American Psychological Association

Journal of Experimental Psychology:
Learning, Memory, and Cognition

ISSN: 0278-7393

http://dx.doi.org/10.1037/xlm0000841

The Emotional Recall Task: Juxtaposing Recall and Recognition-Based Affect Scales

Ying Li
Max Planck Institute for Human Development, Berlin, Germany

Annasya Masitah and Thomas T. Hills
University of Warwick

Existing affect scales typically involve recognition of emotions from a predetermined emotion checklist. However, a recognition-based checklist may fail to capture sufficient breadth and specificity of an individual's recalled emotional experiences and may therefore miss emotions that frequently come to mind. More generally, how do recalled emotions differ from recognized emotions? To address these issues, we present and evaluate an affect scale based on recalled emotions. Participants are asked to produce 10 words that best described their emotions over the past month and then to rate each emotion for how often it was experienced. We show that average weighted valence of the words produced in this task, the Emotional Recall Task (ERT), is strongly correlated with scales related to general affect, such as PANAS, Ryff's Scales of Psychological Well-being, the Satisfaction with Life Scale, Depression Anxiety and Stress Scales, and a few other related scales. We further show that the Emotional Recall Task captures a breadth and specificity of emotions not available in other scales but that are nonetheless commonly reported as experienced emotions. We test a general version of the ERT (the ERT general) that is language neutral and can be used across cultures. Finally, we show that the ERT is valid in a test-retest paradigm. In sum, the ERT measures affect based on emotion terms relevant to an individual's idiosyncratic experience. It is consistent with recognition-based scales, but also offers a new direction toward enriching our understanding of individual differences in recalled and recognized emotions.

Keywords: emotion, free recall, memory, well-being

"How people recall and estimate their moods is an important component of people's self-concepts and how they conceptualize their lives" (Thomas & Diener, 1990, p. 292).

New affect scales often originate when limitations are identified in existing affect scales (Lucas, Diener, & Larsen, 2003; McDowell & Praught, 1982; Thompson, 2007; Watson, Clark, & Tellegen, 1988). Because all existing affect scales are recognition-based, previously identified limitations have often involved complaints that the list of terms on which participants base their emotional judgments "do not capture the range of people's experienced emotions" (Diener et al., 2009). In other words, the emotions that people experience are not those on the recognition scale. Recognition scales require that people reinterpret their emotions in relation to emotions they may not have experienced or that may not readily come to mind in day-to-day experience. Though existing recognition-based scales have attempted to deal with this problem using a variety of elegant and principled methods (e.g., Diener et al., 2010), a scale based on recalled emotions would offer an important point of comparison across as well as a more enriched indication of people's affect across a broad range of emotions. Moreover, such a scale, by revealing where it is not predictive of people's recognized emotions, would offer insight into how emotions are accessed and the dimensionality of recalled versus recognized emotions. Perhaps most important, however, a recall-based scale might simply be unreliable or fail to capture affect measured by recognition-based scales. These too would be important findings and lead to useful new questions about recognized versus recalled emotions. With this in mind, in this article we introduce a recall-based affect scale, the Emotional Recall Task, and compare it with a number of currently popular recognition-based affect scales. Before introducing the Emotional Recall Task, we first briefly discuss the need for a recall-based emotional scale motivated by the history of research on emotional dimensionality. We then explain the potential differences in the memory literature between recall and recognition as they apply to emotions.

The Specificity and Breadth of Emotional Dimensions

A brief historical overview of the many approaches to dimensionalizing emotional experience shows two things: this is a long-standing topic and there has been little historical consensus on exactly what and how many dimensions are important. The history

Editor's Note. Marc Brysbaert handled this manuscript as a Guest Associate Editor.—ASB

Ying Li, Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany; Annasya Masitah and Thomas T. Hills, Department of Psychology, University of Warwick.

Ying Li and Annasya Masitah are joint first authors of this article.

Study 2 was preregistered on OSF: https://osf.io/48qwx.

Correspondence concerning this article should be addressed to Ying Li, Center for Adaptive Rationality, Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany. E-mail: li@mpib-berlin.mpg.de

of speculations about human emotions dates back to at least Aristotle's *Nicomachean Ethics* (Broadie & Rowe, 2002), which lists 11 different emotions, including 'pity' and 'emulation' (the act of copying another individual's behavior). Darwin (1872), taking an evolutionary approach, attempted to classify emotions in relation to their adaptive value, and in addition to high and low valence emotions, included such dimensions as surprise, meditation, and shyness. Looking across cultures, Ekman (1992) proposed a set of "natural kinds" for emotions, similar to that of Darwin's, which included anger, fear, disgust, sadness, happiness, and surprise. Wundt (1905) proposed that emotions largely fell along three dimensions: valence, arousal, and tension. Osgood, Suci, and Tannenbaum (1957) found further support for a similar three dimensions of meaning (pleasantness, control, and arousal) using what is now called the semantic differential, which used a factor analysis of a large number of scales evaluating people's responses to various items. Russell (1980) proposed a further reduction in this dimensionality with the circumflex model of emotion, which suggests emotions are distributed in a two-dimensional space, with arousal and valence as independent dimensions. However, a more recent approach based on principal components analysis found evidence for a further emotional dimension, unpredictability (Fontaine, Scherer, Roesch, & Ellsworth, 2007).

The discrepancies and agreements across this diversity of emotional primitives potentially stem from a number of sources. One key source is emotional granularity (Tugade, Fredrickson, & Feldman Barrett, 2004). Emotional granularity refers to an individual's ability to discriminate between different emotions. For example, a person with high (as opposed to low) emotional granularity would tend to express their emotions using more distinct words, like *exuberant* (as opposed to *happy*). A key individual difference identified in previous work is that people with less emotional granularity are more likely to focus on valence and may simply report degree of positivity or negativity, such as "very happy" (see Russell, 2003; Russell & Barrett, 1999). *In other words, the way people experience and define affect systematically varies across individuals, and this may lead to difficulties in establishing a consensus on the basic dimensionality of emotion.*

If people experience emotional dimensionality in different ways, this throws existing affective scales into question. This is because the most popular approach to measuring emotions is to ask people about their ability to recognize how much they felt each of a set of emotions provided on a predetermined checklist in the recent past. Such recognition-based scales make two overarching assumptions. The first is that people will be able to identify their own emotions in relation to the words provided in the checklist. This we call the assumption of emotional *specificity*. The second is that the checklist will adequately cover a person's experience of emotions. This we call emotional *breadth*.

To put the ideas of emotional specificity and breadth in context, let us consider what is arguably the most widely used recognition-based checklist, the Positive and Negative Affect Schedule (PANAS; for review see: Diener et al., 2010). Because PANAS presents emotional stimuli, it necessarily frames respondents' emotional experiences in relation to emotions which may or may not be specific to the emotions respondents actually felt (e.g., Diener et al., 2009). PANAS focuses on a closed set of words, some of which are not generally considered as emotions (*strong*, *alert*, *inspired*, *determined*, and *active*), while common emotion words (*happy* and *sad*) are excluded. Four of the terms in PANAS focus on anxiety, and there are few low-arousal terms (Diener et al., 2009). Thus, PANAS's breadth is potentially narrower than the full emotional range that respondents experience.

Though PANAS is only one example, its potential problems of breadth and specificity are likely to be common to recognition-based scales more generally. Moreover, it may also suffer from order and priming effects (e.g., Hansen & Shantz, 1995; Wang, Busemeyer, Atmanspacher, & Pothos, 2013). For example, being reminded of a forgotten emotion may make that emotion more salient than it otherwise would be in day-to-day experience.

One way to overcome these problems is to allow individuals to freely recall emotions they have recently experienced (e.g., in the last month). Because emotional terms are highly salient in free-recall tasks (Altarriba & Bauer, 2004) and emotional experiences are often remembered with better quality (Kensinger & Schacter, 2008), the experience of an emotion may be easily recalled. Moreover, the recollection of emotional memories in a free-recall task may be a better indicator of general emotional states and well-being than recognition-based scales because they reflect the emotional pathways laid down in the associative memory network (Bower, 1981), which plays a substantial role in the recollection of experience.

