

6 The old classics: measures of distinct states

The previous chapters described the nature and the severity of some of the problems associated with the measurement of affect, mood, and emotion in health-behavioral research (Chapter 1), outlined a framework for defining and distinguishing between these three constructs (Chapter 2), and summarized certain issues of fundamental significance for the conceptualization and assessment of affective constructs, namely whether these states should be considered as distinct entities or as positioned along dimensions (Chapter 3) and whether pleasure and displeasure are independent or polar opposites (Chapter 4). This review culminated in a proposal for a three-tiered system for justifying the selection of a measure (Chapter 5).

In this and the remaining chapters, the background provided so far will be used in critiquing some of the most frequently used measures in health-behavioral research. Emphasis is placed on the conceptual bases and the developmental histories of the measures, the significance and implications of which are all too often overlooked. The aim of this presentation is to sensitize researchers to approach each measure critically, raise awareness of the relative strengths and limitations of each measure, and offer key references for further study.

The Multiple Affect Adjective Check List

The Multiple Affect Adjective Check List (Zuckerman & Lubin, 1965a) was one of the first self-report measures designed to assess transient affective states as opposed to stable traits. Furthermore, it was primarily aimed at the general population rather than clinical groups. These characteristics made the Multiple Affect Adjective Check List a pioneering development in the study of affective phenomena. Its development history is an interesting example of the distinct-states approach described in Chapter 3.

The predecessor of the Multiple Affect Adjective Check List was the Affect Adjective Check List (Zuckerman, 1960), a 21-item check

list measure of anxiety consisting of 11 anxiety-present (e.g., *afraid, fearful*) and 10 anxiety-absent items (e.g., *calm, cheerful*). The anxiety score increased for every anxiety-present item checked and for every anxiety-absent item left unchecked. In the last sentence of the 1960 article, Zuckerman noted that the same general format could also be used to develop other tests, with depression and hostility being “two obvious possibilities” (p. 462). So, a few years later, the first version of the Multiple Affect Adjective Check List appeared, combining scales for anxiety, depression, and hostility (Zuckerman, Lubin, Vogel, & Valerius, 1964). Readers should note how fundamentally different this distinct-states approach is from approaches based on dimensional views of the global affective space. Essentially, the Multiple Affect Adjective Check List was developed as a measure of three distinct states that its developers considered important, without making any claims whatsoever about covering a global domain of content (such as affect, mood, or emotion).

The added Depression scale consisted of 20 depression-present (e.g., *blue, discouraged*) and 20 depression-absent items (e.g., *alive, enthusiastic*) that were all different from those used in the Anxiety scale. Similarly, the Hostility scale consisted of 16 hostility-present (e.g., *angry, bitter*) and 12 hostility-absent items (e.g., *agreeable, amiable*). Several additional items were retained for further testing, for a total of 132 items. Zuckerman and Lubin (1965b) provided normative data.

A problem that quickly became apparent was that the scales were highly intercorrelated and lacked discriminant validity, possibly because the check list format increased the influence of acquiescent response sets (i.e., a person with a tendency to check many items would receive lower negative affect scores). When factor analysis became more readily available, the 132-item pool was factor analyzed (Zuckerman, Lubin, & Rinck, 1983) and a new structure emerged. Specifically, anxiety-present, depression-present, and hostility-present items formed three separate factors, while the positively worded items formed two factors (one named Positive Affect and the other Sensation Seeking). However, given the strong intercorrelations among factors within each category, Zuckerman et al. (1983) proceeded to also merge Anxiety, Depression, and Hostility into a Dysphoria (DYS) factor and Positive Affect (PS) and Sensation Seeking (SS) into a combined PASS factor.

A subsequent independent factor analysis confirmed that the items formed separate factors consisting of the positively and the negatively worded items, respectively (Gotlib & Meyer, 1986). The influence of response sets was again presented as the most likely explanation, leading to the following warning: “investigators who use the Multiple Affect

Adjective Check List scales as measures of affect should be aware of the potential problems with this questionnaire” (p. 1165). The hierarchical structure (five first-order factors and two second-order factors) formed the basis for the revised edition of the Multiple Affect Adjective Check List, the Multiple Affect Adjective Check List – Revised (Lubin et al., 1986; Zuckerman & Lubin, 1965a), which comprised 66 scored and 66 filler items. However, an independent factor analysis again yielded just two factors, separating the positively and the negatively worded items (Hunsley, 1990).

The Multiple Affect Adjective Check List has been used in thousands of studies in many areas of psychology. However, its popularity has declined in recent years. Nevertheless, it was important to review it here to illustrate two crucial points. First, the history of the Multiple Affect Adjective Check List essentially parallels the conceptual developments in the field of affective psychology, highlighting the progression from distinct-states to dimensional and, eventually, to hierarchical models. Second, to set the background for the review of newer measures, it was important to remind readers of the history of one of the most popular measures of affect. In particular, it is worth pondering whether a measure initially intended to assess three distinct constructs (anxiety, depression, hostility) can be considered an encompassing measure of the global domain of “affect.” Even when it is scored in terms of the higher-order dimensions of DYS and PASS, a measure initially developed to target three distinct states will always be limited by the scope and representativeness of its item pool. Factor analysis cannot uncover the structure of a content domain; it can only help uncover the structure of a given item pool.

The Profile of Mood States

The Profile of Mood States is one of the most popular self-report measures, not only in health-behavioral research but in psychology in general. However, few of the researchers using the Profile of Mood States offer specific reasons for their selection (other than its enormous popularity) and few seem fully aware of the history of this measure.

