analyses. Participant details for each study are provided in our Methods section.

empathic tendency (Interpersonal Reactivity Index (IRI); Davis et al., 1993; Questionnaire of Cognitive and Affective Empathy (QCAE); Reniers et al., 2011) and positive empathy (Positive Empathy Scale (PES)) (Morelli et al., 2015). For affect and mental health, we measured general affect (Positive and Negative Affect Schedule (PANAS); Watson et al., 1988; Short Affect Intensity Scale (SAIS); Geuens & De Pelsmacker, 2002), depression (Beck Depression Inventory (BDI); Beck et al., 1988), and anxiety (Generalized Anxiety Disorder (GAD); Spitzer et al., 2006). For emotion regulation, we measured reappraisal and suppression frequency (Emotion Regulation Questionnaire (ERQ); Gross & John, 2003) and coping (Coping Orientation to Problems Experienced (COPE); Carver et al., 2013). For expressivity, we measured negative/positive expressivity (Berkeley Expressivity Questionnaire (BEQ); Gross & John, 1997; Emotional Expressivity Questionnaire (EEQ) ; Kring et al., 1994), difficulty identifying and describing feelings (Toronto Alexithymia Scale (TAS); Bagby et al., 1994), and mindfulness as it relates to describing one’s feelings (Five Facet Mindfulness Questionnaire (FFMQ); Baer et al., 2006).

Study 1

Item Development. We began by generating 15 items across four prospective subscales that were theoretically relevant to affect labeling as described in our introduction, referencing existing scales for emotion regulation such as the ERQ and IRQ for structure and length of items (Gross & John, 2003; Williams et al., 2018). We generated four items corresponding to affective awareness (e.g. “When I am upset, I can pinpoint what is bothering me”), seven items corresponding to affect labeling tendency (e.g. “I get my feelings "off my chest" by using words to describe them”), and four items corresponding to affect labeling capacity (e.g. “I can usually describe how I feel in great detail”) (see Table 1 for full list of items).

We generated thirteen additional items across four theoretical constructs to explore the relationship between affect labeling and broader emotional awareness and regulatory needs (Supplemental Table 1). These included five items corresponding to strength of emotional experiences (e.g. “It does not take much to make me feel very upset”), four items corresponding to need for emotional cognition (e.g. “I try to figure out where my feelings are coming from”), four items corresponding to aspirational affect labeling (e.g. “I’d like to be better at knowing what my emotions are”), and nine items corresponding to affective expression (e.g. “I frequently write about my feelings”). Since we do not consider these items to be core components of affect labeling, they are not included in our measurement analyses.

Participants. We recruited 330 participants through mTurk to complete the initial questionnaire via Qualtrics. After cleaning the data according to our exclusion criteria, our sample consisted of 200 participants. This sample was approximately: 8% 18-24 years old, 49% 25-34 years old, 26% 35-44 years old, 12% 45-54 years old, 5% 55-64 years old, and 1% 65-74 years old; 51% female, 49% male, and <1% transgender; 75% Caucasian/White, 6% Asian/Asian American, 8% Hispanic/Latino, 8% Black/African American, and the remaining selected another identity or multiracial.

Exploratory Factor Analysis. To assess the general factor structure of our measure and determine the optimal number of initial factors for future studies, we conducted a series of EFA models to examine item-factor correlations. We evaluated three possible solutions (2-4 factors). While we anticipated the existence of three factors — mapping onto each of the theorized subscales — we wanted to test for the possibility that two of our theorized subscales could be explained by a single factor (2 factor solution) or that a theorized subscale possibly comprised of two separate factors (4 factor solution). We followed existing guidelines such that we only retained factors with at least four items loading with a minimum $r \sim 0.5$ (DeVellis, 2016). We monitored fit statistics as guidelines, but did not decisively use them, in isolation, to reject or accept one factor structure over another. Importantly, our focus in study 1 was to determine whether the theorized subscales mapped onto distinct latent factors. Whether these latent factors are common factors in a correlated factor model, or group-factors in a bifactor model is not relevant for this study: the goal was to determine whether there was some kind of coarse latent factor structure that warranted the three theorized subscales.

Item Response Theory. After identifying the scale's factor structure, we conducted unidimensional, exploratory IRT analyses on item-level data from each subscale (corresponding to a latent factor) to assess the validity of each item. IRT is an item-level analysis whose fundamental goal is to estimate individuals' position on a latent trait (Θ) and relate latent trait scores to the likelihood of endorsing an item using exponential modeling (Morizot et al., 2007; Reise & Henson, 2003; Toland et al., 2017). In other words, IRT quantifies the association between a latent trait (e.g., a facet of affect labeling) and responses on a survey item, allowing us to assess the items that best measure a particular construct.

IRT analyses yield parameters for *location* (or *difficulty)* (denoted with *b*) — the latent trait level at which the probability of endorsing an item is .50 — and *discrimination* (denoted with *a*) — the strength of association between Θ and responses on a given item. Because our novel measure is scored along a polytomous 5-point scale, we have four location parameters (b_{1-4}), each indicating the position on Θ corresponding to a .5 chance of rating an item at a particular scale value or a higher value. From these parameters, response (association between Θ and an item), information (an item's ability to differentiate individuals at different trait levels conditional on Θ), and standard error of measurement (measurement error conditional on Θ) curves can be constructed for each item and aggregated into scale-level totals.