Recalled Versus Recognized Emotions

Memory can be divided up into an effortful recollection-based process (recall) and less effortful familiarity-based process (recognition, see Raaijmakers & Shiffrin, 1981). Recall is the process of retrieving the details linked with a previous experience, while recognition is the process of identifying whether or not details presented to mind are present in memory. A principal difference between recall and recognition is therefore the retrieval stage of memory, which is not present in recognition-based scales (Anderson & Bower, 1972; Bahrick, 1970; Estes & Da Polito, 1967; Kintsch, 1970). In addition, several clinical studies have described cases where individuals have intact recognition memory but impaired recall memory, or vice versa, which suggest these processes may be controlled by different areas of the brain (Delbecq-Derouesné, Beauvois, & Shallice, 1990; Hanley, Davies, Downes, & Mayes, 1994).

The distinction between recognition and recall is therefore based on cognitive and neural differences and this is may influence the kinds of emotions that come to people's minds in day-to-day experience and therefore their responses in different affect paradigms. For example, Tulving and Pearlstone (1966) observed that more memories may be available by recognition than by recall. At first glance, this appears to be a benefit to recognition-based scales. But this potentially comes with a cost of overlooking the emotions that most frequently come to mind—because they are not on the recognition list. The availability heuristic refers to the well-documented observation that people often use the ease with which memories come to mind as indicators of their frequency and importance of occurrence (Pachur, Hertwig, & Steinmann, 2012; Schwarz, Bless, & Bohner, 1991; Tversky & Kahneman, 1973). As such, emotions that come to mind easily are likely to be those most frequently experienced. In addition, previous studies have found that the effort involved in recall may be a better cue to the accuracy

of a memory. For example, Robinson and Johnson (1996) found that a recall-based measure of eyewitness memory led to a better confidence-accuracy correlation, indicating that recall provided additional information that was lost in assessments based only on recognition (see also Koriat & Goldsmith, 1996). This is potentially a problem for recognition-based scales.

Procedure

The challenge we set forth here is to create a recall-based affect scale and compare it with existing recognition-based scales. Several studies have demonstrated the ability to use a recall-based scale in assessing the Big Five personality traits. Their participants were asked to describe their personality using 10 adjectives. Participants' personality scores were then obtained from the average correspondence between these adjectives and the Big Five personality factors (Claeys, De Boeck, Van Den Bosch, Biesmans, & Böhrer, 1985; Van Rensbergen, Kuppens, Storms, & De Deyne, 2015). Following this work, we constructed the Emotional Recall Task (ERT), which asks participants to recall and rate recent emotions. Participants are asked to use 10 words that describe their emotional experiences over the past month. Next, participants are asked to rate how frequently they experienced each of these emotional experiences over the past month. The purpose of this step is allow participants to indicate the strength with which they experienced each emotional experience. The ERT score is then computed by taking the average valence ratings of emotion words weighted by their experienced frequency. Similar to the PANAS, the ERT measures affect by collecting information on a list of emotion terms and then compresses the information into one single dimension of valence for comparison across individuals. The key difference is that the emotion space the PANAS probes into is predetermined by its emotion checklist, while for the ERT it is determined by a less constrained search through emotional memory *that may reflect individual differences in the way people experience emotions*. Consequently, compared to the PANAS, the valence of the ERT is potentially a more individual specific measure.

In Study 1, we compared the ERT against existing recognition-based metrics. In Study 2, we present a more generalizable version of the ERT (the ERT general) to show that it can be easily administered, making it ideal for assessments across languages, ages, cultures and contexts. In Study 3 we show that test–retest reliability of the ERT is on par with existing recognition-based scales. All studies have received ethical approval from the Department of Psychology at the University of Warwick.

Study 1: Comparison and Validation of Recall and Recognition-Based Scales

Study 1 compares the ERT with several standard recognition-based scales. The central goal in this study is to evaluate construct validity of the ERT. The ERT encourages people to actively search their memory for emotions they have experienced. External validity was evaluated using correlations between the ERT and other scales measuring similar concepts (convergent validity) or different concepts (divergent validity). More specifically, we assessed convergent validity by testing correlations between the ERT and other scales that directly measure aspects of general affect, including the Positive and Negative Affect Schedule (PANAS, Watson, Clark, & Tellegen, 1988), the Scale of Psychological Well-Being (SPWB, Ryff & Keyes, 1995), the Satisfaction with Life Scale (SWLS, Diener, Emmons, Larsen, & Griffin, 1985), and the four well-being questions from the Office of National Statistics (ONS, Tabor & Stockley, 2018). We also test correlations between the ERT and concepts that are separate from, but not unrelated to, general affect such as depression, stress, and anxiety using the Depression Anxiety and Stress Scales (DASS-21, Lovibond & Lovibond, 1995) and the Beck Depression Inventory (BDI, Beck, Steer, & Brown, 1996). In addition, testing construct validity against these scales provides convenient comparisons with existing affect scales since they were often validated against a similar set of measures.

Method

Participants. 130 participants were recruited from Amazon Mechanical Turk. They are based in the United States and reported as native English speakers. We excluded 4 participants from the analysis because they failed to follow instructions. This left us with 126 participants.

Procedure. The questionnaire was administered on Qualtrics. Following the consent form, participants were taken to a web page and provided with the following instruction: "Please use 10 English words to describe feelings you have experienced during the past month." After that, a second page appeared presenting the 10 words participants just produced in a randomized order, with an instruction to "use the slider bar to indicate how often you have experienced each of these feelings during the past month." The slider ranged from 0 (*not often at all*) to 100 (*very often*). All participants filled out the ERT first to avoid being primed with words from other scales. Following this, they were presented with the following scales:

The Positive and Negative Affect Schedule (PANAS, Watson et al., 1988) consists of two 10-item mood scales. It was developed to provide a brief measure of positive and negative affect. The 20 PANAS items were derived from a principal component analysis of Zevon and Tellegen's (1982) 60-item mood checklist. Respondents are asked to rate the extent they experienced each emotion within a specific time frame, with reference to a 5-point scale that ranges from 'very slightly or not at all' to 'very much.' Different time frames (e.g., "right now," "today," "during the past few days," "during the past week," "during the past few weeks," "during the past year," "in general") have been used with the PANAS. In the present study we set time frame to "during the past month."

The Ryff Scale of Psychological Well-Being (SPWB, Ryff & Keyes, 1995) is a theoretically grounded instrument that specifically focuses on measuring multiple facets of psychological well-being. These facets include the following: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance. Individuals respond to various statements and indicate on a 6-point Likert scale on how true each statement is of them. Higher scores on each scale indicate greater well-being on that dimension. We used the 18-item version in the current study.

The Diener Satisfaction with Life Scale (SWLS, Diener et al., 1985) is a short 5-item instrument designed to measure global

cognitive judgments of satisfaction with one's life as a whole. The scale does not assess satisfaction with life domains such as health or finances but allows subjects to integrate and weight these domains in whatever way they choose.

The ONS-4 was developed by the Office for National Statistics of U.K. to assess personal well-being using 4 measures that capture 3 types of well-being: evaluative, eudemonic and experience (Tabor & Stockley, 2018). These measures ask people to evaluate their overall life satisfaction, worthiness of things they do, happiness, and anxiety. It was first added to the Annual Population Survey (APS) in April 2011 and has been used in many surveys across the UK.

The Depression Anxiety and Stress Scales (DASS-21, Lovibond & Lovibond, 1995) consists of three 7-item self-report scales that measure depression, anxiety, and stress correspondingly. Each items was rated on a 4-point scale.