What seems to have been the earliest draft of what later became known as the Profile of Mood States appeared in a study designed to assess the effects of tranquilizers (meprobamate and chlorpromazine) in a sample of male veterans receiving psychotherapy on an outpatient basis (Lorr, McNair, Weinstein, Michaux, & Raskin, 1961). The measure, which did not yet have a name, was described as a collection of “55 common adjectives describing feeling states” (p. 383) grouped

in five categories (anxiety, hostility, depression, inertia, and activity). In a subsequent publication with a similar aim (i.e., investigating the effects of chlordiazepoxide), the measure (which still did not have a name) was described as a “feeling and attitude scale, which has been shown to be sensitive to change in several studies” (Lorr, McNair, & Weinstein, 1963, p. 261). The number of items had been increased to 60, each accompanied by a 4-point response scale (“not at all,” “a little,” “quite a bit,” “extremely”). The measure purportedly assessed “six moods,” namely Tension-Anxiety, Anger-Hostility, Depression, Vigor, Fatigue-Inertia, and Thinking-Confusion.

The now famous initials POMS appeared for the first time in 1964, although they did not originally stand for Profile of Mood States but rather for Psychiatric Outpatient Mood Scale (McNair & Lorr, 1964). The stated goal of the developers of the POMS was to “construct and develop a useful method for identifying and assessing mood states in psychiatric outpatient populations” (p. 620). They now hypothesized six mood states “on the basis of clinical observation and a review of previous studies” (p. 620). Those six were Tension, Anxiety, Anger, Depression, Vigor, and Fatigue.

It is very important for current and prospective users of the Profile of Mood States to understand that the item pool was composed of items drawn from various adjective lists, a dictionary, and a thesaurus with the purpose of matching them to the six targeted mood states (which were slightly different from those in the present-day six-factor structure). Furthermore, the items were retained only if “four psychologists concurred in judging [them] to be descriptive of the six hypothesized mood states” (McNair & Lorr, 1964, p. 621). In other words, as was the case with the Multiple Affect Adjective Check List, there was again no intention to develop an item pool that would reflect the global content domain of mood. The goal of McNair and Lorr was to assess certain specific states they deemed of interest for the study of psychiatric outpatients.

It is also important to note that the reference to “previous studies” as the basis for hypothesizing six factors seems to point to the highly influential analyses by Nowlis and Nowlis (1956). These investigators were studying the effects of drugs but also hoped to develop a measure “that is applicable to a wide variety of situations not involving drugs” (p. 352) and to discover “generally applicable categories” of mood states (p. 354). They had originally hypothesized that they would uncover four bipolar dimensions (level of activation, level of control, social orientation, and hedonic tone), but, when their empirical analysis was completed (Green & Nowlis, 1957), they found eight unipolar and correlated factors

(concentration, aggression, pleasantness, activation-deactivation, egotism, social affection, depression, and anxiety). When Nowlis (1965, 1970) later developed the Mood Adjective Check List, the number was raised to 12 (aggression, anxiety, surgency, elation, concentration, fatigue, social affection, sadness, skepticism, egotism, vigor, and non-chalance). Clearly, neither the eight nor the twelve factors agreed fully with the six factors proposed by McNair and Lorr (1964). Thus, their insistence on these six factors was driven more by their clinical judgment and experience than by previous empirical findings.

McNair and Lorr (1964) conducted a series of factor analyses with data gathered from neurotic psychiatric outpatients, during which some items were deleted and new ones added. In the process, the Tension and Anxiety items were found to form a single factor. A Confusion factor did not emerge initially but did emerge after new items were added (presumably, the interest in this factor was due to confusion being an important variable when investigating the effects of psychotropic drugs). For researchers who have questioned whether "confusion" is a mood state, it is important to note that McNair and Lorr (1964) were also unsure whether this factor represented "cognitive inefficiency, a mood state, or both" (p. 624). Furthermore, a weak Friendliness factor appeared unexpectedly out of items added to strengthen the Anger factor. Confidence in a five-factor grouping was later increased based on findings from factor analyses by other investigators but on similar item pools (Lorr et al., 1967). The five factors were labeled Depression, Vigor-Activity, Fatigue-Inertia, Tension-Anxiety, and Anger-Hostility. A sixth factor "variously called Concentration, Thoughtful, Confusion, and Clear Thinking" (p. 89) was also considered possible. However, when 62 items given to samples of undergraduate students (for the first time to a nonclinical sample) were factor analyzed, eight rather than five or six factors emerged. The additional factors were Cheerful, Thoughtful (similar to Confusion), and Relaxed-Composed.

Interestingly, Lorr et al. (1967) examined whether these eight factors could be ordered in a circular fashion, as would be predicted by the circumplex model. Indeed, they found some support but also some discrepancies (e.g., the correlation between Inert-Fatigued and Composed-Relaxed was negative). They concluded that "while the present data offer some support for the hypothesis, further trial will be necessary before a circular order can be established" (p. 94). Furthermore, based on a visual inspection of the scale intercorrelations, they noted that even the eight scales did not suffice to cover the entire circle without leaving large gaps: "it is evident that at least one mood variable is

missing between Composed-Relaxed and Inert-Fatigued” and possibly another one “between Energetic and Angry-Irritable” (p. 94).