Study 2

Participants. We recruited 392 participants through a combination of a University of California Los Angeles (UCLA) subject pool and an in-lab study to complete the questionnaire via Qualtrics. After cleaning the data according to our exclusion criteria, our sample consisted of 293 participants. This sample was approximately: 90% 18-24 years old and 1% 25-34 years old, and 9% did not respond; 68% female, 23% male, and 9% did not respond; 23% Caucasian/White, 31% Asian/Asian American, 18% Hispanic/Latino, 3% Black/African American, and the remaining selected another identity or multiracial.

Confirmatory Factor Analysis. We considered two theoretically possible models that would describe the factor structure of the ALQ: (a) a correlated factors model under which each subscale loads onto a latent factor and latent factors corresponding to each subscale are allowed to correlate with each other; and (b) a bifactor model, in which all items load onto a *general* factor and a single group factor corresponding to each item's respective subscale. Cross-loadings between subscales are not allowed to exist in the bifactor model (i.e., are set to zero). The general factor is thought to explain the correlation between group factors. Correlated factor models are abundant in psychology, including affective science (e.g. Williams et al., 2018), and it is theoretically consistent with the affect labeling construct. Bifactor models are becoming increasingly common in psychology due to their purported ability to enhance content validity and thus adequately describe multidimensional constructs (Reise, 2012; Rodriguez et al., 2016). This model is also theoretically consistent with the construct of affect labeling, as one could argue there are generalized affect labeling propensities as well as specialized facets. Thus, a correlated factors model and bi-factor model were both appropriate to test here.

Following recommendation for evaluating model fit, we examined the following indices associated with these models: root mean square error of approximation (RMSEA), standardized root-mean-square residual (SRMR), Tucker-Lewis Index (TLI), and Comparative Fit Index (CFI) (Bentler & Bonett, 1980; Hu & Bentler, 1999; Steiger & Lind, 1980). SRMR and RMSEA values of 0.08 and under are considered acceptable, such that smaller values suggest less unexplained variance, and greater model fit (Browne & Cudeck, 1992). TLI and CFI values of 0.90 and higher are considered acceptable, with larger values suggesting a smaller difference between actual model and null model chi-squared values (Bentler & Bonett, 1980; Tucker & Lewis, 1973), and thus greater model fit (Byrne, 1994; Hu & Bentler, 1999). Research suggests that these indices are better metrics of model fit than chi-square likelihood ratio statistics, which compare the model in question to a perfect model (MacCallum, 1990), thereby running greater risk of falsely rejecting an acceptable model (Hakstian & Cattell, 1982; Humphreys & Montanelli, 1975). The bifactor model had additional considerations. Bifactor models are biased towards better fit, yet a seemingly well-fitting bifactor model does not necessarily indicate the bifactor model adequately describes the underlying factor structure (Bornovalova et al., 2020). Therefore, we examined the pattern of factor loadings between the general factor and the group factors. In order to be considered valid, the general factor must have loaded strongly (e.g., .55+) on (nearly) all of the items, and the group factors must also have had appropriately high loadings (.4+) (Bornovalova et al., 2020).

Item Response Theory. Item response theory was subsequently used to again confirm the final items. Notably, items at this stage were kept if they provided a unique advantage in aggregating composite scores, due to the additive nature of item response and information functions. For example, some items with relatively flatter information curves were kept because the range at the extremes of the latent trait were greater than items that might have had higher curves but a narrower range.

Study 3.

Participants. For study 3, we pooled across study 1 and 2 participants recruited through a combination of mTurk and the UCLA student population who completed additional questionnaires, for a total of 359 participants. After cleaning the data according to our exclusion criteria, our sample consisted of 227 participants. This sample was approximately: 7% 18-24 years old, 43% 25-34 years old, 22% 35-44 years old, 11% 45-54 years old, 4% 55-64 years old, <1% 65-74 years old, and the remaining did not respond; 45% female, 43s% male, and <1% transgender or did not respond; 66% Caucasian/White, 5% Asian/Asian American, 7% Hispanic/Latino, 7% Black/African American, and the remaining selected another identity or multiracial.

Analyses. To establish convergent and discriminant validity, we correlated participants' ALQ subscale scores (affective awareness, affect labeling capacity, affect labeling tendency) and ALQ total score with their responses to each outcome measure. ALQ total and subscale scores were determined by summing across the corresponding items. We used the spearman correlation method, pairwise deletion of cases, and the Holm method for multiple comparisons correction (i.e. a conservative method for controlling for family-wise error rate).

Results

Study 1.

Exploratory Factor Results. Results yielded an exploratory factor solution with three factors. All subscales had at least four items, but the tendency subscale had 3 additional items. The highest loadings for each subscale ranged from Λ = .57 - .87 (affect labeling tendency), Λ = .72 - .85 (affect labeling capacity), and Λ = .46 - .77 (affective awareness). Affect labeling tendency explained 46% of the common variance, affect labeling capacity explained 32% of the common variance, and affective awareness explained 22% of the common variance. The *TLI* was 0.954, *RMSEA* was .066, and $\chi^2(63) = 31.012$. Notably, while fit statistics for the four-factor solution were slightly better (*TLI* = 0.980, *RMSEA* = .044, $\chi^2(51) = 15.425$), the pattern of factor loadings did not indicate strong differentiation between the third and fourth factor, suggesting an over-factored solution. Fit statistics for the two-factor solution were demonstrably worse than both the 3 and 4 factor solutions (*TLI* = 0.903, *RMSEA* = .097, $\chi^2(51) = 93.851$), and some items exhibited a modest degree of cross-loading between the factors, suggesting an under-factored solution.

