UNIVERSITY OF CALIFORNIA
RIVERSIDE

Going Through the Emotions: The Measurement of Emotions with the Brief Affect Measure

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Psychology

by

Julia C. Revord

June 2021

Dissertation Committee:
Dr. Kate Sweeny, Chairperson
Dr. Sonja Lyubomirsky
Dr. David Funder

The Dissertation of Julia C. Revord is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE DISSERTATION

Going Through the Emotions: The Measurement of Emotions
with the Brief Affect Measure

by

Julia C. Revord

Doctor of Philosophy, Graduate Program in Psychology
University of California, Riverside, June 2021
Dr. Kate Sweeny, Chairperson

Many measures of affect exist, but they are used inconsistently in psychological studies (Weidman, Steckler, & Tracy, 2016). The present aimed to create a short affect measure that reconciles different views of emotion. The aim was a measure that represents all four quadrants of the affective circumplex, reduces redundancy, and retains good measurement properties. A list of items was created from previous measures and studies of affect, then refined manually. In Study 1a, participants described emotional episodes, then recalled the extent to which they felt each of the 354 feelings. In Study 1b, research assistants answered questions about each of the feeling items, rating them on a series of dimensions including pleasantness and arousal. These ratings were used to sort words into the four quadrants of the affective circumplex. An exploratory factor analysis on the data from Study 1a revealed two factors. The 20 highest-loading items from each quadrant were included in Study 2. After eliminating redundant items and adding theoretically important items, 76 items were included in Study 2. Study 2 was used to further narrow down items using an EFA, as well as other indices such as study

frequency, Google Ngram frequency, and reliability. Using these metrics, three items from each quadrant were chosen for the final 12 item scale, called the brief affect measure (BAM). The scale revealed high convergent validity with other affect scales, good discriminant validity, and high internal consistency reliability. Future studies are needed to further validate this measure in a new sample.

Table of Contents

List of Tables

List of Figures

Aims

The aims of this dissertation are to develop and introduce a new scale for measuring affect and to explore the relationship between different feeling states. In service of these goals, sub-goals are to briefly review some popular existing scales and address relevant emotion theories.

Literature Review

Existing Theories of Affect and Emotions

In order to understand the existing scales that assess affect and discrete emotions, it is helpful to understand the research traditions that spawned them. This task is especially difficult in a field so massive and diverse as the field of emotion research. Thus, I will not conduct a comprehensive review, but instead I will sample some key definitions and theories.

Although the terms affect, emotion, feeling, and mood are often used interchangeably, there are some distinctions (Ketai, 1975). These distinctions are often used inconsistently, and the definitions are fuzzy. Generally, "moods" imply mental states that are longer-lasting, while "affect" and "emotion" tend to refer to shorter boosts, which are more immediate and intense (Ketai, 1975). What constitutes an emotion is not completely agreed upon, as it seems to include a heterogeneous mix of states. Russell and Barrett (1999) define a *prototypical emotional episode* as a complex combination of core affect, reaction behavior, attention, appraisal and attributions, in conjunction with all of the bodily events and the subjective sensation. In their definition, core affect is the fuzzy

subjective quality of an experience as positive or negative, and as arousing or relaxing (Russell & Barrett, 1999). Overall, *feeling* seems to be the broadest term describing subjective experiences. Emotions may be a subset of feelings, or perhaps feelings are a component of the complex experience that is an emotion.

Two dominant approaches of emotion are especially relevant to measurement of affect. These approaches are discrete theories, like the theory of functionalism, and dimensional theories, like the theory of constructed emotion. Both of these approaches stem from divergent traditions originating in the 19th century and today can provide a picture of where the field stands in regard to defining and measuring emotion.

Discrete Theories of Emotion

The theory of functionalism, and related theories of discrete emotions, can be traced back to Charles Darwin. In 1859, Darwin introduced the theory of natural selection in *On the Origin of Species.* Darwin published *The Expression of Emotions in Man and Animals* in 1872. In this text, he began to speculate about possible origins of facial expressions and other non-verbal displays that can be observed in human beings—both young and old—and in animals. Retrospectively, *The Expression of the Emotions in Man and Animals* is centrally important to modern emotion theory, largely because its central tenets were resurrected, supported, and adopted in the mid-late 1900s by Paul Ekman (b. 1934). One of the main contributions of *The Expression of the Emotions in Man and Animals* is what today is referred to as a functionalist perspective of emotions or simply functionalism.

The backbone of functionalism is as follows: If individuals who displayed traits that interfered with their ability to survive and/or reproduce were less likely to have offspring, then these maladaptive traits would have slowly been removed from the gene pool, and adaptive traits would have spread through a relatively larger portion of the population over multiple generations. This process would result in a human genome today composed of features—both physical and mental—that promoted (or at least did not prohibit) individuals' survival, reproduction, or behaviors that increased the likelihood of the survival of their offspring. Functionalist theories raise the following question: Why were individuals with a given feature (here, those who experienced a particular emotion) more likely to survive and produce viable offspring? These theories ask both what adaptive function emotions served on the ancestral plane and, tangentially, what function (if any) they serve now (e.g., Keltner & Gross, 1999).

Work stemming from the functionalist tradition has tended to focus on the existence of discrete emotions, or emotions as distinct experiences (e.g., "anger," "joy," "fear"). Ekman's work—and the work of his colleagues—has largely assumed the existence of such emotions and as such has focused on mapping features of this terrain (e.g., Ekman, 1992a, 1992b; Ekman, Friesen, & Ancoli, 1980). For example, much of Ekman's work has focused on how emotions are displayed via specific facial expressions, stemming from the Darwinian belief that communicating via expression was one of the evolutionary functions of emotions (Ekman, Friesen, & Ancoli, 1980). Ekman and his colleagues created the Facial Action Coding System (FACS), which identifies particular muscles used to display specific emotions (Ekman & Friesen, 1977). Furthermore, he and

his academic progeny have sought to identify universal basic emotions across cultures (Ekman & Friesen, 1971; Ekman et al., 1985). This quest has generated numerous papers identifying and describing specific emotions, including possible causes, appraisal components, and associated action tendencies, or the response that specific emotions prepare individuals to carry out (e.g., Frijda, 1987).

Identifying the action tendencies of specific emotions has been one of the primary tasks of researchers who take a functional perspective (e.g., Shiota et al., 2014). In general, negative emotions have been addressed first, whereas attention to positive emotions has only begun more recently, largely in response to theories addressing negative emotions.

Dimensional Theories of Emotion

Although discrete emotions form the basis of one important research tradition in the psychology of emotion, another semi-independent line of inquiry has also driven emotion research. This line of research focuses less on the qualities that define distinct, individual emotions, and more on themes that unite broad dimensions of emotional experience, including pleasantness-unpleasantness, certainty, controllability, and agency (Smith & Ellsworth, 1985). Recently, the theory of constructed emotions has succinctly brought together findings from neuroscience, evolutionary psychology, and cognitive psychology into an integrated theory that describes how the brain creates an internal model of the world based on statistical regularities in the world (both internal and external) and tests this model against incoming sensory information (Barrett, 2016). This tradition of dimensional views also began in the 19th century, primarily with the work of

Herbert Spencer (b. 1820), who was himself a contemporary of Darwin. In 1855, he published *Principles of Psychology,* in which he asserted that emotions are merely dimensions of consciousness (Spencer, 1899).

Dimensional views have garnered much empirical support through the years. Namely, two robust dimensions reliably emerge when the structure of affect is analyzed statistically (Watson & Tellegen, 1985). These dimensions have appeared across multiple studies in various forms. For example, sometimes the two dimensions emerge as part of a three-factor (or more) structure rather than the more common two-factor structure. Most commonly, the two factors emerge as independent unipolar dimensions, but they are often given different labels when rotated differently by researchers (Feldman Barrett & Russell, 1998). In one incarnation of the factors, they are two unipolar dimensions: pleasant-unpleasant (i.e., valence) and the degree to which an emotion is characterized by high or low energy (i.e., arousal). When rotated in a factor analysis, one dimension ranges from high-energy pleasant emotions (e.g., enthusiasm) to low-energy unpleasant feelings (e.g., sadness), while the other ranges from high-energy negative emotions (e.g., anger, anxiety) to low-energy positive emotions (e.g., calm). These rotated dimensions are called positive affect and negative affect (Watson, Clark, & Tellegen, 1998; Watson & Tellegen, 1985). Together, these factors make up an affective circumplex, which is a two-dimensional emotional space in which all emotions fall (e.g., Russell, 1980).

This tradition of focusing on emotions not as objective features of reality but rather as dimensions of consciousness has continued with the modern theory of constructed emotions. The basic logic of the theory of constructed emotions is laid out by

Lisa Feldman Barrett (2016). She explains that brains mainly exist to direct internal resources to help an organism grow, survive, and reproduce. This process is referred to as allostasis. Much of allostasis proceeds unconsciously (as is the case with the autonomic nervous system, immune system, endocrine system, and others). However, certain tasks the brain must take on are not as straightforward and require conscious awareness and decision making. In order to make decisions, the brain creates an internal model of the world, and the body in it. This internal model is called embodied simulation. The internal model of the world is not necessarily an accurate, complete description of one's surroundings. Rather, it is based on the features of the world that are most relevant to meeting the body's physiological needs.

This selective focusing tends to happen in a species-specific way. The internal model is constructed in a way that is very similar to how other types of learning unfold, using a process called statistical learning. Statistical learning, typically studied by cognitive developmentalists, happens when certain things frequently co-occur (Aslin & Newport, 2012). A classic example of statistical learning is when a baby begins to identify their parents. Certain features (eyes, mouth, nose) are always present, while other features are only sometimes present (glasses, hairstyles). Thus, the baby learns that the eyes, mouth, and nose are essential parts of their model of their parents. Statistical learning explains why babies may be disturbed when their model of something in the world does not match up with their current experience—for example when a father shaves his beard for the first time. The brain creates an internal model of statistical

regularities about what is happening in the outside world, but also about what is happening in the body's internal world (a process known as interoception).

According to the theory of constructed emotion, when incoming signals from the internal and external world arrive, the brain uses its past experiences to generate a prediction about what will happen. The brain then compares this prediction with the external world, and the differences (prediction errors) are used to update the internal model. It categorizes incoming sensations and prediction errors to guide action (Barrett, 2016).

According to the theory of constructed emotion, emotions are generated in response to the brain's prediction based on past situations when that emotion was felt, which in turn are based on how a given situation affects allostasis. Emotions are not based on specific sensory input but rather on the occurrence of a pattern in a situation that matches one's (learned) concept of that emotion. This concept can be based on the goal of the situation, and thus the same emotion can appear across situations that are rather different. The definition of an emotion in the brain is composed of a series of related situations (Barrett, 2016).

In this view, discrete emotions are not unshakeable categories with which humans are born; rather, they are a series of concepts that have been constructed over the lifetime of a human as the individual has interacted with various situations and learned to name the specific pattern of sensory input. The implications of this view are that cultural context plays a large role in shaping what a person thinks of as a discrete emotion, because the linguistic categories that exist in one's culture are the categories that are

recognized as discrete emotions. Another, less obvious implication is that one's ability to name emotions is limited to the specific situations that one has had the opportunity to classify. Thus, it is possible that the emotions that exist are fluid (like language) in the sense that they stretch to describe the range of human experience. As one example, with the advent of social media, certain emotions may now be possible to feel that could not exist before. On the other hand, it is possible that a set number of situations have been evolutionarily important (like maintaining social ties and status, avoiding physical harm, accruing resources), and in the ever-expanding pool of new situations, the same plotlines (and their associated emotions) play out.

The Intersection of Discrete and Dimensional

Both of these views of affect—a dimensional view and a discrete view—are important. They both have shaped the field. Although certain papers suggest that these two views are in conflict, they both have a role in answering different questions.

One reason that dimensional theories and the discrete emotion theories have diverged so dramatically might be that they are using drastically different criteria for empirically assessing the reality of emotions, based on different philosophies of science corresponding to the relevant historical and cultural zeitgeist. In the mid-20th century, the scientific community experienced a backlash against psychologists who relied on introspection as a method. The upshot of this backlash was the view that introspection and the study of the subjective inner mind were non-scientific and that the only acceptable way to study psychology was by observing behavior (i.e., behaviorism). Behaviorism presents a bit of a challenge to the study of a subjective phenomenon like

feelings, and the discrete and the dimensional views diverged further in order to dodge this challenge. However, in the midst of this era, research on affect became more popular and eventually led the psychologist Sylvan Tomkins to announce in 1981 that "the next decade or so belongs to affect" (Tomkins, 1981, p. 314; cited in Watson & Tellegen, 1985).

The discrete/functionalist view grew in popularity as its practitioners categorized emotions mainly based on observable behaviors, like vocal expressions and facial expressions. Eventually, these objective behaviors were matched up with more subjective dimensions of emotions like their common eliciting appraisals (Smith & Ellsworth, 1985). The dimensional views did not abandon the subjective but rather attempted to take a more empirical approach towards the subjective. Instead of merely asking individuals to qualitatively describe their subjective experiences, researchers attempted to limit the room for human error in subjectivity by applying statistics to quantify subjective phenomena. For example, studies from the 1960s and 1970s yielded a number of dimensions, including evaluation, activity, and potency, by factor analyzing responses using the semantic differential technique, in which participants rate affective concepts on a number of scales like fair-unfair, hot-cold, and hard-soft (Averill, 1975; Osgood, 1969).

In other studies, participants rated adjectives based on measures of pleasure, arousal, and dominance (e.g., Russell & Mehrabian, 1977). These dimensions of pleasant/unpleasant, activation/sleep, and controlled/uncontrolled (a dimension in which words that encompass controlled, deliberate states, like *determination, scorn*, and *cynical bitterness*, rank high, and words that denote uncontrolled, unintentional states, like

anxiety, surprise, and *awe*, rank low) have been replicated in dimensional studies of facial expressions (Osgood, 1966). The first two dimensions, pleasantness and arousal, have consistently emerged, but the third has been more elusive and thus largely dropped over time. One critique of the third dimension is that the apparent replication of these dimensions across methods may be partially due to labeling, rather than true correlation (Russell, 1978). Other dimensions have also emerged, including control/potency/dominance, depth of experience, and locus of causation (Russell, 1978). Indeed, the lack of clearly replicable dimensions may be further evidence to support discrete emotion approaches.