The Beck Depression Inventory (BDI, Beck et al., 1996) measures severity of depression in normal and psychiatric populations. The questionnaire was developed from clinical observations of attitudes and symptoms occurring frequently among depressed psychiatric patients and infrequently in nondepressed psychiatric patients. The questionnaire contains 21 questions, each ranging on a scale from 0 to 3.

Quantification of responses to the ERT. In Study 1 we use the affective norms of Warriner, Kuperman, and Brysbaert (2013) to retrieve the valence for each word participants produce. The Warriner et al. norms is an extended version of Bradley and Lang's (1999) Affective Norm for English Words (ANEW), providing affective ratings for almost 14,000 English words. Each word was rated by around 20 participants on a scale from 1 (*unpleasant, calm, controlled*) to 9 (*pleasant, excited, in control*). The valence ratings of words were centered by subtracting 5, so that the scale ranged from −4 to 4 with negative values corresponding to negative valence and positive values to positive valence. This prevents high frequency low valence emotions from have the same impact as low frequency high valence emotions. To infer valence for each participant, we average valence ratings of recalled words, weighted by the participant's self-reported frequency of experience. For comparison, we also computed ERT scores based on the arousal and dominance ratings from Warriner et al. (2013). These were computed for each participant's words using the same approach as for valence, using the frequency-weighted average of the ratings for each of the recalled words.

Over all participants, there were 139 words that could not be transformed into valence ratings because these words are not included in the norm database (Warriner et al., 2013). Though we get around this issue in Study 2, here we tackled this issue by extrapolating affective ratings using the existing norm database. Following Hollis, Westbury, and Lefsrud (2017), we trained a linear regression model that predicts human judgment of valence, arousal and dominance using the vector representations of words, and then use this model to extrapolate the valence, dominance and arousal of those 139 words not included in the Warriner et al. norms. Hollis et al. found this approach predicts affective norms more accurately than the k-nearest neighbors approach (estimate valence of a word by averaging valence of its semantic neighbors). To do this, we used the most prominent of word vector algorithms—Google's skip-gram model with negative sampling (Mikolov, Chen, Corrado, & Dean, 2013; Word2vec, 2013). This model uses a neural network to map each word to the context it appears. It can effectively capture semantic relationship between words such as *man* to *woman* is like *king* to *queen*.

Results

Participants produced 466 unique words and 64% were mentioned only once. Our analysis shows participants tended to first recall emotions they experienced more frequently, with less frequent words produced later in the sequence (Figure 1A). This supports the notion that the most frequently experienced emotions are the ones most likely to be recalled. Emotions produced earlier in the recall sequence were also produced faster than later words (Figure 1B). Figure 1C shows that the valence of emotion words is bimodally distributed, suggesting that people tend to recall more non-neutral emotions than neutral emotions despite the observation that the vast majority of most peoples' emotion experiences are fairly mundane (Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004).

Breadth and specificity. How is the ERT different from the PANAS in terms of emotional specificity and emotional breadth? Figure 2A shows that few PANAS terms are among the most

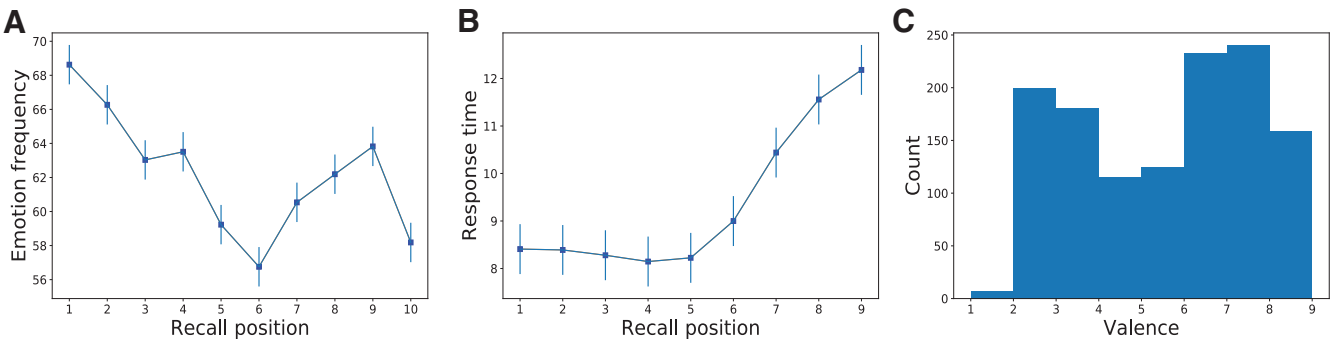

Figure 1. The frequency, response time, and valence of words produced in the ERT. (A) The average frequency of experiencing the reported ERT emotions in each recall position. (B) The average time (in seconds) spent on generating ERT emotion words in each recall position. (C) The distribution of valence values for all terms produced in the ERT. See the online article for the color version of this figure.

