In D. Kahneman, E. Diener, & N. Schwarz (Eds.) (1999).
Well-being: Foundations of hedonic psychology (pp. 40-60). New York: Russell Sage.

3 Measurement Issues in Emotion Research

Randy J. Larsen and Barbara L. Fredrickson

We open this chapter on measurement issues with the recommendation that researchers construct a working definition of emotion(s) that best fits their research agenda prior to selecting measures. We then discuss issues that cut across all types of emotion measurement, such as timing and context, as well as reliability and validity. Next, we provide a selective review of specific measurement techniques, touching on self-reports of subjective experience, observer ratings, facial measures, autonomic measures, brain-based measures, vocal measures, and responses to emotion-sensitive tasks. Our aim in this selective review is to highlight some specific strengths, weaknesses, and measurement issues associated with different types of emotion measures. Finally, because emotions are only probablistically linked to emotion measures, we also recommend that, to the extent possible, researchers collect and cross-reference multiple measures of emotion.

EXPERIENCES OF psychic pain and pleasure, and the limitless variations on this hedonic theme, define the domain of emotions. The content of a person's emotional life strongly influences his or her judgments of the quality of that life. In addition, a person's emotional engagement with the "stuff" of life defines the "wantability" and utility of that stuff for building quality into life, for deciding to do one thing instead of another, and for being satisfied with the outcomes of his or her choices. To be sure, quality of life goes far beyond just feeling more pleasant than unpleasant emotions in one's life over time. Nevertheless, as pointed out in the preface to this book, we can approach an understanding of quality of life by considering some of its lower-level components and building blocks, such as emotions.

How might the study of emotions help in understanding quality of life? We have space to give only a few examples. One question concerns the relation between pleasure and pain, between the positive and negative emotions. Are the conditions that give rise to pleasant emotions simply the opposite of those that produce unpleasant emotions? Can circumstances that bring about pleasantness cancel unpleasant states, and vice versa? Should we think of pleasure and pain as end points on a continuum, or as completely separate and independent dimensions? Emotions can be thought of as both inputs into processes that contribute to quality of life and as outcomes that provide feedback as to how those processes are working. Another line of inquiry concerning quality of life would address the habituation of emotional responses. Good and bad things happen to everyone. And we know that people habituate to the good and the bad at different rates. Are there ways to potentiate habituation to unpleasant events, and ways to slow habituation to positive events? Do different components of the emotional response (such as bodily reactions and subjective feelings) habituate differently? Another emotional topic useful in understanding quality-of-life concerns situational and individual differences in emotional responding. For example, what are the conditions under which most people are likely to experience joy or suffering? Why is it that many episodes of joy and happiness occur in the context of a pending tragedy that has been averted? Similar questions may be applied to individual differences. Why is it that some people are easily made anxious and fearful, whereas others are less vulnerable to these unpleasant emotions? People differ in their thresholds for evoking emotions as well as in the magnitude of their emotional responses to the same events. Because emotions contribute to quality of life, understanding these individual and situational differences in emotional responding may contribute to understanding quality of life.

These few research questions should make it clear that there are many lines of inquiry about emotions that may be important in understanding quality-of-life. Empirical inquiry requires measurement, and so the editors of this book asked us to address the assessment of emotions. This is a daunting task, even if we were to simply list, in how-to fashion, all the different ways emotions have been measured in the research literature. We have chosen instead to focus more on measurement issues than on specific measures per se. Although we review some specific measurement techniques, our review is not meant to be exhaustive. Instead we cover examples that are meant to be illustrative and that we use as vehicles for discussing the strengths, weaknesses, and implications of certain techniques for assessing emotions. Let us first, however, address some issues that should be considered before turning to specific techniques.

Prolegomena to Measurement Issues

What, exactly, is an emotion? Emotion researchers do not fully agree on the answer to this basic question (compare, Ekman and Davidson 1994). In fact, Kleinginna and Kleinginna (1981) identified over ninety different definitions. Yet this does not mean that research on emotions is stalled at the starting gates, haggling over definitions. It does mean, however, that researchers should begin by articulating their own ***working definition of emotion(s)*** in planning and communicating their work. We recommend this step for two reasons. First, it can limit the possible misinterpretations of your results. Second, and perhaps more critically, it can make choosing among various emotion measures an easier task. Working definitions are appropriate because they imply revision and refinement over time as new findings about the nature of emotion inevitably emerge.

Our own working definition of emotions draws from a systems perspective, identifying emotions as multifaceted processes that unfold over time. Emotions are manifest in multiple channels, and the channels themselves are loosely coupled and interact in a complex way (Venables 1984). These channels span both psychological and physiological domains, including subjective experience, facial action, central and peripheral nervous system activation, cognitive or information-processing changes, and behavioral action tendencies. In the ideal case, charting emotions entails assessing organized changes across these multiple components simultaneously. Data streams obtained from these multiple domains may converge on the underlying construct of emotion and increase confidence that we can fathom its presence and magnitude. Yet even with multiple, synchronized measures, the underlying psychological construct of "emotion" remains some inferential steps away from the more tangible data it can produce. In other words, we view an emotion as an inferred construct and caution against purely operational (or reductionistic) definitions of emotion. The term "emotion" carries surplus meaning beyond any set of emotion measures.

Other issues researchers need to consider in constructing working definitions are whether they conceptualize emotions as (a) discrete and/or dimensional (b) states and/or traits and (c) event-related and/or diffuse (for discussions of these issues, see Frijda, this volume; Lazarus 1991; Morris, this volume). While the latter two of these issues may impinge primarily on the expected intensity and duration of emotion experiences, the first issue can impinge directly on emotion measurement. Discussions about whether emotions operate as two or three general dimensions, or as seven or more separate and distinct categories of experience, go back more than one hundred years (Darwin 1872/1965) and continue to this day (for reviews, see Izard 1993; Lazarus 1991). One widely espoused dimensional view of emotion is represented by the circumplex model (Russell 1980; Watson and Tellegen 1985; for a review, see Larsen and Diener 1992). This model posits that emotions conform to a circular or radex arrangement with the coordinates of this circular space representing valence and arousal: emotions that are similar to each other (for instance, anger, distress) are close to each other on the circumference of the circle, whereas emotions that are so-called opposites (for example, happiness, sadness) are 180 degrees away from each other. In contrast, proponents of discrete views (Ekman 1992; Izard 1977; Lazarus 1991) hold that dimensional views often blur meaningful distinctions between adjacent emotions (for example, fear versus anger versus disgust). More recently, research by Feldman-Barrett (1995; in press) suggests that individuals reliably differ in whether they describe their affective states as discrete or as dimensional. The measurement issue embedded within this dialogue concerns specificity: while measures that fit the discrete emotions views can be reduced to a dimensional arrangement post hoc, the converse is rarely possible. For this reason, researchers should consider a priori whether distinctions between specific negative or positive emotions are likely to have an impact on their theoretical and empirical agenda.

In sum, *whether* and *how* the measurement issues and types of emotion measures discussed in this chapter apply to any given research agenda follows from the working definition of emotion(s) adopted within that research agenda. For this reason, we recommend that those embarking on emotions research first consider what they take emotions to be.

Measurement Issues

Before turning to descriptions of specific domains of measurement, we discuss a series of issues that cut across all types of emotion assessment. In addition to discussing the traditional measurement issues of reliability and validity, we also discuss the issues surrounding timing and context that can be particularly vexing in emotion research.

Timing

Emotions take time. They are dynamic processes that unfold, linger, and then dissipate over time—sometimes gradually, other times rapidly. Moreover, emotions involve a cascade of different response systems, and each may have its own time of onset and duration. For instance, if you were to be startled by a sudden loud noise, like a car horn, you would blink your eyes in about forty milliseconds, your heart rate would begin to accelerate in about five hundred milliseconds, your sweat glands would become active after two to three seconds, and a hormonal response might occur minutes later. Emotions also change in character depending on the temporal vantage point from which they are viewed. In real time, for instance, emotions can implicate multiple physiological systems, whereas in retrospect, these same bodily changes become less evident, often dropping out of measurement range altogether. Likewise, an open empirical question is whether perhaps in real time the subjective experience of emotions may be quite nuanced, best represented by specific emotion terms, whereas in retrospect, a single valance dimension (good-bad) may adequately represent this same experience (see Kahneman, this volume).

One critical measurement issue is how to isolate the targeted emotion episode. When does it start and when does it end? Identifying these moments with precision can greatly increase researchers' chances of observing emotion-related changes. Imprecision at this stage can, in effect, dilute the targeted emotional episode within a wash of emotion-irrelevant moments (Levenson 1988).

A second critical measurement issue is how to ensure that purported measures of emotion have sufficient *temporal resolution* to capture the dynamic aspects of the concept under study. Some markers of emotion—for example, an increase in cardiac output—might span only a minute or less. The subjective experience of emotional arousal, however, might last much longer. Thus, readings of cardiac output taken once every fifteen minutes have only a remote chance of capturing an emotion-related change, whereas self-report measures taken in that same time span might successfully capture some of the emotional effect. If working definitions identify the targeted emotion concept as a quick-changing state, then measures should be appropriately fine-grained, exhibiting a temporal resolution that is smaller (ideally much smaller, to provide reliable aggregate measures) than the expected duration of the emotion-related change.

A third issue concerns the *temporal proximity* of emotion measures to the emotion experience. Measures obtained on-line or *during* an emotion experience are perhaps feasible more often than is recognized. This is certainly true for measures obtained from video records and through physiological recording devices, but perhaps no less so for measures obtained via self-report (see discussion later in the chapter). Emotion measures obtained concurrently with emotion experience maximize validity and accuracy (with the exception of measures extracted at the cost of disrupting the emotion experience). When concurrent measures are not feasible or practical, lagged measures that minimize the latency between emotion experience and emotion measurement should be sought. Except when memory for emotion is the target of study, the shorter the latency, the better the measure (Levenson 1988).