However, the method of quantifying subjective dimensions of emotions with the longest lasting impacts was achieved by asking participants to indicate only whether or not they felt a specific feeling in a given period of time (reviewed briefly in Watson & Tellegen, 1985). By factor analyzing these results based on which feelings co-occurred, dimensional researchers effectively created categories of associated emotions. Both of these approaches (categorizing emotions based on measurable physical outcomes and using statistical techniques to aggregate which emotions tend to co-occur) dodged the problem of subjectivity, but neither is necessarily the optimal way to categorize emotions.

It is possible that simply mapping measurable behavioral outcomes (including vocal expressions, facial expressions, or even measures like vagal tone, hormones/neurotransmitters, or brain scans) is not the best way to categorize emotions, because variability on these outcomes may not cut nature at its joints. For example, imagine if human activities were categorized solely based on the behavior of the arms

and hands. The arm does many repetitive motions throughout the day that are made up of basic component movements. Playing the piano and typing would look rather similar—as would typing to chat with a friend and typing up a dissertation. However, sorting these motions using context or meaning (such as the goal of the action) would conclude that they are quite different. The problem with this approach is that human bodies are multi-use tools. Similar to how brain regions are often implicated in multiple seemingly unrelated tasks, observable behaviors may be implicated in multiple unrelated emotions.

Additionally, limiting what is defined as an emotion to what can be observed behaviorally fails to explain considerable variance in behavior because it does not fully explain the variance in human interpretation of situations. Certain emotions have not been found to be universal because they do not carry a universal human physical signature. Alternately, some feelings are not associated with a distinct facial signature or vocalization, but they may still be universal (for example, hunger), or uniquely explain variance in behavior (for example, jealousy).

On the other hand, co-occurrence may not be the best way to categorize emotions either. Consider that a number of categories, when measured solely based on co-occurrence, do not make sense. For example, in biology, each ecosystem contains herbivores and carnivores that regularly co-occur but are likely more related to herbivores and carnivores in other ecosystems. In fact, similar animals would be *less* likely to co-occur, as they would be competing for access to the same resources. The competitive exclusion principle, or Gause's law, states that two species occupying the same ecological niche cannot coexist at balanced numbers (Hardin, 1960). One example of this

phenomenon is that of modern humans and Neanderthals. Another, more mundane example of how co-occurrence does not indicate similarity is how a freshwater turtle is far more related to a sea turtle than to freshwater fish, although the habitats of freshwater and sea turtles are completely exclusive. Co-occurrence also does not always work in the short-term, as being nocturnal does not necessarily indicate similarity with other nocturnal animals. A nocturnal barn owl is more similar to the diurnal northern hawk owl than it is to a bat. Not all animals that live in the sea are fish. Dogs live mainly with humans, but neither humans nor cats are their closest relative, and households may or may not have other dogs.

However, a factor analysis of the co-occurrence of all animals would likely find certain robust characteristics that consistently determine if animals co-occur. For example, a six-factor solution would likely match with the continents, or latitude and longitude would consistently predict animal co-occurrence; a two-factor solution might distinguish between land and water. The fact that a structure that predicts co-occurrence can be reliably found (like valence and arousal for emotions) does not necessarily mean that those dimensions are the most meaningful for categorization. An interesting thought experiment would consider what outcome needs to be factor analyzed (rather than co-occurrence) to yield the structure that we see in many categories that exist in the world (such as animals, flavors and flavor profiles, religions, dog breeds, types of moving vehicles, family structure, types of disease, etc.).

Especially in psychology, in which humans try to study their own minds with a degree of impartiality, there is merit to blinding oneself to opinions and interpretation, as

was pushed by behaviorism. It can be tempting to strive for objectivity by using statistics or by only noting observable behaviors. Unfortunately, due to the general lack of consensus, neither of these methods has yet entirely sliced nature at its joints in terms of defining and categorizing emotions. These two approaches have both been useful and enlightening but perhaps are asking and answering different questions from one another.

The theory of constructed emotion describes how emotions are formed as concepts, and how they are deployed in response to sensory input that the brain has learned based on past experience is either helpful or unhelpful to the animal. This theory answers the question of *how* emotions are learned and formed in a general sense. However, the theory of constructed emotion does not examine *what* the specific emotions are and how they are subjectively experienced. Perhaps the theory does not account for discrete emotions because the logical extension of the theory of constructed emotion is that the specific emotions formed are, like color categories, cultural artifacts. In other words, the cultural context has taught individuals to label certain patterns of the internal milieu and external circumstances as specific emotions, but these boundaries may be arbitrary, as evidenced by the fact that many discrete emotions are not universal (Averill, 1980).

However, even if discrete emotions have particular boundaries because human culture has labeled them as such, they are still worth studying for a number of practical reasons. Using an analogy with food, the theory of constructed emotion provides a compelling template for how humans experience flavor, such that they experience specific molecules, which the brain recognizes and responds by creating flavors.

However, this process addresses an entirely different question from *which* flavors the human tongue can taste, which is a different question from which flavors appear universally across cuisine in different parts of the world and which do not. Another interesting question is *why* the human brain rewards specific flavors over others—and finally, an applied question would be how best to modulate cravings in a world where instant gratification is readily available and health dysfunction due to overeating is abundant. Similarly, many different questions about emotions that could inform human psychology are not answered by the theory of constructed emotion.

Instead of studying "joy," emotion researchers may well be studying "the phenomenon known to Westerners in this time as joy." However, if this phenomenon explains significant variance in human psychology and behavior, then it may well be worth studying. Fundamentally, all studies suffer from the same problem: They are limited in scope, and thus in validity. Often this limit appears in the form of a small sample size or a distinct cohort. However, the fundamental problem is that single instances cannot be generalized to become universal laws—but there is a reverse truth, which is that universal laws cannot be discovered without measuring many inherently biased specific instances, and observing where there is and is not variance.

Practically, one reason to study discrete emotions is that they are useful when communicating about self-reported emotions. If emotions are created by the brain simulating a scenario and subjectively matching up the particulars of a specific scenario to create meaning, then their meaning is primarily subjective. Thus, the best way to measure that subjective meaning is by using subjective measures such as self-report.

Also, many studies that are not primarily focused on emotions (or even psychology) measure emotions using self-report. One of the most salient applied settings is the clinical context. Even if the specific flavors and labels for emotion are culturally specific, it is useful to have words to measure and communicate about the emotional space.

The measure being developed here straddles the line between discrete and dimensional approaches. It is made up of discrete emotions, as with all self-report emotion or affect measures. However, the words that are included in the initial subject pool are much broader than in previous work. If emotions have not been clearly defined, then it does not make a great deal of sense to limit which feelings can be considered emotions based on the gut feeling of a particular researcher or on previous traditions. Here, the words included attempt to sample a space that encompasses all cases in which the brain classifies a pattern of sensory input into a specific feeling state used to direct behavior. Rather than implicitly introducing subjectivity (for example, by predetermining which feelings could be considered emotions, by choosing terms solely based on co-occurrence, or by naming the factors myself), my goal was to explicitly embrace and quantify subjectivity, and then consider all of this information, as well as utility, in the selection of items for the measure.

Existing Scales and Measures

Over the years, numerous measures have been developed to conceptualize and measure affect (see a list of items here), with perhaps the most famous being the Positive and Negative Affect Schedule (PANAS; Watson & Clark, 1999). However, despite being widely known, such measures are not ubiquitously employed to measure momentary

discrete emotions in psychological research at the frequency that might be expected. Thus, regardless of researchers' conclusions about why such measures are inadequate, the net conclusion is that they fundamentally are not successful measures because they are not being used. There are multiple possible reasons that measures are not being used, which may stem from concerns with each measure. I conducted a broad review of multiple common affect measures (items here), conducted informal conversations with researchers about which affect measure they use, and tracked down those measures.

In informal discussions, researchers named many reasons for not using particular validated measures. Many measures are too long to include in a short study that is not primarily about affect. Other measures are short enough but do not include items that cover all four areas of the two-dimensional affective space. One main problem is that short measures tend to underrepresent low-arousal positive feelings, which have been shown to be important in cross-cultural research, especially in collectivist cultures. Some researchers also reported that they were interested in studying a particular corner of the affective space for specific studies and felt that the emotions that they were interested in were not well-represented by a single scale.

Other scales include items that clearly express affective dimensions but are not very important to researchers as discrete emotions (e.g., "activated"). In this vein, some scales seem to have achieved high reliability by eliminating variance in the items. (Note, however, that this choice to eliminate variance might be the result of a theoretical bent in structural equation modeling that conceptualizes latent variables as underlying factors driving the overlap between different questions, when in the case of affect measures, a

wide variety of emotions are likely felt throughout daily life, and the composite of these emotions comprises one's emotional life.) In practice, this approach gives rise to a scale that does not ask if the respondent felt sad, anxious, and angry and then extrapolates affect from those responses, but instead achieves high reliability by asking the same broad, nonspecific questions in multiple nonspecific ways (e.g., "unhappy," "upset"). Although this method ultimately results in the researcher having a very clear and reliable view of someone's general negativity or positivity, it obscures any nuance. The researcher knows, via five different questions, that someone felt "bad" but does not know any further detail about the nature of the badness. In a short survey, any question takes up precious space and time, so five questions to ask whether someone felt "bad" are not necessarily a prudent use of survey space. The high reliability achieved by asking essentially the same general question multiple times almost adds insult, by indicating that they truly are redundant.

A related problem cited is that existing measures can include double- or triple-barreled items that clump together multiple items of different frequencies (e.g., a single item asking "did you feel sad, anxious, or down today?"). Finally, a nonobvious but nontrivial issue is that many scales produced by psychologists are now primarily used for commercial sales to companies and are thus hidden behind a paywall, rendering them largely inaccessible to many researchers.

To summarize, my informal discussions yielded the impression that many researchers were foregoing the use of existing validated measures of affect. Interestingly, although researchers did not agree upon a common validated measure, most described a

similar solution. Indeed, I explicitly asked multiple well-published emotion researchers what measure they used and received roughly the same response, which was that each is using a "hacked-together" measure that sampled from all four corners of the quadrant (i.e., high arousal positive, low arousal positive, high arousal negative, low arousal negative) and included the target emotions in which each researcher is most interested.

One recent paper more formally reviewed the state of measures in emotion research (Weidman, Steckler, & Tracy, 2017). The authors reviewed a representative sample of articles published in the journal *Emotion* over a 10-year span (i.e., 2001-2011; $n = 781$) to ascertain what practices were commonly used to measure self-reported distinct, momentary emotional states. Their results, based on 147 studies with 356 separate measurement instances of momentary affect, indicated that momentary affect was measured inconsistently across studies and that the emotions measured did not match up well with current theories of emotion. In fact, across 356 instances in which discrete momentary emotion was measured, 246 (69%) were categorized as entirely *impromptu* (i.e., the scale was developed for the specific measurement, without any sort of systematic scale development or reference to previous scale development), 27 (7.6%) were categorized as *cited impromptu* (i.e., scales that cited a previous study in which they were impromptu), 30 (8.4%) were categorized as *cited existing altered* (i.e., scales based on a systematically developed scale, which were modified for this specific study), 43 (12%) were categorized as *cited existing unaltered* (i.e., scales that appeared in their previously systematically developed form and had undergone no changes for use in the study), and 10 (3%) were unreported measurement instances.

For various reasons, the authors noted, such use of impromptu and altered scales is likely to contribute to inconsistent and unreliable measurement of emotions, which are difficult to compare and integrate across studies. One of the main problems of the approach to measurement characterized by choosing seemingly relevant emotions and simply asking about them is that it is unclear how various items relate to one another. For example, some words (e.g., joy and happiness) could be pure synonyms; some could be special types of another word (e.g., happiness and schadenfreude); and some could be nonsynonyms or nonnesting but mistaken as such. Interestingly, many of the discrete emotions measured by researchers in studies were *not* those that appear on historical scales. Thus, Weidman and his colleagues (2017) vehemently called for a measure to be created that is based on a more comprehensive taxonomy of emotions, which includes states that previously may have been ignored by researchers depending on how they defined emotion (i.e., based on facial or vocal expressions, physiology, etc.).

Project Overview

The purpose of this project is to create a new scale to assess affect and to further explore specific interconnections between emotions. Scale construction followed the internal consistency (i.e., factor analytic) approach laid out by Simms (2008). Additional consideration was granted to construct validity and classical test theory, informed by modern psychometric methods. This method is largely data driven (i.e., much of the creation of the scale and the factors rely on an initial factor analysis that identifies factors and factor loadings), but whether an interpretable and useful scale can be gleaned from

such analyses relies heavily on a degree of researcher interpretation—both in selecting the items and identifying (i.e., labeling) appropriate factors as subscales.

General Method

Overview

This project had many parts. Some portions of the process are important but did not involve data collection from participants. Other portions involved one or more datasets. Here, each step in the endeavor has its own section, divided based on the purpose of each one.

Pre-Study: Creating an Item Pool

Generating Candidate Items

The primary tasks during the beginning stages of this project (i.e., the substantive validity phase) were to clearly define the constructs to be measured and create a representative item pool (Simms, 2008). Many proponents of the internal consistency approach to scale creation recommend that the researcher begins by developing a formal definition of the construct, and then creating the items accordingly. There are multiple papers that describe the process of item writing (e.g., Clark & Watson, 1995; Kaplan & Saccuzzo, 2005; Simms & Watson, 2007). However, one of the main impediments to this project was the lack of an agreed-upon definition of emotion or affect (see Aims). Thus, I elected to be as inclusive as possible in my creation of an item pool, which was constructed by harvesting previous measures and adding additional candidate items.

I found as many existing measures as I could (see Table 1) and stripped their items. I also included items that were being used by researchers but were not included in validated measures (see Weidman et al., 2017). Additionally, I included feeling states that were physical and mental, which might not have been considered in previous measures. Finally, a series of recent papers by Cowen and Keltner (2017) purport to have mapped out the entire emotion space into 27 discrete emotions, and thus these emotions were included. However, notably, certain self-referent emotions were missing—possibly because their method was to sample emotional responses to internet gifs, which, with rare exception, are *not* self-referent. Finally, in an attempt to build an inclusive item pool and to answer further questions beyond the scope of this dissertation, I included some items that have not been traditionally considered to be emotions. These included words that encompassed other states that could be described as "feelings" because they are associated with a specific internal sensation (e.g., horny, queasy, orgasm, drunk, thirsty).

