1 Documenting the breadth and depth of the problem

> Measurement issues are very frequently ignored, or treated cavalierly, almost mindlessly, in research reports. Measures seem to be used because they are "there," because someone else has used them, because nothing "better" is available. One cannot help but be amazed at the naive faith invested in what are at best crude measures by researchers who exhibit healthy skepticism, care, and sophistication with respect to other aspects of their studies. (Pedhazur & Pedhazur Schmelkin, 1991, p. 28)

A growing number of behavioral scientists across a broad range of disciplines now recognize the central role of affect, mood, and emotion in human behavior in general, and in health behavior in particular (see Figure 1.1). For example, reference to these constructs is made at an increasing rate in the literatures on eating behavior and food choices (e.g., Lutter & Nestler, 2009; Macht, 2008; Moore & O'Donohue, 2008; Stroebe, Papies, & Aarts, 2008), the causes of the obesity epidemic (e.g., Kishi & Elmquist, 2005; Rolls, 2007), the addictive effects of drugs of abuse (e.g., Baker, Piper, McCarthy, Majeskie, & Fiore, 2004; Bechara, 2005; Koob, 2008; Robinson & Berridge, 2008), the initiation and cessation of cigarette smoking (e.g., Carmody, Vieten, & Astin, 2007; Schleicher, Harris, Catley, & Nazir, 2009), the antecedents and consequences of drinking alcohol (e.g., Gilman, Ramchandani, Davis, Bjork, & Hommer, 2008; King, de Wit, McNamara, & Cao, 2011; McKinney, 2010; McKinney & Coyle, 2006), the effects of sleep and the predictors of sleep disruptions (e.g., McCrae, McNamara, Rowe, Dzierzewski, Dirk, Marsiske, & Craggs, 2008; Walker, 2009), and the effects of exercise and the reasons behind physical inactivity (Ekkekakis, Parfitt, & Petruzzello, 2011; Rhodes, Fiala, & Conner, 2009; Williams, 2008). In these diverse literatures, affect, mood, and emotion are treated as independent variables (e.g., predicting the health behavior), as dependent variables (e.g., studies on the effects of the health behavior on depression), or as mediators and moderators of various behavioral interventions for a wide range of outcomes.

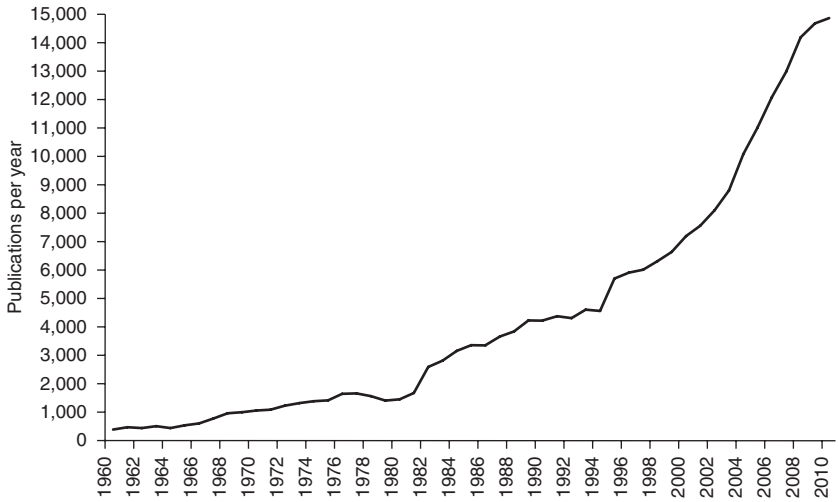

Figure 1.1. The number of entries in the PsycINFO™ database over the past 50 years (1960–2010) that include the keywords "affective," "mood*," or "emotion*." The number of entries has increased from almost zero to nearly 15,000 per year.

It is, of course, unrealistic to dissect all issues specific to the various research areas of health-behavioral research in which affect, mood, and emotion have now become focal topics. Nevertheless, the measurement problems identified, the conclusions drawn, and the recommendations issued are relevant to a very wide swath of behavioral research, including health psychology, behavioral medicine, preventive medicine, applied gerontology, and many others.

By all indications, the range of research areas in which the relevance of affective constructs is being explored is constantly expanding. This means that many new investigators, with limited or no prior experience in affect, mood, or emotion theory and research, will probably enter the fray in the coming years. Under growing pressure to seek funding and publish, they will need a rapid introduction to the theories surrounding these constructs and the measurement options available. Unfortunately, if the "newcomers" attempt to shorten their period of induction to this field by simply replicating some of the views and practices now prevalent in this literature, the prognosis for the future growth and productivity of this research does not seem promising.