Context

Emotions occur within the broader psychological context of subjective and bodily experience. Other features of this context can no doubt impinge upon emotion measurement. A study aimed at inducing a specific emotion in all participants (anger or sadness, for instance) may find that the success of the induction depends on contextual factors that vary from individual to individual, such as ambient mood (for example, irritable or depressed mood), emotion-related personality traits (for example, hostility or pessimism), recent life events (such as perceived personal injustices or losses), or preexisting arousal (did the participant just drink four cups of coffee?). Diurnal, circadian, and circaseptum influences on mood might also alter emotion experience. If the researchers' aim is to create a comparable emotional state across all participants, then contextual influences such as these might be considered noise. There are two defenses against "noise-producing" constructs: hold constant or limit nuisance variance, or measure it. Success at the first strategy comes with familiarity in a research area and good experimental design. Success at the second strategy allows researchers to determine which participants might be extreme outliers in terms of ambient mood or recent life events, and/or how emotion-related personality traits covary with the phenomena under study. Yet one researcher's noise is another researcher's data: the extent to which neighboring aspects of subjective experience, such as emotions, moods, traits, and nonspecific arousal, influence one another is the target of study for several research programs.

Reliability

Many researchers think of measurement reliability as a high test-retest correlation. As a measurement concept, reliability in fact refers to the degree to which observed scores reflect the "true" amount of the construct being measured. Because we never have access to "true" scores, we can only estimate reliability. For certain psychological constructs, a test-retest correlation is a good estimate of reliability. Test-retest is an appropriate way to estimate reliability for between-subjects constructs (traits), where the variance of interest is between participants and we assume there will be little or no meaningful within-participant variance. Intelligence is a good example of a between-subject construct: we assume that, for any single individual, intelligence is stable and not easily changed, at least not over a few weeks or months. As such, reliable measures of intelligence demonstrate high test-retest correlations.

Emotion, however, is more typically construed as a within-subject construct (a state), and we assume that it may change quickly and frequently within any single individual. To complicate matters, emotion can be a between-subjects construct as well, where the variance of interest might be differences between individuals in their responses to identical emotion-provoking events. Because emotion is a complicated state-trait construct, we cannot use simple test-retest correlations as estimates of measurement reliability.

A second way to estimate reliability is through internal consistency estimates, such as coefficient alpha, or odd-even item composite correlations. These are actually measures of item homogeneity: they assess the degree to which the various items are measuring the same underlying construct (though alpha is, in part, inflated by scale length) (compare, Clark and Watson 1995). Because many self-report emotion measures are factor-analytically constructed, internal consistency or item homogeneity is built in during the scale construction process. Internal consistency analysis is thus one way to estimate reliability, and it works equally well for both state and trait measures. However, internal consistency estimates of reliability work only for multi-item scales. Single-item measures, which are very popular in emotion research, simply cannot be examined in terms of internal consistency.

What, then, is the researcher using single-item measures to do? One approach is to bypass reliability concerns altogether and focus instead on concerns about validity. This is reasonable because measurement reliability (in the sense of the proportion of variance in the observed scores that is attributable to true score variance) sets the upper bound on validity correlations. In other words, a measure cannot correlate with external validity criteria higher than it can correlate with itself. As such, valid measures are de facto reliable. Clearly, a researcher who passes up reliability concerns treads on thin ice. Nevertheless, strong evidence for validity, with multiple converging methods and replicated patterns of association, can add credibility to the claim that a particular measure is reliable.

Reliability is most important in interpreting failures to refute the null hypothesis. For example, if a study is completed and no predicted effects are found, three obvious reasons must be entertained: the theory is wrong, the measures used are not reliable, or some auxiliary conditions of the study were not met (for more detailed discussion, see Meehl 1978). If a study fails and the researcher is confident that the measures used are reliable, then the researcher must question the theory or look for something that might have gone wrong with the procedures (including data management and analysis). It is precisely in such circumstances (null findings) that reliability evidence is crucial.

Validity

Emotions, we have argued, are theoretical constructs that are only probabilistically linked to observable indicators. As such, the term "emotion" has surplus meaning: even though it may be represented by many different measures, emotion is not equivalent, nor can it be reduced to, any single measure. This underscores the importance of construct validity in understanding the scientific meaning of emotion terms (Cronbach and Meehl 1955).

In construct validity, meaning is given to a scientific term (such as "emotion") by the nomological network of assertions in which that term appears. Our theories and measurement models guide us in building a network of associations around the construct of emotion. In construct validation, theory testing and measurement development proceed in tandem. Each link in the network adds to the scientific meaning of the term. Some links refer to positive associations (convergent validity), and some refer to negative or null associations (discriminant validity). In addition, some links specify the conditions under which emotions are likely to be evoked (predictive validity).

The total collection of relationships built up around the construct of "emotion," or around specific emotions, creates a mosaic of research findings. When enough pieces of the network are in place, we "get the picture." That is, when enough information is available about what something is, what it isn't, and what it predicts, we begin to have the feeling that we "understand" it. This is not to say that our understanding of an emotion is complete at this point. Construct validity is always unfinished, and things are always "true until further notice." Nevertheless, even though there are always new links to be added to the network of associations surrounding a construct, there comes a point where we reach some consensual agreement about the scientific meaning of a construct, such as an emotion.

Again, because emotions implicate multiple channels or component systems (for example, facial action, autonomic activity, subjective experience, action tendencies), the question arises about whether we should expect strong convergence among measures of these different components. Most researchers hold the view that components of emotion are loosely coupled systems that interact in a complex way (see Frijda, this volume). Clearly, the various response systems have multiple tasks beyond indexing emotions. For instance, the autonomic nervous system responds to metabolic demands and maintains the delicate balance of homeostasis, facial muscles are used for communication and eating, and conscious experience follows the streams of thought. Although emotions may bring the disparate component systems into some synchronization, total convergence among measures is neither expected nor required for construct validity. In fact, discrepancies between component measures can represent challenges to existing theories and may provide insights into how emotion systems work. Moreover, for some researchers, discrepancies between component measures of emotion are used to index emotional dissociation or repression (see Bonanno et al. 1995; Newton and Contrada, 1992).

Perhaps the strongest evidence for validity is when the theory of the particular emotion can be used to generate predictions about the conditions under which that emotion will be evoked, or the type of persons for whom that emotion will be most easily evoked. Couple this with measurement theory and knowledge of specific measures of emotion, and very specific predictions may be generated. We turn now to a consideration of specific measures in the emotions domain.

Types of Emotion Measures

Self-Reports of Subjective Experience

Self-report measures of emotion are widely used and form a broad range of assessment instruments. These measures rely on participants not only to experience their emotions but also to reflect accurately their phenomenal awareness through the use of rating scales or adjective checklists. Proponents of self-report assume that participants are in a privileged position to monitor, assess, and integrate information about their own emotions. Through self-report measures, the participant has the opportunity to express, in some integrated and standardized format, a good deal of information that only he or she has access to.

Although there are a great many instruments, substantial similarities can be found among them. Rather than conduct an exhaustive review, we instead concentrate on a few exemplars and highlight common measurement themes and issues. Additional instruments are reviewed in MacKay (1980) and Stone (1995).

Single-Item Measures A technique with a good deal of face validity is simply to ask research participants to rate how they are/were feeling on a single emotional construct. That construct might be a global affective dimension ("How unpleasant are you feeling?") or a specific emotion ("How angry do you feel?"). And the response scale might be unipolar ("not at all angry" to "extremely angry") or bipolar ("unpleasant" to "pleasant"). Response options are often Likert-type scales, with five-, seven-, or nine-point formats. The advantages of single-item measures are that they are simple to construct, easily understood by participants, and brief to administer. Plus, virtually any emotion term can anchor a single-item scale, making this self-report technique indispensable for researchers targeting specific, discrete emotions (Ekman, Friesen, and Ancoli 1980; Gross and Levenson 1993). The disadvantages are the same as those encountered whenever measurement is extremely brief: concerns about the ratio of error variance to true variance, representativeness, domain specificity, and sampling error. Despite these disadvantages, single-item measures are very popular in the experimental and survey literatures, where brevity is important.

An important variant on this technique is to make the response scale a visual analog of the digits representing response options. Such Visual Analog Scales (VAS) typically present the participant with a horizontal line separating two opposing adjectives. Participants are asked to place a mark on the line describing how they are/were feeling along that dimension. Researchers have also used VAS methods with unipolar response options; the line is anchored with "not at all" to "extremely much" for a specific emotion construct (sad, for instance). A related technique is to make the question itself an analog of the construct being assessed. For example, the participant might be presented with a series of five cartoon faces, going from a neutral expression on one face to an extreme frown on another, and he or she is asked to circle the face that most represents how she or he is/was feeling. This has the advantage of being useful with participants for whom adjectives might not be meaningful, such as very young children or participants from different linguistic cultures.

A recent single-item questionnaire measure, called the Affect Grid, has been introduced by Russell and colleagues (Russell, Weiss, and Mendelsohn 1989). Based on the circumplex model of emotion (Larsen and Diener 1992; Russell 1980; Watson and Tellegen 1985), the Affect Grid is composed of a nine-by-nine matrix. Emotion adjectives are placed at the midpoints of each side of the grid, as well as at the four corners. These adjectives are (starting in the high-arousal, pleasant quadrant, and proceeding clockwise) excitement, pleasantness, relaxation, sleepiness, depression, unpleasantness, stress, and high arousal. Participants are instructed to place a check within the cell of the grid that best reflects how they are/were feeling along the pleasantness and arousal dimensions. The developers of this scale report that its performance is similar to that of other longer and more cumbersome measures of pleasure and arousal. In addition, Russell and his colleagues (1989) report that this measure is sensitive to manipulations designed to alter participants' levels of pleasantness and arousal. One advantage of this measure is that it may be administered many times without fatigue.

Multiple-Item Measures Representing a large class of assessment instruments, the majority of multi-item measures consist of lists of adjectives describing emotional states. Some measures are checklists: the participant is instructed to simply check all those emotions that he or she is/was feeling. Other measures are rating tasks: the participant is instructed to rate each adjective for the degree to which he or she is/was feeling that particular emotion. The numerous multi-item instruments are essentially variations on these response themes; differences have to do primarily with response scales, the number and nature of the emotion adjectives, the scoring and scale names, and the instructions that accompany the self-report tasks.

