ELSEVIER

Psychology of Sport and Exercise 2 (2001) 205–232

Psychology
OF SPORT AND EXERCISE

www.elsevier.com/locate/psychsport

Analysis of the affect measurement conundrum in exercise psychology. III. A conceptual and methodological critique of the Subjective Exercise Experiences Scale

Panteleimon Ekkekakis [a,*], Steven J. Petruzzello [b]

[a] *Department of Health and Human Performance, Iowa State University, 253 Barbara E. Forker Building, Ames, IA 50011, USA*

[b] *Department of Kinesiology, University of Illinois at Urbana-Champaign, Urbana-Champaign, IL, USA*

Received 7 October 1999; received in revised form 23 May 2001; accepted 22 June 2001

Abstract

Background and purpose. The general purpose of this series is to examine the controversy that surrounds the measurement of affect in the context of acute exercise. The present paper focuses on the conceptual underpinnings and the methodological steps that were followed in the development of an increasingly popular measure, namely the Subjective Exercise Experiences Scale (SEES) (J Sport Exercise Psychol, 16 (1994) 163). Emphasis is placed on how conceptual ambiguities may influence methodological decisions and, ultimately, the content and structure of a measure.

Methods. From a conceptual standpoint, attention is given to the delineation and demarcation of the content domain of the scale, the decision to adopt a dimensional conceptualization of affect, the notion of 'subjective experiences unique to exercise', and, mainly, whether positive and negative affect are independent constructs or the opposite ends of a single bipolar dimension. From a methodological standpoint, the analysis focuses on the process of item selection and content validation, and the exploratory and confirmatory factor analyses.

Results and conclusions. Conceptual and empirical evidence is reviewed suggesting that: (a) there is ambiguity regarding the nature and the limits of the content domain of the SEES; (b) the notion of 'subjective experiences unique to exercise' presents considerable logical problems; and (c) positive and negative affect are not necessarily independent constructs. In conclusion, the analysis underscores the importance of conceptual groundwork and clarity as prerequisites for methodological implementation.

Keywords: Structure of affect; Positive affect; Negative affect; Independence; Bipolarity

* Corresponding author. Tel.: +1-515-294-8766; fax: +1-515-294-8740.
E-mail address: ekkekaki@iastate.edu (P. Ekkekakis).

1469-0292/01/$ - see front matter
PII: S1469-0292(01)00022-X

Introduction

In the introduction to the present series (Ekkekakis & Petruzzello, 2000) and in previous work (Ekkekakis, Hall, & Petruzzello, 1999; Ekkekakis & Petruzzello, 1999), we have commented extensively on the controversy that surrounds the measurement of affect in the context of acute exercise. We also noted that, because unsound measurement "can send science on false leads and wasteful detours into fads and fancies" (Bass, 1974, p. 870), measurement problems warrant close scrutiny and, if possible, rapid resolution.

This paper extends our analysis by focusing on the Subjective Exercise Experiences Scale (SEES; McAuley & Courneya, 1994). The SEES is a 12-item self-report measure of the "subjective experiences that are unique to the exercise domain" (p. 165). These items are organized in three subscales, namely 'Positive Well-Being' (PWB), 'Psychological Distress' (PD), and 'Fatigue.' Together with the Exercise-induced Feeling Inventory (EFI; Gauvin & Rejeski, 1993), which we examined in a previous paper (Ekkekakis & Petruzzello, 2001), the SEES has been characterized as an "auspicious beginning to a solution" (McAuley & Rudolph, 1995, p. 90) to the problem of measuring affective responses to exercise. Since its publication in 1994, the popularity of the SEES has been rising and the scale is currently being used not only by its developers, but also by independent researchers (e.g. Blanchard, Rogers, Spence, & Courneya, 2001; Lox & Treasure, 2000; Parfitt, Rose, & Markland, 2000; Rudolph & Butki, 1998; Rudolph & Kim, 1996; Watt & Spinks, 1997). Furthermore, together with the EFI, the SEES formed the basis for the development of a hybrid measure, the Physical Activity Affect Scale (Lox, Jackson, Tuholski, Wasley, & Treasure, 2000). In the first review to focus on the measurement of affect in the context of exercise, Gauvin and Spence (1998) noted that, although a number of questions remain to be investigated, the preliminary indications regarding the validity and reliability of the SEES seem promising. Generally positive conclusions were also reached by two confirmatory factor analytic investigations (Lox & Rudolph, 1994; Markland, Emberton, & Tallon, 1997).