The purpose of this introductory chapter is to document some of the specific problems that plague the assessment of affect, mood, and emotion in many areas of behavioral research dealing with human health.

Hopefully, by shedding light on some of the pitfalls and fallacies that are now so widespread in the literature, this chapter will act as a stimulus for researchers to approach the measurement of affect, mood, and emotion from a different, more critical perspective.

Newcomers, beware: brace yourselves 'cause this ain't gonna be easy!

Let us consider an imaginary, yet realistic, scenario. Let us assume that you are a behavioral scientist (perhaps with a background in clinical, social, health, or exercise psychology) working with an interdisciplinary group of investigators. You find yourself a few days before the deadline for a grant application to a major funding agency. The topic of your application deals with the effects of a physical activity intervention on alcohol dependence. After reading your draft, an experienced colleague suggests that you incorporate a measure of self-efficacy as a possible mediator of the effects of physical activity. You promptly identify some key references on self-efficacy from the recognized authority on this topic (Bandura, 1977, 1997, 2001) and then construct a self-efficacy instrument by carefully following the step-by-step instructions in the aptly titled "Guide for constructing self-efficacy scales," which was developed by the same authoritative figure (Bandura, 2006). The reviewers of the grant application confirm that your choice of measure is consistent with literature-wide conventions, express no concerns, and your application is funded.

This scenario could have followed a very similar course if the colleague had suggested any number of other social-cognitive constructs that have become the mainstays of behavioral interventions in recent decades, such as attitudes, social norms, behavioral intentions, goal orientations, or behavioral regulations. Each of these constructs is typically linked to one authoritative source, is embedded within one well-known and well-defined theoretical framework, and is operationally defined by one measure that has emerged as the de facto literature-wide standard.

Now let us suppose that the wise colleague suggested incorporating a measure of "affect." If your psychological training was like most others' (except until recently), then you probably never took a course exclusively devoted to affect, mood, or emotion. At best, you might have taken a course on "motivation and emotion," with ideas such as "drive" or "reinforcement" that seem outdated and irrelevant to your research. You are also fairly certain that the "affect" to which your colleague was referring does not resemble anything in your clinical psychology course or your copy of the *Diagnostic and Statistical Manual of Mental Disorders*.

The tens of thousands of references that come up after the first exploratory database searches also do not appear helpful, as they seem to refer to a hodgepodge of different things. No single author emerges as a central figure, no conceptual model seems to be cited more frequently than others, and no single measure appears to have risen to the status of de facto standard. Nervousness creeps in as you begin to realize that this is an area unlike most others. It is vast, it is diverse, and, as becomes apparent after reading a few articles, it is immersed in confusion and controversy. So what can you possibly do given the time constraints? Can you really navigate the maze of this literature, cut through the jargon, and articulate an intelligent argument for selecting a particular measure?

The honest, albeit intensely unpopular, answer is that you cannot. The harsh reality is that the study of affect, mood, and emotion will challenge a researcher more than most other topics due to the sheer amount and complexity of information that one needs to master before being ready to make a meaningful contribution. This is clearly not an area of a singular authority figure, a singular theory, or a singular measure. Consequently, researchers accustomed to "off the shelf" or "plug-and-play" measurement solutions will quickly be overwhelmed, fall easy victim to uninformed advice, and, perhaps more important, reproduce more misinformation into an already confusing literature. On the other hand, the dedicated and patient scholar entering this field will discover a truly fascinating wealth of ideas and an area of study that has intrigued humanity since the days of Aristotle.

The aim of this book is to serve as a rudimentary guide to the measurement of affect, mood, and emotion for researchers working in the broad field of health behavior. Given the challenges inherent in this task, the objective is not to provide an all-encompassing reference but rather a springboard for more focused study. To accomplish this goal, the book follows a systematic approach, from examining key theoretical issues to reviewing specific measures. In the process, an effort is made to highlight the most influential conceptual and psychometric works in this field. Most important, the analysis is critical, identifying unanswered questions, issues of concern, and unsettled points of debate, in addition to points of convergence and consensus.

How bad is the situation, really?