One of the first adjective rating scales formally constructed was the 130-item Mood Adjective Check List (MACL) (Nowlis and Green 1957). Despite the name, the MACL is not literally a checklist: the participant is asked to rate how he or she felt at the time the emotion adjective was read on the following scale: "definitely felt it," "slightly," "cannot decide," "definitely not." Based on factor analytic studies, thirty-six items were selected for a short form of the MACL (Nowlis 1965). Scoring results in twelve factor scores: aggression, anxiety, surgency, elation, concentration, fatigue, social affection, sadness, skepticism, egotism, vigor, and nonchalance. Other researchers propose a simple positive-negative valence scale scoring (Stone 1981). The MACL has not become a widely used measure, perhaps because it was never published in a journal format or by a test publisher. The original version (Nowlis and Green 1957) was in an unpublished naval technical report, and the later shortened version (Nowlis 1965) appeared in a chapter in an edited book (Tomkins and Izard 1965).

A subsequent affect checklist has since eclipsed the MACL in popularity: Zuckerman and Lubin's (1965) Multiple Affect Adjective Check List (MAACL). It is very similar to the MACL in length: the MACL has 130 items, and the MAACL has 132. Moreover, many of the items are the same on the two inventories. Despite these similarities, the MACL has languished whereas the MAACL went on to become the most widely used self-report emotion assessment instrument in the psychological literature (Larsen and Sinnett 1991). Perhaps the critical ingredient to the MAACL's success was that it was distributed by Educational and Industrial Testing Service (EITS), a professional test publisher. It comes with a user manual, complete with annotated references, developmental history, psychometric properties, scoring keys, and multiple answer sheets. Other reasons for its popularity might be the checklist format, which makes administering the MAACL much faster than the MACL. And finally, the MAACL has only three subscales, compared to twelve on the MACL. The parsimony associated with the more global scales is probably appealing to many, although more specific scales have their uses as well. The three scale scores on the MAACL are depression, anxiety, and hostility. These three scales are highly intercorrelated and appear to lack discriminant validity. Gotlib and Meyer (1986) factored the original MAACL items and reported two factors, which they labeled positive and negative affect, consistent with the labels proposed by Watson and Tellegen (1985) a few years earlier.

In 1985 Zuckerman and Lubin published a revised version of the Multiple Affect Adjective Check List (MAACL-R). The revision mainly concerns the scoring format, which now allows for several pleasant emotion scores as well as global positive and negative affect and sensation-seeking. While the new scoring format appears better in some respects (it conforms to more recent factor analytic studies of emotion ratings), the new format is not without problems. For example, table 10 in the MAACL-R manual reports that the sensation-seeking scale has a coefficient alpha of .09 in a large (over one thousand) sample.

This is a good point at which to pause and take up the issue of response formats. The MAACL and its revision are in the form of checklists: the subject merely indicates the presence or absence of a particular emotion by checking a box. Some researchers have argued that checklists are particularly susceptible to response sets and other forms of nonrandom error. Almost three decades ago, Bentler (1969) argued against using checklists in psychometric assessment. More recently, Green, Goldman, and Salovey (1993) demonstrated that checklist mood assessments contain significant nonrandom error, and, "like Bentler (1969) before us, we advise caution when researchers analyze data obtained with a checklist format" (1036). A related issue concerns nonbalanced or asymmetric Likert response options, which were popular on early mood assessment inventories. Research on this issue is adequately reviewed by Mackay (1980) and will not be repeated here.

In 1967 Thayer published the Activation-Deactivation Adjective Check List (A-D ACL). Based on his own theory of activation, arousal, and affect, the A-D ACL contains adjectives that primarily refer to valenced arousal states, such as energetic, lively, active, sleepy, tired-tense, clutched-up, fearful-jittery, calm, quiet, and at rest. Participants rate the adjectives on a four-point scale, from "definitely do not feel" to "definitely feel." There are several factor-analytic-based scoring strategies, although the most widely used strategy results in two scores: energetic arousal (which is high-arousal positive affect) and tense arousal (which is high-arousal negative affect). Research with the A-D ACL is reviewed in Thayer (1986).

In 1977 Izard introduced the multi-item Differential Emotions Scale (DES) aimed at assessing multiple discrete emotions. Respondents are asked to rate (on a five-point scale) how much they are/were experiencing various discrete emotions by rating clusters of three emotion words (for example, scared/fearful/afraid, angry/irritated/annoyed, glad/happy/joyful). The original DES has since been modified to distinguish between self-conscious emotions (Mosher and White 1981).

One of the more recent introductions in this long line of mood adjective rating scales is the Positive Affect Negative Affect Schedule (PANAS) (Watson, Clark, and Tellegen 1988). The PANAS is based on the circumplex model of affect (Russell 1980; Watson and Tellegen 1985; Larsen and Diener 1992). Of the eight potential scores derivable from the circumplex model, the PANAS focuses on two positive affect (PA) (high-arousal pleasant), and negative affect (NA) (high-arousal unpleasant). The PANAS contains ten items on each of the two scales. The items are mood adjectives and are rated on a five-point scale, labeled as "not at all or slight," "a little," "moderately," "quite a bit," and "very much." The PA and NA scales were constructed to be uncorrelated, and conforming with the theoretical model positing the independence of positive and negative affect, they generally are.

Much of the work with the PANAS has been correlational, and the scales correlate with external variables in ways that imply validity. For example, extraversion correlates with frequent reports of PA, and neuroticism correlates with frequent reports of NA. Few studies have used the PANAS in experimental research. Larsen and Ketelaar (1991) used the PANAS in an experiment wherein pleasant and unpleasant moods were induced using guided imagery techniques. They found that the positive induction increased PA but did not lower NA, and that the negative induction increased NA but did not lower PA. This differential sensitivity to positive and negative emotion inductions supports the construct validity of the PANAS. The PANAS has not been free from criticism, however. The reader may refer to Larsen and Diener (1992) for a discussion of potential problems and misinterpretations of the circumplex model and of the PANAS as a measure of that model.

Adding the Temporal Dimension to Self-Reports
Self-report measures of emotions, whether single-item or multiple-item, require research participants to report *globally* on an emotional episode that extended over time. For instance, researchers might ask, "How pleasant or unpleasant was your visit to the dentist?" or, ". . . your experience of childbirth?" or, "How much fear did you feel while watching this film?" This measurement strategy is often used without recognizing that the mental processes respondents must invoke to supply these global self-reports can introduce distortion and measurement error. Specifically, providing a global self-report implicates both memory processes (respondents recall the targeted episode) and aggregation processes (respondents in some manner combine their multiple and often varied momentary experiences into an overall report). Both of these mental processes may obscure or misrepresent dynamic changes in emotion as experienced over time. For instance, Fredrickson and Kahneman have documented that people's global reports for extended emotional episodes draw highly from the momentary affect experienced at the most intense and final moments of the episode (called the peak-end rule; see Kahneman, this volume), with the duration of the emotional experience largely neglected (Fredrickson and Kahneman 1993; Kahneman et al. 1993; for related issues, see Thomas and Diener 1990).

Real-Time Ratings One way to circumvent some of the problems inherent in global reports is to collect real-time ratings of emotion. In recent years, several such techniques have been developed. The general strategy across these techniques is to collect self-reports of subjective experience on a moment-by-moment basis, either *on-line* as the emotion is first experienced or *retrospectively* as the temporal dimension of the original episode is "replayed" while real-time momentary self-reports are collected. A sample of momentary self-report measures is described later.

Conceptually, the most basic real-time self-report measure can be viewed as a single-item measure (as described earlier) with a temporal dimension added. Using either a rotating dial or a sliding meter, respondents are instructed to adjust a pointer as often as necessary so that it always reflects how they are feeling each moment throughout an extended episode. Several researchers have described continuous "rating dials" of this sort (Fredrickson and Kahneman 1993; Fredrickson and Levenson, in press; Gottman and Levenson 1985; Bunce, Larsen, and Cruz 1993). Like single-item measures more generally, rating dials may use either bipolar ("very negative" to "very positive") or unipolar verbal anchors ("no sadness at all" to "extreme sadness") and either Likert-type or visual analog scales.

In addition to capturing the ebb and flow of emotional experience over time, continuous rating dials also automate data acquisition. The dial itself is connected to a potentiometer or rheostatic resistor that controls the voltage output from a common nine-volt battery (much like a dimmer switch controls the amount of electricity going to a lighting fixture). The electrical output from the dial is then monitored by an analog-to-digital (A/D) data-acquisition device to record continuously respondents' self-reports. Properly calibrated, the amount of electricity at the recording output is a direct representation of the respondent's moment-by-moment self-report.

When the demands of an experimental protocol are low (for example, viewing emotional film clips), research participants can use a rating dial to provide continuous self-reports of emotion "on-line" during the actual emotional episode (Fredrickson and Kahneman 1993; Fredrickson and Levenson, in press). In contexts where such on-line measurement would be too cumbersome or disruptive (for example, during actual social interaction), participants can use a rating dial to provide continuous, *retrospective* self-reports of their emotional experience, so long as the temporal dimension of the original experience is "replayed" during the rating procedure. In studies of emotions in marital interaction, for instance, Gottman and Levenson (1985) obtained continuous self-reports of emotion experience using a video-recall technique. In an initial session, they had married couples discuss an area of conflict in their marriage while their conversation and nonverbal displays were recorded on video and each spouse's autonomic reactions were recorded with physiological sensors. In subsequent individual sessions (again with video and physiological recording), spouses each independently viewed the videotape of their conversation and used a bipolar rating dial to indicate how positive or negative they were feeling each moment *during the actual interaction*. Validating this video-recall technique, Gottman and Levenson (1985) reported that each spouse's autonomic activity during the later rating session patterned that evident during the actual marital interaction, suggesting that viewing the videotaped conversation was sufficient to re-create (to some degree) the affect experienced during the original episode.

One drawback of the momentary self-report measures described thus far is that they limit self-reports to just one or two dimensions. Certainly it is *technically* feasible to create a whole bank of rating dials, perhaps one to reflect each of several discrete emotional states (anger, fear, sadness, disgust, attraction/love, enjoyment, contentment, and so on—akin to adding a temporal dimension to Ekman's [1992] various single-item scales or Izard's [1977] DES). The limiting factor, however, would be the respondent's ability to track the ebb and flow of multiple discrete emotions simultaneously, in real time. One way around this obstacle would be to collect self-reports for multiple emotions using multiple iterations of the video-recall technique. Such a strategy, however, would no doubt push the limits of participants' cooperation and/or induce fatigue.