Item Response Theory Results. IRT analyses were conducted using the mirt() function from R's 'mirt' package (Chalmers, 2012). Exploratory unidimensional graded response models were fit on each subscale, using maximum likelihood (expectation-maximization algorithm; BFGS optimizer). Importantly, latent trait values were assumed to follow a normal distribution ($\Theta \sim N(0, 1)$). From Table 2, we can see that *b* values ranged from ~-2.0 to ~2.0 across the subscales, indicating that the items covered most of the range typically observed on latent variables (-2.5 to 2.5) in this field of research. The *a* parameters ranged between 1.6 and 4.7, indicating very high discrimination relative to other similar measures in this area of research (Morizot et al., 2007). Supplementary Figure 2 shows the aggregate item response curves (top set) and aggregate test information by standard error curves (bottom set) for each subscale. Visual inspection of this figure indicates that the subscales are reasonably discriminating and provide information across a wide range of Θ. By contrast, they also show that each scale appears to be ill-suited for measuring individuals who score exceptionally high or low on the latent trait (more than 2 standard deviations above/below the mean). After inspecting each item's response and information curves, we decided to drop the extra items from the tendency subscale because they evinced poor psychometric properties, leaving 4 items for each subscale (12 items total).

Study 2.

Confirmatory Factor Analysis Results. Results of the models are displayed in Tables 3 (correlated factors) and 4 (bifactor). Both models fit the data well (correlated factors: $CFI = 0.963$, $TLI = 0.952$, $\chi^2(51) = 91.48$, $RMSEA = 0.053$, $SRMR = 0.049$; bifactor: $CFI = 0.976$, $TLI = 0.961$, $\chi^2(33) = 57.39$, $RMSEA = 0.050$, $SRMR = 0.040$).

While the bifactor model evinced better fit across our four metrics, the pattern of factor loadings suggests that the bifactor model is not suitable. Specifically, the highest loadings for the general factor all came from the capacity subscale and the rest of the general factor loadings were quite low (general factor loadings for capacity subscale ranged from .62 - .79; all other general factor loadings ranged from .29 - .46). Similarly, the group factor loadings for capacity were quite low (ranging from .17 - .40) whereas the group factor loadings for awareness (ranging from .50 - .60) and tendency were relatively high (ranging from .55 - .64). By contrast, the factor loadings in the correlated factors model ranged from .60 - .78 (excluding one loading at .39). Therefore, we concluded that the factor structure of the ALQ most closely matched a correlated factors solution.

We subsequently sought to test whether each subscale's factor structure, in isolation, was unidimensional. For awareness, the loadings ranged from $\Lambda = 0.38 – 0.78$ ($\chi^2(2) = 14.313$; $TLI = 0.829$; CFI = 0.943; $RMSEA = 0.146$, $SRMR = 0.048$). For capacity, the loadings ranged from $\Lambda = 0.66 – 0.78$ ($\chi^2(2) = 2.561$; $TLI = 0.995$; $CFI = 0.998$; $RMSEA = 0.031$, $SRMR = 0.014$). For tendency, the loadings ranged from $\Lambda = 0.62 – 0.76$ ($\chi^2(2) = 1.104$; $TLI = 1.008$; $CFI = 1.000$; $RMSEA = 0.000$, $SRMR = 0.010$). These results suggest the three facets of affect labeling (awareness, capacity, tendency) measured by novel ALQ are indeed unidimensional in isolation.

In terms of reliability, ω-total = 0.87, suggesting that, overall, the ALQ is a reliable multi-dimensional composite. Individually, each subscale evinced acceptable levels of reliability, indexed by Guttman's Lambda 6 (G6): G6-awareness = 0.65, G6-capacity = 0.77; G6-tendency = 0.75).

Item Response Theory Results. Confirmatory IRT analyses were conducted using the mirt() function from R's 'mirt' package (Chalmers, 2012). Confirmatory graded response models were fit on each subscale, using maximum likelihood via an expectation-maximization algorithm (BFGS optimizer). Latent trait values were again assumed to follow a normal distribution ($\Theta \sim N(0, 1)$). From Table 2, we can see that b values ranged from ~-2.5 to ~2.5 across the subscales, indicating that the items covered most of the range typically observed on latent variables (-2.5 to 2.5) in this field of research. The *a* parameters ranged between 0.9 and 2.7, indicating very high discrimination relative to other similar measures in this area of research (Morizot et al., 2007). Supplementary Figure 3 shows the aggregate item response curves (top set) and aggregate test information by standard error curves (bottom set) for each subscale. Visual inspection indicates that the subscales are reasonably discriminating and provide information across a wide range of Θ. By contrast, plots also show that each scale appears to be ill-suited for measuring individuals who score exceptionally high or low on the latent trait (more than 2 standard deviations above/below the mean). Notably, certain items from each subscale did not necessarily contain ideal item response or information curves, but certain features of the curves (e.g., range on the information curve) were better than the other items in the set. Given the additive nature of IRT properties, we retained the items to bolster the subscale composite.