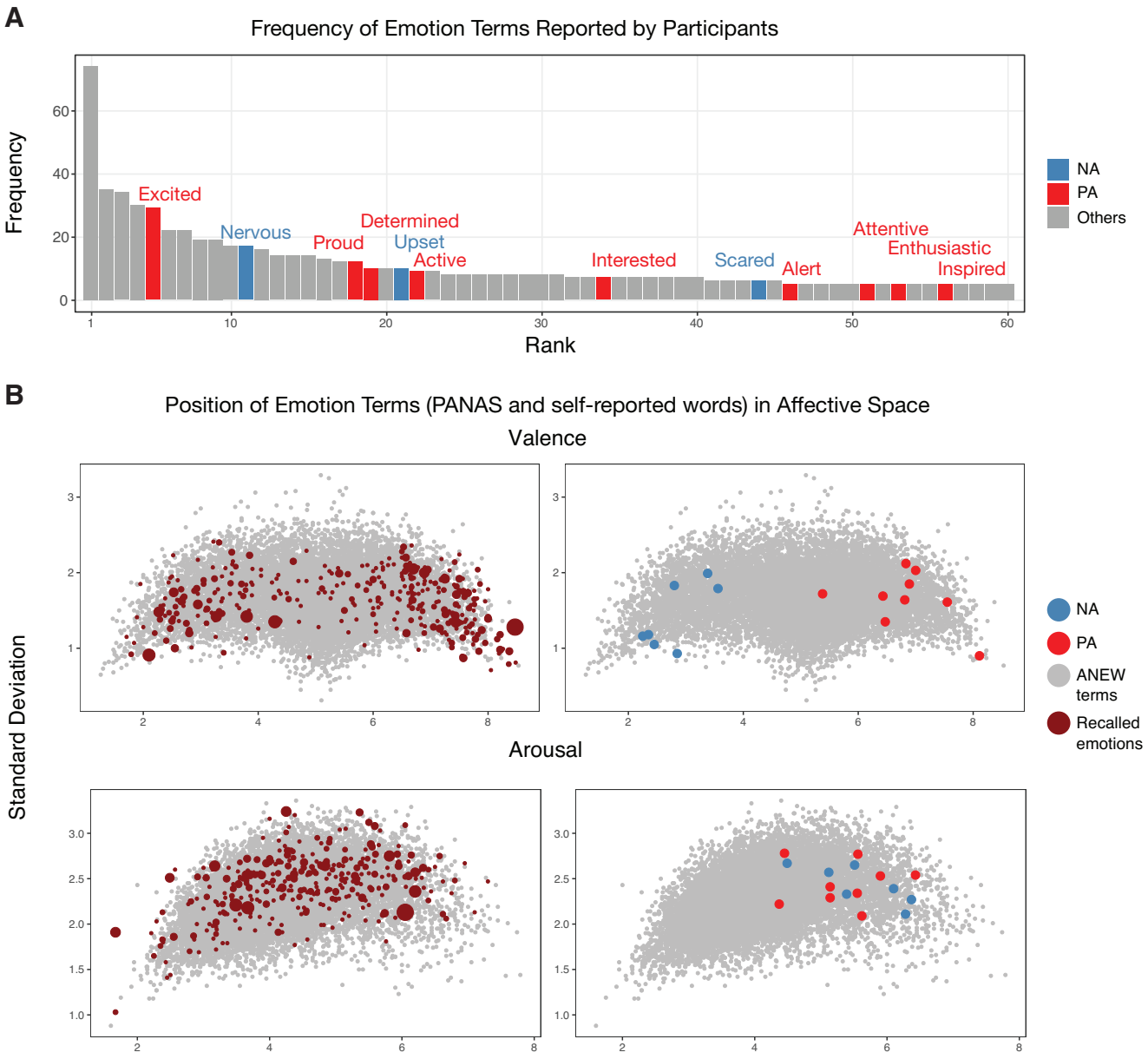

Figure 2. Emotional breath and specificity of the ERT and the PANAS. (A) The frequency of words recalled in the ERT and where the PANAS words are located in the ERT frequency ranking (highlighted in red and blue respectively for positive affect and negative affect). (B) PANAS and ERT terms located along the dimensions of valence and arousal. The *x*-axis is the mean valence or arousal rating and the *y*-axis is the standard deviation of these ratings. Higher standard deviation indicates larger degree of disagreement among those rating the words in the norms. Each gray dot represents one word from the existing affective norm database (Warriner et al., 2013). See the online article for the color version of this figure.

popular words that people used when describing their past feelings. Only 1 PANAS term (*excited*) appeared among the top 10 most frequently recalled emotions. This may of course be a feature of PANAS, but it also suggests that recalled emotions may not typically take the form of the PANAS list. More than half of the words in PANAS are produced by fewer than 10% of our participants.

Figure 2B compares the distribution of PANAS terms and the ERT terms on the affect space of valence and arousal. It shows that ERT terms distribute across the entire arousal space while the PANAS contains no low arousal emotion terms. Moreover, although both scales cover the extreme ends of valence space, PANAS has few neutral terms. This supports previous concerns (e.g., Diener et al., 2009) that it is likely to be challenging for a recognition-based list to capture the emotional breadth and specificity that individuals feel.

Convergent validity. A good emotion scale should be able to predict related constructs. We first analyzed the relation

between the ERT and PANAS. The pairwise correlation coefficient of PA, NA and the ERT can be found in Table 1 alongside other scales. The ERT correlates with both PA and NA ($r = .56$, $p < .001$ for PA and $r = -0.59$, $p < .001$ for NA). Consistent with previous studies of PANAS, we found the PA and NA components of PANAS are independent of each other, $r = -.14$, $p = .12$, a design feature of PANAS.

We further explored the discrepancy between the ERT and the PANAS by examining participants whose emotional states were inconsistent between the two measures. Figure 3A shows how participants' ERT scores are related to their PANAS scores. As Table 1 indicates, there is a negative relationship between ERT and the NA scale of PANAS, and there is a positive relationship between ERT and the PA scale of PANAS. However, several individuals are particularly noteworthy. In the ERT, Participant 15 (ID number = 15) generated a number of negative emotion terms, and no positive terms, and reported experiencing each of the negative terms with high frequency (Figure 3B2). This participant's ERT is in the bottom 5% of all ERT scores. Yet he or she reported extremely low negative affect in the PANAS scale (Figure 3A left), as if they were experiencing no negative emotions. Similarly, Participant 72 recalled 8 positive emotions, 1 neutral emotion, and 1 negative emotion (Figure 3B4). But the same participant's PANAS score suggests the participant experienced little positive affect. Participants 75 and 66 (Figure 3B1 and B3) show similar discrepancies between recalled and recognized emotions.

Correlation with other related constructs. To further test the validity of the ERT, we compared it with the PANAS on how well it predicted related constructs. Table 1 shows that the ERT performs at least as well as the PANAS in predicting the 3 wellbeing-related constructs (Diener, Ryff and ONS4), and 2 depression measures (BDI and DASS). In particular, the ERT has higher correlations for all additional scales than does PA for the PANAS scale. On the other hand, the NA of PANAS performs better in predicting ONS anxiety, BDI Depression, and DASS Depression, Anxiety, and Stress. This may not be surprising since, as noted above, 4 out of 10 terms in the NA portion of the PANAS scale are anxiety related (Diener et al., 2009). Nonetheless, though the correlations are marginally better or worse, the correlations are generally high across all scales, indicating that the ERT is well-positioned with respect to existing scales.

Arousal and dominance. Arousal was not significantly correlated with any related constructs (Table 1). Though as we note in the introduction it was proposed by Wundt, among others, as an emotion dimension, it is clearly not typical of the affect measured in any of the scales evaluated here. In contrast, dominance shows moderate correlations with related constructs. Due to the strong correlation between the valence rating and dominance rating of English words ($r = .717$, $p < .0001$, Warriner et al., 2013), it is difficult to disentangle the effect of dominance on well-being from valence.

Sensitivity analysis. As noted above, we compute the ERT scale as a weighted average of word valence weighted by experienced frequency. For completeness, we also present two additional methods: the unweighted averaged valence and the scaled averaged valence. The unweighted average is the average of the raw valence scores. The scaled average is computed by dividing the difference between a word's valence and the valence of the most negative word by the valence difference between the most positive word and the most negative word one produced. Both of these measures correlate less well with other measures than the unweighted ERT (see Table 2). Given that the purpose of frequency weighting is to reduce the impact of words recalled out of semantic knowledge instead of actual feelings, our result suggests a weak accessibility bias in the ERT (e.g., providing emotions words that are accessible but less frequently experienced). The valence scaled by the range of variance shows much weaker correlations with related constructs than the average valence weighted by frequency (see Table 2).

To evaluate whether 10 words in the ERT are sufficient, we performed a sensitivity analysis to show how correlational strength between the ERT and other constructs change in relation to the number of emotion terms included. Figure 4 left shows that the correlation generally improves when more words are included by their recall sequence. This improvement plateaus roughly between the fifth and the 10th word depending on the specific scale one uses for comparison. Figure 4 right shows that removing the extreme values (most valenced words) weakens the ERT's predicting power of related constructs.

Semantic reconstruction of the PANAS score. Following van Rensbergen et al. (2015), we also evaluated to what extent the PANAS measure could be reconstructed using its semantic similarity with the ERT. We use Google's word2vec (Word2vec, 2013) to extract the semantic similarity between the PANAS and ERT words. An individual's score for each PANAS word is computed by the summation of its semantic similarity with the 10 ERT responses weighted by their experienced frequency. Semantic similarity ranges from 0 (*no shared semantic similarity*) to 1 (*identical*). We forced semantic similarity to 0 if it was smaller than 0.45, which is the semantic similarity between *excited* and *nervous*—the largest value among semantic similarities that compare one positive PANAS word with one negative PANAS word. Not doing this only worsened the resulting correlations. The PA and NA of the reconstructed PANAS score (Reconstructed PA and Reconstructed NA in Table 1) was weakly correlated. Overall, the reconstructed PANAS correlates with related constructs, but less well than the PANAS or ERT. This result suggests that ERT and the PANAS capture a different subsample of the emotion space, and that mapping an emotion space specific to individual's experience onto the more balanced PANAS space might be counterproductive in describing overall affect.