A more reasonable way around this obstacle is to use a hybrid technique, introduced by Rosenberg and Ekman (1994) and called "cued review," which derives partially from Gottman and Levenson's (1985) video recall technique. In cued review, participants are instructed to stop the video replay at moments when they remember having felt an emotion during the original episode. They then use a multiple-item emotion report form to rate what they remember feeling at that precise moment. For example, Rosenberg and Ekman collected Likert-type ratings for eight emotion terms: anger, contempt, disgust, embarrassment, fear, happiness, sadness, and surprise. After completing a given rating form, participants then restarted the video and, whenever they remember having felt a change in emotion (in either degree or type) during the original episode, they stopped the playback again and completed another emotion report form for that moment. This procedure was then repeated for the entire emotion episode. Because the emotion report forms can contain separate ratings for multiple, discrete emotions, the cued-review technique uncouples momentary self-reports from unidimensional scales or two-dimensional circumplex models. The resulting self-report data, although momentary, are neither continuous nor equally spaced. Providing validity for this technique, Rosenberg and Ekman found that momentary reports of specific emotions obtained through cued review coincided in time with facial indications of the same specific emotions.

Advantages to these automated techniques include the ease of administration, the on-line nature of the recording, the ability to record continuously for long time periods, and the lack of data entry concerns (provided the output is read by computerized A/D equipment). The major disadvantage is the need for specialized equipment and the fact that the participant is literally tied down by the device (though this could change if radio telemetry or on-board memory could accompany a rating device that the participant could carry during the reporting period). Moreover, it seems likely that continuously monitoring the participant's emotions may lead to a form of fatigue or be so intrusive that it actually alters his or her emotions. These issues remain open questions for researchers.

Evaluation of Self-Report Methods Self-report methods are perhaps the most efficient and easiest techniques for measuring emotions. Even so, they rely on the assumption that research participants are both *able* and *willing* to observe and report on their own emotions. Often fused with the assumption that participants are able to report their emotions is the corollary assumption that self-reports are in fact the *best* source of information about an individual's emotional experience. Each of these assumptions, however, can be questioned. For example, if some emotional episodes are either outside of phenomenal awareness or not represented in working memory, participants will be unable to perceive or recognize the feeling state accurately and, as a consequence, unable to provide accurate self-reports. Of course, some would question whether an unperceived emotion is an emotion at all. Without fully entering the debate about the existence of unconscious emotions, it seems possible that a person might "have" an emotion in a nonverbal channel (for example, autonomic activation or action tendency) yet never label that experience and hence not perceive it as an emotion at all (Tranel and Damasio 1985). Moreover, some persons may repress emotional experiences, resulting in biased or incomplete memory for emotions (Newton and Contrada 1992). There is some evidence that the repressive coping style works by preventing emotions from being encoded into memory (Cutler, Bunce, and Larsen 1996). Certainly, using measures in addition to self-report would be important in these instances.

Other issues underlying the assumption that participants are able to observe and report on their emotions are more practical. Certain populations, for various reasons, may have meager comprehension of semantic information. Very young children are one example. Other populations, like the very old, may not have the concentration or attention span to complete a lengthy self-report measure like the MAACL. Responses to self-report questionnaires may not provide accurate estimates of emotional states in such samples. Measurement accuracy may be similarly jeopardized when rating scales are used with participants whose principal language is not the one in which the instrument is presented. Translation is always questionable, given cultural variation in the experience, comprehension, and linguistic expression of emotion. Cultural psychologists have further argued that some cultures have emotions, or emotion terms, that are not identifiable in other cultures (see, for example Mesquita and Frijda 1992). For all of these reasons, self-report scales should be brief and easy to comprehend, and researchers should attend to cultural, demographic, and contextual factors that might compromise accurate responding.

Turning to the second assumption—that participants are *willing* to report on their emotions—the main issue is one of response sets, where responses to items may contain noncontent variance. That is, the participants' responses might reflect something that is not contained in the questionnaire itself. The most frequently discussed response set is socially desirable responding: the participant responds to the items in a manner that creates a positive impression or makes him or her appear to possess mostly positive attributes. People may be motivated to deny undesirable attributes or emotions and to endorse positive ones. One way to control social desirability responding is to measure it using a social desirability measure (for example, the Marlowe-Crowne scale) and partial it out in statistical analyses. Some researchers, however, question this approach (Diener, Smith, and Fujita 1995), primarily on the basis of the validity of social desirability measures.

A different response set is extreme responding: a participant may be motivated to use scale end points or large numbers in describing his or her emotions. While some researchers have written about this, the few studies done on extreme responding on emotional trait questionnaires have not found much evidence that this is a problem (Larsen and Diener 1987). Other researchers have argued that even a small amount of extreme responding can introduce systematic distortions that particularly affect the covariance structure of a set of ratings (Bentler 1969). The effect of extreme responding would be to attenuate negative correlations between polar opposite terms. A recent discussion and demonstration of correlated error in affect ratings is provided by Green, Goldman, and Salovey (1993). These authors demonstrate the utility of multiple measures of emotion—something we also recommend—in accounting for random and nonrandom (response bias) measurement error.

Another potential problem with self-report concerns the effects of repeated assessments. One issue is measurement reactivity, the idea that the actual process of measurement alters the thing being measured. Administering an emotion adjective rating scale multiple times may in fact create or alter the emotional state of interest. A second issue is measurement independence. Researchers often want to assess emotion frequently during an experiment, especially in within-subject designs. Ideally, each measurement occasion is independent from the last. The only way to achieve this in a repeated-measures experiment would be to remove the previous experience with the self-report items from the participant's memory prior to each new assessment. Because this is not possible, one potential effect of repeated emotion measurement is stereotypic responding (Stone 1995): participants settle into a response profile that does not change much across the assessment occasions. This can be assessed by examining standard deviations across assessment occasions.

Observer Ratings of Emotion

With sufficient information available, virtually any self-report measure described in the previous section might also be collected from a third-person perspective. Such observer reports might be obtained from "expert" observers of the target person's emotional experiences (such as a spouse, best friend, or therapist) or simply from strangers without any special training. The key is to provide the observer-rater with emotion-relevant information about the target person's experience—written accounts, audiotaped or transcribed dialogue, video recordings or photographs of facial behavior, or some combination of these data. Upon reviewing these data, observers make judgments about the likely emotional state of the target person (including type and/or intensity) either globally or at a particular moment. It is critical to note, however, that observer reports like these represent *social attributions* about a target person's emotional state and should be cross-validated against other emotion measures. Like attribution processes more generally, attributions about emotion are constrained by the information available or biased by an observer's self-serving tendencies. (This is perhaps most true when an observer-rater has a close relationship with the target person.)

A conceptually related method of obtaining observer reports is to use specially trained observers to code emotions. One example of this method is the Specific Affect Coding System (SPAFF) developed by Gottman and his colleagues to study emotions in marital interaction (Krokoff, Gottman, and Hass 1989; Gottman 1993). This system separates the emotion evident in marital exchanges into specific positive and negative categories. The SPAFF positive affect categories are interest, affection, humor, validation, and excitement/joy. The SPAFF negative affects are anger, belligerence, domineering, contempt, disgust, tension/fear/worry, sadness, whining, and defensiveness. Similar to observers without special training making attributions about emotions, when making emotion ratings SPAFF coders consider a gestalt of information, including verbal content, voice tone, context, facial expression, gestures, and body movement. This is what makes the SPAFF a "cultural informants" coding system rather than a physical features coding system (for example, Ekman and Freisen's [1978] Facial Action Coding System [FACS], described later). What sets SPAFF coders apart from other observers is (a) their special training in recognizing important physical markers of emotion in the face and voice, and (b) the pace of their coding. Although Gottman has divised an "Affect Wheel" to use the SPAFF system in real-time interactions, more commonly SPAFF requires microanalyses of video recordings—six to ten hours of coding for each fifteen minutes of dyadic interaction.

A key advantage of observer reports is that they are often unobtrusive and can track naturalistic social exchanges. And when no special training is required of observers, they are also inexpensive and fast measures. Gottman (1993) argues that gestalt approaches to coding emotions circumvent the assumption hidden within physical features coding systems (such as FACS) that different emotion components or channels combine additively to create emotional meaning. Among the disadvantages of observer coding systems like SPAFF is the intensive training required of the coders. Moreover, SPAFF has been developed specifically to study marital interactions and may not suit other types of interactions, like those between friends or coworkers or intergenerational relationships. Studying emotions in these other interpersonal contexts may require new codes altogether.

Other recent studies have shown the utility of using relatively untrained informants to provide observer reports. For example, Watson and Clark (1991) asked subjects to sign up for their study with some friends or acquaintances. Among well-acquainted peers, they found mostly significant correlations between self-report and peer-reported emotions—for example, .52 for sadness, .49 for positive affect, .40 for fear, and .31 for hostility. Similar untrained peers were used by Diener, Smith, and Fujita (1995) and Lucas, Diener, and Suh (1996), with similar convergence results. Such findings bolster the view that, although emotions are thought to be private, there is nevertheless a public aspect that can be tapped through trained and even untrained observers.

Facial Measures of Emotion

Coding Systems One of the most comprehensive and widely used systems for coding emotion in the face is the Facial Action Coding System (FACS) (Ekman and Friesen 1975, 1978). The FACS consists of forty-six anatomically based "action units" (AUs). Each AU refers to a specific observable change in the face. For example, AU 1 raises the inner brows, AU 9 wrinkles the nose, and AU 12 raises the outer lip corners. The system describes all possible movements in the skin of the face observable to the naked eye. There is an extensive training and certification system for learning the FACS (contact the Human Interaction Lab at the University of California at San Francisco). This self-paced training program involves learning about the muscular and appearance basis of each AU, extensive exposure to the forty-six AUs and their combinations in photos and videotape, instructions for producing the AUs with one's own face, and rules for specifying minimal changes for scoring and combining AUs. It requires about forty hours of initial training to achieve acceptable reliability (Ekman and Friesen 1975).

Facial coding is useful in measuring emotion to the extent that overt, spontaneous facial changes accompany people's emotional responses. Yet because facial muscles are also enervated by the voluntary nervous system, observable facial action is not simply a direct readout or "expression" of underlying emotional states. Emotion-related facial actions, for instance, may be controlled through inhibition, exaggeration, or masking. Nevertheless, FACS has been very useful in studies of emotion. For example, FACS codes can reliably distinguish so-called genuine smiles ("Duchenne smiles"), which are spontaneous expressions of positive emotion, from so-called deceptive smiles "non-Duchenne smiles"), which are often deliberate attempts to appear as if positive emotion is being felt when it is not (Ekman, Friesen, and O'Sullivan, 1988). A recent edited volume describes a range of research programs in which FACS has been a critical ingredient (Ekman and Rosenberg 1997).