Despite these highly interesting observations, the version of the POMS that became commercially available in 1971 under the new name “Profile of Mood States” was offered with a six-factor structure comprising Tension-Anxiety, Depression-Dejection, Anger-Hostility, Vigor-Activity, Fatigue-Inertia, and Confusion-Bewilderment (McNair et al. 1971). Soon thereafter, Spielberger (1972b) began his critical review of the Profile of Mood States with what is clearly its most serious limitation (and yet a point largely absent from discussions of the Profile of Mood States in the applied psychology literature to date):

The POMS test manual suggests that the POMS factors emerged from the test construction procedures. This is somewhat misleading in that, for the initial item pool and subsequent additions to the pool, adjectives with semantic connotations that reflected the factors the authors wished to assess were selected. Put differently, factor analysis was employed to refine the mood dimensions the authors wished to assess rather than to sort out or map the mood domain. While the authors have been successful in defining six mood factors, a different set of mood dimensions might have resulted if other adjectives had been included in the item pool in addition to those selected by the authors on a priori grounds as fitting their conceptual definition of the mood dimensions they wished to measure. (p. 387)

Years later, Thayer (1989) made a similar point, criticizing the Profile of Mood States for grouping items without “sufficient theoretical guidance” and for proposing “distinctions among mood dimensions” without offering a theoretical model for “the ways in which these factors interact” (p. 16). Current and prospective users of the Profile of Mood States are urged to consider these warnings. From a practical perspective, these experts are suggesting that one cannot generalize from data collected with the Profile of Mood States to the global domain of “mood,” only to the six distinct states this measure includes (see Figure 6.1). The reason is that no theoretical or empirical basis supports the idea that the particular assortment of factors that make up the Profile of Mood States provides an encompassing representation of this global domain of content. One can easily think of mood states not included in the Profile of Mood States that might be of interest to investigators in the domain of health-behavioral research (e.g., cheerfulness or joviality, peacefulness or serenity). Recall, for example, that Relaxed-Composed was identified in earlier analyses by Lorr et al. (1967) but was not included in the commercial version of the Profile of Mood States.

Independent investigations of the factor structure of the Profile of Mood States have yielded inconsistent results. Norcross, Guadagnoli,

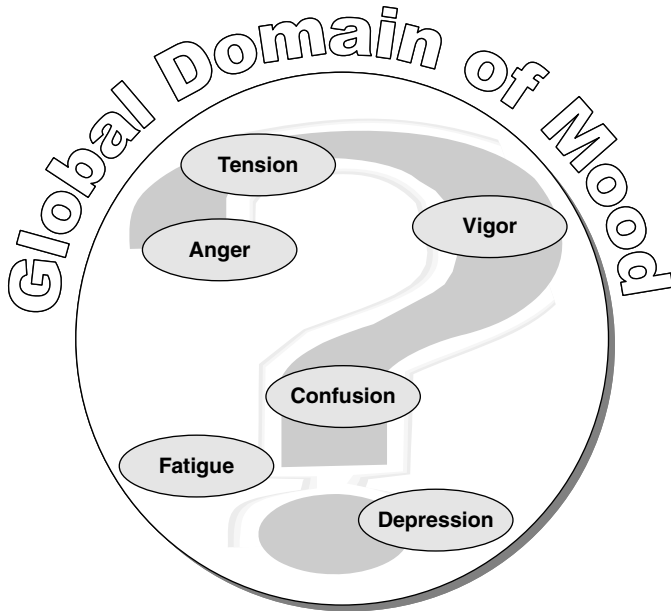


Figure 6.1. There is no conceptual argument or empirical evidence that the six distinct states assessed by the Profile of Mood States can provide comprehensive coverage of the global domain of mood. Therefore, extrapolations from assessments based on the Profile of Mood States to the global domain of mood are unwarranted.

and Prochaska (1984), examining responses from a sample of psychiatric outpatients and a sample of adult smokers, found that the Anger-Hostility, Vigor-Activity, and Fatigue-Inertia factors were reproduced in both samples. However, the Tension-Anxiety and Confusion-Bewilderment factors were not reproduced in either sample and the Depression-Dejection factor was reproduced in the outpatient sample but not in the smoker sample. Because of some very high inter-factor correlations (up to 0.81), the authors warned that “caution is recommended in the separate scoring and interpretation of these scales” (p. 1277). Boyle (1987) found evidence of nine factors, including the standard six plus a second Depression factor, a Friendliness factor, and an Arousal-Alertness factor (each accounting for approximately 2 percent of the variance). Focusing on the loadings of individual items, Reddon, Marceau, and Holden (1985) found that, while the factors had satisfactory homogeneity and internal consistency, less than two-thirds of the items were discriminantly valid in any given sample (i.e., had

higher loadings on their own scales than on irrelevant scales) and less than half of the items were consistently (i.e., across different samples) discriminantly valid. Thus, these authors noted that “possibly the test constructor, clinician, or researcher in the mood domain may or may not wish to consider a more articulated assessment of moods than that available through the standard form of the POMS” (p. 257).

It is also important to consider the evolution of the Profile of Mood States after its initial commercial version and the developments that led to the publication of the newer, bipolar version. Prompted by Osgood’s (1962) famous analyses of the structure underlying affective meaning, which revealed bipolar dimensions, Bentler’s (1969) demonstration that measurement error tends to conceal bipolarity, and Meddis’ (1972) influential work on the role played by unipolar response formats in the same regard, Lorr and Shea (1979) started exploring the question “are mood states bipolar?” For example, they noted that because the Profile of Mood States uses a unipolar response format, its scales “are more subject to extreme response bias” (p. 469). They postulated that nine states (cheerful, energetic, grouchy, anxious, dejected, tired, agreeable, composed, and confidence/potency) would exhibit circular ordering, similar to a circumplex. An advantage they saw in this type of modeling was that it would provide “a definition of the universe of content” (p. 469), something clearly missing from the Profile of Mood States.