Study 2: The ERT General

One issue with the ERT is that it uses valence norms to compute the valence for each recalled emotion. This has two problems. First, this assumes that different people share the same valence ratings for emotional words. As shown in Figure 2B, this assumption is false. Words have nonzero standard deviations in their valence and arousal ratings. Second, referring to an English database limits the ERT's generalizability to other cultural and social groups where affective norms are not available. In addition, even when a database on affective norms is available, participants may produce words that are not found in the affective norms. Although we can use the machine learning algorithm introduced in the previous section to replace those words with semantically similar

Table 1
Correlation Table Between All Measures

Affect scale	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
ERT normed																
1. Valence																
Variations of the ERT																
2. Arousal	0.22*															
3. Dominance	0.59***	0.82***														
4. Reconstructed_PA	0.57***	0.31***	0.39***													
5. Reconstructed_NA	−0.67***	0.13	−0.27**	−0.21*												
PANAS																
6. PA	0.56***	0.16	0.24**	0.41***	−0.43***											
7. NA	−0.59***	0.07	−0.27**	−0.39***	0.49***	−0.14										
SWLS																
8. SWLS	0.66***	0.15	0.34***	0.32***	−0.53***	0.57***	−0.26**									
SPWB																
9. SPWB	0.59***	0.05	0.28**	0.41***	−0.49***	0.43***	−0.57***	0.54***								
ONS																
10. Life satisfaction	0.75***	0.14	0.39***	0.36***	−0.59***	0.64***	−0.43***	0.87***	0.61***							
11. Life worthiness	0.69***	0.05	0.27**	0.36***	−0.55***	0.61***	−0.39***	0.72***	0.68***	0.83***						
12. Happiness	0.71***	0.06	0.33***	0.36***	−0.56***	0.53***	−0.47***	0.63***	0.56***	0.78***	0.71***					
13. Anxiety	−0.53***	0.06	−0.26**	−0.32***	0.55***	−0.17	0.73***	−0.25**	−0.53***	−0.39***	−0.36***	−0.52***				
BDI																
14. Depression	−0.7***	0.11	−0.25**	−0.45***	0.56***	−0.36***	0.75***	−0.44***	−0.64***	−0.58***	−0.55***	−0.58***	0.67***			
DASS																
15. Depression	−0.7***	0.07	−0.27**	−0.45***	0.56***	−0.38***	0.68***	−0.45***	−0.69***	−0.62***	−0.66***	−0.65***	0.63***	0.87***		
16. Anxiety	−0.4***	0.01	−0.22*	−0.39***	0.25**	−0.03	0.65***	−0.05	−0.47***	−0.2*	−0.24**	−0.27**	0.58***	0.64***	0.72***	
17. Stress	−0.55***	0.03	−0.26**	−0.35***	0.48***	−0.21*	0.71***	−0.29***	−0.54***	−0.47***	−0.48***	−0.53***	0.63***	0.75***	0.83***	0.81***

Note. BDI = Beck Depression Inventory; DASS =Depression Anxiety and Stress Scales; ERT = Emotional Recall Task; PANAS = Positive and Negative Affect Schedule; SPWB = Scales of Psychological Well-Being; SWLS = Satisfaction with Life Scale; ONS = Office of National Statistics.
* $p < .05$. ** $p < .01$. *** $p < .001$.

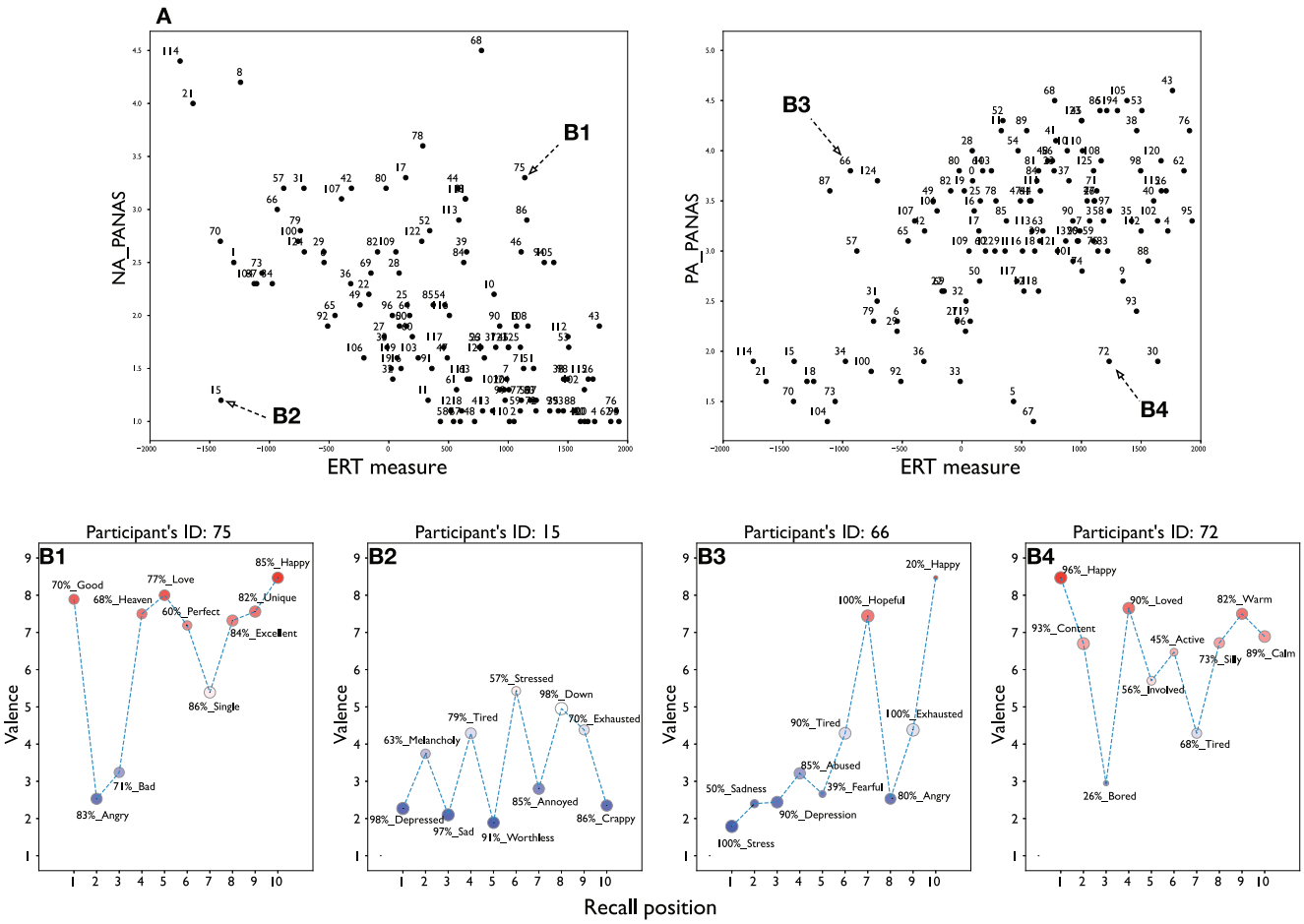

Figure 3. Discrepancy between the ERT measure of emotion and the PANAS. (A) The correlation between ERT measures and NA and PA of the PANAS (B1–B4). The sequence of 10 words produced by the four participants identified in A. The frequency of experience (in % and also indicated by dot size) is provided next to each emotion word produced. Color shows word valence (blue = positive, red = negative). ERT = Emotional Recall Task. See the online article for the color version of this figure.

words, this introduces additional computational costs and assumptions (Mandera, Keuleers, & Brysbaert, 2015).

Study 2 demonstrates a general solution this problem that both simplifies the ERT and generalizes its use to a broader set of communities. This method (henceforth the ERT-general) retrieves valence by asking participants to rate the valence of the words they produced. We use the term *general* to indicate the ease with which the test can be readily applied to wider range of social groups and contexts. In the ERT-general, participants produce 10 emotion words to describe what they have felt in the past month, and then they rate these emotions for how often they have felt each of them. Finally, they rate each emotion for its valence on a scale from 1 to 9. Below we compare this with the norms-based ERT described in Study 1 (henceforth ERT-normed).