The reason for including a wider span of feelings was that some feelings (e.g., happy, sad, disgusted, satisfied) have traditionally been thought of as emotions and others have not (e.g., impulsive, pondering, anticipation, unfocused), but there is no systematic rule that distinguishes these two types of feelings. For example, emotions are not clearly defined by whether they have a distinct biological or physical signature because some well-recognized emotions do and some do not. Emotions are not distinguished by their interpersonal nature (e.g., as tools for communication), as some can be felt alone and others cannot. Emotions are not distinguished by being exclusively mental sensations, as there are well-established physical feelings associated with pleasure, tension, fear, and so

forth. Indeed, arousal—one of the two primary underlying affective dimensions—might be best thought of as a physical sensation. Additionally, encompassing all mental sensations would entail including words like "concentration," "focus," "drunk," or "high,", which are likely not well-described as emotions. Indeed, some emotions are distinguished from one another purely on semantic meaning, which is a cognitive construct. One question that I asked to guide whether to include an item is as follows: If one were to wake up in the body of someone feeling this emotion, with no context, how much context would one need to pinpoint it? For example, "envy" and "jealousy" are very similar, differing only on a person's cognitive understanding of their environment. It is possible that some words are subtypes of other feelings, distinguished based on semantic meaning. For example, schadenfreude may be a specific type of pleasure that comes from a very specific source (i.e., the downfall of one's rivals). Shame may be a type of pain, distinguished by specific cognitions that one has about the situation.

One candidate for what constitutes an emotion is that it is a derivative of pleasure or pain, created by the brain in response to an internal or external situation. This approach would rule out words like "concentration," which has no valence, but it would also rule out words like "apathy" and "neutral" and include words like "hunger" and "pain." Additionally, many items on scales such as the PANAS are neither positive nor negative, such as "activated."

From a functionalist view—that emotions exist in order to drive evolutionarily helpful behavior, signal to the organism about threats/rewards, or prepare the organism to handle a specific situation—most feeling states would be viable candidates for an affect

measure. Indeed, no consistent rule, when applied, describes the subset of feelings that are currently considered to be emotions.

One reason that I decided to include a wide variety of feeling states was to determine whether there were quantifiable distinctions between "emotions" and "non-emotions." In order to use emotion-ness as a variable, and to test what predicts it, the pool must have items that are both emotions and non-emotions.

Narrowing Down the Item Pool

After the items were compiled, they needed to be reduced to a reasonable number for participants to consider in one sitting. I decided to reduce the original 1217 items to roughly 354 items. To this end, I carefully considered the semantic meaning and usage of each word one by one. Other individuals, including a set of graduate students, undergraduate students, and faculty, also reviewed the words.

I tried my best not to trim the item pool based on my own a priori theories about the definition of affect, nor what constitutes an emotion. Rather, I (and my colleagues) narrowed the item pool to roughly 354 words based on criteria that mainly prioritized clarity and consistency, while trying to maintain breadth. When possible, I retained words or phrases while transmuting them to better fit with others in the same set or prompt. Specific examples are listed below.

Double-Barreled and Multiple Meanings

First, I eliminated word phrases that were multiple words squashed together. For example, the Differential Emotions Scale-Modified (M-DES; Fredrickson, Tugade, Waugh & Larkin, 2003) presents respondents with sets of three similar words clustered

together (e.g., “angry, irritated, annoyed” and “glad, happy, joyful”). Although this approach has obvious advantages, I separated the clusters into their own items because my scale aimed to comprise single words to avoid myriad problems with double-barreled (or, in this case, triple-barreled) items. Other examples of word splits from other scales were “happy surprise,” “sad anger,” “sad disgust,” “distressful confusion,” “embarrassed relief,” “happy disgust,” and “triumphant anger.”

Other words were inadvertently unclear because they had multiple meanings, either in general or when used specifically with the phrase “I feel/felt.” For example, the phrase “I feel cold” has both a physical and emotional meaning. “I felt distasteful” could mean that one feels disgust towards someone else or towards oneself. “I feel lost” could be either a literal description of misunderstanding one's own location or a description of a sense of sadness. “I feel positive” could mean “I am sure” or “I am feeling optimistic.”

A subcategory of multiple meanings are words that have a different meaning when they are used to describe a trait of a person than when they are used to describe a feeling. In English, the phrase “I feel” can be used to describe a fleeting emotional state or one's beliefs about one's traits or actions. For example, to describe a person’s character as “unpleasant” is different from describing oneself as having unpleasant feelings or being worried that one is being unpleasant to others. Likewise, “I feel dull” could mean “I feel bored” or “I feel that I am a dull person.” “I feel bright” could refer to a person who feels temporarily cheery or has the belief that they are a smart person.

Too Wordy/Did Not Fit the Word Stem

Word phrases were also eliminated for being too wordy. For example, “Things feel pretty dull right now” (Feldman Barrett & Russell, 1998) was eliminated, as it did not fit with the word stem “I feel ___” and the format of the rest of the items. Other examples of this wordiness include “wide awake” (shortened to “awake”) and “bone-chilling terror” (substituted with “terrified”).

Reading Level/Esoteric

Words were also eliminated based on their reading level and understandability. This happened throughout multiple stages in the process, as different samples of pilot participants found different words difficult, and the final item list had to include words accessible to broad populations. In a first pass, I eliminated the few words with which I (a college graduate who was raised in the United States and majored in English) was less familiar. These included “surgency,” “blasé,” “blithe,” “disconsolate,” “ennui,” and “indolent” (which could easily be confused with “insolent”). Next, preliminary surveys on the Prolific research participation site asked participants to consider the words in the sentence “I feel [word]” and rate how well they felt that they understood the meaning. This survey revealed a number of words that were poorly understood, including “dejected,” “exuberant,” “jovial,” “listless,” “affable,” “detesting,” “insolent,” and “indignation.” I did not necessarily eliminate such words outright; however, when pitted against synonyms, more accessible words were selected. When participants began to complete the survey, I also gave them the opportunity to indicate that they did not

understand words. After they were assigned a word, they were allowed to indicate that they were not sure that they knew what the word meant, up to five times.

Some words were eliminated due to concerns about their usage being outdated. Examples of words with esoteric or outdated usages included “gay,” “dragged out,” “woebegone,” “grit,” “keyed up,” “clutched up,” and “full-of-pep.”

Adjectives vs. Nouns

Often, words appeared on different scales in different forms (e.g., “fear” and “afraid”). In a semi-arbitrary process, I decided to select the adjective form of words when possible. Often, the adjective forms were shorter versions. This approach was often helpful in eliminating a certain degree of ambiguity. Take, for example, the word “love.” When used in the sentence “I feel love,” it is unclear whether people are referring to the feeling of love that they feel towards another person or the feeling that another person feels toward them. Thus, “love” was split into the words “loved” and “loving”; “admiration” was split into “admired” and “admiring”; “bitterness” turned into “bitter”; “hate” and “hatred” into “hateful”; “playfulness” into “playful”; “queasiness” into “queasy”; and so forth. One exception to this rule, however, occurred when the adjective form was rare or may hinder understanding (for example, I kept the noun “grief” instead of the more unusual “grieved”, and “anticipation” instead of “anticipatory”).

Non-Feeling States

One exception was due to a different reason for elimination: Some words—for example, those connoting states versus traits—were not feeling states per se. Although many adjectives are both state and trait words, some adjectives describe long-term traits

that are not feeling states. Examples included “extraverted” (which also happened to be poorly understood by participants), “smart,” “conscientious,” “bad-tempered,” and “egotistic.” Feeling states of these words tend to indicate that a person momentarily had a belief about themselves, rather than a distinct, temporary feeling.

Synonyms: Thinning the Herd

In order to narrow the item pool to a manageable size, I deleted certain highly overlapping synonyms. Usually, this meant that a search for the definition of the word included a different word that was also on the list. In this case, I would choose the word that I felt was better understood, based on pilot data from Prolific indicating how unfamiliar participants were with words, commonality ratings from online dictionaries, and personal familiarity. For example, “sedate” was deleted in lieu of “calm”; “sunk” was deleted in lieu of “depressed”; “woeful” was deleted in lieu of “sad”; and “befuddled” was deleted in lieu of “confused.” Some words were simply extreme states of other words. In this case, I deleted the more extreme word. For example, “fuming” was deleted in lieu of “angry,” as “fuming” could be captured by “very angry,” but “angry” could not be captured as well by “somewhat fuming.”

Non-Feeling States: Actions

Certain words were eliminated because they were descriptions of actions or ways of acting rather than feelings. For example, “defiant,” “accommodating,” “giggling,” “animated,” “sulky,” “stubborn,” “snarky,” “boastful,” and “domineering” were removed for this reason. In this category, I especially paid attention to words that are typically descriptions of a person’s *behavior*, which has an accompanying feeling state. The

question that I posed for this purpose was how accurate an external judge could be. For example, whether or not someone was cooperative could be judged entirely without asking a person how they felt, thus this was judged to be a behavior rather than a feeling. Another example involves the words "chuckling" or "giggling," which are actually behaviors that indicate the feeling amusement. If an external judge rated someone as chuckling, but the person rated themselves as not chuckling, the two sources would be considered equally valid. On the other hand, if an external judge rated someone as amused, but the person rated themselves as *not* amused, the person would be considered a more valid source of information, so this would not be rated as a behavior. To list some other examples, "hostile" and "snarky" are the behaviors associated with feeling anger or contempt; "animated" is the behavioral cue of feeling excited or happy; "easygoing" is the behavior associated with being calm; "perseverance" is the behavior associated with being determined; "stubborn" is a way of acting associated with the cognition that one is right, or feeling superior; "friendly" is a way of treating others that may or may not indicate positive feelings; "grimacing" is usually an indication of feeling pain; and "silent" is a behavior associated with a number of different feeling states.

Non-Feeling States: Cognitions or Situational Stimuli

Analogous to non-feeling states that are actions, some words were eliminated because they tended to indicate truths or beliefs about a situation (or oneself) which may prompt feelings but are not necessarily themselves feelings. Words in this category included "ignored," "dumb," "dominant," "fortunate," "sense of failure," "unready," "exalted," "in danger," and "feeble." For example, a person may judge that they are in

danger, but this is a cognition about the situation that may or may not inspire fear, excitement (i.e., jumping out of a plane), or apathy (e.g., someone speeding or handling a poisonous snake may not be bothered by the danger).

A specific case in which such categorization might be problematic involves a word describing an objective fact about the situation, which also is a feeling. Take, for example, the word "alone." It is plausible that participants who were objectively alone, but not feeling lonely, might still rate themselves as feeling alone. I thus retained such words.

Choosing Between Synonyms

In some cases, multiple synonyms existed for a particular feeling state, and in order to narrow the list I had to choose between them. In such cases, I opted to delete the one that was less common or that could be defined as "very [other word]." For example, "agony" is an intense version of "pain." One can feel a little pain, or a lot of pain, but one cannot feel a little agony. Thus, pain was chosen for the scale because it included agony. Other examples of deleted items include "forlorn," which was defined by Merriam Webster as "bereft, forsaken, sad and lonely"; "furious," which is essentially "very angry"; "mortified," which is essentially "very embarrassed"; and "revulsion" was deleted for "disgusted." Essentially, in order to shorten the list, I attempted to eliminate words that were specific or rarely-felt subsets of other words, or words that are more rarely used (e.g., "jaded" was deleted in lieu of "cynical" and "satiated" in lieu of "satisfied").

Discussion

The creation of the item pool was unique, as typically scale creation involves writing new items, whereas this item pool was crafted by narrowing down an unmanageable number of potential items. The final item pool contained 354 total words, which were used in Study 1a and 1b.

Study 1

Overview of Study 1

The aim of Study 1 was to define the structure of the item pool and to develop a preliminary measure. Ideally, this item pool should be deep, balanced, and comprehensive, with all four quadrants of the arousal-valence circumplex represented. To that end, I collected two types of data. Study 1a includes data from three samples, in which participants rated the degree to which they felt each feeling state. Study 1b includes data from research assistants, rating various features of each feeling state.

Study 1a: Affect Questions

Overview of Study 1a

Study 1a was modeled after traditional studies used to create measures of affect. This study was conducted in order to define the structure of the words by identifying a set of prominent factors and to develop a preliminary item pool that could be used to form the scale. The goal of this study was to obtain a sample of responses in which participants indicated the extent to which they felt certain things in order to factor analyze these responses to determine which feelings hung together, statistically speaking.

Typically, studies of this kind ask participants to rate the extent to which they feel a given emotion or sensation within a given time period. For example, the PANAS can be used with seven different time instructions, asking participants how they feel/have felt/felt in the present moment, today, during the past few days, during the last few weeks, during the past year, or in general/on average (Watson & Clark, 1999; Watson, Clark & Tellegen, 1988). These time instructions break down into roughly two categories: during one moment (right now) or over a span of time (today, past few days, week, past few weeks, year, and in general). Using these instructions in order to discover which emotions hang together can be problematic. When participants rate how they feel "in the present moment," their responses are by definition limited to how they feel during a very specific and not particularly generalizable experience—that is, how they feel as they are answering survey questions about how they feel. Although it is possible that participants feel particularly strong positive or negative emotions during that experience, it is unlikely. Thus, the amount of variance to analyze is limited by the question itself. If a scale is created based mainly on instances in which strong feelings are present to a low to moderate degree, it could greatly limit the validity of the measure when these feelings are present to a high degree. Put more simply, a scale of feelings normed on a sample of participants experiencing stunted feelings might not be valid in a sample of participants experiencing feelings in more typical contexts.