In a sample using a unipolar response format, after partialling out an extreme response style score, four of the five factors that emerged were bipolar (Composed-Anxious, Energetic-Tired, Agreeable-Angry, Optimistic-Pessimistic), whereas only Cheerful remained unipolar. In a second sample using a bipolar response format, three of the six factors that emerged were bipolar (Composed-Anxious, Energetic-Tired, Agreeable-Angry) and three were unipolar (Confident, Dejected, Cheerful). Lorr and Shea (1979) also found some support for circular ordering and wondered whether the dimensions of Pleasantness-Unpleasantness and Arousal would be found to explain this phenomenon. Lorr et al. (1982) followed up on these findings, this time hypothesizing five bipolar factors (Composed-Anxious, Agreeable-Hostile, Energetic-Tired, Elated-Depressed, Clear-thinking-Confused). Using a slightly modified version of the Profile of Mood States (with the addition of items measuring friendliness and elation), they performed a factor analysis on data from psychiatric patients. After partialling out the variance due to extreme response bias from the interitem correlations, they found evidence for four of the five hypothesized bipolar factors (Elated-Depressed, Agreeable-Hostile, Energetic-Tired, Clear-thinking-Confused). The fifth factor (Composed-Anxious) could

not emerge as bipolar for the simple reason that there was a very unbalanced representation of the two poles in the item pool, with nine items representing the Anxious pole but only one representing the Composed pole. The authors also mentioned that, in a separate study, they derived a bipolar Confident-Unsure factor. Thus, since the 1980s, the bipolar version of the Profile of Mood States includes the following factors: Composed-Anxious, Agreeable-Hostile, Elated-Depressed, Confident-Unsure, Energetic-Tired, and Clearheaded-Confused.

Continuing to work on this topic and convinced of the bipolar nature of mood states, Lorr developed a measure that used the semantic differential format (Lorr & Wunderlich, 1988). The new measure was named the Feeling and Mood Scales and later the Semantic Differential Mood Scale to avoid any interference with the commercial success of the Profile of Mood States. Lorr and Wunderlich (1988) hypothesized six bipolar factors (Cheerful-Dejected, Energetic-Tired, Relaxed-Anxious, Good-natured-Grouchy, Confident-Unsure, and Excited-Bored) and found factor-analytic evidence for five (Cheerful-Dejected, Relaxed-Anxious, Confident-Unsure, Energetic-Tired, and Good-natured-Grouchy), whereas items hypothesized to form the Excited-Bored factor were grouped with Cheerful-Dejected.

Closing this long cycle, Lorr et al. (1989) made one more attempt to identify the “five or six mood states [that] can account for most of the inter-individual variation in affect” (p. 156). The attempt, however, was unsuccessful. Lorr et al. hypothesized six bipolar factors: Relaxed-Anxious, Agreeable-Grouchy, Joyful-Depressed, Confident-Unsure, Energetic-Tired, and Excited-Apathetic. With what is now known about the content and structure of the affective domain (see Chapters 1–3), it is clear that some of the hypothesized factors cover the same space (e.g., from high-activation pleasant to low-activation unpleasant affect: Joyful-Depressed and Energetic-Tired), whereas others refer to cognition rather than affect (i.e., Confident-Unsure). So the resultant factor structure did not agree with the hypothesis (bipolar factors for Hostile Depression versus Satisfied Pleasure, Energetic versus Fatigued/Drowsy, Composure versus Anxiety, Confidence versus Self-doubt, and a unipolar Active Arousal factor). Important, however, even without the benefit of a guiding theoretical framework, most of the factors were found to be bipolar after the removal of variance due to extreme response bias.

Furthermore, Lorr et al. (1989) examined the first two principal components after a Varimax rotation and found that one represented Pleasant versus Unpleasant Affect and the second represented Energetic Arousal versus Drowsy Fatigue. Although this solution did

not align with either Russell's (1980) or Watson and Tellegen's (1985) model due to the unbalanced composition of the item pool, the important conclusion is that, after almost 30 years of exploring the nature of mood, Lorr et al. (1989) produced data that converged with the universal two-dimensional model of affect presented here (for more on the endorsement of the circumplex model, see Lorr, 1989, 1997).

Given this history, it might seem surprising that even the original unipolar version of the Profile of Mood States, long rendered outdated by the theoretical developments in affective psychology, continues to be used with great regularity in health-behavioral research. The main (if not the only) argument put forth by researchers for their choice of measure is that the Profile of Mood States has been very popular. For example, as noted in Chapter 1, in exercise studies, researchers have claimed that they chose this measure because it "has been used extensively to assess the acute effects of exercise on mood" (Hoffman & Hoffman, 2008, p. 359), or "has been widely utilized in exercise research" (Bryan et al., 2007), or is "one of the most frequently used mood measures in sport and exercise psychology" (Johansson et al. 2008, p. 201), or is "historically the most frequently used measure in exercise and mood state studies" (Hansen et al. 2001, p. 269).