Study 3.

Convergent and Discriminant Relationships Results. We report spearman's correlation coefficients, corrected for multiple comparisons, between the three ALQ subscales, as well as the ALQ total score, with the relevant validity measures (Table 5).

We describe significant correlations between 0.2 and 0.4 to be weak, between 0.4 and 0.6 to be moderate, and above 0.6 to be strong.

In terms of personality, greater affect labeling was weakly associated with lower social anxiety via the SIAS, lower neuroticism via the BFI, greater extraversion, agreeableness, conscientiousness, and openness via the BFI, and greater self-esteem via the RSE. It was not related to social desirability via the MCSDS.

In terms of empathic tendency, greater affect labeling was weakly associated with greater perspective-taking via QCAE, lower peripheral and proximal responsivity via QCAE, higher empathic concern via IRI, and greater positive empathy via PES. It had no relationship to emotion contagion or simulation via QCAE, or to personal distress, perspective-taking, or fantasy via IRI.

In terms of affect/mood, greater affect labeling was weakly associated with higher positive affect via PANAS and SAIS, lower negative affect via PANAS, lower depression via BDI, and lower anxiety via GAD. It had no relationship to negative affect via SAIS.

In terms of emotion regulation, greater affect labeling was strongly associated with lower suppression via ERQ, and greater emotional social support via COPE. It was moderately associated with greater reappraisal via ERQ, and greater positive reinterpretation and growth, venting, and instrumental social support via COPE.

In terms of expressivity, greater affect labeling use was strongly associated with lower difficulty describing feelings via TAS, and greater mindfulness-describing via FFMQ. It was also moderately associated with greater negative and positive expressivity via BEQ, impulse strength via BEQ, and lower difficulty identifying feelings and externally oriented thinking via TAS.

Discussion

In the present collection of studies, we developed and validated a questionnaire for measuring individual differences in affect labeling: the ALQ. This questionnaire allowed us to (a) examine the different components or facets of affect labeling as an emotion regulation strategy; (b) assess how affect labeling relates to social behavior, relationships, and emotion regulation more broadly; and (c) provide a tool for measuring variability in affect labeling in future research. We identified three factors contributing to affect labeling: (i) affective awareness; (ii) affect labeling tendency; and (iii) affect labeling capacity. By decomposing affect labeling into these three facets, we gain insight into the different processes that fold into labeling emotions in everyday life and identify several points of intervention for improving regulatory outcomes. Additionally, we found that greater affect labeling was associated with increased emotional expressivity, social support, and use of other emotion regulation strategies (i.e. reappraisal, coping). Thus, our results provide insight into how affect labeling tracks with various socio-emotional metrics of well-being, establishing a broader picture of how different regulatory processes interact with each other and everyday behaviors.

These data shed light on the psychological structure underlying affect labeling, while validating a measure of affect labeling for future use. Future research might use this scale to explore how variability on the ALQ tracks with well-being on a day-to-day basis through ecological momentary assessments or daily diary studies (Ford et al., 2018). Additionally, it would be interesting to investigate how the ALQ tracks with linguistic markers in free-flowing text through natural language processing or other text analysis approaches (Fan et al., 2019). Finally, one might unpack how the different facets of affect labeling interact with each other across development, for example by exploring how the emergence of affective awareness influences the tendency and capacity to use language to regulate one's emotions (Nook et al., 2017).

Importantly, research suggests that different ways of labeling – for example labeling features of emotional stimuli as opposed to one's own emotional states – are associated with different regulatory outcomes (McRae et al., 2010). Our questionnaire focuses on labeling one's own emotional states since this is likely a form of affect labeling that people commonly employ in everyday life. Future work can extend this questionnaire to consider different ways of affect labeling and assess the degree of overlap between these approaches. Additionally, while we assess how *often* and how *well* people label their emotions, we do not consider how *fine-grained* peoples' emotion labels are, or the variety of emotion words they use to identify and process their feelings. Research suggests that such emotion differentiation can be an important predictor for regulatory outcomes (Barrett et al., 2001; Kircanski et al., 2012). Thus, emotion knowledge and associated introspection processes can be considered alongside the broader features of affect labeling presented in this research.

Ultimately, this work provides a jumping point to explore several questions surrounding how affect labeling unfolds in daily life and how it interacts with different social and emotional behaviors. The present work not only motivates such future research, but also provides a framework for understanding the different components that contribute to affect labeling as an emotion regulation strategy, alongside a validated measurement tool that can facilitate our ability to more widely examine affect labeling across contexts, populations, and lab paradigms.

Statements and Declarations

The authors declare no conflicts of interest. The material presented in this manuscript has not been published and is not under consideration elsewhere. All procedures were approved by the local IRB committee (IRB#15-000693) and the research was performed to ethical standards as laid down in the 1964 Declaration of Helsinki. Informed consent was obtained from all participants.

Author Contributions

This research was planned by JBT and MDL. Data collection and preliminary analyses were conducted by JBT. Analyses were refined and results were prepared for publication by RSS and JFGM. The manuscript was drafted by RSS and JFGM, and revised by all authors.