On the other hand, instructions that include other possible moments—i.e., non-present-moment instructions—all include swaths of time. This approach could create problems for a factor analysis aiming to discover which emotions co-occur at the same

point in time. For example, it might be highly unlikely that, in the present moment, a person feels simultaneously hostile and loving. However, over the course of a week, these two emotions could easily co-occur. Another possibility is that if the prompt considers long swaths of time, the resultant scale will capture personality dimensions rather than temporary states. It might be that people feel positive and negative emotions simultaneously, but when looking across time we lose the granularity to observe this phenomenon, instead detecting a specific person's general tendency to feel specific emotions. Thus, a person high in neuroticism might report feeling guilty, scared, hostile, ashamed, and nervous, artificially causing these different negative feelings to cling together in a factor analysis in a way that they might not in moment-to-moment experiences.

In order to ameliorate both of these problems, in this study I decided to capture momentary experience during the experience of emotions. To that end, I derived 58 key emotions from existing surveys that were the central, overlapping emotions from these surveys, as well as important emotions from a recent paper (Cowen & Keltner, 2017). The point of this list was to capture a diverse range of human experiences. These words were as follows: active, admiration, adoration, appreciation of beauty, afraid, alert, amused, angry, anxious, ashamed, attentive, awed, awkward, bored, calm, confused, contented, craving, depressed, determined, disgusted, distressed, elated, empathic pain, entranced, excited, fatigued, guilty, happy, horrified, hostile, inspired, interested, irritable, jittery, joyful, lethargic, nervous, nostalgic, proud, relaxed, relieved, romantic,

sad, satisfied, scared, serene, sexual desire, stressed, strong, surprised, tense, upset, grateful, hopeful, and loving.

Participants

Study 1 included data from three different samples.

Sample 1: Subject Pool. The first sample included 360 participants recruited from the University of California, Riverside (UCR) psychology subject pool, ranging from 18 to 28 years old (M_{age} = 19.55, *SD* = 1.47). There were slightly more women than men in this sample (57.22% female, 41.94 % male, 0.28 % other; 0.55 % did not report gender). Like the campus student population, this sample was quite diverse; 85.83% reported a single ethnicity (38.61% Asian, 34.44% Hispanic, 7.78% White, 0.83% Black/African American, 0.28% American Indian or Alaskan Native, 0.28% Native Hawaiian, 3.61% other; all wrote in Middle Eastern), and 10.83% of participants reported multiple ethnicities—most commonly Hispanic & White (6.67%), Asian & White (1.94%), Black/African American & White (1.39%), Asian & Black (0.83%), and Asian & Hispanic (0.83%). Two participants declined to indicate any ethnicity (0.56%).

Sample 2: Prolific. The second sample included 92 participants from the website Prolific. This sample was quite different from the subject pool. Ages ranged from 18 to 57 (M_{age} = 26.94 , *SD* = 7.92). There were more men in this sample than women (48.91% male, 38.04% female, 2.17% other; 11.96% declined to indicate gender). The majority (79.35%) indicated that they were a single ethnicity (58.70% White/Caucasian, 9.78% Hispanic, 5.43% Asian, 2.17% Black/African American, 3.26% other); 6.52 % indicated multiple ethnicities, and 14.13% did not indicate any ethnicity. Participants were based in

many different countries: 25% indicated they were in the US, 9.78% indicated that they were based in the UK, 11.96% declined to answer the question, and 53.26% indicated that they were elsewhere (countries included Australia, Canada, Chile, Czech Republic, Denmark, Italy, Finland, Spain, Germany, Greece, Italy, Mexico, New Zealand, Ontario, Poland, Portugal, Spain and Sweden).

Sample 3: Volunteers. The third sample included 39 volunteer participants who were research assistants blind to the nature of this study. The ages of these research assistants (who were all college students or recent graduates) ranged from 19 to 30 (M = 21.68, SD = 2.22). This sample was predominantly women (5.12% male, 94.87% female, 2.56% did not indicate gender). The ethnic makeup of this sample was more reflective of the UCR population: The majority (74.35%) reported a single race (41.03% Hispanic, 20.51% Asian, 7.69% White/ Caucasian, 2.56% Black/ African American, and 2.56% other); 12.82% reported two or more ethnicities, and 7.69% did not report any ethnicity. The majority (92.31%) were in the United States; 7.69% did not report their location.

Measure

After giving demographic information, all participants (N = 491) were assigned one of the 58 target feeling states and then asked to "take a moment and think of a recent experience in which you felt [assigned word] to a moderate to high degree. Ideally, this experience will still be vivid enough in your memory that you can describe the situation and remember how you felt during it." Participants then affirmed that they had a clear memory of a time in which they felt the assigned feeling state to a moderate to high

degree. If they could not think of a specific instance or did not know what the word meant, they were assigned a new word (up to five times).

Next, participants were asked more questions about the instance during which they felt their assigned feeling state. This step served to address substantive research questions but also to help participants bring to mind a clear memory of the instance. They were asked when the feeling took place and then asked a series of open-ended questions about the experience. First, they gave a brief description of the situation, then described why they felt that way, including any specific causes. Next, participants were asked to describe what it felt like in their body to experience the feeling, then how it felt in their mind (emotionally or cognitively). Finally, they were asked about how they wanted to react to the feeling, and then how they in fact reacted. After the open-ended questions, participants were asked on Likert scales about how often they felt this feeling, how intense it usually was, and whether it was typically pleasant or unpleasant.

Once participants had answered these questions, they entered the main portion of the survey. They had to pass an instruction check by indicating that they understood that they were answering questions about how they felt *during the experience they just described.* They were also warned about how long the next portion would take.

During the final portion of the survey, participants were presented serially with each of 354 words, in a random order. For each word they were asked, “During the experience you just described, during which you felt [assigned word], to what extent did you feel [new word]?” (with responses ranging from 1 = *very slightly or not at all* to 5 = *extremely*, with an option for “I don't know what this word means”). They were also

given optional check boxes to indicate if they felt the new feeling state before or after feeling their originally-assigned feeling state.

Finally, participants were informed that their answers to the final questions would not affect their compensation. They were then given a chance to indicate whether they felt, for any reason, that their data were not high-quality enough to use and asked for any additional comments for the researchers.

Study 1b: Coding of Features

Study 1b was conducted to complement Study 1a. Typically during the creation of scales, an exploratory factor analysis yields factors (i.e., specific groups of items that tend to move together, sometimes loading positively and sometimes negatively onto the factor). One challenge of factor analyses is that identifying and naming the factors is often left up to the researcher, which leaves considerable room for subjectivity. The process of identifying factors involves the researcher looking at the items that load highly onto a factor and trying to identify similarities between them.

In order to make the process of classifying groups of words easier and less subjective, I decided to classify words in other ways to determine more objectively which features correlated highly with each factor. To that end, I created a coding scheme for research assistants to rate features of each feeling state in the dataset. Coded features included:

- the extent to which the feeling is experienced in one's body vs. one's mind
- the extent to which the feeling is an emotion
- how positive or negative the feeling is typically

- whether the feeling is high or low energy
- typical duration, frequency, onset, and dissipation of the feeling
- target and cause of the feeling (i.e., self, other person, non-person, no cause)
- experience for others nearby when one experiences the feeling
- whether the hallmark of the feeling is avoiding something bad, receiving something bad, missing out on something good, or receiving something good
- whether the feeling is typically past-, present-, or future-oriented
- if the feeling is related to reputation, competence, power, or connection
- typical urges associated with the feeling (e.g., toward action/inaction, creativity, approach/avoidance)
- social ramifications of experiencing the feeling

I employed the help of 55 undergraduate coders who each coded a single word at a time selected from a list of 354. The modal time to respond to the entire survey for a given word was roughly 19 minutes. The mean was much higher (roughly 124 minutes) due to a positive skew. That is, a few responses had an unusually long response time, likely due to the survey-taker pausing in the middle of responding and failing to close the tab; however, due to the nature of time data, there were no corresponding outliers with an extreme negative time. Each of the words were rated between 3 and 8 times (M = 3.21, *Mode* = 3), resulting in a total of 1545 ratings. The entire list and these results are beyond the scope of this data, but are available upon request.

Results of Studies 1a and 1b

Exploratory Factor Analysis

Because the goal of this set of analyses was to select items for my scale, I first conducted an exploratory factor analysis (EFA) on the data from Study 1a.

Due to the large number of items (356), the smaller samples were not large enough to run independent factor analyses, as the number of observations must be greater than the number of items. Thus, I combined all three samples into one (n = 539), and ran a single EFA with oblique rotation. The scree plot (Cattell, 1966) revealed that two factors were capturing most of the variance (see Figure 1). This result was especially interesting because despite the unusual items included (e.g., thirsty, drunk, engrossed), the two-factor solution found in previous studies was replicated.

EFA loadings were used in conjunction with research assistant ratings from Study 1b in order to trim down items for the preliminary scale. Past studies have selected the highest-loading items, without ensuring equal sampling from all four quadrants of the circumplex. In order to combat this problem, I used multiple questions from the research assistant ratings. For each feeling state, they indicated how typical it was for each state to feel pleasant, and, when it did feel pleasant, the extent to which it was pleasant (1 to 5 Likert-type scales). They responded to equivalent questions replacing pleasant with unpleasant, low-energy, and high-energy.

The "typical feeling" and "extent of feeling" questions were highly intercorrelated (average r = .91), and the correlation between opposites (i.e., pleasant/unpleasant, high energy/low energy) were correlated on average at r = -.75. Thus, these rating groupings were combined into two dimensions of pleasant-unpleasant and high energy-low energy. These dimensions were each standardized, and items were divided into four categories

(crossing valence and arousal) based on whether they were above or below average in each category. Interestingly, these categories were not equally sized; rather, they reflected a category distribution more typical to affect scales. There were 135 pleasant/high-arousal items ($M_{pleasant}$ = 3.94/5, $M_{high\text{-}arousal}$ = 3.98/5), 39 pleasant/low-arousal items ($M_{pleasant}$ = 3.75/5, $M_{high\text{-}arousal}$ = 2.39/5), 58 unpleasant/high-arousal items ($M_{pleasant}$ = 1.67/5, $M_{high\text{-}arousal}$ = 3.78/5), and 124 unpleasant/low-arousal items ($M_{pleasant}$ = 1.72/5, $M_{high\text{-}arousal}$ = 2.12/5).

I then compared these four categories of words to the results of the factor analysis. The factors did not map cleanly onto the circumplex quadrants; they clearly identified valence, but the arousal dimension did not emerge as clearly. In order to select items, I included 12 items from each category that had the highest loadings for each factor. I then removed some highly redundant words and re-inserted some unique words that have been of special interest to researchers (e.g., grateful, awed, loved, amused, jealous, afraid, jittery, tired, bored, hopeless, weary, rejected, confused). Tables 2-5 present the measure at this stage in its development.

Discussion

Studies 1a and 1b aimed to identify a manageable number of items (i.e., widely useable) with equal representation across the valence-arousal circumplex for a new measure of affect. To this end, these studies combined participant reports of co-occurrence between feeling states (Study 1a) with coder-rated characteristics of feeling states (Study 1b) to identify a maximally useful measure of affect. Although these questions (i.e., which emotions co-occur, and what factors emerge?) has been asked

multiple times in the creation of other affect measures, this study differs from previous studies in two main ways. First, prior studies have usually addressed either the emotions that people feel while completing a scale (i.e., "right now") or have collapsed emotions across time (e.g., "over the past week"). Second, the item pool has been far narrower and excluded many potentially interesting feeling states.

At this stage, the aim of item selection becomes about achieving a delicate balance between retaining items with high factor loadings and achieving good construct validity—that is, actually measuring what the scale is designed to—as well as other types of validity. Thus, I aimed to select high-loading items for each factor that still represented all four quadrants of the original circumplex. Using this abbreviated 76-item scale, the next step was to run a follow-up study to assess internal reliability and several types of validity (i.e., face validity, construct validity, content validity, and criterion validity).

Study 2: Validation

Method

This study included data from 334 participants recruited from the UCR subject pool, ranging from 18 to 35 years old (M_{age} = 19.39, SD = 2.10). There were slightly more women than men in this sample (57.69 % female, 46.41 % male, 0.60 % other; 0.30 % did not report gender). Like the campus student population, this sample was quite diverse; 82.04% reported a single ethnicity (40.42% Asian, 34.43% Hispanic, 5.39% White, 1.20% Black/African American, 0.60 % other); 12.87% of participants reported multiple ethnicities, most commonly Hispanic & White (4.19%) and Black/African

American and White (1.50%); and 5.09% of participants declined to indicate any ethnicity.

Validation Measures

Big Five Inventory II Short Form. The Big Five Inventory II (BFI-2-S; Soto & John, 2017) was developed to measure facets in each of the Big Five personality domains. Each of its 30 items include a short description (e.g., "__is outgoing, sociable", "__can be cold and uncaring"). Participants are instructed to rate the extent to which the characteristic applies to them, on a scale ranging from *disagree strongly* to *agree strongly.* In this sample, Cronbach's alpha for each personality trait was acceptable (Conscientiousness = 0.74, Agreeableness = 0.76, Neuroticism = 0.84, Openness = 0.65, Extraversion = 0.79).

Positive and Negative Affect Schedule – Expanded Form. The Positive and Negative Affect Schedule (PANAS - X; Watson & Clark, 1999; Watson, Clark & Tellegen, 1988) was developed based on the structure of two independent dimensions of affect. The PANAS-X consists of 60 adjectives. Participants are instructed to indicate the extent to which they felt a particular way during some given frame of time (here, we used the past week) on a 5-point scale ranging from 1 = *Very slightly or not at all* to 5 = *Extremely*. Cronbach's alpha was quite high across negative words (0.95) and positive words (0.95), when calculated including only the positive and negative affect items and excluding "other affective states" (items mapping onto shyness, fatigue, serenity, and surprise).

Affect Adjective Scale. The Affect Adjective Scale (AAS; Diener & Emmons, 1985) is a 9-item measure of emotions (happy, worried/anxious, pleased, angry/hostile, frustrated, depressed/blue, joyful, unhappy, and enjoyment/fun), which participants rated on a 7-point scale from 1 = *Not at all* to 7 = *Extremely Much*. Some commonly added adjectives were included beyond the 9-item version of the measure (peaceful/serene, dull/bored, relaxed/calm). Cronbach's alpha was high for this 12-item scale both for positive emotions (0.92) and negative emotions (0.85)

Procedure

For the preliminary items for the brief affect measure, participants were asked to indicate the extent to which they "have felt this way during the past week" on a five-point scale ranging from 1 = *Very slightly or not at all* to 5 = *Extremely.* After completing these items, participants completed the BFI II short form, the PANAS, the Affect Adjective Scale, and finally demographics.