The first step toward positive reform is realizing that the current situation deviates from an optimal standard. Presumably, all researchers

want to do the best work possible and improve their chances of producing meaningful and valid results. In most research areas, this is ensured mainly by (a) the quality of training provided at the undergraduate and postgraduate levels, (b) to some extent, the relatively manageable size and straightforward nature of most topical literatures, and (c) the system of peer review. Put differently, research generally functions well because researchers are adequately trained in their particular area of study, the areas themselves have a reasonable size and relatively few controversial aspects, and, if something goes awry, a knowledgeable and alert peer reviewer will probably catch it before it goes to print.

Unfortunately, most of these safeguards seem absent in areas of health-behavioral research dealing with affect, mood, and emotion. First, at most institutions, formal courses devoted to emotional and affective phenomena are a very recent development. Furthermore, it must be recognized that both undergraduate and postgraduate curricula in most areas of psychology reflect the current zeitgeist. This means that graduates are considered adequately trained if they have learned the theories in line with the current paradigm, which is still heavily influenced by cognitivism. If the contents of textbooks in health psychology or health promotion, for example, are any indication, students must be well versed in such topics as social-cognitive theory, the theory of planned behavior, or the transtheoretical model. On the other hand, how many students or recent graduates of health behavior programs are aware of the landmark theories that have defined the fields of affect, mood, and emotion research over the last century?

Second, the theoretical and empirical literature on affect, mood, and emotion is extremely convoluted, reflecting more controversy than consensus. Although many researchers will find this diversity of ideas fascinating, for the majority it will probably act as a deterrent. Especially considering the lack of a previous classroom-based or textbook-guided introduction to this field, the time and effort required to gain a firm understanding of this literature will probably seem prohibitive.

Third, judging from the quality of measurement choices in many published articles, the effectiveness of the peer review process appears limited. This is due to the fact that most reviewers are typically no more educated about the conceptualization and assessment of affect, mood, and emotion than most authors.

The unfortunate outcome of this breakdown of the system is that misguided practices and erroneous claims have seeped into the published literature. Once this started happening, the problems were exacerbated. The culprit is the well-known tendency to reproduce practices

and claims that have appeared in the secondary literature instead of undertaking the incomparably more demanding task of plunging into the huge and unwelcoming primary literature to perform one's own independent and critical analysis of the issues. Obviously, the longer this trend continues unabated, the worse the situation will get.

With this as background, it is probably fair to say that the measurement of affect, mood, and emotion in the field of research dealing with health behaviors is presently, by and large, far from optimal. Certainly, the persistent rift between theory and application will always result in a "phase delay" before conceptual advances trickle down to the measurement practices followed in any field of applied research. However, arguably, in this field, there is yet to be clear evidence of a trickle-down effect. Instead, there seems to be a disconnect from theoretical and psychometric advances in affective psychology. So a lot needs to be done to get things moving in the right direction.

As a referee of a measurement-related manuscript once wrote as part of a five-line review culminating in the recommendation to "reject unconditionally," measurement issues are "arcane" and "of interest to only a few." Evidently, mysterious forces operating over the past few years have made it acceptable for researchers to select measures without taking the time to study and understand them. Scrutinizing measurement issues seems to have become something that is looked upon by some as a banality.

Other investigators, perhaps the majority, may still recognize the importance of measurement issues but, discouraged by the size and complexity of the literature, seek shortcuts. Graduate students often ask their advisors or other researchers to resolve their conundrums for them: "Which measure should I use for my thesis or dissertation?" This is sometimes followed by "Should I use *X*, I've been told it's good" or "I was thinking of using *Y* because of reason *Z*" (but, alas, "reason *Z*" usually represents a false or irrelevant premise). Similarly, grant proposals and manuscripts submitted for review to scientific journals frequently contain glaring mistakes on the theory and measurement of affective constructs. In some cases, these mistakes echo false statements that have appeared in the secondary literature. In other cases, the mistakes are peculiar, often remarkably creative, misconstruals of conceptual positions, presumably the result of the authors' cursory or uncritical inspection of the primary sources.

At this point, it seems that research investigating the role of affect, mood, and emotion in health behaviors is at a critical juncture. As these constructs become focal topics in research agendas, researchers must decide what standards they consider acceptable. Requiring all authors

to become thoroughly familiar with the conceptual basis and psychometric properties of each measure of affect, mood, or emotion before they use it would be highly desirable, but this standard seems idealistic and unattainable at this point. On the other hand, it is fairly clear that, if the threshold of acceptability is continuously lowered, the overall quality of the research will suffer. Improving the quality, persuasiveness, and overall impact of a line of research is not just about designing larger clinical trials, following meticulous randomization procedures, or blinding the outcome assessors to group allocation. It is also about selecting constructs and measures of those constructs that can stand up to the strictest standards of theoretical and psychometric scrutiny and about articulating and documenting the rationale behind these selections.