Data Sharing and Open Practices Statement

This data was collected in 2016 without an explicit data sharing policy specified in the IRB approved materials. Thus, we do not have permission to publicly share data. However, we have made our analyses publicly available via OSF (osf.io/fj3cz), and our data has been uploaded to OSF (time-stamped prior to submission of this manuscript) and can be shared with researchers upon request. Access to these materials can been shared with reviewers via a link that allows sharing without compromising reviewers' anonymity. These studies were not pre-registered.

-

Tables

Table 1. Full list of items tested to develop Affect Labeling Questionnaire (ALQ)

Affective Awareness	
ALQ Item 27	I frequently know what emotion I am feeling as I feel it.
ALQ Item 32	Often I know I am feeling an emotion, but cannot figure out exactly what it is.
ALQ Item 34	I know how things tend to affect me emotionally.
ALQ Item 37	When I am upset, I can pinpoint what is bothering me.
Affect Labeling Tendency	
ALQ Item 1	Putting my feelings into words is an important step for me in managing my emotions.
ALQ Item 5	Using my words to describe my feelings makes me feel them less strongly.
ALQ Item 7	When I want to feel less upset, I will describe how I am feeling.
ALQ Item 10	Labeling my feelings with words helps me understand them.
ALQ Item 12	I get my feelings 'off my chest' by using words to describe them.
ALQ Item 15	I lessen the impact my emotions have on me by expressing them in words.
ALQ Item 18	I control my emotions by putting my feelings into words.
Affect Labeling Capacity	
ALQ Item 3	It is easy for me to find an appropriate word to capture how I am feeling.
ALQ Item 9	I can usually describe how I feel in great detail.
ALQ Item 16	I can find a way to describe my feelings in words even when the feeling is very strong.
ALQ Item 19	It takes a lot of effort to convert my feelings into words.

Note. Bolded headings indicate subscale names. Item numbers correspond to original numbering in our questionnaires. Items highlighted in grey were not retained in the final scale.

Table 2. Parameter estimates of exploratory and confirmatory IRT analyses based on subscale

Item	*a*		*b1*		*b2*		*b3*		*b4*	
	Exp	Con	Exp	Con	Exp	Con	Exp	Con	Exp	Con
Awareness										
ALQ Item 27	3.6	2.6	-2.2	-2.4	-1.7	-0.8	-1.1	-0.0	0.6	1.7
ALQ Item 32	2.3	1.6	-2.8	-2.5	-1.7	-0.7	-1.1	0.3	0.6	2.4
ALQ Item 34	2.5	0.9	-3.2	-5.4	-2.1	-3.0	-1.5	-1.3	0.8	2.9
ALQ Item 37	2.8	2.0	-1.8	-2.9	-1.2	-1.1	0.8	-0.0	-	2.0
Capacity										
ALQ Item 3	4.3	2.4	-2.0	-1.7	-1.1	-0.3	-0.5	0.5	0.8	2.0
ALQ Item 9	4.5	2.7	-1.7	-1.7	-0.9	-0.1	-0.2	0.5	0.9	1.7
ALQ Item 16	4.2	2.0	-2.0	-2.2	-1.0	-0.7	-0.5	0.4	1.1	2.1
ALQ Item 19	2.5	1.8	-1.9	-1.7	-1.0	-0.2	-0.4	0.6	1.0	2.6
Tendency										
ALQ Item 1	2.4	-	-1.7	-	-0.7	-	-0.2	-	1.5	-
ALQ Item 5	1.5	-	-1.7	-	-0.3	-	0.7	-	2.5	-
ALQ Item 7	3.4	2.3	-1.5	-2.2	-0.3	-0.5	0.3	0.1	1.8	2.0
ALQ Item 10	2.9	-	-1.6	-	-0.7	-	-0.1	-	1.4	-
ALQ Item 12	3.0	2.3	-1.4	-2.2	-0.6	-0.8	-0.2	-0.2	1.5	1.6
ALQ Item 15	3.6	1.8	-1.3	-2.4	-0.3	-0.6	0.4	0.4	1.7	2.6
ALQ Item 18	5.4	2.5	-1.4	-2.1	-0.2	-0.4	0.3	0.5	1.9	2.3

Note. ‘Exp’ column contains exploratory results, ‘Con’ column contains confirmatory results; *a* refers to the discrimination parameter; b_j refers to the location/difficulty parameters. Items with hashed entries in the ‘con’ column are those which were dropped from Study 1 to Study 2; the hashed entry for *b4* in item 37’s ‘exp’ column reflects the fact that the highest response option was not endorsed by any participants in that sample. See Table 1 for items corresponding to each question number.

Table 3. Confirmatory factor analysis results for the three correlated factors model

Item	Awareness	Capacity	Tendency
ALQ Item 27	0.76 (.059) [.65 - .88]	-	-
ALQ Item 32	0.60 (.061) [.48 - .72]	-	-
ALQ Item 34	0.69 (.060) [.57 - .81]	-	-
ALQ Item 37	0.39 (.064) [.27 - .52]	-	-
ALQ Item 3	-	0.75 (.054) [.65 - .86]	-
ALQ Item 9	-	0.77 (.054) [.67 - .88]	-
ALQ Item 16	-	0.71 (.055) [.60 - .82]	-
ALQ Item 19	-	0.65 (.057) [.54 - .76]	-
ALQ Item 7	-	-	0.71 (.057) [.60 - .82]
ALQ Item 12	-	-	0.73 (.056) [.62 - .84]
ALQ Item 15	-	-	0.62 (.059) [.50 - .73]
ALQ Item 18	-	-	0.77 (.055) [.66 - .88]