Full use of the FACS provides exhaustive real-time description of facial action. It also demands a lot of time and effort. For example, FACS-scoring videotaped faces requires about one hour of coding for each minute of videotape (depending, of course, on the density of facial action). For many research questions less fine-grained codings of facial expressions may be reasonable, and several researchers (including Ekman and his colleagues) have developed selective, emotion-specific, and/or global systems for coding facial action (for example, EMFACS [Emotion FACS]—see Fridlund, Ekman, and Oster, 1987; MAX [for Maximally Discriminative Facial Movement Coding System] by Izard 1979; for global coding systems, see Gross and Levenson 1993; Kring and Neale 1996).

Electromyography Facial measures of emotion may also be obtained using physiological measures of muscle contractions. The neural activation of the striated muscles in the face (and elsewhere in the body) produce muscle action potentials that can be detected using electromyography (EMG). EMG recordings are obtained using two electrodes placed over the muscle bundle of interest. The electrical signal given off by the muscle during contraction is on the order of a few to a few hundred microvolts, though facial muscle contractions in typical lab settings rarely exceed eighty microvolts. The amount of electrical activity detected over the muscle is directly related to the number of motoneuronal pools involved in the contraction. Detailed descriptions of facial electromyographic technique may be found in Cacioppo and Tassinary (1990).

The muscles typically assessed using EMG are the corrugator supercilia (responsible for the furrowed brow that occurs with many unpleasant emotions) and the zygomaticus major (responsible for pulling the corner of the mouth back and up, toward the ear). Other muscles, such as those responsible for wrinkling the nose during disgust, are also sometimes assessed. Evidence for the validity of facial EMG suggests that this is an effective technology for assessing both the valence and intensity of affective responses (Cacioppo et al. 1986). Moreover, EMG techniques can assess neuromuscular actions that are too small to generate visible changes in the face (Cacioppo et al. 1986). As such, EMG may be more sensitive to emotion in the face than is FACS, albeit in many fewer locations. Such sensitivity has a downside, however, in that electrical signals from sites other than the muscle of interest may also be detected during EMG assessments. Researchers interested in measuring emotions with facial EMG should seek training in electophysiological measurement and/or collaborators with appropriate expertise.

Autonomic Measures of Emotion

Emotions are often closely tied to urges to act in specific ways, be it to strike out against a competitor, escape imminent danger, or be near a loved one. Many emotion theorists view the link between emotions and action tendencies as part and parcel of the definition of emotions (Frijda 1988; Lazarus 1991), and arguably, it is this association that makes emotions "embodied" (Lazarus 1991), evident in both somatic nervous system activity and, when emotions are intense and/or prolonged, in autonomic nervous system (ANS) activity (Cacioppo et al. 1993).

Although a wide range of theorists have tried to describe the precise relationships between emotions and ANS activity, these can be distilled into those who argue that distinct emotions are associated with distinct ANS activity (for example, Averill 1969; Levenson, Ekman and Friesen 1990) and those who argue that distinct emotions are associated with undifferentiated ANS activity (for example, Cannon 1927; Mandler 1975; Schacter and Singer 1962). Although empirical support for autonomic specificity across emotions has been observed in multiple studies and in multiple laboratories, the cumulative data are mixed and therefore remain inconclusive (for reviews, see Cacioppo et al. 1993; Levenson 1992; Zajonc and McIntosh 1992).

Psychophysiological Inference The state of the science, then, does not support the use of autonomic measures (used singly or in combination) to index or infer specific emotions. That is, we cannot distinguish anger from fear, or disgust from anger (or any emotion from any other emotion), solely using autonomic measures. Moreover, even though emotions (when sufficiently intense) reliably yield ANS changes, neither can we distinguish emotional states from non-emotional states solely using autonomic measures. As Cacioppo and Tassinary (1990) put it: "When a physiological event differentiates the presence versus absence of a particular psychological element, one may infer the *absence* of this psychological element given the nonoccurrence of the physiological event, but one cannot infer anything about the *presence* of the psychological element given the occurrence of the physiological event" (24). This is so because ANS activity can (and does) index a range of psychological events, including but certainly not limited to emotions. These include attentional states, such as orienting to novel stimuli (Graham and Clifton 1966; Lacey et al. 1963), anticipated or actual somatic activity (Obrist et al. 1970), respiration (Porges, 1995), as well as individual differences (Levenson, 1983).

At present, then, how can autonomic measures be used to index emotions? The answer is simple: in combination with other (non-ANS) measures of emotion. In other words, the same caveat that we have applied to all other measures discussed in this chapter also applies to autonomic measures: any single measure of emotion is imperfect and incomplete. Researchers can be more confident about the presence of an emotion to the extent that multiple measures provide independent and converging evidence of that emotion.

Fruitful Autonomic Measures Of the dozens of different autonomic measures that have been used to measure emotions over the last several decades, some have been more fruitful than others. One set of measures indexes *electrodermal* activity (skin conductance is currently the accepted and most reliable measure), another set of measures index *respiratory* activity, and perhaps the broadest set of measures index *cardiovascular* activity. Within this last and largest set, measures range from gross, end-organ responses (for example, heart rate, diastolic and systolic blood pressure) to measures of the various underlying hemodynamic processes responsible for these end-organ responses (such as cardiac output, stroke volume, and total peripheral resistance; interested readers should see Sherwood [1993] and Sherwood et al. [1990] for information on impedance cardiography). Still other measures link respiratory to cardiovascular activity (for example, Respiratory Sinus Arrhythmia [RSA], a purported measure of cardiac vagal tone; Grossman, van Beek, and Wientjes 1990; Porges 1995). Researchers interested in incorporating autonomic measures into their empirical projects should seek out special training and/or experienced collaborators.

In addition to issues of theory and inference, there are many practical issues to keep in mind when considering the use of autonomic measures. First, autonomic measures vary widely in how invasive they are. By consequence, some autonomic measures might elicit emotions in and of themselves. On the less invasive end are measures of pulse rate and skin conductance that simply require that sensors be placed on a participant's fingers. Impedance cardiography, by contrast, uses band electrodes that circle a participant's neck and chest in several locations. To have these sensors attached, participants must disrobe partially, a requirement that is likely to elicit subjective reactions. Measures of blood pressure are often invasive in another way: most use pressurized cuffs on either the upper arm or the finger. The pressure in these cuffs can draw attention and sometimes even pain, which, again, can elicit emotional reactions in and of itself. Invasiveness of this sort certainly complicates emotion measurement. Second, autonomic measures have typically greatly restricted participants' mobility simply because autonomic signals are carried by wires that connect sensors to amplifiers and recording devices (for example, computers). Because lengthy wires and body movement can sometimes increase measurement noise, participants are often on a rather short tether, being required to remain seated and largely immobile. While ambulatory autonomic monitors have been available for some time, their reliability does not yet match laboratory-based measures and may still be subject to movement artifacts. And third, the temporal resolution of various autonomic measures varies widely. Although autonomic measures are increasingly available on a continuous basis, some measures (for example, RSA, impedance cardiography) require somewhat longer durations for reliable measurement (perhaps one minute), epochs that may be longer than the duration of any given emotion episode (see earlier discussion of timing).

Brain-Based Measures of Emotion

In the last decade or so, researchers have begun to refine neurophysiological measures of emotion. Scalp-recorded brain electrical activity, or electroencephalogram (EEG), can index patterns of anterior asymmetries that distinguish specific emotion states as well as individual differences in affective style (for a review, see Davidson 1993). For instance, Davidson and his colleagues have demonstrated that approach-related positive emotions are associated with left anterior activation whereas withdrawal-related negative emotions are associated with right anterior activation. Other, more localized measures of emotion-related changes in the brain are on the horizon as well, including PET scans and functional MRI (see, for example, Lang et al. 1998). Many of the same inferential and measurement issues discussed with respect to autonomic measures of emotion also apply to measures of brain activity. Again, researchers interested in measuring emotion-related changes in the brain should seek out special training and/or experienced collaborators.

Vocal Measures of Emotion

Although most information conveyed by vocalization derives from verbal content (language use), voice stylistics (such as pitch, loudness, tone, quality, timing) can convey much information about a speaker's emotional state because vocalization is a *bodily* process sensitive to emotion-related changes in the broader bodily context (for example, muscle tension, respiration rate, and blood pressure). As such, changes in a speaker's emotional state often yield quite noticeable changes in voice stylistics.

Emotion-related vocal changes have been assessed using both low-tech and high-tech means. The low-tech path is to listen (with or without special training) to audiotaped speech samples and evaluate them on emotional terms. The high-tech path is to have these same audiotapes digitized and analyzed by electro-acoustic equipment and/or digital computers that decompose the speech sound waves into a set of acoustic parameters. We describe each of these measurement techniques in turn.

Several "decoding" studies have tested the abilities of untrained listeners to recognize correctly or infer speakers' emotional states (for an early review, see Scherer 1986; van Bezooijen 1984). Typically, these studies ask actors to read standard or meaningless sentences in manners that convey specific emotional states, such as anger, fear, disgust, joy, sadness, even contempt, pride, love, and jealousy. These speech samples (sometimes content-filtered) are then played for naive listeners who must choose the intended emotional state from a list of forced-choice alternatives. After correcting for chance guessing and sampling error, recognition rates across these studies are about 50 percent, which is four to five times what would be expected by chance (Pittam and Scherer 1993). Recognition rates across cultures (and languages) remain higher than chance, consistent with the claim of universal vocal patterns of emotion (van Bezooijen, Otto, and Heenan 1984). Nonetheless, some emotions are easier to recognize than others: sadness and anger are best recognized, whereas disgust, contempt, and joy are least recognized (Pittam and Scherer 1993; van Bozooijen et al. 1983). Moreover, analyzing confusions in these decoding studies reveals that the arousal level of a speaker's emotional state is better transmitted by vocal cues than is the evaluative component (positivity or negativity) (Apple and Hecht 1982; van Bozooijen et al. 1983).