Results

I began by narrowing the item pool. I first eliminated words that were very infrequently indicated. To this end, I initially considered using the overall mean for each feeling state included in my measure. However, that approach would have confounded frequency with intensity. Thus, I decided to instead count the number of "1s" indicated for each state, indicating that the word was felt *very slightly or not at all.* This was collected from both the study 2 dataset, but also from the study 1a dataset. The frequency of such responses was not equal for all four quadrants (see Tables 1-4 for the frequency of *very slightly or not at all* responses for each item). Using these frequencies, I

eliminated words from each of the four categories (i.e., valence/arousal quadrants) until each category had 14 words left.

One issue with this method is that the frequency of words might have been largely dictated by the circumstances of the sample; as college students nearing the end of the quarter, their highest reported words included tired, stressed, bored, anxious, overwhelmed, worried, grateful, happy, loved, sleepy, safe, and generous. On the other hand, their less frequent words in each category included relaxed, relieved, at ease, refreshed, awed, rejuvenated, rejected, miserable, despair, resentful, jealous, and hateful. Thus, I did not want to eliminate too many words using this method.

To this end, I included another frequency metric. The Google Books Ngram Viewer charts the frequency of words appearing in books by year, tracking the usage of words in the language (Lin et al., 2012; Michel et al., 2010). The most recent year with data available is 2019. These numbers are visible in Tables 2-5. Across all words, the Ngram frequencies correlated with the frequency of the *very slightly or not at all* responses at $r = -.24$. This negative correlation made sense, as the Ngram frequencies increased as words were mentioned more frequently, and the *very slightly or not at all* responses decreased as words were reported more frequently.

Next, I aimed to eliminate words that were overly redundant. Even though this approach would likely reduce the internal reliability of the scale, it was crucial to include a representative range of items. The survey included one redundant word (happy), which had an autocorrelation of $r = .84$. This fortuitous error provided a point of reference for a definition of redundancy. I identified pairs of items that had correlations over .70 and

inspected them for similarity. Table 6 displays the item pairs with correlations of over .70. Notably, most of these pairs were positive, high-arousal feeling states (delighted, joyful, happy, uplifted, pleased, cheerful, and blissful were all highly intercorrelated).

Finally, I ran an exploratory factor analysis on the items (see Table 7 for a factor analysis on all initial items) to examine which items loaded the most highly onto the factors. Using a balance all of these metrics, I selected 12 words for the final scale (see Table 8 for factor loadings from an EFA of only the final items). Some items were dropped for low alpha reliabilities, redundancy, and lower relevance.

Internal Reliability

The subscales of what I have dubbed the Brief Affect Measure (BAM) showed high internal reliability. The three high-arousal negative items (frustrated, anxious, overwhelmed) had a Cronbach's alpha of 0.83. The three low-arousal negative items (sad, uneasy, unhappy) had an alpha of .84. All six negative items had an alpha of .88.

The three high-arousal positive items (happy, grateful, enthusiastic) had an alpha of .75. The three low-arousal positive items (comfortable, calm, relieved) had an alpha of .76. All six positive items had an alpha of .84.

Discriminant and Convergent Validity

Discriminant and convergent validity were tested using the secondary measures included in Study 2. The Affect Adjective Scale (AAS) and Positive and Negative Affect Schedule (PANAS) were used to test convergent validity, and the Big Five Inventory 2 Short Form (BFI-2-S) was used to test discriminant validity against the BAM.

My brief affect measure showed good convergent validity for both the negative and positive items. The AAS negative items correlated with the negative items of the BAM at r = .83 (p < .001) and with the positive items at r = -.55 (p < .001). The AAS positive items correlated with the negative items of the brief affect measure at r = .82 and with the negative items at r = -.53 (p < .001). The PANAS negative items were correlated with the negative items of the brief affect measure at r = .83 (p < .001) and with the positive items at r = -.49 (p < .001). The PANAS positive items were correlated with the positive items of the brief affect measure at r = .77 (p < .001) and with the negative items at r = -.40 (p < .001).

Similarly, in addition to correlating with the expected BFI-2-S subscales (i.e., N & E), the BAM showed good discriminant validity with the scales of the BFI-2-S that were not expected to be correlated with affect. Taken together, the BFI-2-S traits each showed expected correlations with the brief affect measure. The BFI Conscientiousness scale was correlated with the BAM negative items at r = -.28 (p < .001) and the positive items at r = .33 (p < .001). This mapped closely onto findings from a 2020 meta-analysis, which found that conscientious correlates with negative affect at r = -.25 and positive affect at r = .35 (Anglim et al., 2020). The BFI Agreeableness scale was correlated with the BAM negative items at r = -.08 (p = .14) and the positive items at r = .22 (p < .001). Agreeableness is not traditionally considered to be closely related to affect, and indeed, the correlations in this study were weaker than the meta-analytic findings relating agreeableness with negative affect at r = -.25 and with positive affect at r = .19 (Anglim et al., 2020).

Neuroticism was most strongly related to scores on the BAM; it correlated with negative items at $r = .65$ ($p < .001$), and positive items at $r = -.53$ ($p < .001$). This finding again mirrored the meta-analytic findings, which were $r = .56$ and $r = -.34$, respectively (Anglim et al., 2020). Although the correlations in the present study are stronger, they do not discount discriminant validity because negative affect and positive affect are theoretically very strongly related to negative emotionality. Openness had very low correlations with scores on the BAM; it was correlated with positive affect at $r = .05$ ($p = .39$) and negative affect at $r = -.05$ ($p = .37$). These correlations were lower than meta-analytic findings of $r = .20$ and $r = -.05$, respectively, but this finding supported discriminant validity as openness is not theoretically related to positive or negative affect (Anglim et al., 2020). Finally, extraversion was moderately related to scores on the BAM; it correlated with negative affect at $r = -.19$ ($p < .001$) and positive affect at $r = .36$ ($p < .001$). These findings were similar to the meta-analytic findings of $r = -.21$ and $r = .44$, respectively (Anglim et al., 2020). For a full correlation matrix between the BFI facets, PANAS, AAS, and brief affect scale positive high, positive low, negative high, and negative low items, see Table 9.

Discussion

Many short measures of affect are available in the literature. A primary reason that many such measures have not been widely adopted may be that they do not cover all four quadrants of the widely accepted affective circumplex (i.e., high arousal positive emotions, low arousal positive emotions, high arousal negative emotions, low arousal negative emotions), or they contain affective states that are too general. The present

research pursued the goal of creating a short affect measure that represents all four quadrants of the affective circumplex, reduces redundancy, and retains good measurement properties.

To this end, previous measures of affect and lists of words from previous affect studies were collected, and the items from them were combined. This list was manually refined based on a number of criteria, such as eliminating double-barreled items, rarely used items, non-feeling states, and redundant items, resulting in a list of 354 feeling states.

In Study 1a, multiple samples of participants were prompted to think of a time when they felt specific feelings, then recalled the extent to which they felt each of the 354 feeling states during that time. In Study 1b, research assistants coded each of the feeling states on a number of dimensions, including the pleasantness or unpleasantness of each feeling and whether the feeling is characterized by high or low energy. The coding data were normed, and each word was rated as above or below average on pleasantness and above or below average on energy. These ratings were used to sort words into four categories corresponding to the four quadrants of the affective circumplex. An exploratory factor analysis on the data from Study 1a revealed two factors. The highest-loading items from these two factors were included in Study 2. These items were selected in order to meet a quota of items from each of the four quadrant categories. The resultant item pool contained 76 items.

Study 2 aimed to validate the measure and provide some indices of internal reliability. In this study, participants indicated the extent to which they had felt each of

the feelings over the past week. An exploratory factor analysis was conducted on all 76 items, which again yielded two factors. In order to eliminate redundant items, inter-correlations between items were inspected, and those with very high correlations were not both included in the final scale. In an effort to include only items that were frequently felt in day-to-day life, data from Google Ngram was included to capture usage in the lexicon. In addition, the extent to which items endorsed by Study 2 participants was considered, in combination with the Ngram data. Using all of these metrics, as well as researcher judgment, three items from each of the four quadrants were chosen.

The resultant scale, dubbed the brief affect measure (BAM) has 12 items. Correlations with other affect scales and the BFI-2-S revealed high convergent validity and discriminant validity, respectively. The scale also demonstrated high internal consistency reliability. The final scale comprises a measure that includes items from all four quadrants and has good measurement properties—thus fulfilling the aims of this dissertation.

Future directions for this scale are to replicate these finding in an independent sample, as well as to validate the scale with different samples (i.e., other than college samples), with other time instructions, and using different affect induction procedures. This measure was developed with participants rating the extent to which they have felt "this way" during the past week, but it would be useful to validate this measure with other instructions used with the PANAS or other measures, such as in the present moment, today, during the past few days, during the last few weeks, during the past year, or in general/on average (Watson & Clark, 1999; Watson, Clark & Tellegen, 1988). A

longer version of this scale could also be developed with the possible additions of other subscales (e.g., transcendent emotions, other-focused emotions, self-focused emotions).

Constraints

One constraint inherent to the development of the measure is that the range of emotions ultimately considered for the scale was limited by the method in Study 1a of asking participants to think of a specific time they felt a target feeling, which could be influenced by memory biases. Similarly, Study 1b asked coders to consider feeling states generally, not in a specific incident, which may bring to bear other biases. I used different approaches in different phases of scale development in an effort to balance the pros and cons of each approach. Nonetheless, no method is perfect.

Another constraint stems from the length of the original study. Participants might have been unmotivated to try hard, or have found the first study egregiously long. This burden was partially offset by the fact that the words were randomized, so even if participants gave up in the middle and began randomly clicking, it would not systematically affect the results. Rather, it would just introduce noise. Again, the issue of participant burden was balanced against the need for comprehensive responses across a large array of feeling states.

Another notable constraint that is especially relevant to the discussion of emotions includes culture. As discussed, culture has been central to discussion of discrete emotions (i.e., Ekman et al., 1987). Namely, the elicitation, categorization and expression of emotions differs substantially across cultures (Barrett et al., 2019). This paper falls into a common trap of focusing on a relatively small proportion of the world's population that

lives in Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies (Henrich, Heine, & Norenzayan, 2010). The samples in this study were more diverse than many of the typical samples of college sophomores, as the University of California Riverside comprises a diverse population including students with ethnic backgrounds from non-western countries, and the Prolific sample included individuals from all over the world (over 50% indicated that they were in neither the US nor the UK, although many were from other western cultures). Despite seeking diversity, the samples used here were still all united by their ability to speak English, meaning that they all have a strong link to the western world, and the university sample was still more educated than the majority of the world's population. Thus, the emotions selected and felt by these participants will still largely reflect WEIRD populations. It is possible that this specific set of emotions is more commonly felt in the day-to-day experiences of such populations than in other populations. For example, afraid and anger were both infrequently felt by this population, and explained less variance than did more mild terms such as anxious, frustrated, or irritated. The lower frequency of the more extreme terms might reflect the emotional reaction to an environment where immediate life-and-death situations are relatively infrequent, and where open displays of rage-provoking violence are relatively rare. Not only may other populations may experience different emotions more regularly, but they may also categorize emotions differently altogether. Different cultures have different culturally constructed word which are not present in other languages. For example, the word *Heimat* is present in German, referring to a "deep-rooted fondness towards a place to which one has strong feelings of belonging" (Anthes, 2016).

Perhaps, because emotions are not entirely universal, a truly inter-cultural measure is an infeasible goal. Thus, this measure provides a first step of creating a single measure for a single sample. This measure may be expanded or adapted to fit other samples, including different groups of English-speakers (perhaps, who are less educated), non-English speakers, or those from non-Western cultures. Or, it is possible that for each new culture, a new measure will need to be created to best capture the feelings created by the day-to-day experiences of the members of that culture.

Conclusion

In conclusion, this study contributes a practical measure for researchers who would like to use a short affect scale that captures the full valence-arousal circumplex. This measure attempts to strike a balance between respecting existing traditions (i.e., discrete and dimensional), making a measure that is actually useful to researchers, and making a measure that is data-driven. Importantly, this measure will be freely available to researchers. Time will tell whether I have been successful at creating an appealing measure, but my findings suggest that at the very least, the BAM succeeds in meeting my original aims.

References

Anglim, J., Horwood, S., Smillie, L. D., Marrero, R. J., & Wood, J. K. (2020). Predicting psychological and subjective well-being from personality: A meta-analysis. *Psychological Bulletin*, *146*(4), 279-323. https://doi.org/10.1037/bul0000226

Anthes, E. (2016, May 12). *The Glossary of Happiness.* The New Yorker: Annals of Technology. https://www.newyorker.com/tech/annals-of-technology/the-glossary-of-happiness

Aslin, R. N., & Newport, E. L. (2012). Statistical learning: From acquiring specific items to forming general rules. *Current Directions in Psychological Science*, *21*(3), 170–176. https://doi.org/10.1177/0963721412436806

Averill, J. R. (1975). A semantic atlas of emotional concepts. *JSAS Catalog of Selected Documents in Psychology*, *5,* 330.

Averill, J. R. (1980). A constructivist view of emotion. In R. Plutchik & H. Kellerman (Eds.), *Theories of emotion* (pp. 305–339). Elsevier. https://doi.org/10.1016/B978-0-12-558701-3.50018-1

Barrett, L. F. (2016). The theory of constructed emotion: An active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*, 1-23. https://doi.org/10.1093/scan/nsw154

Barrett, L. F., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74,* 967–984. https://dx.doi.org/10.1037/0022-3514.74.4.967

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, *20*, 1-68, https://www.doi.org/10.1177/1529100619832930

Boyle, G. J. (1992). Multidimensional Mood State Inventory (MMSI). Department of Psychology, University of Queensland, St. Lucia, Queensland, Australia (Revised 2012).

Bradburn, N. M. (1969). *The structure of psychological well-being*. Aldine.

Brim, O. G., & Featherman, D. L. (1998). Surveying midlife development in the United States. Unpublished Manuscript.