Note. Factor correlations: Φ(awareness, capacity) = 0.629, Φ(capacity, tendency) = 0.500, Φ(awareness, tendency) = 0.24; $\chi^2(51)$: 91.48, CFI: 0.963, TLI: 0.952, RMSEA: 0.053, SRMR: 0.049

Table 4. Confirmatory factor analysis results for the bi-factor model

Item	G-Factor	Awareness	Capacity	Tendency
ALQ Item 27	0.43 (.069) [.29, .56]	0.60 (.080) [.45, .76]	-	-
ALQ Item 32	0.40 (.066) [.27, .53]	0.50 (.074) [.35, .64]	-	-
ALQ Item 34	-	-	-	-
ALQ Item 37	0.46 (.069) [.32, .59]	0.51 (.076) [.36, .66]	-	-
ALQ Item 3	0.71 (.068) [.60, .83]	-	0.17 (.109) [-.05, .38]	-
ALQ Item 9	0.73 (.096) [.55, .92]	-	0.40 (.316) [-.21, 1.02]	-
ALQ Item 16	0.79 (.088) [.62, .97]	-	-0.17 (.384) [-.92, .59]	-
ALQ Item 19	0.62 (.068) [.49, .75]	-	0.19 (.117) [-.04, .42]	-
ALQ Item 7	0.32 (.063) [.20, .44]	-	-	0.64 (.059) [.53, .76]
ALQ Item 12	0.35 (.063) [.23, .47]	-	-	0.64 (.059) [.53, .76]
ALQ Item 15	0.29 (.068) [.15, .42]	-	-	0.55 (.063) [.42, .67]
ALQ Item 18	0.43 (.063) [.31, .56]	-	-	0.63 (.058) [.51, .74]

Note. Model would not converge unless item 34 was removed; $\chi^2(33)$: 57.39, CFI: 0.976, TLI: 0.961, RMSEA: 0.050, SRMR: 0.040

References

Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment*, *13*(1), 27–45. https://doi.org/10.1177/1073191105283504

Bagby, R. M., Parker, J. D. A., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia scale-I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*(1), 23–32. https://doi.org/10.1016/0022-3999(94)90005-1

Bainbridge, T. F., Ludeke, S. G., & Smillie, L. D. (2022). Evaluating the Big Five as an organizing framework for commonly used psychological trait scales. *Journal of Personality and Social Psychology*, *122*(4), 749–777. https://doi.org/10.1037/PSPP0000395

Barrett, L. F., Gross, J., Christensen, T. C., & Benvenuto, M. (2001). Knowing what you're feeling and knowing what to do about it: Mapping the relation between emotion differentiation and emotion regulation. *Cognition and Emotion*, *15*(6), 713–724. https://doi.org/10.1080/02699930143000239

Beck, A. T., Steer, R. A., & Carbin, M. G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clinical Psychology Review*, *8*(1), 77–100. https://doi.org/10.1016/0272-7358(88)90050-5

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*(3), 588–606. https://doi.org/10.1037/0033-2909.88.3.588

Bornovalova, M. A., Choate, A. M., Fatimah, H., Petersen, K. J., & Wiernik, B. M.

(2020). Appropriate Use of Bifactor Analysis in Psychopathology Research: Appreciating Benefits and Limitations. In *Biological Psychiatry*. https://doi.org/10.1016/j.biopsych.2020.01.013

Browne, M. W., & Cudeck, R. (1992). Alternative Ways of Assessing Model Fit. *Sociological Methods & Research*, *21*(2), 230–258. https://doi.org/10.1177/0049124192021002005

Burklund, L. J., David Creswell, J., Irwin, M. R., & Lieberman, M. D. (2014). The common and distinct neural bases of affect labeling and reappraisal in healthy adults. *Frontiers in Psychology*, *5*(MAR), 1–10. https://doi.org/10.3389/fpsyg.2014.00221

Byrne, B. M. (1994). Testing for the Factorial Validity, Replication, and Invariance of a Measuring Instrument: A Paradigmatic Application Based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*, *29*(3), 289–311. https://doi.org/10.1207/s15327906mbr2903_5

Carver, C., Scheier, M., The, J. W.-M. I. D. for, & 2013, U. (2013). COPE inventory. *Measurement Instrument Database for the Social Science.*

Chalmers, R. P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software*, *48*, 1–29. https://doi.org/10.18637/jss.v048.i06

Comrey, A. L. (1988). Factor-Analytic Methods of Scale Development in Personality and Clinical Psychology. *Journal of Consulting and Clinical Psychology*, *56*(5), 754–761. https://doi.org/10.1037/0022-006X.56.5.754

Constantinou, E., Van Den Houte, M., Bogaerts, K., Van Diest, I., & Van den Bergh, O.

(2014). Can words heal? Using affect labeling to reduce the effects of unpleasant cues on symptom reporting. *Frontiers in Psychology*, *5*(JUL), 807. https://doi.org/10.3389/FPSYG.2014.00807/BIBTEX

Crowne, D. P., & Marlowe, D. (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology*, *24*(4), 349–354. https://doi.org/10.1037/h0047358

Davis, M., Fabes, R., Eisenberg, N., Psychology, L. E.-D., & 1993, U. (1993). *Interpersonal reactivity index: Simplified and shortened*.