Critiques of the Profile of Mood States in health-behavioral research are rare and have concentrated on issues of secondary importance, such as its length and the inclusion of only one positive scale, namely Vigor (Berger & Motl, 2000; Gauvin & Rejeski, 1993; Guadagnoli & Mor, 1989; McAuley & Courneya, 1994). At this stage of knowledge development, the argument that a measure is used because it has been used before cannot be deemed compelling. Researchers who select the Profile of Mood States for a particular study should explain this decision in light of the issues raised in the literature. Furthermore, it is important that researchers avoid generalizations from the six distinct states to the global domain of mood, as such extrapolations are clearly unjustified by the history, the scope, and the structure of this measure.

The State-Trait Anxiety Inventory

The State-Trait Anxiety Inventory (Spielberger et al., 1970; Spielberger, 1983) was based on Spielberger's highly influential theory of state and trait anxiety developed in the 1960s and 1970s (Spielberger, 1966, 1972a). According to this theory, state anxiety is defined as a "transitory psychobiological emotional state or condition that is characterized by subjective, consciously experienced thoughts and feelings relating to tension, apprehension, nervousness, and worry that vary in intensity and

fluctuate over time” (Spielberger & Reheiser, 2004, p. 70). In contradistinction, trait anxiety is defined as “relatively stable individual differences in anxiety proneness as a personality trait” or “differences in the strength of the disposition to respond to situations perceived as threatening with elevations in state anxiety” (pp. 70–71). Spielberger’s theory was influenced by the advent of “trait by situation” interactionism in personality psychology, as well as the rise of appraisal-based approaches in the study of emotion. The theory thus predicts that high-trait-anxious individuals will tend to interpret a larger proportion of situations as entailing some type of threat than low-trait-anxious individuals. The perception of threat will depend on appraising a higher level of threat in each situation, a lower level of coping capabilities, or both.

The State-Trait Anxiety Inventory includes two scales, each consisting of 20 items: one assesses state anxiety (with items such as “I am worried” and “I feel frightened”) and the other assesses trait anxiety (with items such as “I worry too much over something that really doesn’t matter” and “I lack self-confidence”). The state-anxiety items are accompanied by a four-point scale of intensity, ranging from “not at all” to “very much so.” The trait anxiety items are accompanied by a four-point scale of frequency, ranging from “almost never” to “almost always.” For example, the state-anxiety scale has been used in exercise psychology to investigate the anxiolytic effects of single bouts of activity (e.g., Bodin & Martinsen, 2004), whereas the trait anxiety scale has been used as a measure of the effects of exercise training studies lasting for weeks or months (e.g., DiLorenzo et al., 1999).

The original version of the State-Trait Anxiety Inventory was published in 1970 and designated as Form X-1 for the state version and Form X-2 for the trait version (Spielberger et al., 1970). A revision, designated as Form Y-1 (for state anxiety) and Form Y-2 (for trait anxiety), was published in 1983 (Spielberger, 1983). The purpose of the revision was to replace “several items with poor psychometric properties” (Vagg, Spielberger, & O’Hearn, 1980, p. 212). For example, the item *anxious* was replaced by *frightened* because some respondents interpreted it to mean “eager” (Spielberger, 1985, p. 12). Furthermore, Spielberger (1983) noted that the development of Form Y was driven by the desire to make the questionnaire more “consistent with theoretical refinements in our concept of anxiety” by giving “greater emphasis to the cognitive or ‘worry’ aspects of anxiety than the original items” (p. 12).

Primarily because it incorporated the important conceptual distinction between states and traits and its items were appropriate for use with respondents without clinical levels of anxiety, the State-Trait

Anxiety Inventory became (and continues to be) exceptionally popular. However, it has also received a fair amount of criticism. Current and prospective users of this questionnaire should be aware of the limitations discussed in the literature and these should inform their measurement decisions. Critics have primarily focused on two important issues, namely the factor structure and the content overlap with depression.

Factor-analytic results of both Form X and Form Y have been inconsistent in terms of the nature and composition of the resultant factors but have been consistent in showing that the state and trait anxiety scales contain more than one factor each (Bernstein & Eveland, 1982; Caci, Baylé, Dossios, Robert, & Boyer, 2003; Endler, Magnusson, Ekehammar, & Okada, 1976; Mook, Kleijn, & van der Ploeg, 1991; Mook, van der Ploeg, & Kleijn, 1992; Sherwood & Westerback, 1983; Vagg et al., 1980).

The most commonly reported finding has been the separation of the positively worded (or “anxiety-absent”) and negatively worded (or “anxiety-present”) items into different factors. Initially, Spielberger et al. (1970) selected items intended to differ in terms of their “item intensity specificity.” This psychometric concept refers to a property of different items to discriminate better at one level of anxiety than at another. According to Spielberger (1985), in constructing the state-anxiety scale, “the main goal was to measure a continuum of increasing intensity on which low scores indicated feeling calm and serene, intermediate scores were associated with moderate levels of tension and worry, and high scores reflected intense fear, approaching terror and panic” (p. 11). For example, the item “I feel rested,” which is positively worded (i.e., denotes the absence of anxiety), was selected because it was thought more sensitive to fluctuations in state anxiety at the low end of the state-anxiety continuum than at the high end. In actuality, an empirical investigation of the location of the items along the state and trait anxiety continua using Rasch modeling revealed several weaknesses, including parts of the continua not represented by any items, parts covered by multiple items, and several badly fitting items that misrepresented the respondents’ level of anxiety (Tenenbaum, Furst, & Weingarten, 1985). Nevertheless, in general, the anxiety-absent items were located near the bottom and the anxiety-present items were located near the top of the continua.