Table 2
Comparison of Different Approaches to Aggregate Free-Recall Responses

	PANAS				ONS				BDI	DASS		
Affect scale	PA	NA	Diener	Ryff	LifeSatisfy	LifeWorthy	Happy	Anxiety	Depression	Depression	Anxiety	Stress
ERT_V_weighted	0.56	−0.59	0.66	0.59	0.75	0.69	0.71	−0.53	−0.7	−0.7	−0.4	−0.55
ERT_V_unweighted	0.52	−0.52	0.65	0.55	0.71	0.63	0.66	−0.49	−0.65	−0.63	−0.34	−0.47
ERT_V_scaled	0.37	−0.29	0.47	0.40	0.48	0.39	0.44	−0.30	−0.48	−0.45	−0.22	−0.29

Note. BDI = Beck Depression Inventory; DASS = Depression Anxiety and Stress Scales; ERT = Emotional Recall Task; PANAS = Positive and Negative Affect Schedule; ONS = Office of National Statistics.

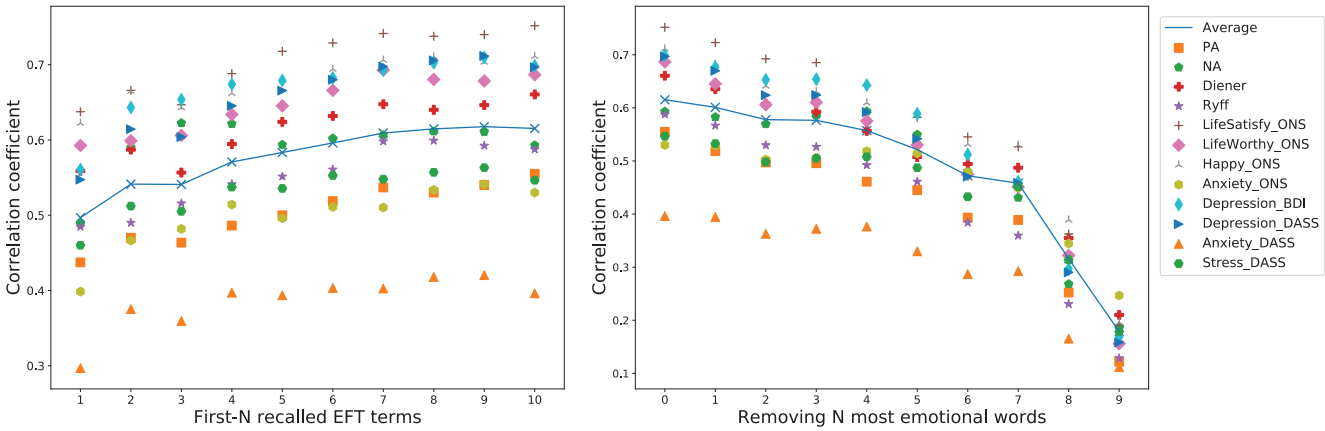

Figure 4. Sensitivity analysis between the ERT measure and other constructs in relation to increasing number of the ERT words included (in recall order). ONS = Office of National Statistics; BDI = Beck Depression Inventory; DASS = Depression Anxiety and Stress Scales. See the online article for the color version of this figure.

Method

Participants. The tasks were presented to 200 native English speakers recruited from Amazon Mechanical Turk (MTurk). Four participants were removed because they failed to follow instructions. The study was preregistered on OSF (Li, Masitah, & Hills, 2019).

Procedure. The procedure is the same as Study 1 except after reporting 10 emotion words and their frequency of experience, participants were additionally required to evaluate the valence of each word. Specifically, they were asked to "evaluate each of these feelings on the scale of unpleasant (1) to pleasant (9)." We provided the following additional descriptions on the scale: on the left-hand side, you feel *happy, pleased, satisfied, contented, hopeful*; on the right-hand side, you feel *unhappy, annoyed, unsatisfied, melancholic, despair*, or *bored*. If the feeling was completely neutral, neither *happy* nor *sad*, move the bar to the middle.

Results

Table 3 shows that results from the ERT-normed and the ERT-general are highly correlated, $r = .97$, $p < .001$. As in Study 1, both versions of the ERT perform similarly in predicting measures of well-being and depression, while consistent with Study 1 the negative component of PANAS better predicts anxiety and stress. This not only strongly suggests that the Warriner et al. norms can be used accurately to capture emotional affect in the ERT, but also indicates that the ERT can be used reliably without the need for using valence norms or machine-learning. Indeed, the ERT-general can be easily computed on the back of an envelope.

Study 3: Test–Retest Reliability

General affect is likely to remain relatively stable over time. In this study we compare results from three different samples over various periods of time to demonstrate that the test–retest reliability of the ERT is similar to that of recognition-based scales.

Method

Participants. We examined test–retest reliability of the ERT-general on three samples. The first sample consists of 119 undergraduate students from the University of Warwick. They were asked to complete the survey twice separated by an interval of 2 weeks. Seven failed to complete the first or second test. The remaining 112 students completed both test and retest and are included in the analysis below. The second sample consists of 90 undergraduate students from the University of Warwick who were required to complete the survey 5 times with an intersession interval of 2 weeks for the first four sessions, and 9 weeks between the fourth and the last session. Twenty students dropped out in the third session, ten out in the fourth session, and seven out in the last session. Students in sample 1 and sample 2 were compensated with course credit. The last sample consists of 115 English-speaking Americans recruited through Prolific. They were asked to complete the survey twice separated by an interval of 2 weeks. 18 participants failed to complete the second session, making the sample size 97.

Measures. Participants in the test-retest study filled out the ERT-general, the PANAS, Diener's SWLS and the ONS. The first Warwick sample completed the BDI-21 and the other two samples completed the DASS. These are as described in the Study 1.

Results

Table 4 summarizes the test–retest reliability for each sample. Our results suggest that the ERT has test–retest reliability similar to that of the PANAS and other scales. The 5-session study on the Warwick Sample 2 shows that for both the PANAS and the ERT the test-retest correlations decrease over time from 2 weeks to 15 weeks. When time is less than 6 weeks (comparing Session 1 with Session 2, 3, and 4), there is little difference between the ERT and other scales in test–retest reliability. At 15 weeks the ERT is less reliable, which is most likely a promising result. This suggests the

Table 3
Correlations Between ERT-Normed, ERT-General, and All Related Constructs

Affect scale	1	2	3	4	5	6	7	8	9	10	11	12	13
ERT													
1. ERT normed													
2. ERT general	0.97***												
PANAS													
3. PA	0.75***	0.74***											
4. NA	−0.69***	−0.7***	−0.51***										
SWLS													
5. SWLS	0.73***	0.72***	0.71***	−0.54***									
SPWB													
6. SPWB	0.72***	0.69***	0.71***	−0.62***	0.78***								
ONS													
7. Life satisfaction	0.77***	0.75***	0.73***	−0.56***	0.92***	0.82***							
8. Life worthiness	0.73***	0.72***	0.73***	−0.53***	0.83***	0.84***	0.9***						
9. Happiness	0.82***	0.8***	0.73***	−0.6***	0.79***	0.8***	0.86***	0.84***					
10. Anxiety	−0.64***	−0.64***	−0.51***	0.73***	−0.49***	−0.6***	−0.53***	−0.52***	−0.6***				
BDI													
11. Depression	−0.69***	−0.68***	−0.59***	0.66***	−0.69***	−0.75***	−0.72***	−0.73***	−0.75***	0.57***			
DASS													
12. Depression	−0.72***	−0.7***	−0.61***	0.64***	−0.72***	−0.78***	−0.78***	−0.78***	−0.78***	0.58***	0.87***		
13. Anxiety	−0.44***	−0.46***	−0.3***	0.5***	−0.41***	−0.46***	−0.43***	−0.45***	−0.45***	0.57***	0.64***	0.63***	
14. Stress	−0.64***	−0.64***	−0.45***	0.72***	−0.52***	−0.58***	−0.53***	−0.55***	−0.6***	0.68***	0.71***	0.75***	0.72***