DeVellis, R. (2016). *Scale development: Theory and applications*. Sage publications. https://books.google.com/books?hl=en&lr=&id=9yfYCwAAQBAJ&oi=fnd&pg=PP1&dq=DeVellis,+R.+F.+(2012).+Scale+development:+Theory+and+applications+(Vol.+26).+Thousand+Oaks,+CA:+Sage+Publications.&ots=HoJESmYbhp&sig=9WaQZcmxcylyEHv7JYfPzyccvpo

Fan, R., Varol, O., Varamesh, A., Barron, A., van de Leemput, I. A., Scheffer, M., & Bollen, J. (2019). The minute-scale dynamics of online emotions reveal the effects of affect labeling. *Nature Human Behaviour*, *3*(1), 92–100. https://doi.org/10.1038/s41562-018-0490-5

Ford, B. Q., Lam, P., John, O. P., & Mauss, I. B. (2018). The psychological health benefits of accepting negative emotions and thoughts: Laboratory, diary, and longitudinal evidence. *Journal of Personality and Social Psychology*, *115*(6), 1075–1092. https://doi.org/10.1037/pspp0000157

Geuens, M., & De Pelsmacker, P. (2002). Developing a Short Affect Intensity Scale. *Psychological Reports*, *91*(2), 657–670. https://doi.org/10.2466/pr0.2002.91.2.657

Greenburg, L. S. (2001). *Emotion-focused therapy: Coaching clients to work through their feelings*. American Psychological Association.

Gross, J. J. (1998). Sharpening the Focus: Emotion Regulation, Arousal, and Social Competence. *Psychological Inquiry*, *9*(4), 287–290. https://doi.org/10.1207/s15327965pli0904_8

Gross, J. J., & John, O. P. (1997). Revealing Feelings: Facets of Emotional Expressivity in Self-Reports, Peer Ratings, and Behavior. *Journal of Personality and Social Psychology*, *72*(2), 435–448. https://doi.org/10.1037/0022-3514.72.2.435

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

Gross, J. J., Richards, J. M., & John, O. P. (2006). Emotion Regulation in Everyday Life. In *Emotion regulation in couples and families: Pathways to dysfunction and health.* (pp. 13–35). https://doi.org/10.1037/11468-001

Hakstian, A. R., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research*, *17*(2), 193–219. https://doi.org/10.1207/s15327906mbr1702_3

Hariri, A., Bookheimer, S., & Mazziotta, J. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport*, *11*, 43–48. https://search.proquest.com/openview/f937d0207c67fcbc58c94c48dff1836f/1?pq-origsite=gscholar&cbl=18750&diss=y&casa_token=o4ZVBzczq3UAAAAA:hxhUTUwdnqN3CmO26zvkoF1RGf8-wbISxDc4ljC0XG9oKXTXN6_eeEFa63TNqw_-

BMIyQTh0DQ

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

Humphreys, L. G., & Montanelli, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, *10*(2), 193–205. https://doi.org/10.1207/s15327906mbr1002_5

Jaso, B. A., Kraus, N. I., & Heller, A. S. (2021). Identification of careless responding in ecological momentary assessment research: From posthoc analyses to real-time data monitoring. *Psychological Methods*. https://doi.org/10.1037/met0000312

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of Personality: Theory and Research*, *2*(510), 102–138. https://books.google.com/books?hl=en&lr=&id=iXMQq7wg-qkC&oi=fnd&pg=PA102&dq=John,+O.+P.,+%26+Srivastava,+S.+(1999).+The+Big-Five+trait+taxonomy:+History,+measurement,+and+theoretical+perspectives+.+In+L.+A.+Pervin+%26+O.+P.+John+(Eds.),+Handbook+of+person

Kircanski, K., Lieberman, M. D., & Craske, M. G. (2012). Feelings into words: Contributions of language to exposure therapy. *Psychological Science*, *23*(10), 1086–1091. https://doi.org/10.1177/0956797612443830

Koole, S. L., & Rothermund, K. (2011). “I feel better but I don’t know why”: The psychology of implicit emotion regulation. *Cognition and Emotion*, *25*(3), 389–399. https://doi.org/10.1080/02699931.2010.550505

Kring, A. M., Smith, D. A., & Neale, J. M. (1994). Individual Differences in Dispositional Expressiveness: Development and Validation of the Emotional Expressivity Scale. *Journal of Personality and Social Psychology*, *66*(5), 934–949. https://doi.org/10.1037/0022-3514.66.5.934

Leweke, F., Leichsenring, F., Kruse, J., & Hermes, S. (2011). Is alexithymia associated with specific mental disorders? *Psychopathology*, *45*(1), 22–28. https://doi.org/10.1159/000325170

Lieberman, M. D., Inagaki, T. K., Tabibnia, G., & Crockett, M. J. (2011). Subjective Responses to Emotional Stimuli During Labeling, Reappraisal, and Distraction. *Emotion*, *11*(3), 468–480. https://doi.org/10.1037/a0023503

MacCallum, R. C. (1990). The need for alternative measures of fit in covariance structure modeling. *Multivariate Behavioral Research*, *25*(2), 157–162. https://doi.org/10.1207/s15327906mbr2502_2