Listener ratings of vocal emotion have been particularly useful in research on the emotional trait of hostility, perhaps because many people may be unwilling or unable to identify themselves as hostile on self-report instruments (for information on the Interpersonal Hostility Assessment Technique, or IHAT, see Barefoot 1992; Haney et al. 1996). Strikingly, hostility assessed through the voice has been shown to relate better to adverse health outcomes than hostility assessed through self-report (Barefoot et al. 1994).

Comparing low-tech and high-tech means of detecting vocal expressions of emotion, Scherer (1986) has noted a paradox: "Whereas judges seem to be rather accurate in decoding emotional meaning from vocal cues, researchers in psychoacoustics and psychophonetics have so far been unable to identify a set of vocal indicators that reliably differentiate a number of discrete emotions" (143–44). Among the set of acoustic parameters typically gathered using high-tech means are (a) fundamental frequency, or F_0 (the rate at which the vocal folds vibrate, perceived as overall voice pitch); (b) minute perturbations in F_0 (two indices—"jitter" and "shimmer"—assess cycle-to-cycle variations in the frequency and amplitude, respectively, of F_0); (c) intensity (energy values indexed in decibels, perceived as loudness); and (d) speech rate or tempo (for a more complete listing of acoustic parameters, see Scherer 1986). In digitally analyzed emotional speech, these acoustic parameters tend to covary and cluster into two types: one pattern combines high and variable F_0, high intensity, and fast tempo and marks the high-arousal emotions such as joy, anger, and fear. A second pattern combines low and stable F_0, low intensity, and slow tempo and marks the low-arousal emotions such as sadness, boredom, and contempt. To date, then, high-tech acoustical analyses effectively identify the arousal level associated with different emotional states, even when emotions are quite mild (for an excellent example, see Bachorowski and Owren 1995). Even so, these high-tech measures fall short of identifying the particular emotions experienced. Perhaps most notably, positive and negative emotional states often remain undistinguished. Scherer (1986; see also Pittam and Scherer 1993) has argued, however, that given untrained listeners' abilities to infer speakers' specific emotional states, patterns across acoustic parameters ought to distinguish between discrete emotions—in theory. Empirical support for Scherer's claim is still lacking.

When considering whether to pursue vocal measures of emotion, researchers should keep the following in mind. First, vocal indicators of emotion are not always present. Unlike facial and autonomic measures, voice is not a continuous variable for the simple reason that people do not speak continuously. Second, positive and negative emotions often go undistinguished by sound-wave analyses and are even sometimes confused by human raters. Vocal measures, then, are perhaps best used in conjunction with other measures of emotion. And third, much like the face and self-report, the voice can reflect both emotional/physiological "push" and sociocultural "pull" effects (Scherer 1989). That is, beyond reflecting internal physiological states, vocalization also reflects ritualized communication patterns, impression management, and coping styles. Disentangling push effects from pull effects is rarely an easy task.

Vocal measures of emotion are perhaps most useful either when voice is only one strand of the emotion-related data available (for example, audiotaped interviews, perhaps via telephone) or when the experimental situation disallows more invasive or obtrusive bodily measures (for example, physiological sensors on the skin, visible video cameras). Certainly, high-tech vocal measures will advance in step with the currently building knowledge about the acoustic parameters associated with emotional speech. Once again, training and/or expert collaborators are recommended for sound-wave analysis.

Emotion-Sensitive Tasks

A variety of tasks have been shown to be sensitive to affective states. These emotion-sensitive tasks can be defined as tasks on which response or performance differences are, at least in part, a function of emotional state (Mayer 1986; Mayer and Bremer 1985; Mayer, Mamberg, and Volanth 1988). While many of these emotion-sensitive tasks started out as independent and dependent variables in experimental paradigms, many are now used as manipulation checks. That is, many researchers see the links between these tasks and emotions as reliable enough to use responses to these tasks as indicators that an emotion has been induced through some manipulation.

Cognitive Appraisals Many emotion researchers hold that emotions result when individuals appraise the meaning of a particular situation or event in certain ways. Different emotions are distinguished by different appraisals. For example, if an individual perceives that she was wrongly treated by someone, and she makes the further appraisal that the other did so *willingly*, the resulting emotion is likely to be anger. Moreover, a relatively small set of appraisal dimensions (eight or ten) are thought to account for a large proportion of emotional experiences. This appraisal model implies that one way to measure emotions is to measure cognitive appraisals of specific situations or events. Smith and Ellsworth (1985; 1987) identified a series of cognitive appraisal dimensions that distinguish between specific emotions. Because appraisals target people's interpretations of *situations*, they circumvent some of the problems associated with self-reports of subjective experience. As such, measuring a person's appraisals of a situation may indirectly reveal his or her emotional state. For example, if a subject said that some hurtful event was caused by someone else on purpose, then we might infer that he or she is angry.

Action Tendencies Many lines of research have converged on a central notion that some form of action tendency (sometimes called action readiness, action disposition, or behavioral activation) is a central component of emotional experience. Physiological reactions (such as increased heart rate in fear) are often seen as preparatory for action readiness (the tendency to flee that comes with fear). Other specific action tendencies are associated with other specific emotions (such as withdrawal and sadness or striking out and anger; see Fridja [this volume] for a more detailed discussion of action tendencies). The probability of participants engaging in such actions, or saying that they would like to engage in such actions, has been related to specific affective states (Frijda, Kuipers, and ter Schure 1989; see also Fredrickson et al., in press).

A related judgment task is to ask participants how much they would like to engage in various behaviors, such as talk with a good friend, engage in some exercise, or have a pleasant meal. Teasdale, Taylor, and Fogarty (1980) have used this task and found it sensitive to depressed mood. This task supposedly works because sadness is related to the action tendency to withdraw. A related concept is that, when depressed, people often lose interest in activities that formerly gave them pleasure. Depressed mood is thought to be associated with depressed psychomotor function. As such, tasks involving coordinated psychomotor movements should be sensitive to sadness or unpleasant affective states. Writing speed, for example, is a popular psychomotor task thought to be influenced by depressed mood. Velten (1968) used this task as a non-self-report measure of affect in the validation study of his "Velten mood induction procedure." Participants' writing speed was significantly slower after reading the depression Velten statements, compared to writing speed following the elation Velten statements. Other psychomotor tasks that have been used in emotion research include letter cancellation and smooth pursuit motor tasks. Performance speed is most often the variable sensitive to depressed emotional state, and the effect appears unipolar (for example, pleasant moods do not necessarily increase psychomotor speed).

Performance Measures One category of emotion-sensitive performance measures consists of various judgment tasks. One popular judgment task is to have participants make probability estimates of the likelihood of various good and bad events. For example, participants may be asked the probability of being killed in a tornado, dying in an airplane crash, or contracting cancer in their lifetime. It has been shown that persons in unpleasant emotional states overestimate the probability of such bad events (Johnson and Tversky 1983). Moreover, the events do not have to be self-referential to be sensitive to affective states (Cunningham 1988). Ketelaar (1989) showed that people in a good mood also overestimate the probability of pleasant events, such as the probability of the economy improving over the next year, or the probability of a good friendship lasting an entire life.

Several theoretical explanations for why such judgments should be sensitive to mood have been offered (see, for example, Mayer 1986). Chief among these are mood-congruent recall effects, spreading activation models, and category boundary shifts. When using such tasks to assess affective states, it is important that participants understand probability ratings. Providing an example or two is helpful (for example, "The probability of tossing a coin and having it land heads up is 50 percent"). Some questions might be phrased in terms of "what percent of the population . . ." Because percentages are often skewed, some data transformation may be in order. Also, because some participants have no idea about percentages but will nevertheless provide estimates, it may be advantageous to normalize data within participants across probability estimates.

Another useful performance task is to ask participants to generate associations to positive, neutral, and negative stimuli. For example, write down as many words as come to mind in sixty seconds when you hear each of the following stimulus words: happy, disappointed, generous, destroy, peace, pain, and so on. Mayer and Bremer (1985) showed that performance on this task correlates with naturally occurring mood. Ketelaar (1989) showed that such a word association task correlates with self-reported mood following pleasant or unpleasant mood inductions. Seidlitz and Diener (1993) used a variation on this task: participants were asked to recall as many happy experiences from their lives as they could in a given time period. Participants higher on trait happiness recalled more pleasant experiences, in the same time period, than participants lower on trait happiness. Teasdale and his colleagues (Teasdale and Fogarty 1979; Teasdale and Russell 1983) have also demonstrated that temporary emotion inductions influence recall of pleasant and unpleasant events in predictable ways.

A second category of performance measures involves information-processing parameters. Reaction times in lexical decision tasks, for example, have been shown to be sensitive to affective states (Challis and Krane 1988). The participant's task here is to judge whether a string of letters presented on the computer screen represents a word or a nonword. On each trial, the letters that come on the screen represent a nonword, an emotion word (for instance, *anger*), or a neutral word (such as *house*). Participants in positive affective states are quicker and sometimes more accurate at judging positive words as words than participants in neutral states, and vice versa for unpleasant moods (Niedenthal and Setterlund 1994).

A variation that also involves information-processing is to present participants with incomplete word stems and ask them to add letters to complete the word. Word stems are selected so that they can be completed as an emotion term or as a neutral term. For example, ANG__ could be completed as ANGER, ANGLE, ANGEL, or ANGLO. A related technique is the use of homophones (words that sound alike but have different meanings). With this technique, the subject hears the word (*die* or *dye*, for example) and is asked to write it. Participants in an unpleasant mood are more likely to write or complete the word stems in a manner congruent with their mood (Halberstadt, Niedenthal, and Kushner 1995).

Startle Potentiation Another emotion-sensitive task relies on a very simple behavior—the startle reflex. The startle is characterized by a rapid shutting of the eyes (blink), pulling the chin down, and a rapid inhalation. It is a defensive response, and its protective value (shutting the eyes) is obvious. The startle reflex is also easy to elicit through the application of a sudden and loud acoustic stimulus. Because it is a reflex, it is not easily controlled, although like many reflexes, adaptation occurs with repeated stimulation.

Startle potentiation refers to an increase in the startle response brought about by an emotional state (Vrana, Spence, and Lang 1988). For example, if a person is in an aversive or unpleasant state when the startle stimulus is emitted, the startle blink will be faster and stronger than if the participant were in a neutral emotional state. If a person were already anxious, an augmented defensive response makes evolutionary sense. Researchers typically measure the muscle contraction responsible for the blink during the startle response. They can then score the blink for latency (time from the startle sound to the onset of the muscle contraction) as well as for magnitude (the force of the muscle contraction producing the blink).