Buss, A. H., & Durkee, A. (1957). An inventory for assessing different kinds of hostility. *Journal of Consulting Psychology, 21(4),* 343–349. https://doi.org/10.1037/h0046900

Buss, A. H, & Murray, E. N. (1965). Activity level and words connoting mood. *Perceptual and Motor Skills*, *21*, 684-686. https://doi.org/10.2466/pms.1965.21.3.684

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research, 1*(2), 245-276. https:/doi.org/10.1207/s15327906mbr0102_10

Cowen, A. S., & Keltner D. (2017). Self-report captures 27 distinct categories of emotion bridged by continuous gradients. *PNAS, 114*(38), E7900-E7909. https://doi.org/10.1073/pnas.1702247114

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, 7, 309-319. https://dx.doi.org/10.1037/1040-3590.7.3.309

Cohen, T. R., Wolf, S. T., Panter, A. T., & Insko, C. A. (2011). Introducing the GASP scale: A new measure of guilt and shame proneness. *Journal of Personality and Social Psychology, 100*(5), 947–966. https://doi.org/10.1037/a0022641

Curran, J. P. & Cattell, R. B. (1975). Manual for the eight state questionnaire (8SQ). Champaign, Illinois: Institute for Personality and Ability Testing.

Darwin, C. (1859). *On the origin of species by means of natural selection, or preservation of favoured races in the struggle for life.* John Murray.

Darwin, C. (1872). *The expression of the emotions in man and animals.* John Murray.

Diener, E. & Emmons, R. A. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology*, *47*, 1105-1117. https://doi.org/10.1037/0022-3514.47.5.1105

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*(2), 143-156. https://doi.org/10.1007/sll205-009-9493-y

Ekman, G. (1955). Dimensions of emotion. *Nordisk Psykologi*, *7*(3-4), 103-112. https://doi.org/10.1080/00291463.1955.10780038

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

Ekman, P. (1992b). Facial expressions of emotion: New findings, new questions. *Psychological Science, 3,* 34-38.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17,* 124–129. https://doi.org/10.1037/h0030377

Ekman, P., & Friesen, W. V. (1978). *Facial action coding system*. Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Ancoli. (1980.). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39*(6), 1125-1134. https://doi.org/10.1037/h0077722

Ekman, P., Friesen, W. V., O'Sullivan, M., Chan, A., Diacoyanni- Tarlatzis, I., Heider, K., . . . Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology, 53,* 712–717. https://doi.org/10.1037/0022-3514.53.4.712

Fredrickson, B. L., Tugade, M. M., Waugh, C. E., & Larkin, G. R. (2003). What good are positive emotions in crises? A prospective study of resilience and emotions following the terrorist attacks on the United States on September 11, 2001. *Journal of Personality and Social Psychology, 84*(2), 365–376. https://dx.doi.org/10.1037/0022-3514.84.2.365

Frijda, N. H. (1987). Emotion, cognitive structure, and action tendency. *Cognition and Emotion*, *1*(2), 115–143.

Gross, M. M., Crane, E. A., & Fredrickson, B. L. (2010). Methodology for assessing bodily expression of emotion. *Journal of Nonverbal Behavior, 34,* 223-248. https://doi.org/10.1007/s10919-010-0094-x

Hardin, G. (1960). The competitive exclusion principle. *Science*, *141*(3409), 1292-1297. https://www.jstor.org/stable/1705965

Harmon-Jones, C., Bastian, B., & Harmon-Jones, E. (2016). The discrete emotions questionnaire: A new tool for measuring state self-reported emotions. *PLoS ONE. 11*(8): e0159915. https://doi.org/10.1371/journal.pone.0159915

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2-3), 111-135. https://www.doi.org/10.1017/s0140525xx10000725

Heuchert, J. P., & McNair, D. M (2012). Profile of Mood States, 2nd Edition: POMS 2. North Tonawanda, NY: Multi-Health Systems, Inc. https://doi.org/10.1177/0734282913505995

Irvin, F. S. (1967). Adjective check list correlates of positive and negative sentence completion responses. *Journal of Clinical Psychology*, *23*(4), 419-421. https://doi.org/10.1002/1097-4679(196710)23:4<419::aid-jclp2270230405>3.0.c0;2-w

Izard, C. E., Libero, D. Z., Putnam, P., & Haynes, O. M. (1993). Stability of emotion experiences and their relations to traits of personality. *Journal of Personality and Social Psychology*, 64, 847-860. https://doi.org/10.1037/0022-3514.64.5.847

Izard C. E., Dougherty F. E., Bloxom, B. M., & Kotsch, N. E. (1974) *The differential emotions scale: A method of measuring the subjective experience of discrete emotions.* Vanderbilt Univ. Press: Nashville, Tenn.

Jacobs Bao, K. (2012). The course of well-being in romantic relationships: Predicting positive affect in dating participants. *Psychology, 3*(12A), 1091–1099. https://doi.org/10.4236/psych.2012.312A161

Kaplan, R. M., & Saccuzzo, D. P. (2005). *Psychological testing: Principles, applications, and issues* (6th ed.). Belmont, CA: Thomson Wadsworth.

Keltner, D., & Gross, J. J. (1999). Functional accounts of emotions. *Cognition & Emotion*, *13*(5), 467–480. https://doi.org/10.1080/026999399379140

Ketai, R. (1975). Affect, mood, emotion, and feeling: Semantic considerations. *American Journal of Psychiatry, 132*(11), 1215–1217. https://doi.org/10.1176/ajp.132.11.1215

Laurent, J., Catanzaro, S. J., Joiner, T. E. J., Rudolph, K. D., Potter, K. I., Lambert, S., … Gathright, T. (1999). A measure of positive and negative affect for children: Scale development and preliminary validation. *Psychological Assessment*, 11(3), 326-338. https://doi.org/10.1037/1040-3590.11.3.326

Lin, Y., Michel, J., Aiden, E., Orwant, J., Brockman, W., & Petrov, S. (2012). Syntactic Annotations for the Google Books Ngram Corpus. *Proceedings of the 50th Annual Meeting of the Association for Computational Linguistics.*

Lubin, B., & Zuckerman, M. (1999). *Manual for the MAACL-R: Multiple Affect Adjective Checklist-Revised.* San Diego, CA: Educational and Industrial Testing Service.

Lyubomirsky, S., & Lepper, H. S. (1999). A measure of subjective happiness: Preliminary reliability and construct validation. *Social Indicators Research, 46,* 137–155. https://doi.org/10.1023/A:1006824100041

Matthews, G., Jones, D. M., & Chamberlain, A. G. (1990). Refining the measurement of mood: The UWIST Mood Adjective Checklist. *British Journal of Psychology, 81,* 17-42.

Matthews, G., Hillyard, E. J., & Campbell, S. E. (1999). Metacognition and maladaptive coping as components of test anxiety. *Clinical Psychology and Psychotherapy*, *6*, 583-626.

Matthews, G., Hillyard, E. J., & Campbell, S. E. (2000). Personality and emotion: Cognitive science perspectives. In S.E. Hampson (Ed.), *Advances in personality psychology, 1*, 199-237.

McCullough, M. E., Emmons, R. A., & Tsang, J.-A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology*. 82(1), 112-127. https://doi.org/10.1037//0022-3514.82.1.112

McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual: Profile of mood states*. San Diego, CA. Educational and Industrial Testing Service.

Michel, J-B., Shen, Y.K., Aiden, A.P., Veres, A., Gray, M.K., Brockman, W., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A., & Aiden, E. L. (2010). Quantitative analysis of culture using millions of digitized books. *Science*, *331*(6014), 176-182.

Nowlis, V. (1965). Research with mood adjective checklist. In S. S. Tomkins & C. E. Izard (Eds.), *Affect, cognition, and personality: Empirical studies*. Springer-Verlag

Osgood, C. E. (1966). Dimensionality of the semantic space for communication via facial expressions. *Scandinavian Journal of Psychology*, *7*, 1-30. https://doi.org/10.1111/j.1467-9450.1966.tb01334.x

Osgood, C. E. (1969). On the whys and wherefores of E, P, and A. *Journal of Personality and Social Psychology*, *12*(3), 194–199. https://doi.org/10.1037/h0027715

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*(6), 1161–1178. https://doi.org/10.1037/h0077714

Russell, J. A. (1978). Evidence of convergent validity on the dimensions of affect. *Journal of Personality and Social Psychology*, *36*(10), 1152-1168. https://doi.org/10.1037/0022-3514.36.10.1152

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*(5), 805–819. https://doi.org/10.1037/0022-3514.76.5.805

Russell, J. A., & Mehrabian, A. (1977). Evidence for a three-factor theory of emotions. *Journal of Research in Personality*, *11*(3), 273-294. https://doi.org/10.1016/0092-6566(77)90037-X

Schnall, S., Roper, J., & Fessler, D. M. T. (2010). Elevation leads to altruistic behavior. *Psychological Science, 21*(3), 315-320. https://doi.org/10.1177/0956797609359882

Shiota, M. N., Neufeld, S. L., Danvers, A. F., Osborne, E. A., Sng, O., & Yee, C. I. (2014). Positive emotion differentiation: A functional approach. *Social and Personality Psychology Compass*, *8*(3), 104–117. https://doi.org/10.1111/spc3.12092

Simms, L. (2008). Classical and modern methods of psychological scale construction. *Social and Personality Psychology Compass*, *2*, 414-433. https://doi.org/10.1111/j.1751-9004.2007.00044.x

Simms, L. J., & Watson, D. (2007). The construct validation approach to personality scale construction. In R. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 240-258). New York: Guilford.

Sjöberg, L., Svensson, E., & Persson, L.-O. (1979). The measurement of mood. *Scandinavian Journal of Psychology, 20*(1), 1-18. https://doi.org/10.1111/j.1467-9450.1979.tb00677.x

Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology*, *48*(4), 813-838.

Soto, C. J., & John, O. P. (2017). Short and extra-short forms of the Big Five Inventory-2: The BFI-2 and BFI-2-XS. *Journal of Research in Personality*, *68*, 69-81.

Spencer, H. (1899). *The Principles of psychology by Herbert Spencer, Volume II.* D. Appleton and Company.

Spielberger, C.D., Ritterband, L. M., Sydeman, S. J., Reheiser, E. C., & Unger, K. K. (1995). Assessment of emotional states and personality traits: measuring psychological vital signs. In J. N. Butcher (Ed.), *Clinical personality assessment: Practical approaches* (pp. 42-58). New York: Oxford University Press.

Spielberger, C. D., & Reheiser, E. C. (2009). Assessment of emotions: Anxiety, anger, depression, and curiosity. *Applied Psychology: Health and Well-being, 1,* 271-302. https://doi.org/10.1111/j.1758-0854.2009.01017.x

Thayer, R. E. (1986). Activation-Deactivation Adjective Check List (AD ACL): Current overview and structural analysis. *Psychological Reports, 58,* 607-614.

Thayer, R.E. (1989). *The biopsychology of mood and arousal.* New York: Oxford University Press.

Timmermans, T., Van Mechelen, I., & Nezlek, J. B. (2009). Individual differences in core affect reactivity. *Personality and Individual Differences, 47,* 510-515. https://doi.org/10.1016/j.paid.2009.05.002

Tomkins, S. S. (1981). The quest for primary motives: Biography and autobiography of an idea. *Journal of Personality and Social Psychology, 41,*306-329.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*(2), 219-235. https://doi.org/10.1037/0033-2909.98.2.219

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

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the positive and negative affect schedule-expanded form.* https://doi.org/10.17077/48vt-m4t2

Weidman, A. C., Steckler, C. M., & Tracy, J. L. (2016). Jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. *Emotion, 17*(2), 267-295. https://doi.org/10.1037/emo0000226

Wendt, G. R., Cameron, J. S., & Specht, P. G. (1962). Chemical studies of behavior: VI. Placebo and Dramamine as methodological controls, and effects on moods, emotions, and motivations. *The Journal of Psychology: Interdisciplinary and Applied*, *53*(2), 257-279. https://doi.org/10.1080/00223980.1962.9916571

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic/nomothetic analysis. *Journal of Personality and Social Psychology, 43*(1), 111-122. https://doi.org/10.1037/0022-3514.43.1.111

Zuckerman, M., & Lubin, B. (1965). The Multiple Affect Adjective Check List. San Diego, CA: Educational and Industrial Testing Service.

Zuckerman, M., & Lubin, B. (1985). Manual for the Revised Multiple Affect Adjective Check List. San Diego, CA: Educational and Industrial Testing Service.

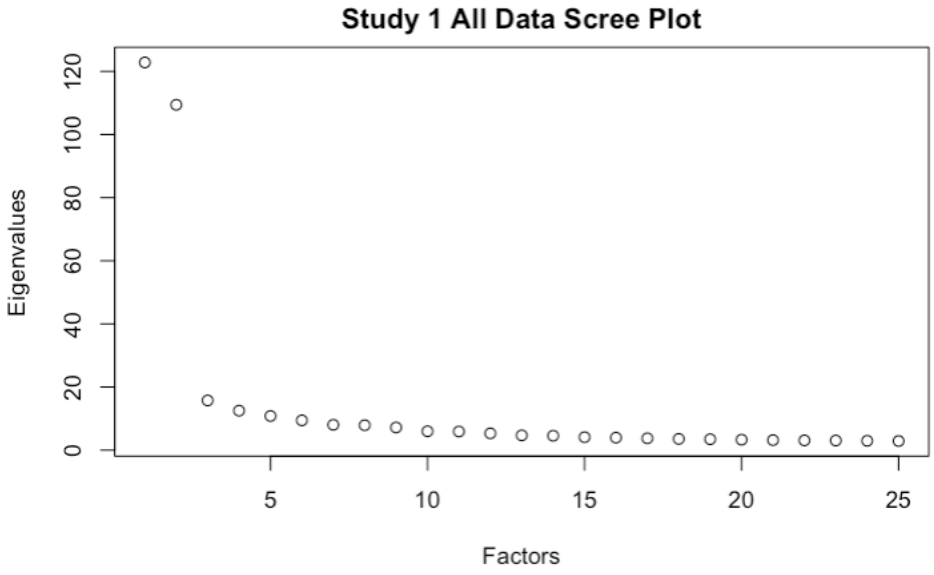

Figure 1. **Study 1a Scree plot of factors in combined data.**

Table 1
Common Affect Measures and Other Sources of Items

Source	Measure Name	Description
Boyle (1992, revised 2012)	Multidimensional Mood-State Inventory (MMSI)	Multi-part measure including different scales for various mood states, including Arousal-Alertness, Anger/Hostility, Neuroticism, Extraversion, Curiosity. Participants indicate on a scale from 1 = *Not at All* to 4 = *Very Much So* the extent to which they feel various things related to each mood state. For example, for the curiosity scale asks " do you feel ___" , with items including " like asking questions" , " curious about new developments in science and arts" , " like improving your general knowledge" , " like reading newspapers" , " like imagining yourself as an investigative reporter" , " like doing quizzes, crosswords, and puzzles" .
Bradburn (1969)	Bradburn Affect Balance Scale	This 10-item scale asks participants to indicate over the past few weeks the extent to which they felt a series of items, including " particularly excited or interested in something" , " so restless you couldn' t sit long in a chair" , " bored" . Respondents check whether the item is " applicable" or " not applicable.
Brim & Featherman (1998)	Midlife Development Inventory (MIDI)	This 12-item measure was specifically developed for the MIDUS survey. Participants indicate on a Likert scale from 1 = *None of the time* to 5 = *All of the Time* how much of the time in the past 30 days the extent to which they felt various things. Items mapped onto Negative Affect (" so sad nothing could cheer you up" , " nervous" , " restless or fidgety" , " hopeless" , " that everything was an effort" , " worthless") and Positive Affect (" cheerful" , " in good spirits" , " extremely happy" , " calm and peaceful" , " satisfied" , " full of life").