The head-in-the-sand approach: choosing a measure without providing a rationale

By far the most frequently encountered problem is pretending there is no decision-making process involved in choosing a measure. In such cases, reference to the issue of measurement is made for the first time in the Methods section, where the instruments simply appear "out of the blue," unaccompanied by a rationale to support their selection (often as a laundry list). When this approach is applied to a topic characterized by such diversity of constructs, theories, and measurement options, one can easily appreciate its fundamental inadequacy. The omitted information is of paramount importance to readers trying to evaluate the reasoning behind crucial methodological decisions.

The following excerpt from a published article is both typical, in that it is very similar to text used in hundreds of other published articles, and somewhat atypical, in that it was published in one of the most prestigious and highly selective journals in the field of health-behavioral research, as evidenced by its top-tier impact factor and extremely high rejection rate for submitted manuscripts:

> *Measures of mood*. We included two measures of mood, one a domain general and well-established scale and the second a newer, exercise-specific measure. The Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971) uses bipolar adjective scales to assess mood. Four POMS dimensions (vigor, tension, depression, elation) were utilized in this research. The POMS is a domain general measure but has been widely utilized in exercise research … The Physical Activity Affect Scale (PAAS; Lox, Jackson, Tuholski, Wasley, & Treasure, 2000) assesses exercise-induced feeling states of positive affect, negative affect, tranquility, and physical fatigue. The PAAS was developed

in response to concerns about the lack of exercise-specificity of measures like the POMS and shows adequate internal consistency and discriminant validity among the factors.

This excerpt exemplifies several practices that have become commonplace in this literature. Perhaps the most striking element is that conceptual considerations are conspicuously absent. Nothing is mentioned about the theoretical basis upon which these measures were built or why these particular theoretical frameworks were deemed most appropriate for this study. In lieu of a conceptual rationale, reference is made to other, more superficial features: (a) one measure is older and more extensively used while the other is newer and (b) one was developed for use in a variety of contexts whereas the other was developed for use specifically in the context of exercise. However, closer analysis reveals that several important pieces of information are missing.

First, why was "mood" selected as the most appropriate construct to target in this study? According to an undergraduate textbook on emotion theory and research, "the term 'mood' refers to a state that typically lasts for hours, days, or weeks, sometimes as a low-intensity background" and, furthermore, "moods are often objectless, free-floating" (Oatley, Keltner, & Jenkins, 2006, p. 30). Given this definition, the relevance of mood to a study aimed at investigating the immediate response to a brief session of physical activity among participants without a mood disorder is not entirely obvious, so readers would benefit from an explanation.

Second, although the Profile of Mood States and the Physical Activity Affect Scale are both listed as measures of "mood," one is labeled a measure of "affect." As will be explained in the next chapter, "mood" and "affect" are not synonymous terms. So readers might want to know why it was deemed necessary in the context of this study to assess both constructs and, secondarily, why it was deemed conceptually justified, given their differences, to subsume both measures under the rubric "measures of mood."

Third, there is considerable ambiguity regarding the use of the Profile of Mood States, ultimately making it impossible for readers to decipher which items were presented to respondents. There are two versions of the Profile of Mood States, an older unipolar version (McNair et al., 1971), and a newer bipolar one (Lorr, McNair, & Heuchert, 2003). Which one was used in this study is not clear because the reference given is for the unipolar version of 1971, yet the authors noted that the version they employed "uses bipolar adjective scales to assess mood." Moreover, it seems reasonable to suggest that, since the researchers had

to choose between two conceptual alternatives (or indeed opposites), namely unipolar states versus bipolar dimensions, readers would benefit from knowing the considerations upon which this important decision was based.