The researcher most responsible for developing this technique in humans is Peter Lang (see, for example, Lang, Bradley, and Cuthbert 1990). Lang has demonstrated startle potentiation for unpleasant emotions, as well as a slowing-down of the startle during positive emotions, compared to neutral states. This effect had been well documented in animals for decades. Christopher Patrick has studied individual differences in startle potentiation, with an emphasis on psychopaths (Patrick 1994; Patrick, Cuthbert, and Lang 1994). Psychopaths are thought to be deficient in fear and other self-regulating negative emotions. Patrick's research shows that psychopaths do not show the expected pattern of startle potentiation to fear or anxiety stimuli, even though they do show the expected slowing of startle during positive emotions found in normal samples.

The strengths of the startle potentiation technique are that it is a nonverbal, nonvoluntary, and extremely fast measure of internal affective state. This measure, however, appears limited to assessing the pleasantness-unpleasantness or approach-avoidance dimension of affective state. In addition, the laboratory equipment and expertise necessary to employ this technique represent a heavy cost to the researcher. Nevertheless, this emotion measure appears promising, and researchers who think they might benefit from employing this technique should consider training and/or collaborating with a psychophysiologist who is set up to analyze eyeblinks.

Conclusions

Emotion measures come in many forms and, in our opinion, should be used in many forms. Perhaps most important, no single emotion measure can serve as the "gold standard" for other emotion measures. Each measurement type has its strengths and its weaknesses, and each in isolation provides only an incomplete picture of emotion processes. So, to the extent that emotions invoke changes across numerous channels or component systems, data streams from those various channels should be collected in synchrony. Cross-referencing multiple measures of emotion increases researchers' chances of pinpointing emotions and discerning their precursors and effects.

In this chapter, we have reviewed the many classics of emotion measurement, alongside some relatively new, cutting-edge measures. When choosing and employing these measures, researchers should consider the various issues underlying emotion measurement discussed throughout this chapter. With an appreciation of these measurement issues, researchers should feel comfortable adding novel, theoretically derived measures to the mix of those that they collect. Any list of valid emotion measures will surely need updating with time as new links are added to the network of associations surrounding the construct of emotion.

References

Apple, W., and Hecht, K. (1982). Speaking emotionally: The relation between verbal and vocal communication of affect. *Journal of Personality and Social Psychology, 42,* 864–75.

Averill, J. R. (1969). Autonomic response patterns during sadness and mirth. *Psychophysiology, 5,* 399–414.

Bachorowski, J., and Owren, M. J. (1995). Vocal expression of emotion: Acoustic properties of speech are associated with emotional intensity and context. *Psychological Science, 6,* 219–24.

Barefoot, J. C. (1992). Recent developments in the measurement of hostility. In H. S. Friedman (Ed.), *Hostility, coping, and health* (pp. 13–31). Washington, D.C.: American Psychological Association.

Barefoot, J. C., Patterson, J. C., Haney, T. L., Cayton, T. G., Hickman, J. R., and Williams, R. B. (1994). Hostility in asymptomatic men with angiographically confirmed coronary artery disease. *American Journal of Cardiology, 74,* 439–42.

Bentler, P. M. (1969). Semantic space is (approximately) bipolar. *Journal of Psychology, 71,* 33–40.

Bonanno, G. A., Keltner, D., Holen, A., and Horowitz, M. J. (1995). When avoiding unpleasant emotions might not be such a bad thing: Verbal-autonomic response dissociation and mid-life conjugal bereavement. *Journal of Personality and Social Psychology, 69,* 975–89.

Bunce, S. C., Larsen, R. J., and Cruz, M. (1993). Individual differences in the excitation transfer effect. *Personality and Individual Differences, 15,* 507–14.

Cacioppo, J. T., Klein, D. J., Berntson, G. G., and Hatfield, E. (1993). The psychophysiology of emotion. In M. Lewis and J. M. Haviland (Eds), *Handbook of emotions* (pp. 119–42). New York: Guilford.

Cacioppo, J. T., Petty, R. E., Losch, M. E., and Kim, H. S. (1986). Electromyographic activity over facial muscle regions can differentiate the valence and intensity of affective reactions. *Journal of Personality and Social Psychology, 50,* 260–68.

Cacioppo, J. T., and Tassinary, L. G. (1990). Inferring psychological significance from physiological signals. *American Psychologist, 45,* 16–28.

Cannon, W. B. (1927). The James-Lange theory of emotions: A critical examination. *Psychological Review, 38,* 281–95.

Challis, B. H., and Krane, R. V. (1988). Mood induction and the priming of semantic memory in a lexical decision task: Asymmetric effects of elation and depression. *Bulletin of the Psychonomic Society, 26,* 309–12.

Clark, L. A., and Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7* (special issue: Methodological issues in psychological assessment research), 309–19.

Cronbach, L. J., and Meehl, P. (1955). Construct validity in psychological tests. *Psychological Bulletin, 52,* 281–302.

Cunningham, M. R. (1988). What do you do when you're happy or blue? Mood, expectancies, and behavioral interest. *Motivation and Emotion, 12,* 309–31.

Cutler, S. E., Bunce, S. C., and Larsen, R. J. (1996). Repressive coping style and its relation to daily emotional experience and remembered emotional experience. *Journal of Personality, 64,* 379–405.

Darwin, C. (1872/1965). *The expression of the emotions in man and animals.* Chicago: University of Chicago Press.

Davidson, R. J. (1993). The neuropsychology of emotion and affective style. In M. Lewis and J. M. Haviland (Eds.), *Handbook of emotions* (pp. 143–54). New York: Guilford.

Diener, E., Smith, H., and Fujita, F. (1995). The personality structure of affect. *Journal of Personality and Social Psychology, 69,* 130–41.

Ekman, P. (1992). An argument for basic emotions. *Cognition and Emotion, 6,* 169–200.

Ekman, P., and Davidson, R. (1994). *The nature of emotion.* New York: Oxford University Press.

Ekman, P., and Friesen, W. (1975). *Unmasking the face.* Englewood Cliffs, N.J.: Prentice-Hall.

———. (1978). *Facial Action Coding System.* Palo Alto, Calif.: Consulting Psychologists Press.

Ekman, P., Freisen, W. V., and Ancoli, S. (1980). Facial signs of emotional experience. *Journal of Personality and Social Psychology, 39,* 1125–34.

Ekman, P., Friesen, W. V., and O'Sullivan, M. (1988). Smiles when lying. *Journal of Personality and Social Psychology, 54,* 414–20.

Ekman, P., and Rosenberg, E. L. (Eds.) (1997). *What the face reveals: Basic and applied studies of spontaneous facial expressions using the Facial Action Coding System (FACS).* New York: Oxford University Press.

Feldman, L. (1995). Variations in the circumplex structure of emotion. *Personality and Social Psychology Bulletin, 21,* 806–17.

Feldman-Barrett, L. (in press). Discrete emotions or dimensions: The role of valence focus and arousal focus. *Cognition and Emotion.*

Fredrickson, B. L., and Kahneman, D. (1993). Duration neglect in retrospective evaluations of affective episodes. *Journal of Personality and Social Psychology, 65,* 45–55.

Fredrickson, B. L., and Levenson, R. W. (in press). Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. *Cognition and Emotion*.

Fredrickson, B. L., Roberts, T., Noll, S. M., Quinn, D. M., and Twenge, J. M. (in press). That swimsuit becomes you: Sex differences in self-objectification, restrained eating, and math performance. *Journal of Personality and Social Psychology*.

Fridlund, A. J., Ekman, P., and Oster, H. (1987). Facial expressions of emotion. In A. W. Siegman and S. Feldstein (Eds.), *Nonverbal behavior and communication* (2nd ed., pp. 143–223). Hillsdale, NJ: Lawrence Erlbaum.

Frijda, N. H. (1988). The laws of emotion. *American Psychologist, 43,* 349–58.

Frijda, N. H., Kuipers, P., and ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology, 57,* 212–29.

Gotlib, I., and Meyer, J. (1986). Factor analysis of the Multiple Affect Adjective Check List: A separation of positive and negative affect. *Journal of Personality and Social Psychology, 50,* 1161–65.

Gottman, J. M. (1993). Studying emotion in social interaction. In M. Lewis and J. M. Haviland (Eds.), *Handbook of emotions* (pp. 475–87). New York: Guilford.

Gottman, J. M., and Levenson, R. W. (1985). A valid measure for obtaining self-report of affect. *Journal of Consulting and Clinical Psychology, 53,* 151–60.

Graham, F. K., and Clifton, R. K. (1966). Heart rate change as a component of orienting response. *Pschological Bulletin, 65,* 305–20.

Green, D. P., Goldman, S. L., and Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology, 64,* 1029–41.

Gross, J. J., and Levenson, R. W. (1993). Emotional suppression: Physiology, self-report, and expressive behavior. *Journal of Personality and Social Psychology, 64,* 970–86.

Grossman, P., Van Beek, J., and Wientjes, C. (1990). A comparison of three quantification methods for estimation of respiratory sinus arrhythmia. *Psychophysiology, 27,* 702–14.

Halberstadt, J. B., Niedenthal, P. M., and Kushner, J. (1995). Resolution of lexical ambiguity by emotional state. *Psychological Science, 6,* 278–82.

Haney, T., Maynard, K. E., Houseworth, S. J., Scherwitz, L. W., Williams, R. B., and Barefoot, J. C. (1996). The Interpersonal Hostility Assessment Technique: Description and validation against the criterion of coronary artery disease. *Journal of Personality Assessment, 66,* 386–401.

Izard, C. E. (1977). *Human emotions*. New York: Plenum Press.

Izard, C. E. (1979). *Emotions in personality and psychopathology*. New York: Plenum.

Izard, C. E. (1993). Four systems for emotion activation: Cognitive and noncognitive processes. *Psychological Review, 100,* 68–90.

Johnson, E. J., and Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of Personality and Social Psychology, 45,* 21–31.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., and Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4,* 401–5.

Ketelaar, T. (1989). Examining the circumplex model of affect in the domain of mood-sensitive tasks. Master's thesis, Purdue University.