Spielberger’s position regarding these findings was that they do not necessarily indicate multidimensionality:

While the consistent finding of anxiety-present and anxiety-absent factors could be interpreted as evidence of the multidimensionality of the [state anxiety]

and [trait anxiety] scales, it seems more plausible to interpret these factors as resulting from either "item-method variance" (i.e. the STAI consists of two distinctive types of items, anxiety-present and anxiety-absent), or to "item-intensity specificity" (i.e. the anxiety-present items appear to be more effective in measuring higher levels of state and trait anxiety, whereas the anxiety-absent items are more sensitive at lower levels. (Vagg et al., 1980, p. 213)

However, other researchers have been more critical, since it is indeed uncommon for a supposed unidimensional construct to be represented by multiple factors, regardless of any effects of item-method variance or item-intensity specificity. For example, Bernstein and Eveland (1982), on the basis of their confirmatory factor analysis, concluded that the "worst feature" of the State-Trait Anxiety Inventory "is that differences between positive and negative items are approximately as great as the differences between State and Trait items" (p. 372). They aptly noted that, although the separation of anxiety-present and anxiety-absent items is "interesting in its own rights, ... it should not obscure the appropriateness of the state-trait distinction" (p. 372). In other words, method-related factors should not preclude the straightforward interpretation of the factors meant to be substantive.

Adding further evidence that this is problematic, Mook et al. (1991) showed that, when the items of the State-Trait Anxiety Inventory and those from a depression questionnaire were factor analyzed together, the distinction that emerged was between symptom-present and symptom-absent items rather than between anxiety and depression items. Especially if the formation of separate factors can be attributed to the influence of such factors as denial and social desirability (e.g., Bonke, Smorenburg, van der Ent, & Spielberger, 1987; Mook et al., 1991), the presence of multiple factors has clear implications for validity and, therefore, cannot be seen as a benign phenomenon.

Of particular relevance to many of the studies in health-behavioral research, in which the respondents are young, healthy, and non-anxious college students, is the observation that, while respondents score very low in anxiety-present items (indicating a low degree of anxiety), scores on anxiety-absent items tend to cover a wide range (Bonke et al., 1987). It has often been noted that healthy college students give total scores only a few points above the bottom of the range (Dishman, 1995) and rate anxiety-present items (e.g., *frightened*), in particular, so close to the bottom of the range that their variance approaches zero (Ekkekakis et al., 1999). This phenomenon raises the possibility that anxiety-present and anxiety-absent items indeed reflect different sources of variance.

The second major criticism directed toward the State-Trait Anxiety Inventory relates to its conspicuously high content overlap

with depression. Although anxiety and depression frequently occur as comorbid conditions (Mineka, Watson, & Clark, 1998; Watson & Kendall, 1989), a measure of anxiety should remain as “uncontaminated” by depression as possible. Replacing items that seemed more relevant to depression than anxiety (e.g., “I feel blue,” “I feel like crying”) was one of the main reasons cited for revising Form X of the State-Trait Anxiety Inventory and developing Form Y (Spielberger, 1985; Vagg et al., 1980). However, the revision did not seem to solve the problem, since Form Y (both the state and trait scales) continues to show very strong overlap with measures of depression (Bieling, Antony, & Swinson, 1998; Endler, Cox, Parker, & Bagby, 1992; Gotlib, 1984). It is clear that some of the remaining items also reflect depression more than anxiety (e.g., “I feel like a failure”).

In the late 1980s, Watson and Clark (Clark & Watson, 1991; Watson, Clark, & Carey, 1988) began developing a so-called tripartite model to distinguish between symptoms of anxiety and depression. According to this model, negative affect is common to both anxiety and depression, low positive affect and anhedonia are specific to depression, and physiological tension and hyperarousal are specific to anxiety. Empirical tests of the tripartite model have been supportive (Mineka et al., 1998) and the model has stimulated renewed efforts to improve the discriminant validity of measures of anxiety and depression. However, having been developed and revised prior to the emergence of the tripartite model, the State-Trait Anxiety Inventory has not incorporated this important conceptual advance. According to Bieling et al. (1998):

The construction of the STAI, in both its original and revised forms, predated recent advances in the understanding of the relationship between anxiety and depression. Indeed, the items of the trait scale that represent the “anxiety-absent” dimension may possibly reflect depression (i.e. low positive affect) rather than anxiety. Based on content, these items appear to assess high levels of dysphoric mood and harsh self-judgements, both of which are more typically associated with depression than anxiety (e.g. *I wish I could be as happy as others seem to be; I feel like a failure*). (pp. 778–779)

The weaknesses of the State-Trait Anxiety Inventory brought to the forefront in other areas of psychology have not attracted much attention within the domain of health-behavioral research. The popularity of the questionnaire remains very high. As is the case with the Profile of Mood States, the selection of the State-Trait Anxiety Inventory is typically justified precisely on the basis of its prior popularity by suggesting that extensive prior use is tantamount to validation or de facto evidence of psychometric merit. For example, as noted in Chapter 1,

in exercise psychology, authors have chosen to use the State-Trait Anxiety Inventory because it is the “most widely used measure of anxiety in exercise research” (Knapen et al., 2009) or the “most widely used anxiety measure in exercise studies” (Youngstedt, 2010), it has been “employed in nearly 50% of studies on exercise and anxiety” (Motl et al., 2004, p. 98), or is “the most often employed measure of state anxiety in the exercise literature” (Bartholomew & Linder, 1998, p. 208).