Table 1 (Cont.)

Source	Measure Name	Description
Buss & Durkee (1957)	Hostility-Guilt Inventory	This 60-Item inventory includes 60 regular-coded items and 15 reverse-coded items. Items are organized in groups of 5-10 onto sub-scales targeting different facets of hostility and guilt (Assault, Indirect, Irritability, Negativism, Resentment, Suspicion, Verbal, and Guilt). Items comprise True/False statements related to that facet (e.g., Indirect: "I sometimes spread gossip about people I don't like", "I sometimes pout when I don't get my own way"; Suspicion: "I know that people tend to talk about me behind my back", "I tend to be on my guard with people who are somewhat more friendly than I expected").
Buss & Murray (1965)	-	This paper include a list of 90 words used to identify intensity and activity (e.g., grieved, morbid, blue, doleful, lethargic, bland, blithe, earnest, giddy, gay)
Cohen, Wolf, Panter, & Insko (2011)	Guilt and Shame Proneness Scale (GASP)	This 16-item measure includes a number of situations that individuals may encounter in their day-to-day life. Participants rate each item on a Likert scale from *1 = Very Unlikely* to *7 = Very Likely.* There are four subscales (Guilt-Negative Behavior Evaluation (NBE), Guilt-Repair, Shame-Negative-Self-Evaluation (NSE), and Shame-Withdraw), each of which has four items. Example items include "You lie to people but they never find out. What is the likelihood that you would feel terrible about the lies you told?" (Guilt-NBE), and "Your home is very messy and unexpected guests knock on your door and invite themselves in. What is the likelihood that you would avoid the guests until they leave?" (Shame-Withdraw).
Cowen & Keltner (2017)	-	This paper includes 27 emotion states thought to be the dimensions of emotions, (e.g., admiration, adoration, aesthetic appreciation, craving, disgust, empathic pain, interest, sexual desire, triumph).

Table 1 (Cont.)

Source	Measure Name	Description
Curran & Cattell (1976)	Eight State Questionnaire (8SQ)	This battery includes items comprising eight states (anxiety, stress, depression, regression, fatigue, guilt, extraversion, arousal).
Diener & Emmons, (1985)	Affect Adjective Scale (AAS)	9-item scale, comprised of four positive states (Happy, Joyful, Enjoyment/Fun, Pleased) and five negative states (Worried/Anxious, Angry/Hostile, Frustrated, Depressed/Blue, Unhappy), rated on a 7-point scale from *Not Typical at All* to *Very Typical.* Sometimes used with a final item, "Compared to most weeks, how typical has this past week been?" Has been used with low-arousal items (Peaceful/Serene, Dull/Bored, Relaxed/Calm).
Diener et al. (2010)	Scales of Positive and Negative Experiences (SPANE)	12-item measure asking participants to rate items on a 5-point scale from 1 = *Very Rarely or Never* to 5 = *Very Often or Always.* Items include very general positive and negative feelings (i.e., Positive, Negative, Good, Bad, Pleasant, Unpleasant, Happy, Sad, Afraid, Joyful, Angry, Contented). Positive and negative feelings can be calculated separately, and affect balance can be calculated by subtracting negative from positive.
Feldman Barrett, & Russell (1998).	-	In this paper, 20 items are presented as an example of the structure of affect (Activation, alert, excited, elated, happy, pleasant, contented, serene, relaxed, calm, deactivation, fatigued, bored, depressed, sad, unpleasant, upset, stressed, nervous, tense). Items were stripped and added to the item pool. In addition, the Appendix provides multiple items used to measure affect, including adjectives with a 5-point Likert scale (1= *not at all,* 5 = *extremely).* The appendix also includes various statements to indicate feeling states (e.g., "I am having some trouble paying attention" (unpleasant deactivated), and "I'm bothered by something" (unpleasant activated)).

Table 1 (Cont.)

Source	Measure Name	Description
Fredrickson, Tugade, Waugh, & Larkin, (2003)	Differential Emotions Scale-Modified (M-DES)	23-item scale, modified from Izard's DES. 20 items include "I have felt ____", then three related affective words, (e.g., "angry, irritated, annoyed", "awe, wonder, amazement", "grateful, appreciative, thankful"), on a scale from 0 = *Never,* to 4 = *Most of the Time.* The final three items ask which emotions were felt most strongly, whether other emotions were felt (and if so, which, and to what extent).
Gross, Crane, & Fredrickson, (2010)	Feelings Questionnaire	20-item questionnaire, assessing the extent to which an emotion was felt from 0 = *Not at all,* to 4 = *Extremely.* Each includes three items (e.g., "amused, fun-loving, silly", "ashamed, humiliated, disgraced")
Harmon-Jones, Bastian, & Harmon-Jones (2016)	Discrete Emotions Questionnaire	This instrument includes eight distinct states (anger, disgust, fear, anxiety, sadness, happiness, relaxation and desire), with four items each. For example, Anxiety is denoted by the items "dread", "anxiety", "nervous" and "worry". Participants rate the extent to which they feel emotions on a Likert scale from *1 = Not at all* to *7 = An extreme amount.*
Heuchert & McNair (2012); Modified from McNair, Lorr, & Droppleman (1971)	Profile of Mood States (POMS-2)	65-item measure asking participants to rate items on a 5-point Likert scale ranging from 0 = *Not at All,* to 5 = *Extremely.* This scale contains multiple subscales including Anger-Hostility (example item: Furious), Confusion-Bewilderment (e.g., Muddled), Depression-Dejection (e.g., Hopeless), Fatigue-Inertia (e.g., Exhausted), Friendliness (e.g., Friendly), Tension-Anxiety (e.g., Uneasy), Vigor-Activity (e.g., Energetic).
Irvin (1967)	Adjective Check List	This paper includes 29 adjectives as part of a checklist (e.g., liked by others, worrisome, conscientious, good sense of humor, impulsive).

Table 1 (Cont.)

Source	Measure Name	Description
Fredrickson, Tugade, Waugh, & Larkin, (2003)	Differential Emotions Scale-Modified (M-DES)	23-item scale, modified from Izard's DES. 20 items include "I have felt ____", then three related affective words, (e.g., "angry, irritated, annoyed", "awe, wonder, amazement", "grateful, appreciative, thankful"), on a scale from 0 = *Never,* to 4 = *Most of the Time.* The final three items ask which emotions were felt most strongly, whether other emotions were felt (and if so, which, and to what extent).
Gross, Crane, & Fredrickson, (2010)	Feelings Questionnaire	20-item questionnaire, assessing the extent to which an emotion was felt from 0 = *Not at all,* to 4 = *Extremely.* Each includes three items (e.g., "amused, fun-loving, silly", "ashamed, humiliated, disgraced")
Harmon-Jones, Bastian, & Harmon-Jones (2016)	Discrete Emotions Questionnaire	This instrument includes eight distinct states (anger, disgust, fear, anxiety, sadness, happiness, relaxation and desire), with four items each. For example, Anxiety is denoted by the items "dread", "anxiety", "nervous" and "worry". Participants rate the extent to which they feel emotions on a Likert scale from *1 = Not at all* to *7 = An extreme amount.*
Heuchert & McNair (2012); Modified from McNair, Lorr, & Droppleman (1971)	Profile of Mood States (POMS-2)	65-item measure asking participants to rate items on a 5-point Likert scale ranging from 0 = *Not at All,* to 5 = *Extremely.* This scale contains multiple subscales including Anger-Hostility (example item: Furious), Confusion-Bewilderment (e.g., Muddled), Depression-Dejection (e.g., Hopeless), Fatigue-Inertia (e.g., Exhausted), Friendliness (e.g., Friendly), Tension-Anxiety (e.g., Uneasy), Vigor-Activity (e.g., Energetic).
Irvin (1967)	Adjective Check List	This paper includes 29 adjectives as part of a checklist (e.g., liked by others, worrisome, conscientious, good sense of humor, impulsive).

Table 1 (Cont.)

Source	Measure Name	Description
Izard, Dougherty, Bloxom, & Kotsch (1974); Izard, Libero, Putnam & Haynes (1993)	Differential Emotions Scale (DES)	This scale includes various forms, including a 67-item and a 30-item self-report inventory with each item scored on a 5-point Likert scale.
Jacobs Bao, (2012)	Global Mood and Life Satisfaction Sliders	2-item scale, each with sliders ranging from a frowning face (0) to a smiling face (100). Participants indicate how they feel right now, and how satisfied with their life they are.
Laurent et al. (1999)	Positive and Negative Affect Scale-Children (PANAS-C)	30-item measure for children based on the PANAS, including 15 positive words and 15 negative words. Participants indicate how often they have felt items on a Likert scale from 1 = *Very slightly or not at all* to 5 = *Extremely.*
Lubin & Zuckerman (1999); Zuckerman & Lubin (1965, 1985)	The Multiple Affect Adjective Check List (MAACL)	70-item scale comprised of adjectives, asking participants to rate how intensely they felt adjectives (e.g., complaining, critical, cross, enraged, forlorn, free, impatient, lost, mad, mean, pleased, pleasant, sunk, wild, worrying) from 1 = *Not at All* to 7 = *Extremely.*
Lyubomirsky & Lepper (1999)	Subjective Happiness Scale (SHS)	This 4-item measure measures general happiness. Participants indicate their answers on a 1-7 Likert scale for each question. The scale changes for each question. For example, the first question is "In general, I consider myself:" with a Likert scale ranging from *1 = Not a Very Happy Person* to *7 = A Very Happy Person.*

Table 1 (Cont.)

Source	Measure Name	Description
Matthews, Hillyard, & Campbell (1999, 2000)	Dundee Stress State Questionnaire (DSSQ)	This measure differentiates between 11 primary scales, including the three UMACL mood scales (Energetic Arousal, Tense Arousal, and Hedonic Tone), two motivational scales (Intrinsic Interest, Success Striving) and six cognitive scales (Self-Focus, Self-Esteem, Concentration, Confidence and Control, Task-related Cognitive Interference, and Task-Irrelevant Cognitive Interference). These primary scales are part of three higher-order dimensions of Task engagement, Distress, and Worry.
Matthews, Jones, & Chamberlain (1990)	UWIST Mood Adjective Checklist (UMACL)	29-Item Adjective Checklist (i.e., Happy, Dissatisfied, Energetic, Relaxed, Alert, Nervous, Passive, Cheerful, Tense, Jittery, Sluggish, Sorry, Composed, Depressed, Restful, Vigorous, Anxious, Satisfied, Unenterprising, Sad, Calm, Active, Contented, Tired, Impatient, Annoyed, Angry, Irritated, Grouchy) asking participants to rank items from 1 = *Definitely* to 4 = *Definitely Not.* This scale measures three correlated bipolar dimension of Energetic Arousal, Tense Arousal, and Hedonic Tone, with a supplemental dimension of Anger-Frustration.
McCullough, Emmons & Tsang (2002)	The Gratitude Questionnaire- 6 (GQ-6)	This 6-item measure includes items that measure a tendency to feel gratitude. Participants rate items on a 7-point Likert Scale ranging from *1 = Strongly Disagree* to *7 = Strongly Agree*. Two items are reverse coded. Example items include "I have so much in life to be thankful for", "If I had to list everything that I felt grateful for, it would be a very long list", and "Long amounts of time can go by before I feel grateful to something or someone" (R).

Table 1 (Cont.)

Source	Measure Name	Description
Nowlis (1965)	Mood Adjective Checklist (MACL)	This is a 12-item checklist includes the words aggression, anxiety, surgency, elation, concentration, fatigue, vigor, social affection, sadness, skepticism, egotism, nonchalance, rated on a three-point scale with 1 = *definitely applies,* 2 = *slightly applies,* and 3 = *does not apply.*
Russell & Mehrabian (1977)	-	This paper used 151 terms to test a three-factor model (Pleasure, Arousal, and Dominance).
Schnall, Roper, & Fessler, (2010)	Elevation Scale	This 6-item scale assesses elevation and elevation-related feelings (e.g., Uplifted, Optimistic about Humanity, A desire to become a better person).
Sjöberg, Svensson, & Persson (1979)	-	This paper includes 72 items used in a principal components analysis (e.g., indolent, clear-headed, gloomy, sorry, efficacious, active, concentrated, alert, peppy).
Spielberger & Reheiser (2009); Spielberger et al., (1995);	State-Trait Personality Inventory (STPI)	80-Item measure including ten items for each of 8 state and trait scales. Items are rated on a 4-point Likert scale ranging from *Not at All* to *Very Much So*. Example questions include "I feel at ease", "I feel upset", and for the trait version, "I lack self-confidence", "I am a steady person". Based on previous scales, the State-Trait Anxiety Inventory (STAI; Spielberger et al. 1983) and the State-Trait Anger Expression Inventory (STAXI: Spielberger, 1999).
Thayer, (1986, 1989)	Activation-Deactivation Adjective Checklist (AD ACL)	20-item scale comprised mainly of arousal-related words (e.g., active, drowsy, placid, fearful, sleepy, lively, jittery, still, energetic). Items are rated on a unique 4-point scale (vv = *definitely feel,* v = *feel slightly*, ? = *cannot decide,* no = *definitely do not feel*).