In contrast to previously published evaluations of the SEES, the primary focus of the present analysis is not on quantitative psychometric indices. We believe that, when a new measure targets a novel construct or incorporates a novel structure, as is the case with the SEES, the computation of indices of internal consistency and goodness-of-fit should be preceded by another critical step. It is important to remember that measures are not developed in a conceptual vacuum but, instead, reflect the conceptual assumptions adopted by their developers (Stone, 1995; Wallbott & Sherer, 1989). Morever, statistical goodness-of-fit cannot be construed as evidence for whether a structural model is theoretically meaningful (Cole, 1987). Therefore, the first step in evaluating a measure of a novel construct or a measure that proposes a novel structural model should be the examination of its conceptual underpinnings and the correspondence between conceptual modeling and methodological implementation.

In light of this, it is noteworthy that previous psychometric evaluations of the SEES involved the computation of psychometric indices and comparisons against conventional yardsticks, but did not address the theoretical basis of the measure. Similarly, it should be clear that the selection of a measure for use in applied research should be a decision primarily driven by theoretical considerations (e.g. what specific dependent variables a researcher is interested in and why, based

on the experimental conditions or treatments involved). Yet, the researchers who have opted to use the SEES in their studies, as opposed to some other measure of affective constructs, have done so without citing any theoretical reasons. These facts can arguably be perceived as alarming indications that researchers are approaching the measurement of affect in the context of exercise in an atheoretical and uncritical manner. The goal of this paper is to help change this trend by shedding light on several important theoretical issues surrounding the conceptualization and development of the SEES.

The analysis is organized in two parts, the first focusing on the conceptual foundation of the SEES and the second on the methodological steps that were taken for its development. The aim of this approach is to underscore the importance of theoretical postulates for shaping methodological decisions and, ultimately, the content and structure of a new measure. In discussing the conceptual basis of the SEES, the introductory paper of the present series (Ekkekakis & Petruzzello, 2000) is used as the organizing framework and readers are referred to it for additional information.

Conceptual foundation

Nature and limits of the content domain

In the paper describing the development of the SEES, McAuley and Courneya (1994) used a variety of terms to refer to the general content domain of the scale. These terms included 'psychological health', 'psychological or mental health', 'psychological well-being', 'psychological distress', 'emotion', 'affective states', 'subjective experiences', 'exercise-induced experiences', etc. In the introduction, McAuley and Courneya also made references to measures of *mood*, such as the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971), the Activation Deactivation Adjective Check List (AD ACL; Thayer, 1989), and the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988). When direct references were made to the content domain of the SEES, broad terms like 'subjective experiences' or 'subjective responses' (McAuley & Courneya, 1994, p. 165) were used and the SEES was described as "a measure of global psychological responses to the stimulus properties of exercise" (p. 163). Moreover, McAuley and Courneya noted that "although the SEES is not strictly proposed as a measure of emotion, it certainly appears to be a measure of responsivity from which particularized emotional states may emanate" (p. 173).

In his guidelines for scale developers, DeVellis (1991) emphasized that "the boundaries of the phenomenon must be recognized so that the content of the scale does not inadvertently drift into unintended domains" (p. 51). In general, the lack of clarity, specificity, and precise definitions can undermine the validation process. In the case of the SEES, the use of varied and general terminology, in conjunction with the absence of a direct definition of the content domain of the scale, created considerable ambiguity. One factor that appears to have contributed to this ambiguity is the apparent tendency to extend the content domain of the SEES beyond the sphere of affect and into the vast and less easily definable areas of 'subjective experiences' and 'global psychological responses'.