Mattick, R. P., & Clarke, J. C. (1998). Development and validation of measures of social phobia scrutiny fear and social interaction anxiety. *Behaviour Research and Therapy*, *36*(4), 455–470. https://doi.org/10.1016/S0005-7967(97)10031-6

Mcrae, K. (2013). Emotion Regulation Frequency and Success: Separating Constructs from Methods and Time Scale. *Social and Personality Psychology Compass*, *7*(5), 289–302. https://doi.org/10.1111/spc3.12027

McRae, K., Taitano, E. K., & Lane, R. D. (2010). The effects of verbal labelling on psychophysiology: Objective but not subjective emotion labelling reduces skin-conductance responses to briefly presented pictures. *Cognition and Emotion*, *24*(5), 829–839. https://doi.org/10.1080/02699930902797141

Morelli, S. A., Lieberman, M. D., & Zaki, J. (2015). *UCLA UCLA Previously Published Works Title The emerging study of positive empathy*. https://doi.org/10.1111/spc3.12157

Morelli, S. A., Ong, D. C., Makati, R., Jackson, M. O., & Zaki, J. (2017). Empathy and well-being correlate with centrality in different social networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(37), 9843–9847. https://doi.org/10.1073/pnas.1702155114

Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward Modern Psychometrics. *Handbook of Research Methods in Personality Psychology*, 407–423. https://books.google.com/books?hl=en&lr=&id=-VTvN3aPw8sC&oi=fnd&pg=PA407&dq=Morizot,+J.,+Ainsworth,+A.+T.,+%26+Reise,+S.+P.+(2007).+Toward+modern+psychometrics.+Handbook+of+research+methods+in+personality+psychology,+407-423.++&ots=Fyr8IO5FUD&sig=7cCcJaOY

Niles, A. N., Craske, M. G., Lieberman, M. D., & Hur, C. (2015). Affect labeling enhances exposure effectiveness for public speaking anxiety. *Behaviour Research and Therapy*, *68*, 27–36. https://doi.org/10.1016/j.brat.2015.03.004

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour*, *1*(12), 881–889. https://doi.org/10.1038/s41562-017-0238-7

Ogrodniczuk, J. S., Piper, W. E., & Joyce, A. S. (2011). Effect of alexithymia on the process and outcome of psychotherapy: A programmatic review. *Psychiatry*

Research, *190*(1), 43–48. https://doi.org/10.1016/j.psychres.2010.04.026

Pennebaker, J. W. (1993). PUTTING STRESS INTO WORDS: HEALTH, LINGUISTIC, AND THERAPEUTIC IMPLICATIONS. In *Behaviour Research and Therapy* (Vol. 31, Issue 6).

Reise, S. P. (2012). The Rediscovery of Bifactor Measurement Models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., & Henson, J. M. (2003). A Discussion of Modern Versus Traditional Psychometrics As Applied to Personality Assessment Scales. *Journal of Personality Assessment*, *81*(2), 93–103. https://doi.org/10.1207/S15327752JPA8102_01

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

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

Rosenberg, M. (1979). *Conceiving the Self.* Basic Books.

Sahi, R. S., Guassi Moreira, J. F., Torre, J. B., & Lieberman, M. D. (2022). *Affect Labeling Questionnaire*. Open Science Framework. https://osf.io/jctru/

Silvers, J. A., & Guassi Moreira, J. F. (2019). Capacity and tendency: A neuroscientific framework for the study of emotion regulation. *Neuroscience Letters*, *693*, 35–39. https://doi.org/10.1016/J.NEULET.2017.09.017

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

Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the numbers of factors. *Annual Meeting of the Psychometric Society*. https://ci.nii.ac.jp/naid/10012870999/

Tinsley, H. E. A., & Tinsley, D. J. (1987). Uses of Factor Analysis in Counseling Psychology Research. *Journal of Counseling Psychology*, *34*(4), 414–424. https://doi.org/10.1037/0022-0167.34.4.414

Toland, M. D., Sulis, I., Giambona, F., Porcu, M., & Campbell, J. M. (2017). Introduction to bifactor polytomous item response theory analysis. *Journal of School Psychology*, *60*, 41–63. https://doi.org/10.1016/j.jsp.2016.11.001

Torre, J. B., & Lieberman, M. D. (2018). Putting Feelings Into Words: Affect Labeling as Implicit Emotion Regulation. *Emotion Review*, *10*(2), 116–124. https://doi.org/10.1177/1754073917742706

Troy, A. S., Ford, B. Q., McRae, K., Zarolia, P., & Mauss, I. B. (2017). Change the things you can: Emotion regulation is more beneficial for people from lower than from higher socioeconomic status. *Emotion*, *17*(1), 141–154. https://doi.org/10.1037/emo0000210

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*(1), 1–10. https://doi.org/10.1007/BF02291170

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

3514.54.6.1063

Williams, W. C., Morelli, S. A., Ong, D. C., & Zaki, J. (2018). Interpersonal emotion regulation: Implications for affiliation, perceived support, relationships, and well-being. *Journal of Personality and Social Psychology*, *115*(2), 224–254. https://doi.org/10.1037/pspi0000132

Zaki, J. (2020). Integrating Empathy and Interpersonal Emotion Regulation. *Https://Doi.Org/10.1146/Annurev-Psych-010419-050830*, *71*, 517–540. https://doi.org/10.1146/ANNUREV-PSYCH-010419-050830