Research in the context of exercise (Rejeski et al., 1991) has also uncovered an important limitation of the state-anxiety portion of the State-Trait Anxiety Inventory that had not been noticed in other contexts. The State-Trait Anxiety Inventory was developed based on the assumption that, when a state-anxiety response occurs, the various characteristic symptoms of anxiety (i.e., affective, cognitive, behavioral, and physiological) manifest themselves more or less in unison. In fact, this coordinated symptomatology was thought to define the occurrence of a state anxiety “syndrome” (Spielberger, 1972a). As a result, the state-anxiety portion of the State-Trait Anxiety Inventory is scored as a unidimensional instrument despite the fact that it includes items referring to affective (e.g., *I feel pleasant*), cognitive (e.g., *I feel self-confident*, *I am worried*), behavioral (e.g., *I am jittery*), and perceived physiological symptoms (e.g., *I feel calm*, *I am tense*). In most cases (e.g., when facing an academic exam or a job interview), these diverse symptoms do exhibit a coordinated response and do form a relatively tight cluster.

Exercise and other arousal-inducing stimuli (e.g., caffeine, chronotropic drugs), however, are different in a very important respect; they can induce noticeable or even substantial changes in perceived physiological activation independently of an appraisal of threat (which is the defining element of a state-anxiety response in Spielberger’s and other cognitive theories). Rejeski et al. (1991) showed that, near the end of a 20-minute bout of treadmill exercise at 75 percent of heart rate reserve, scores on items indicative of perceived physiological activation (e.g., *calm*, *relaxed*) increased (i.e., participants felt less calm and less relaxed, which are scored as indicating increased state anxiety), whereas scores on items indicative of cognitive components of anxiety (e.g., *worried*) decreased. This divergent pattern of responses suggested that, during exercise, the different items of the State-Trait Anxiety Inventory became indices of different constructs rather than a unitary construct of *state anxiety*. Weakened item intercorrelations led to a collapse of the internal consistency of the scale (alpha coefficient of 0.33). Multiple laboratories have since replicated the same pattern of item responses using different

exercise modalities and intensities (Bartholomew & Linder, 1998; Ekkekakis et al., 1999; Focht & Hausenblas, 2003; Katula, Blissmer, & McAuley, 1999; McAuley, Mihalko, & Bane, 1996).

These data have important implications for the validity of research investigating changes in state anxiety in response to stimuli that cause changes in perceived physiological activation based on the State-Trait Anxiety Inventory. Treatments that *increase* state-anxiety scores may not necessarily be anxiety inducing and, conversely, those that *decrease* state-anxiety scores may not necessarily be anxiolytic. Such changes could in fact reflect, at least in part, increases and decreases in items indicative of perceived physiological activation. Thus, the most likely reason the participants in the study by Rejeski et al. (1991) reported feeling less *calm* and *relaxed* during vigorous exercise was not that they became more anxious (after all, they reported feeling less *worried*) but rather that they had been exercising vigorously. Likewise, when the participants reported feeling more *calm* and *relaxed* 10 minutes after exercise, this probably did not reflect a decrease in state anxiety per se but rather, at least to some extent, the typical postexercise decrease in perceived physiological activation as their bodies returned to homeostasis. The fact that state-anxiety scores derived from the State-Trait Anxiety Inventory increase as exercise intensity increases (Katula et al., 1999) supports this conjecture. Likewise, the fact that the postexercise decrease in state-anxiety scores is delayed as exercise intensity increases (e.g., Raglin & Wilson, 1996) may reflect a delay in physiological recovery rather than a true “delayed anxiolysis,” as the phenomenon has been labeled in the exercise literature (Cox, Thomas, Hinton, & Donahue, 2004; Raglin, 1997). Nevertheless, the widely touted phenomenon of exercise-induced anxiolysis remains a reasonable possibility since items indicative of what Spielberger (1972a) considered the core of the state-anxiety response, namely the cognitive appraisal of threat (e.g., *I am worried*), have been found to decrease with exercise (Ekkekakis et al., 1999).

Besides these inherent limitations of the State-Trait Anxiety Inventory as a measure of anxiety, additional problems have emerged regarding how this measure is used and interpreted by researchers. In many studies in the field of health-behavioral research, the State-Trait Anxiety Inventory is simply misused. First, as noted in Chapter 1, it is not uncommon to see this questionnaire used as a measure of mood, stress, relaxation, general well-being, or “how people feel” in general. Clearly, these uses are incompatible with the stated scope of the measure, which was developed to assess a specifically defined and delimited emotional state (state anxiety) and disposition (trait anxiety).

Second, given that the cognitive appraisal of (primarily ego-related) threat is theorized to be the defining element of the anxiety process in Spielberger's (1972a) theoretical model, using the State-Trait Anxiety Inventory with samples of young, healthy, and non-anxious college students in studies that do not involve any experimental manipulation of anxiety-associated appraisals of threat presents some interpretational challenges. Is one really studying anxiety when the participants are non-anxious and anxiety is not experimentally induced? On a scale ranging from 20 to 80, does a score change from 25 to 22 really constitute evidence of "anxiolysis" with meaningful practical implications? Some researchers have argued that such studies are still instructive as they constitute cases of "analogue research" (Martinsen & Morgan, 1997). However, as commonly defined, analogue research involves providing treatment to individuals who manifest a certain problematic trait, such as anxiety proneness, albeit to a nonclinical degree and/or inducing a problematic state, such as fear, albeit transiently and without the risk of causing trauma to the participants. With few exceptions, in which either the participants are selected to manifest high trait anxiety (e.g., Breus & O'Connor, 1998; Focht & Hausenblas, 2003) or anxiety is transiently elevated by an experimental manipulation (e.g., Acevedo, Dzewaltowski, Kubitz, & Kraemer, 1999; Crocker & Grozelle, 1991; Doan et al., 1995; Evatt & Kassel, 2010), these conditions are not frequently met and, therefore, the label "analogue research" does not seem suitable. Given this situation, a first step before choosing the State-Trait Anxiety Inventory should be to contemplate whether one is genuinely interested in studying anxiety, as this construct is defined and delineated within the conceptual framework upon which this measure was based.