McAuley and Courneya (1994), citing Clore, Ortony, and Foss (1987), noted that "although (perceptions of somatic states) can be classified as affective responses, they may also be representative of perceived physiological activation (i.e. nonmental states) and thereby may be discarded as affects" (p. 165). The problem, again, appears to be one of definition. In the introduction to the present series (Ekkekakis & Petruzzello, 2000), we presented a widely accepted definition, according to which affect is the "irreducible" (Frijda, 1993, p. 383) or "most elementary consciously accessible" (Russell & Feldman-Barrett, 1999, p. 806) experiential component of all valenced (i.e. positive or negative, good or bad) responses, including emotions and moods. This makes affect a broader concept compared to emotions. According to Ortony, Clore, and Foss (1987), "emotions are affective conditions, but not…all affective conditions are emotions" (p. 343) and, according to Batson, Shaw, and Oleson (1992), "of affect, mood, and emotion, affect is the most general" (p. 298).

Based on these clarifications, McAuley and Courneya's (1994) statement that perceptions of somatic states "may be discarded as *affects*" (p. 165, emphasis added) is not accurate. A more accurate statement would be that, according to most theorists, such states would not qualify as *emotions*, because they do not necessarily depend on cognitive appraisal, the main distinguishing feature of emotions (Clore et al., 1987; Lazarus, 1991; Ortony et al., 1987). An examination of the terms that Clore et al. (1987) considered as 'nonmental' (e.g. 'aroused', 'sleepy') shows that these refer to somatic states that do not intrinsically denote some (positive or negative) valence. However, if these states are assigned positive or negative valence in a given context, then, by definition, they would qualify as affective states.

It could be argued, therefore, that resorting to overly inclusive terms, such as 'subjective experiences' and 'global psychological responses', to describe the content domain of the SEES might have been unnecessary and might have contributed to the ambiguity regarding the domain of content targeted by the SEES. The content of the scale does not appear to extend beyond the domain of affect, as the term is commonly defined.

The study of affect in exercise: categories or dimensions?

One of the fundamental decisions that developers of measures of affect have to face is whether it is more appropriate for their goals to adopt a categorical or a dimensional conceptualization of affect (see Ekkekakis & Petruzzello, 2000, for more on this topic). In short, in categorical conceptualizations, affective states are organized in distinct categories consisting of states that bear resemblance to prototypical exemplars (e.g. fear, guilt, pride, etc.; see Lazarus, 1991; Ortony, Clore, & Collins, 1988, for reviews). On the other hand, dimensional conceptualizations are based on the assumption that affective states are systematically interrelated, such that their relationships can be modeled parsimoniously by as few as two basic dimensions (e.g. affective valence and activation; see Larsen & Diener, 1992; Russell 1989, 1997; Tellegen, 1985, for reviews).

Depending on the nature of the research question, categorical and dimensional models exhibit both relative strengths and weaknesses. Categorical models can highlight subtle distinctions between affective states and, thus, may facilitate the investigation of unique situational or cognitive antecedents. This makes categorical models particularly useful for the study of distinct

emotions (Lazarus, 1991; Ortony et al., 1988) but can prove restrictive in descriptive studies, where little is known in advance about the nature, the cognitive substrates, or the direction of the affective responses that are likely to occur. Conversely, dimensional approaches offer the advantages of a broad investigative scope and parsimony (Larsen & Diener, 1992), but lack the ability for fine discriminations between experientially similar yet distinct affective states (Lazarus, 1991). Thus, although they are considered as inadequate templates for the study of emotions (Lazarus, 1991; Russell & Feldman-Barrett, 1999), they do offer an important advantage over categorical models. Specifically, the ability of dimensional models to map affective responses in terms of a few basic dimensions makes them particularly effective in those situations where categorical models prove limited, namely when the aim is to capture and describe an affective response whose exact nature and direction cannot be predicted (Russell, 1989; Stone, 1995).

Given that the present stage of knowledge development in exercise psychology does not permit one to anticipate the kinds of affective responses that are likely to occur in different exercisers under various conditions, we have argued that most descriptive studies would probably benefit from the broad investigative scope afforded by dimensional models (Ekkekakis & Petruzzello, 2000). Of course, it should be kept in mind that the relationship between specificity and comprehensiveness is reciprocal; in order to develop an initial rudimentary understanding of the nature of the affective changes that accompany exercise under various conditions, some sacrifice of specificity is inevitable. In agreement with other authors (Gross 1998, 1999; Rosenberg, 1998; Russell & Feldman-Barrett, 1999), we view the affective domain as a hierarchically organized structure, with a few broad dimensions accounting for the commonalities at the level of basic affect and categorical models accounting for the specificity at the level of distinct emotions. As was noted by Watson and Clark (1997), the important thing to remember is that categorical and dimensional models "are not incompatible or mutually exclusive; rather, they essentially reflect different levels of a single, integrated hierarchical structure" (p. 269).