Table 1 (Cont.)

Source	Measure Name	Description
Timmermans, Van Mechelen, & Nezlek, (2009); modified from Russell, Weiss, & Mendelsohn, (1989)	Modified Affect Grid	A 9x9 two-dimensional grid, with an x-axis tracking pleasant/unpleasant a y-axis tracking arousal/sleepiness. There are eight words to anchor the corners, including unpleasant, pleasant, highly active, sleepy, stressed, excited, depressed, and relaxed. The center (5th) row and column are neutral. This grid is used to indicate how individuals feel during a specific experienced.
Watson & Clark (1999)	The Positive and Negative Affect Schedule Long Form (PANAS-X)	60-item scale including a number of feelings and emotions (e.g., cheerful, sad, active, angry at self, guilty, calm, enthusiastic, attentive, bashful, sheepish, sluggish, shaky) on a scale from 1 = *Very Slightly or Not at All,* to 5 = *Extremely.*
Watson, Clark, & Tellegen (1988)	Brief Measures of Positive and Negative Affect – Short Form (PANAS)	20-item measure asking participants to rank items on a 5-point Likert scale ranging from (1 = *Very Slightly or Not at All* to 5 = *Extremely*). Items are high- and low-arousal, and positive and negative (e.g., interested, irritable, distressed, alert, excited, ashamed, upset, inspired, strong).
Wendt, Cameron & Specht (1962)	-	This paper includes a list of 130 emotions and motivations measured in this study. (e.g., drowsy, lackadaisical, drifting, close-mouthed, skeptical, sexy, kindly). These items were harvested for the item pool.
Zevon & Tellegen (1982)	Mood Checklist	This paper included a 60-item mood checklist with 20 content categories (Attentive, Excited, Proud, Strong, Joyful, Friendly, Tired, Downhearted, Distressed, Angry, Contempt, Revulsion, Angry at Self, Fearful, Guilty, Jittery, Rejected, Shy, Content, Surprised).

Table 2
Frequency Metrics of Positive High-Arousal Words

Word	Study 1 % *Slightly or Not At All*	Study 2 % *Slightly or Not At All*	Ngram Frequency
happy	29.87	6.60	0.00886
grateful	37.11	2.40	0.00215
pleased	38.59	8.38	0.00262
loved	43.41	7.19	0.00699
optimistic	34.88	11.08	0.00053
satisfied	39.52	15.57	0.00260
cheerful	39.15	10.48	0.00074
warmth	43.04	10.78	0.00149
delighted	41.00	9.88	0.00103
excited	40.26	14.67	0.00246
joyful	38.40	11.68	0.00037
lively	39.89	14.37	0.00073
amused	50.65	11.08	0.00075
enthusiastic	41.19	15.57	0.00066
uplifted	39.70	14.37	0.00010
fulfilled	43.04	22.16	0.00098
refreshed	43.23	25.15	0.00020
blissful	43.97	19.16	0.00015
awed	50.28	23.65	0.00012
rejuvenated	45.08	28.14	0.00004
All	M = 41.11 SD = 4.71	M = 14.12 SD = 6.67	M = 0.00168 SD = 0.00232

Table 3
Frequency Metrics of Positive Low-Arousal Words

Word	Study 1 % *Slightly or Not At All*	Study 2 % *Slightly or Not At All*	Ngram Frequency
safe	30.43	3.89	0.00690
calm	32.10	8.38	0.00329
comfortable	34.32	6.89	0.00314
sleepy	43.78	.80	0.00042
relaxed	36.18	17.07	0.00151
generous	53.62	4.79	0.00145
committed		10.78	0.00334
peaceful	38.22	12.57	0.00137
relieved	38.22	17.37	0.00141
compassionate	45.83	5.39	0.00038
At ease	37.48	20.66	0.00045
reassured	44.34	16.47	0.00034
tranquil	39.70	16.47	0.00023
serene	-	15.87	0.00029
All	*M* = 39.52 *SD* = 6.49	*M* = 11.31 *SD* = 6.10	*M* = 0.00175 *SD* = 0.00187

Table 4
Study 2 Frequency Metrics of Negative High-Arousal Words

Word	Study 1 % *Slightly or Not At All*	Study 2 % *Slightly or Not At All*	Ngram Frequency
anxious	26.72	11.08	0.00218
worried	32.28	14.67	0.00315
overwhelmed	27.83	11.38	0.00082
frustrated	34.51	20.36	0.00084
irritated	41.19	21.86	0.00046
uncomfortable	41.37	42.22	0.00152
panic	42.86	44.91	0.00163
distressed	36.36	30.54	0.00043
afraid	47.50	46.41	0.00562
agitated	43.23	24.85	0.00050
on edge	36.18	29.94	0.00016
angry	48.79	51.80	0.00392
alarmed	47.87	39.82	0.00053
dislike	51.02	42.81	0.00053
jittery	42.49	44.61	0.00005
jealous	67.16	66.47	0.00091
devastated	52.32	53.59	0.00033
resentful	53.99	55.39	0.00015
disgraced	63.08	51.50	0.00012
hateful	61.60	69.46	0.00022
All	*M* = 44.92 *SD* = 11.21	*M* = 38.68 *SD* = 16.59	*M* = 0.00120 *SD* = 0.00150

Table 5
Frequency Metrics of Negative Low-Arousal Words

Word	Study 1 % *Slightly or Not At All*	Study 2 % *Slightly or Not At All*	Ngram Frequency
tired	30.98	8.98	0.00333
sad	35.81	30.24	0.00271
stressed	24.68	11.68	0.00111
upset	38.59	37.43	0.00186
disappointed	43.23	33.23	0.00123
confused	47.31	25.75	0.00226
unhappy	38.40	37.43	0.00113
uneasy	34.14	33.83	0.00069
empty	46.94	39.22	0.00505
gloomy	43.04	35.33	0.00054
bored	59.37	17.37	0.00079
powerless	40.26	47.01	0.00041
miserable	46.94	61.68	0.00112
defeated	47.12	50.30	0.00107
weary	42.67	40.42	0.00094
displeased	42.67	44.91	0.00019
hopeless	43.97	52.10	0.00056
pessimistic	39.70	43.71	0.00020
rejected	62.15	60.78	0.00203
despair	50.28	63.17	0.00134
distraught	43.97	53.59	0.00017
All	*M* = 42.96 *SD* = 8.44	*M* = 39.44 *SD* = 15.25	*M* = 0.00137 *SD* = 0.00119

Table 6
Zero-Order Correlations Over r = .70

Words	Study 2	Study 1a (all Datasets)
Satisfied & Fulfilled	.79	.75
Enthusiastic & Excited	.79	.76
Happy & Joyful	.78	.78
Enthusiastic & Cheerful	-	.77
Excited & Delighted	-	.77
Happy & Delighted	.76	.77
Joyful & Cheerful	.76	.80
Displeased & Unhappy	.76	-
Excited & Cheerful	-	.76
At Ease & Relaxed	.75	-
Satisfied & Pleased	-	.75
Joyful & Uplifted	.75	-
Gloomy & Sad	.75	-
Joyful & Delighted	.75	.76
Upset & Unhappy	.75	.72
Happy & Pleased	.74	.72
Unhappy & Sad	.74	-
Happy & Cheerful	.73	.74
Delighted & Pleased	.73	-
Joyful & Pleased	.73	.72
Joyful & Blissful	.73	-
Angry & Irritated	-	.73
Lively & Excited	-	.73
Cheerful & Pleased	.72	-
Joyful & Enthusiastic	.72	.72
Happy & Uplifted	.72	-
Upset & Displeased	.72	.74
Relaxed & Calm	-	.72
Miserable & Unhappy	.72	-
Peaceful & Comfortable	.71	-
Unhappy & Gloomy	.71	-
Tranquil & Calm	-	.71
Excited & Happy	-	.71
Rejuvenated & Uplifted	-	.71
Sad & Upset	-	.71
Enthusiastic & Delighted	-	.70
Upset & Angry	-	.70
Upset & Irritated	-	.70
Lively & Delighted	-	.70
Refreshed & Rejuvenated	.70	-
Excited & Joyful	.70	.73
Stressed & Overwhelmed	.70	-
Frustrated & Irritated	-	.70
Gloomy & Empty	.70	-
Weary & Hopeless	.70	-
Cheerful & Delighted	.70	.77

Table 7
Study 2 Factor Analysis

Item	Factor 1	Factor 2
Positive High-Arousal		
Joyful	.04	.89
Uplifted	.06	.85
Enthusiastic	.02	.81
Fulfilled	-.02	.80
Cheerful	.02	.80
Lively	.00	.79
Warmth	.08	.79
Delighted	-.01	.77
Pleased	-.05	.76
Blissful	-.04	.74
Excited	.02	.74
Satisfied	-.03	.73
Happy	-.15	.72
Rejuvenated	-.03	.71
Refreshed	-.07	.71
Amused	.04	.69
Awed	.09	.66
Optimistic	-.13	.65
Loved	.04	.63
Grateful	.04	.57
Positive Low-Arousal		
Relieved	.01	.73
Reassured	-.04	.72
Serene	-.02	.71
Committed	.10	.64
At Ease	-.22	.63
Compassionate	.13	.62
Generous	.14	.62
Peaceful	-.16	.60
Calm	-.11	.59
Tranquil	-.15	.58
Relaxed	-.17	.57
Comfortable	-.17	.56
Safe	-.09	.43
Sleepy	.29	-.02

Table 7 (cont.)

Item	Factor 1	Factor 2
Negative High-Arousal		
Disgraced	.63	.21
Devastated	.73	.17
Alarmed	.67	.14
Resentful	.61	.12
Hateful	.53	.12
Afraid	.72	.12
Jittery	.59	.11
Angry	.68	.10
Panic	.66	.09
Agitated	.70	.05
Jealous	.43	.05
Anxious	.66	.02
Dislike	.65	.01
Worried	.73	.00
Uncomfortable	.62	-.01
Distressed	.74	-.02
On Edge	.74	-.05
Overwhelmed	.63	-.07
Frustrated	.74	-.08
Irritated	.55	-.14
Negative Low-Arousal		
Despair	.77	.08
Distraught	.83	.07
Confused	.63	.06
Rejected	.64	.02
Defeated	.77	-.01
Uneasy	.75	-.03
Pessimistic	.65	-.06
Weary	.65	-.08
Displeased	.76	-.08
Powerless	.68	-.09
Upset	.73	-.09
Hopeless	.68	-.11
Disappointed	.69	-.11
Gloomy	.74	-.12
Sad	.69	-.13
Stressed	.57	-.13
Bored	.18	-.15
Miserable	.65	-.15
Unhappy	.72	-.16
Tired	.45	-.17
Empty	.55	-.24

Table 8
Final Brief Affect Measure Items EFA Loadings

Item	Factor 1	Factor 2
Positive High-Arousal		
Happy	-0.12	0.71
Grateful	0.17	0.64
Enthusiastic	0.00	0.76
Positive Low-Arousal		
Comfortable	-0.21	0.51
Calm	-0.13	0.57
Relieved	0.03	0.75
Negative High-Arousal		
Frustrated	0.75	-0.04
Anxious	0.75	0.10
Overwhelmed	0.84	0.09
Negative Low-Arousal		
Sad	0.68	-0.11
Uneasy	0.68	-0.10
Unhappy	0.61	-0.22

Table 9
Final Scale Correlation with Other Scales

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
BAM														
1. High Pos	-													
2. Low Pos	**0.66****	-												
3. All Pos	**0.91****	**0.92****	-											
4. High Neg	**-0.38****	**-0.47****	**-0.47****	-										
5. Low Neg	**-0.49****	**-0.52****	**-0.55****	**0.69****	-									
6. All Neg	**-0.47****	**-0.53****	**-0.55****	**0.92****	**0.91****	-								
PANAS														
7. Pos	**0.74****	**0.67****	**0.77****	**-0.34****	**-0.40****	**-0.40****	-							
8. Neg	**-0.45****	**-0.44****	**-0.49****	**0.70****	**0.82****	**0.83****	**-0.28****	-						
AAS														
9. Pos	**0.74****	**0.75****	**0.82****	**-0.47****	**-0.50****	**-0.53****	**0.84****	**-0.44****	-					
10. Neg	**-0.51****	**-0.49****	**-0.55****	**0.74****	**0.79****	**0.83****	**-0.41****	**0.84****	**-0.52****	-				
BFI-2-S														
11. O	0.09	0.00	0.05	-0.06	-0.03	-0.05	**0.12**	-0.04	0.05	-0.07	-			
12. C	**0.33****	**0.27****	**0.33****	**-0.26****	**-0.26****	**-0.28****	**0.39****	**-0.33****	**0.29****	**-0.32****	**0.21****	-		
13. E	**0.40****	**0.26****	**0.36****	**-0.17***	**-0.18****	**-0.19****	**0.50****	**-0.21****	**0.34****	**-0.21****	**0.20****	**0.36****	-	
14. A	**0.24****	**0.17***	**0.22****	-0.03	**-0.12**	-0.08	**0.15***	**-0.20****	**0.16***	**-0.18***	**0.10***	**0.36****	**0.17***	-
15. N	**-0.46****	**-0.51****	**-0.53****	**0.60****	**0.60****	**0.65****	**-0.51****	**0.60****	**-0.51****	**0.68****	-0.01	**-0.36****	**-0.30****	-0.05

Note: BAM = Brief Affect Measure (my scale). PANAS = Positive and Negative Affect Schedule. AAS = Affect Adjective Scale. BFI-2-S = Big Five Inventory 2, Short version. $p < .05$ are in bold, * indicates $p < .01$, ** indicates $p < .001$.