Note. BDI = Beck Depression Inventory; DASS =Depression Anxiety and Stress Scales; ERT = Emotional Recall Task; PANAS = Positive and Negative Affect Schedule; SPWB = Scales of Psychological Well-Being; SWLS = Satisfaction with Life Scale; ONS = Office of National Statistics.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 4
Test-Retest Reliability Correlations Between Affect Scales

Affect scale	Warwick Sample 1	Warwick Sample 2				Prolific
	Interval = 2w (n = 108)	Interval = 2w (n = 90)	Interval = 4w (n = 70)	Interval = 6w (n = 60)	Interval = 15w (n = 53)	Interval = 2w (n = 97)
ERT general	0.42	0.72	0.60	0.61	0.40	0.80
PANAS Positive Affect	0.57	0.75	0.67	0.65	0.55	0.79
PANAS Negative Affect	0.53	0.78	0.54	0.56	0.57	0.81
Satisfaction with Life Scale	0.69	0.81	0.80	0.79	0.69	0.94
ONS Happiness (yesterday)	0.26	0.22	0.29	0.36	n.a	0.60
DASS_Stress	n.a	0.76	0.68	0.79	0.70	0.82
DASS_Depression	n.a	0.78	0.68	0.67	0.74	0.86
DASS_Anxiety	n.a	0.81	0.67	0.85	0.82	0.80
Beck Depression Inventory	0.56	n.a	n.a	n.a	n.a	n.a

Note. BDI = Beck Depression Inventory; DASS = Depression Anxiety and Stress Scales; ERT = Emotional Recall Task; PANAS = Positive and Negative Affect Schedule; ONS = Office of National Statistics. All correlations are significant with $p < .001$.

ERT may capture a less dispositional component of affective states than PANAS and respond more sensitively to changes in affect over longer periods of time. For almost all scales, the prolific sample shows more consistency between two sessions than the Warwick samples, with Warwick Sample 2 more consistent than Warwick Sample 1. This may due to the fact that the data from the Warwick samples were in a first-year Psychology program, with students new to the University experience, and took place near term/semester boundaries.

General Discussion

Recognized and recalled emotions are potentially very different kinds of psychological constructs. However, our results suggest that what comes to mind when one is asked to remember recent emotional experiences is strongly informative of the kinds of responses one is likely to give when asked if they have had certain feelings in the recent past. The ERT, by relying on recalled memory of emotional experience, captures affective states and correlates highly with other commonly used recognition-based measures of affect and well-being. The ERT also shares a similar test–retest reliability with recognition-based scales. Thus, one clear finding of this research is that recognized and recalled emotions are mutually informative.

The performance of a recall-based measure like the ERT may stem from its generalizability. People across various social groups or in various contexts often have systematic differences in their experience and expression of emotions. Scollon, Diener, Oishi, and Biswas-Diener (2004) have identified numerous important emotion terms that are not shared across cultures and are sometimes difficult to translate, such as words like the German *Schadenfreude* (approximately, "enjoyment derived from another's pain"), the Portuguese *saudade* (approximately, "the love/pain that remains when something is absent"), or the Indonesian *kebelet* (approximately, "a feeling of urgent need"). In addition, because PANAS lacks low-arousal terms, young people may score higher on positive affect than the elderly because they are generally more sensation-seeking (Oishi, Schimmack, & Colcombe, 2003). Similarly, studies of emotional responses to low arousal settings, such as natural or spiritual environments, may suffer from a lack of low-arousal terms.

The ERT overcomes this problem by allowing individuals to freely choose emotion terms that readily come to mind. The ERT-general further allows individuals to personally evaluate the terms they produce. In other words, it is a personalized affect scale that adapts its emotion word list to best fit one's experience. This personalization presents a natural trade-off for affective researchers, pitting a finer-grained individually tailored instrument of emotional recall against a more normative experimenter-specified instrument of emotional recognition. Fortunately, our results indicate this is a low-cost trade-off, as the ERT correlations with recognition-based scales are as good as or better than recognition-based scales with one another for most all measures.

As a final point, the ERT offers a number of future directions for emotion research focused around investigating the structure and navigation of emotional memory. The ERT is a kind of category fluency task, where an individual is asked to recall items from a specific category. Studies of category fluency are rich in information about structure and process, allowing researchers to use patterns of recall to reconstruct networks of associations (Zemla & Austerweil, 2018), compare recalled structure with experienced structure (Hills & Segev, 2014), and evaluate individual differences in control processed involved in retrieval (Hills, Jones, & Todd, 2012; Hills, Mata, Wilke, & Samanez-Larkin, 2013). Comparisons with more in vivo experiential methods of affect monitoring, such as the day reconstruction method and experience sampling (Kahneman et al., 2004; Gabriel et al., 2019), also offer opportunities to tease apart the difference between experienced, recalled, and recognized emotional experience. Studies of emotional recall may also help address questions such as why and how individuals with clinical depression perseverate on negative emotions (Gotlib & Joormann, 2010). Thus, while the ERT is a novel measure, we feel it holds great promise.

References

Altarriba, J., & Bauer, L. M. (2004). The distinctiveness of emotion concepts: A comparison between emotion, abstract, and concrete words. *The American Journal of Psychology, 117,* 389–410. http://dx.doi.org/10.2307/4149007

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review, 79,* 97–123. http://dx.doi.org/10.1037/h0033773

Badley, M. M., & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings (Tech. Rep. No. C-1)*. Gainesville: University of Florida, The Center for Research in Psychophysiology.

Bahrick, H. P. (1970). Two-phase model for prompted recall. *Psychological Review, 77,* 215–222. http://dx.doi.org/10.1037/h0029099

Beck, A. T., Steer, R. A., & Brown, G. K. (1996). Beck depression inventory-II. *San Antonio, 78,* 490–498.

Bower, G. H. (1981). Mood and memory. *American Psychologist, 36,* 129–148. http://dx.doi.org/10.1037/0003-066X.36.2.129

Broadie, S., & Rowe, C. (2002). *Aristotle, Nicomachean ethics: Translation, introduction and commentary*. Oxford, England: Oxford University Press.

Claeys, W., De Boeck, P., Van Den Bosch, W., Biesmans, R., & Böhrer, A. (1985). A comparison of one free-format and two fixed-format self-report personality assessment methods. *Journal of Personality and Social Psychology, 49,* 1028–1039. http://dx.doi.org/10.1037/0022-3514.49.4.1028

Darwin, C. (1872). *The expression of the emotions in man and animals*. London, England: Murray. http://dx.doi.org/10.1037/10001-000

Delbecq-Derouesné, J., Beauvois, M. F., & Shallice, T. (1990). Preserved recall versus impaired recognition. A case study. *Brain: A Journal of Neurology, 113,* 1045–1074. http://dx.doi.org/10.1093/brain/113.4.1045

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

Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., & Choi, D. (2009). *New measures of well-being*. Amsterdam, the Netherlands: Springer. http://dx.doi.org/10.1007/978-90-481-2354-4_12

Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. W., Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research, 97,* 143–156. http://dx.doi.org/10.1007/s11205-009-9493-y

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6*(3–4), 169–200. http://dx.doi.org/10.1080/02699939208411068

Estes, W. K., & Da Polito, F. (1967). Independent variation of information storage and retrieval processes in paired-associate learning. *Journal of Experimental Psychology, 75,* 18–26. http://dx.doi.org/10.1037/h0024902

Fontaine, J. R., Scherer, K. R., Roesch, E. B., & Ellsworth, P. C. (2007). The world of emotions is not two-dimensional. *Psychological Science, 18,* 1050–1057. http://dx.doi.org/10.1111/j.1467-9280.2007.02024.x

Gabriel, A. S., Podsakoff, N. P., Beal, D. J., Scott, B. A., Sonnentag, S., Trougakos, J. P., & Butts, M. M. (2019). Experience sampling methods: A discussion of critical trends and considerations for scholarly advancement. *Organizational Research Methods, 22,* 969–1006. http://dx.doi.org/10.1177/1094428118802626