Although McAuley and Courneya (1994) did not directly address the issue of affective categories and dimensions and did not use these terms in their paper, it appears that the SEES was conceptualized as a dimensional measure. Comparing the SEES to the EFI (Gauvin & Rejeski, 1993), a measure based on a categorical conceptualization (Ekkekakis & Petruzzello, 2001), McAuley and Courneya referred to a "hierarchy of psychological responses to exercise participation" (p. 173), characterizing the SEES as a measure of "general psychological responses" (p. 173) and the EFI as a measure which may "represent further underlying structural aspects" (p. 173) of these responses. Although not explicitly stated, the reasoning behind conceptualizing the SEES as a dimensional measure appears to have been based on the belief that there should be a progression from the study of the more general (i.e. affective dimensions) to the more specific (i.e. categories). McAuley and Courneya stated that they view the SEES as assessing responses at a "global level" (p. 173) and, as such, "a starting point perhaps for the examination of the hierarchy of psychological responses to exercise participation" (p. 173). This process, they argued, may eventually lead to the study of "particularized emotional states" (p. 173).

Although the use of terms that are different than those typically found in the literature (i.e. *categorical* and *dimensional*) makes our interpretation somewhat speculative, McAuley and Courneya's statements seem to be consistent with our views on this topic (Ekkekakis & Petruzzello, 1999, 2001). Specifically, for reasons we have presented, we agree that, at the present stage of knowledge development in exercise psychology, dimensional models present substantial advan-

tages over categorical ones for descriptive investigations due to their broad scope and parsimony. Furthermore, we agree with the idea that the affective domain is hierarchically organized and we support the proposition for a systematic progression of research from the general (i.e. the study of affective dimensions) to the specific (i.e. the study of affective categories or specific emotions).

Arguably, a statement of agreement may seem superfluous in the context of a critique. Nevertheless, we believe that the issue is important enough to warrant reiteration and emphasis. A case in point is the recent argument that, because some of the scales of the SEES and the EFI are significantly correlated, the two instruments could be merged (Lox et al., 2000). As explained here, the former measure is assumed to tap general dimensions, whereas the latter is assumed to tap specific states. This is a substantial conceptual incompatibility that makes a merger inappropriate. It should be emphasized that a correlation between constructs from different levels of a theoretically hierarchical domain should be expected and does not constitute grounds for a merger. To draw an illustrative example from another hierarchical model, physical self-worth may correlate with global self-worth, but this does not negate the important conceptual distinction between the two constructs. At least at a conceptual level, the same is true for positive well-being (a construct from the SEES, presumed to reflect a general dimension) and positive engagement (a construct from the EFI, presumed to reflect a specific state); the correlation of 0.783 (61% shared variance) reported by Lox et al. does not negate the fact that the two measures were presumed to reflect different levels of the affective hierarchy.

Measuring 'subjective experiences unique to exercise': logical problems

One of the most oft-cited arguments in support of developing measures of 'exercise-specific' affect is that the older and more general (i.e. non exercise-specific) measures of affective constructs contain items that seem irrelevant to the context of exercise. For example, concentrating their criticism on the PANAS, McAuley and Rudolph (1995) noted the following:

> Those of you who exercise on a regular basis might consider the following question: Who among you experiences either an increase or a decrease in your level of compassion or guilt as a function of exercise? Yet a measure of positive and negative affect that is receiving increasing attention is the [PANAS]… and the items "guilty" and "compassion" appear in this measure [*sic*: the item "compassion" is not included in the PANAS]. In a recent laboratory study, our research group assessed responses on the PANAS during activity; responses to such items (especially "guilty") resulted in very negative emotional responsivity. Individuals were clearly frustrated at having to respond to what they perceived as a nonrelevant emotion (p. 90).