Kleinginna, P. R., and Kleinginna, A. M. (1981). A categorized list of emotion definitions, with suggestions for a consensual definition. *Motivation and Emotion, 5,* 345–79.

Kring, A. M., and Neale, J. M. (1996). Do schizophrenic patients show a disjunctive relationship among expressive, experiential, and psychophysiological components of emotion? *Journal of Abnormal Psychology, 105,* 249–57.

Krokoff, L. J., Gottman, J. M., and Hass, S. D. (1989). Validation of a global rapid couples interaction scoring system. *Behavioral Assessment, 11,* 65–79.

Lacey, J. I., Kagan, J., Lacey, B. C., and Moss, H. A. (1963). The visceral level: Situational derterminants and behavioral correlates of automatic response patterns. In P. H. Kapp (Ed.), *Expression of the emotions in man* (pp. 161–96). New York International Universities Press.

Lang, P. J., Bradley, M. M., and Cuthbert, B. N. (1990). Emotion, attention, and the startle reflex. *Psychological Review, 97,* 377–95.

Lang, P. J., Bradley, M. M., Fitzsimmons, J. R., Cuthbert, B. N., Scott, J. D., Moulder, B., and Nangia, V. (1998). Emotional arousal and activation of the visual cortex: An fMRI analysis. *Psychophysiology, 35,* 199–210.

Larsen, R. J., and Diener, E. (1987). Affect intensity as an individual difference characteristic: A review. *Journal of Research in Personality, 21,* 1–39.

———. (1992). Problems and promises with the circumplex model of emotion. *Review of Personality and Social Psychology, 13,* 25–59.

Larsen, R. J., and Ketelaar, T. (1991). Personality and susceptibility to positive and negative emotional states. *Journal of Personality and Social Psychology, 61,* 132–40.

Larsen, R. J., and Sinnett, L. (1991). Meta-analysis of manipulation validity: Factors affecting the Velten mood induction procedure. *Personality and Social Psychology Bulletin, 17,* 323–34.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Levenson, R. W. (1983). Personality research and psychophysiology: General considerations. *Journal of Research in Personality, 17,* 1–21.

———. (1988). Emotion and the autonomic nervous system: A prospectus for research on autonomic specificity. In H. L. Wagner (Ed.), *Social psychophysiology* *and emotion: Theory and clinical applications* (pp. 17–42). Chichester, Eng.: Wiley.

———. (1992). Autonomic nervous system differences among emotions. *Psychological Science, 3,* 23–27.

Levenson, R. W., Ekman, P., and Friesen, W. V. (1990). Voluntary facial action generates emotion-specific autonomic nervous system activity. *Psychophysiology, 27,* 363–84.

Lucas, R. E., Diener, E., and Suh, E. (1996). Discriminant validity of well-being measures. *Journal of Personality and Social Psychology, 71,* 616–28.

MacKay, C. J. (1980). The measurement of mood and psychophysiological activity using self-report techniques. In I. Martin and P. Venables (Eds.), *Techniques in psychophysiology* (pp. 501–62). New York: Wiley.

Mandler, G. (1975). *Mind and emotion.* New York: Wiley.

Mayer, J. D. (1986). How mood influences cognition. In N. E. Sharkey (Ed.), *Advances in cognitive science* (pp. 290–314). Chichester, Eng.: Ellis Horwood.

Mayer, J. D., and Bremer, D. (1985). Assessing mood with affect-sensitive tasks. *Journal of Personality Assessment, 49,* 95–99.

Mayer, J. D., Mamberg, M. M., and Volanth, A. J. (1988). Cognitive domains of the mood system. *Journal of Personality, 56,* 453–86.

Meehl, P. E. (1978). Theoretical risks and tabular asterisks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology, 46,* 806–34.

Mesquita, B., and Frijda, N. H. (1992). Cultural variations in emotions: A review. *Psychological Bulletin, 112,* 179–204.

Mosher, D. L. and White, B. B. (1981). On differentiating shame and shyness. *Motivation and Emotion, 5,* 61–74.

Newton, T. L., and Contrada, R. J. (1992). Repressive coping and verbal-autonomic response dissociation: The influence of social context. *Journal of Personality and Social Psychology, 62,* 159–67.

Niedenthal, P. M., and Setterlund, M. B. (1994). Emotion congruence in perception. *Personality and Social Psychology Bulletin, 20,* 401–11.

Nowlis, V. (1965). Research with the Mood Adjective Check List. In S. S. Tomkins and C. E. Izard (Eds.), *Affect, cognition, and personality* (pp. 352–89). New York: Springer.

Nowlis, V., and Green, R. (1957). *The experimental analysis of mood.* Technical report, contract no. Nonr-668(12). Washington, D.C.: Office of Naval Research.

Obrist, P. A., Webb, R. A. Sutterer, J. R., and Howard, J. L. (1970). The cardiac-somatic relationship: Some reformulations. *Psychophysiology, 6,* 569–87.

Patrick, C. J. (1994). Emotion and psychopathy: Startling new insights. *Psychophysiology, 31,* 319–30.

Patrick, C. J., Cuthbert, B. N., and Lang, P. J. (1994). Emotion in the criminal psychopath: Fear image processing. *Journal of Abnormal Psychology, 103,* 523–34.

Pittam, J., and Scherer, K. R. (1993). Vocal expression and communication of emotion. In M. Lewis and J. M. Haviland (Eds.), *Handbook of emotions* (pp. 185–97). New York: Guilford.

Porges, S. W. (1995). Cardiac vagal tone: A physiological index of stress. *Neuroscience and Biobehavioral Reviews, 19,* 225–33.

Rosenberg, E. L., and Ekman, P. (1994). Coherence between expressive and experiential systems in emotion. *Cognition and Emotion, 8,* 201–29.

Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39,* 1161–78.

Russell, J. A., Weiss, A., and Mendelsohn, G. A. (1989). The Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology, 57,* 493–502.

Schacter, S., and Singer, J. E. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review, 69,* 379–99.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. *Psychological Bulletin, 99,* 143–65.

———. (1989). Vocal correlates of emotional arousal and affective disturbance. In H. L. Wagner and A. Manstead (Eds.), *Handbook of social psychophysiology* (pp. 165–97). Chichester, Eng.: Wiley.

Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. (1991). Vocal cues in emotion encoding and decoding. *Motivation and Emotion, 15,* 123–48.

Seidlitz, L., and Diener, E. (1993). Memory for positive versus negative life events: Theories for the difference between happy and unhappy persons. *Journal of Personality and Social Psychology, 64,* 654–63.

Sherwood, A. (1993). Use of impedance cardiography in cardiovascular reactivity research. In J. Blascovich and E. S. Katkin (Eds.), *Cardiovascular reactivity to psychological stress and disease* (pp. 157–99). Washington, D.C.: American Psychological Association.

Sherwood, A., Allen, M. T., Fahrenberg, J., Kelsey, R. M., Lovallo, W. R., and van Doornen, L. J. P. (1990). Committee Report: Methodological guidelines for impedance cardiography. *Psychophysiology, 27,* 1–23.

Smith, C. A., and Ellsworth, P. C. (1985). Patterns of cognitive appraisal in emotion. *Journal of Personality and Social Psychology, 48,* 813–38.

———. (1987). Patterns of appraisal and emotion related to taking an exam. *Journal of Personality and Social Psychology, 52,* 475–88.

Stone, A. A. (1981). The association between perceptions of daily experiences and self- and spouse-rated mood. *Journal of Research in Personality, 15,* 510–22.

———. (1995). Measures of affective response. In S. Cohen, R. Kessler, and L. Gordon (Eds.), *Measuring stress: A guide for health and social scientists* (pp. 148–71). New York: Cambridge University Press.

Teasdale, J. D., and Fogarty, S. J. (1979). Differential effects of induced mood on retrieval of pleasant and unpleasant events from episodic memory. *Journal of Abnormal Psychology, 88,* 248–57.

Teasdale, J. D., and Russell, M. L. (1983). Differential effects of induced mood on the recall of positive, negative, and neutral words. *British Journal of Clinical Psychology, 22*, 163–71.

Teasdale, J. D., Taylor, R., and Fogarty, S. J. (1980). Effects of induced elation-depression of the accessibility of memories of happy and unhappy experiences. *Behavior Research and Therapy, 18*, 339–46.

Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports, 20*, 663–78.

———. (1986). Activation-Deactivation Adjective Check List: Current overview and structural analysis. *Psychological Reports, 58*, 607–14.

Thomas, D. L. and Diener, E. (1990). Memory accuracy in the recall of emotions. *Journal of Personality and Social Psychology, 59*, 291–97.

Tomkins, S. S., and Izard, C. E. (1965). *Affect, cognition, and personality: Empirical studies*. New York: Springer.

Tranel, D., and Damasio, A. R. (1985). Knowledge without awareness: An autonomic index of facial recognition by prosopagnosics. *Science, 228*, 1453–54.

Van Bezooijen, R. (1984). *The characteristics and recognizability of vocal expression of emotions*. Dordrecht, the Netherlands: Foris.

Van Bezooijen, R., Otto, S. A., and Heenan, T. A. (1983). Recognition of vocal expressions of emotion: A three-nation study to identify universal characteristics. *Journal of Cross-cultural Studies, 14*, 387–406.

Velten, E. (1968). A laboratory task for the induction of mood states. *Behavior Research and Therapy, 6*, 473–82.

Venables, P. H. (1984). Arousal: An examination of its status as a concept. In M. G. H. Coles, J. R. Jennings, & J. A. Stern (Eds.), *Psychophysiological perspectives* (134–42). New York: Van Nostrand Reinhold.

Vrana, S. R., Spence, E. L., and Lang, P. J. (1988). The startle probe response: A new measure of emotion? *Journal of Abnormal Psychology, 97*, 487–91.

Watson, D., and Clark, L. A. (1991). Self- versus peer ratings of specific emotional traits: Evidence of convergent and discriminant validity. *Journal of Personality and Social Psychology, 60*, 927–40.

Watson, D., Clark, L. A., and Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*, 1063–70.

Watson, D., and Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin, 98*, 219–35.

Zajonc, R. B. and McIntosh, D. N. (1992). Emotions research: Some promising questions and some questionable promises. *Psychological Science, 3*, 70–74.

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

———. (1985). *Manual for the Multiple Affect Adjective Check List—Revised*. San Diego: Educational and Industrial Testing Service.