To complicate matters, in this case, the authors appear to have used only some of the "poles" of the six factors of the bipolar Profile of Mood States. The bipolar Profile of Mood States consists of the following six scales: (a) Composed-Anxious, (b) Agreeable-Hostile, (c) Elated-Depressed, (d) Confident-Unsure, (e) Energetic-Tired, and (f) Clearheaded-Confused (see Figure 1.2). Each scale consists of 12 items, half of which represent one pole and half the other. By choosing to measure only "vigor, tension, depression, elation," the authors in effect limited the universe of content the instrument was intended to assess but provided no explanation for the reasoning behind or the necessity of this decision. Furthermore, there are no scales named "vigor" or "tension" in the bipolar Profile of Mood States (only in the older, unipolar version). It is possible, though uncertain, that these labels refer to the "Energetic" and "Anxious" poles of the Energetic-Tired and Composed-Anxious factors, respectively. More important, despite noting that the version of the Profile of Mood States that was used consists of "bipolar adjective scales," the authors apparently only scored unipolar half-scales. One pair (Elated and Depressed) is theorized to form a single bipolar factor but was scored as two separate unipolar scales. The other two (Vigor, Tension) represent single poles, each possibly extracted from a different bipolar factor. Once again, it appears that, since the instrument was scored and interpreted in a manner different from the way its developers intended, readers should have been informed of the rationale behind these changes.

Fourth, the selection of the Physical Activity Affect Scale appears to have been based on its presumed "exercise specificity." To readers willing to place this claim under appropriate scrutiny, this warrants an explanation. One may wonder, for example, why some items are characterized as more "exercise specific" than others. For example, why are *alert*, *vigorous*, or *lively* (items from the Energetic pole of the bipolar Profile of Mood States) less exercise specific than *enthusiastic*, *energetic*, or *upbeat* (items from the Positive Affect scale of the Physical Activity Affect Scale)? Why are *miserable* and *discouraged* (items from the Negative Affect scale of the Physical Activity Affect Scale) more exercise specific than *dejected* and *discouraged* (items from the Depressed half-scale of the bipolar Profile of Mood States)? While the focus was placed on the (debatable) issue of exercise specificity, it is interesting to point out that

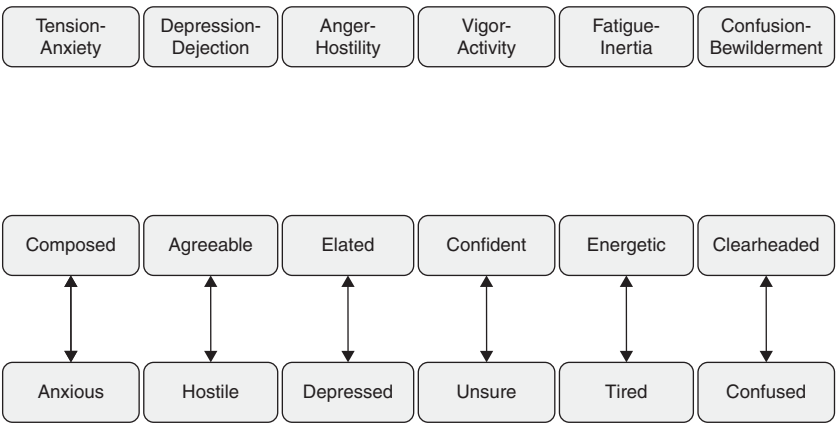

Figure 1.2. The factors of the older, unipolar Profile of Mood States (top row) and the newer, bipolar version of the Profile of Mood States (bottom row).

the conceptual features of the Physical Activity Affect Scale (e.g., the theorized nature, polarity, and relations among the factors) were not mentioned as a consideration that led to its selection.

Readers also frequently encounter studies with similar aims employing measures vastly different from a conceptual standpoint but are given no explanation that could justify these differences. Having no explanation for the different measurement decisions, readers are often left confused and frustrated. For example, commenting on studies investigating the relationship between sleep and affect by using different measures, such as a list of items previously used by Lorr, Daston, and Smith (1967) or the University of Wales Institute of Science and Technology Mood Adjective Checklist (Matthews, Jones, & Chamberlain, 1990), McCrae et al. (2008) noted that this complicates "the interpretation and contextualization of these findings" (p. 43). However, they then proceeded to use yet another measure, the Positive and Negative Affect Schedule (Watson, Clark, & Tellegen, 1988), without providing any explanation or justification for this decision.

Similarly, authors often use combinations of measures without explaining why the use of multiple measures was necessary or how the multiple measures complement each other. For example, one study was designed to examine the effects of two forms of exercise (stationary cycling and martial arts) on "mood" in a sample of individuals with recurrent major depressive disorder (Bodin & Martinsen, 2004). In the Methods section, the authors listed two measures, both reportedly