After deciding that anxiety is their true target, the next step for researchers should be to consider whether Spielberger's model of anxiety, which formed the conceptual basis of the State-Trait Anxiety Inventory, is the most appropriate for their study among alternative conceptualizations (particularly those that have emerged in the four decades that have passed since the State-Trait Anxiety Inventory was first introduced). Clearly, to make this determination, any researcher studying anxiety should have a good grasp of conceptual developments in the anxiety field since that time.

With this in mind, it is very important to point out that, during the "formative" period of the State-Trait Anxiety Inventory (i.e., in the late 1960s and throughout the 1970s), the first models of anxiety that distinguished between cognitive and somatic components began to appear in the test anxiety (e.g., Liebert & Morris, 1967; Morris,

Davis, & Hutchings, 1981) and psychotherapy literatures (e.g., Lehrer & Woolfolk, 1982; Schwartz et al., 1978). Interesting, before the 1983 development of Form Y of the State-Trait Anxiety Inventory, Spielberger fully endorsed the idea that anxiety is a multidimensional construct comprising cognitive and somatic elements. Accordingly, he developed the Test Anxiety Inventory (Spielberger, Golzalez, Taylor, Algaze, & Anton, 1978) as a questionnaire that contained both a “worry” (i.e., cognitive concerns about consequences of failure) and an “emotionality” scale (i.e., perceived reactions of the autonomic nervous system evoked in response to the perception of threat). However, without directly addressing this inconsistency, the State-Trait Anxiety Inventory was never revised to incorporate this crucial conceptual advance. Thus, by continuing to reflect the conceptualization of anxiety as a unidimensional construct, the State-Trait Anxiety Inventory has essentially been rendered obsolete from a conceptual standpoint. Modern measures of anxiety, including the Endler Multidimensional Anxiety Scales (Endler, Edwards, Vitelli, & Parker, 1989; Endler, Parker, Bagby, & Cox, 1991), the Four-System Anxiety Questionnaire (Koksal & Power, 1990), and the State-Trait Inventory for Cognitive and Somatic Anxiety (Grös et al., 2007; Ree et al., 2008), distinguish between cognitive and somatic components. This distinction is clearly more than a subtle refinement, as it has direct relevance to the critique by Rejeski et al. (1991) discussed here. A measure that distinguishes between cognitive and somatic symptoms would avoid the confound identified by these authors, enabling researchers to observe independent or divergent changes in these two components. Thus, it is important to note that, in the first known study to employ the Endler Multidimensional Anxiety Scales in conjunction with a bout of exercise, Blanchard, Rodgers, Bell, Wilson, and Gesell (2002) found that exercise induced significant changes in the autonomic-emotional scale (e.g., “Hands feel moist,” “Heart beats faster,” “Perspire”) but not in the cognitive-worry scale (e.g., “Unable to focus on task,” “Unable to concentrate,” “Feel self-conscious”).

A third problem with how the State-Trait Anxiety Inventory is used in health-behavioral research pertains to the use of shortened versions. This trend may be driven in part by convenience (since short scales are easier to administer repeatedly during an experimental session) and in part by the desire to circumvent the cost associated with using a copyrighted and commercially distributed measure. If a shortened scale is used, it is important to ensure that it is empirically derived (i.e., that published data support the selection of the specific items) and that the items that are used are from Form Y rather than from the now defunct Form X. The reason is that Form X “included several items

with poor psychometric properties” (Vagg et al., 1980, p. 212), which were replaced in Form Y. Researchers are also cautioned to only draw item lists from the original articles reporting the derivation of the shortened scales because item lists that have appeared in some studies do not always match those in the cited original sources.

In closing this critique, it is useful to point out that one issue that complicates the evaluation of the State-Trait Anxiety Inventory by researchers is its undeniable popularity across a wide range of scientific fields and over an extended period of time. This type of extreme popularity creates two significant problems. First, it tends to “automate” the selection of certain measures, creating the false impression that, to measure a certain variable, there is *de facto* only one choice. Popular measures seem like a “safe bet,” unlikely to be questioned by reviewers. When a popular measure exists, regardless of its flaws, the selection of a less popular measure often requires rigorous justification. For example, authors have summarily rejected criticisms against the State-Trait Anxiety Inventory as “not compelling,” urging researchers to continue measuring anxiety “by means of existing measures such as the STAI ... that are known to possess construct validity” (Raglin, 1997, p. 111) since it has been used in a “large body of research” (O’Connor et al., 2000, p. 138).

Second, extreme popularity means that the measure has been historically the basis for a large amount of data. Given the reliance of the research process on the measure, investigators and practitioners working in that field usually accept these data as an established body of evidence. Therefore, questioning the measure that formed the basis of this body of evidence (as one would have to do, directly or indirectly, in supporting and selecting an alternate measure) is often seen as an iconoclastic attack against what is considered the orthodox knowledge base of the field, against previous studies, or even against the scientific credibility of investigators who have used the popular measure. Naturally, most researchers, young and established ones alike, are unlikely to risk creating such undesirable impressions.