Gotlib, I. H., & Joormann, J. (2010). Cognition and depression: Current status and future directions. *Annual Review of Clinical Psychology, 6,* 285–312. http://dx.doi.org/10.1146/annurev.clinpsy.121208.131305

Hanley, J. R., Davies, A. D., Downes, J. J., & Mayes, A. R. (1994). Impaired recall of verbal material following rupture and repair of an anterior communicating artery aneurysm. *Cognitive Neuropsychology, 11,* 543–578. http://dx.doi.org/10.1080/02643299408251985

Hansen, C. H., & Shantz, C. A. (1995). Emotion-specific priming: Congruence effects on affect and recognition across negative emotions. *Personality and Social Psychology Bulletin, 21,* 548–557. http://dx.doi.org/10.1177/0146167295216001

Hills, T. T., Jones, M. N., & Todd, P. M. (2012). Optimal foraging in semantic memory. *Psychological Review, 119,* 431–440. http://dx.doi.org/10.1037/a0027373

Hills, T. T., Mata, R., Wilke, A., & Samanez-Larkin, G. R. (2013). Mechanisms of age-related decline in memory search across the adult life span. *Developmental Psychology, 49,* 2396–2404. http://dx.doi.org/10.1037/a0032272

Hills, T., & Segev, E. (2014). The news is American but our memories are . . . Chinese? *Journal of the Association for Information Science and Technology, 65,* 1810–1819. http://dx.doi.org/10.1002/asi.23082

Hollis, G., Westbury, C., & Lefsrud, L. (2017). Extrapolating human judgments from skip-gram vector representations of word meaning *The Quarterly Journal of Experimental Psychology, 70,* 1603–1619.

Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306,* 1776–1780. http://dx.doi.org/10.1126/science.1103572

Kensinger, E. A., & Schacter, D. L. (2008). Memory and emotion. In M. Lewis, J. M. Haviland-Jones, & L. Feldman-Barrett (Eds.), *Handbook of emotions* (3rd ed., pp.601–617). New York, NY: Guilford Press.

Kintsch, W. (1970). *Learning, memory, and conceptual processes*. New York, NY: Wiley.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490–517. http://dx.doi.org/10.1037/0033-295X.103.3.490

Li, Y., Masitah, A., & Hills, T. (2019, January 16). Emotion recall task. Retrieved from osf.io/xr3c4

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33,* 335–343. http://dx.doi.org/10.1016/0005-7967(94)00075-U

Lucas, R. E., Diener, E., & Larsen, R. J. (2003). Measuring positive emotions. In S. J. Lopez & C. R. Snyder (Eds.), *Positive psychological assessment: A handbook of models and measures* (pp. 201–218). Washington, DC: American Psychological Association. http://dx.doi.org/10.1037/10612-013

Mandera, P., Keuleers, E., & Brysbaert, M. (2015). How useful are corpus-based methods for extrapolating psycholinguistic variables? *The Quarterly Journal of Experimental Psychology, 68,* 1623–1642. http://dx.doi.org/10.1080/17470218.2014.988735

McDowell, I., & Praught, E. (1982). On the measurement of happiness. An examination of the Bradburn scale in the Canada Health Survey. *American Journal of Epidemiology, 116,* 949–958. http://dx.doi.org/10.1093/oxfordjournals.aje.a113497

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv,* 1301.3781.

Oishi, S., Schimmack, U., & Colcombe, S. J. (2003). The contextual and systematic nature of life satisfaction judgments. *Journal of Experimental Social Psychology, 39,* 232–247. http://dx.doi.org/10.1016/S0022-1031(03)00016-7

Osgood, C., Suci, G., & Tannenbaum, P. (1957). *The measurement of meaning*. Urbana: University of Illinois.

Pachur, T., Hertwig, R., & Steinmann, F. (2012). How do people judge risks: Availability heuristic, affect heuristic, or both? *Journal of Experimental Psychology: Applied, 18,* 314–330. http://dx.doi.org/10.1037/a0028279

Raaijmakers, J. G., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88,* 93–134. http://dx.doi.org/10.1037/0033-295X.88.2.93

Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence–accuracy correlation. *Journal of Applied Psychology, 81,* 587–594. http://dx.doi.org/10.1037/0021-9010.81.5.587

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–1178. http://dx.doi.org/10.1037/h0077714

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110,* 145–172. http://dx.doi.org/10.1037/0033-295X.110.1.145

Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology, 76,* 805–819. http://dx.doi.org/10.1037/0022-3514.76.5.805

Ryff, C. D., & Keyes, C. L. M. (1995). The structure of psychological well-being revisited. *Journal of Personality and Social Psychology, 69,* 719–727. http://dx.doi.org/10.1037/0022-3514.69.4.719

Schwarz, N., Bless, H., & Bohner, G. (1991). Mood and persuasion: Affective states influence the processing of persuasive communications. *Advances in Experimental Social Psychology, 24,* 161–199. http://dx.doi.org/10.1016/S0065-2601(08)60329-9

Scollon, C. N., Diener, E., Oishi, S., & Biswas-Diener, R. (2004). Emotions across cultures and methods. *Journal of Cross-Cultural Psychology, 35,* 304–326. http://dx.doi.org/10.1177/0022022104264124

Tabor, D., & Stockley, L. (2018). *Personal well-being in the U.K.: January to December 2017*. London, England: Office for National Statistics.

Thomas, D. L., & Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59,* 291–297. http://dx.doi.org/10.1037/0022-3514.59.2.291

Thompson, E. R. (2007). Development and validation of an internationally reliable short-form of the positive and negative affect schedule (PANAS). *Journal of Cross-Cultural Psychology, 38,* 227–242. http://dx.doi.org/10.1177/0022022106297301

Tugade, M. M., Fredrickson, B. L., & Feldman Barrett, L. (2004). Psychological resilience and positive emotional granularity: Examining the benefits of positive emotions on coping and health. *Journal of Personality, 72,* 1161–1190. http://dx.doi.org/10.1111/j.1467-6494.2004.00294.x

Tulving, E., & Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning & Verbal Behavior, 5,* 381–391. http://dx.doi.org/10.1016/S0022-5371(66)80048-8

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5,* 207–232. http://dx.doi.org/10.1016/0010-0285(73)90033-9

Van Rensbergen, B., Kuppens, P., Storms, G., & De Deyne, S. (2015). *Computationally coding responses of a free-format self-description personality test using word association data* (Unpublished doctoral dissertation). Katholieke Universiteit Leuven. Retrieved from https://lirias.kuleuven.be/1786073?limo=0

Wang, Z., Busemeyer, J. R., Atmanspacher, H., & Pothos, E. M. (2013). The potential of using quantum theory to build models of cognition. *Topics in Cognitive Science, 5,* 672–688. http://dx.doi.org/10.1111/tops.12043

Warriner, A. B., Kuperman, V., & Brysbaert, M. (2013). Norms of valence, arousal, and dominance for 13,915 English lemmas. *Behavior Research Methods, 45,* 1191–1207. http://dx.doi.org/10.3758/s13428-012-0314-x

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,* 1063–1070. http://dx.doi.org/10.1037/0022-3514.54.6.1063

Word2vec. (2013). Retrieved October 18, 2015, from https://code.google.com/p/word2vec/

Wundt, W. (1905). *Fundamentals of physiological psychology*. Leipzig, Germany: Engelmann.

Zemla, J. C., & Austerweil, J. L. (2018). Estimating semantic networks of groups and individuals from fluency data. *Computational Brain & Behavior, 1,* 36–58. http://dx.doi.org/10.1007/s42113-018-0003-7

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43,* 111–122. http://dx.doi.org/10.1037/0022-3514.43.1.111

Received May 9, 2019
Revision received February 26, 2020
Accepted March 3, 2020 ■