Thus, McAuley and Rudolph (1995) questioned the utility of the PANAS based on the argument that it fails to "tap the stimulus properties of the exercise environment" (p. 90). Similar criticisms were directed toward the POMS by McAuley and Courneya (1994). Consequently, one of the primary goals in the development of the SEES was to select items that are relevant to exercise, thus presumably increasing the sensitivity of the measure to exercise stimuli. Importantly, McAuley and Courneya further assumed that exercise brings about responses that are *unique*, stating that the new measure should be "able to assess subjective experiences that are unique to the exercise domain" (p. 165).

The notion of domain-specific measurement has some well-established precedents in sport and exercise psychology and this may increase the willingness of researchers to accept these propositions without much skepticism. There is a substantial difference, however, between an exercise-specific measure of affect and a measure of exercise-specific affect. Until now, in sport and exercise psychology we have been dealing primarily with domain-specific measures, such as measures of *competitive* state anxiety or *sport* goal orientations or *exercise* locus of control. In the case of such measures, the nature and the structure of the psychological construct that is assessed remains essentially unaltered from conceptualizations that were based on extensive previous theorizing and empirical research, typically in general psychology. These measures are made domain-specific by the relatively simple and uncontroversial process of including domain-specific (i.e. sport- or exercise-specific) references in the instructions to the respondents and in the items (e.g. 'I am worried about the upcoming competition' instead of 'I am worried').

Contrary to this straightforward scenario, in the case of a measure of 'exercise-specific' affect, a *novel theoretical construct* is proposed, whose exact nature and structure are still unexplored. Therefore, developing a measure of 'exercise-specific' affect introduces considerable complexity and requires extensive groundwork. Let us examine some of the obstacles.

First, as Stone (1995) has argued, proposing the development of what he called an 'idiosyncratic' measure cannot be based solely on the intention to increase the measure's responsiveness or sensitivity to a particular treatment. Instead, substantive evidence must be provided demonstrating that, due to the properties of the treatment, the content and structure of the affective domain are *uniquely* transformed. This is not an easy task. To substantiate the claim that the affective responses that accompany exercise are *unique*, either in nature or in structure, one would have to systematically examine a variety of exercise and non-exercise stimuli, a variety of participants, and a variety of situations. To our knowledge, the extant empirical evidence is insufficient to support such a conclusion.

Second, the notion that there is a distinct set of "subjective responses that are driven by the stimulus properties of the exercise environment" (McAuley & Courneya, 1994, p. 173) presents some serious problems of generalizability. There is agreement that the affective responses to exercise are the products of complex interactions between the attributes of the exercise stimulus, the physiological and psychological makeup of the participants, and the physical and social environment (Ekkekakis & Petruzzello, 1999). As a result of these interactions, it is not surprising that individual affective responses have been found to exhibit tremendous inter-individual variability, not only in terms of magnitude (Gauvin & Brawley, 1993) but, more importantly, in terms of their nature and direction (Van Landuyt, Ekkekakis, Hall, & Petruzzello, 2000). Therefore, examining the items that may be considered 'relevant' to exercise by one segment of the population or based on one set of experimental conditions is problematic. It is entirely possible that a different set of items may become relevant in a different population or under different experimental conditions. In that case, the 'restricted' measurement instrument is bound to misrepresent (i.e. underestimate or altogether miss) the pattern of affective changes that will emerge.

Third, it is important to remember that, unlike other domain-specific measures, an instrument developed to tap 'exercise-specific' responses, in addition to exercise, is also likely to be used in a variety of non-exercise conditions, such as pre-exercise assessments or various sedentary control or comparison conditions. Using a measure that has been tailored *a priori* to tap only those facets of affective experiences that are likely to be influenced by exercise to the exclusion of all other

facets is tantamount to stacking the cards against all control or comparison treatments. Theoretically, if the measure only taps the experiences that are 'unique' to exercise, no other treatment or condition will ever demonstrate any effects. As we have argued previously (Ekkekakis & Petruzzello, 2000, 2001), showing that exercise produces large effect sizes under such conditions would be an essentially trivial finding, since the measure itself would be specifically engineered to maximize the effect of exercise and to minimize the effect of all other treatments. The logical problems and the potential for bias created by this approach should be evident.

On the other hand, one can easily appreciate McAuley and Courneya's (1994) and McAuley and Rudolph's (1995) concern that some items may indeed be viewed as irrelevant by certain participants under a given set of conditions. This is clearly a vexing problem and its resolution poses a serious challenge. From our perspective, this area could benefit greatly by taking into consideration the important distinctions between the constructs of *emotion*, *mood*, and *basic affect* and their relative positions on the affective hierarchy (for definitions and a discussion, see Ekkekakis & Petruzzello, 2000). As an example, let us examine the item 'guilty' that McAuley and Rudolph (1995) characterized as irrelevant to exercise. Guilt is an *emotion*. As such, it depends on a specific pattern of appraisals (Lazarus, 1991) and, as with all emotions, if this pattern of appraisals is present, this emotion will be elicited. In other words, it is not the context of exercise itself that will determine whether this emotion will or will not be elicited, but rather the appraisal made by a given individual. Therefore, theoretically, one cannot state with certainty that a given stimulus (i.e. exercise) will *never* elicit guilt. Guilt, according to Lazarus' analysis (1991), is elicited when one blames oneself for having transgressed a moral imperative. Although this appraisal is perhaps *unlikely* to occur in the context of exercise, it is not *impossible*. For instance, a habitual exerciser may experience guilt if she decides to exercise rather than spend time with her family. Therefore, the item 'guilty' is not irrelevant to *exercise* per se, but rather to exercise under specific conditions (participants and treatment). From a theoretical standpoint, it cannot be argued that 'guilt' is *de facto* irrelevant to exercise, but rather, because it is an *emotion* that depends on a specific pattern of appraisals, it is perhaps unlikely to be elicited in the context of exercise (unless the experimental treatment is designed to induce its antecedent appraisal).

It should be evident that approaches aimed at classifying affective states as de facto relevant or irrelevant to exercise face some sizeable, perhaps insurmountable, theoretical and practical obstacles. Simply put, it would be impossible to develop a measure that contains all the specific affective states that may be elicited in the context of exercise under all possible conditions. So, is there a viable solution? We believe there may be. In our view, this would require 'zooming out' — focusing not on individual affective states but rather on an integral component of all affective states, regardless of whether they are emotions, moods, or simpler responses that do not have a cognitive basis (e.g. the discomfort associated with strenuous effort). This integral component is *basic affect*. The study of affect, by virtue of the relative simplicity of this construct compared to emotions and moods, affords a more secure starting point of exploration and a more reasonable progression from the general to the specific and from the simple to the complex. We concur with McAuley and Courneya (1994) that there is a "hierarchy of [affective] responses to exercise participation" (p. 173) and that we may, some day, be able to hone in on the "more particularized emotional states" and the "further underlying structural aspects" (p. 173) of these responses.

In conclusion, although the development of domain-specific measures is commonly perceived as a significant advance, it is important for researchers to be aware of the important distinction between a measure of a previously well-developed construct that has domain-specific references and a measure of a novel construct, assumed to be unique to a given domain. The SEES falls in the latter category. This substantially increases the complexity of the theoretical and practical issues that must be resolved before proceeding. As we noted, to our knowledge, there is presently no evidence that there are affective states "unique to the exercise domain" (McAuley & Courneya, 1994, p. 165). There is also no theoretical or empirical basis for characterizing some affective states as de facto relevant or irrelevant to exercise. These are important matters that need to be considered before researchers accept the notion of "exercise-specific" affect.

Why 'Positive Affect' and 'Negative Affect' differ from 'positive affect' and 'negative affect'

In presenting the rationale for the development of the SEES, McAuley and Courneya (1994) noted that previous research on the relationship between exercise and affect had concentrated on negative states, such as anxiety and depression, and not so much on the positive responses that accompany exercise. To some extent, this could be attributed to the fact that the most frequently used measures (e.g. the POMS) are mainly geared toward the assessment of negative, rather than positive, states. Based on the belief that health is not simply the absence of negative symptoms, but should also encompass the promotion of well-being, McAuley and Courneya sought to develop the SEES as a balanced measure that would assess both positive and negative responses. This is a reasonable goal and one can easily appreciate the need for balance. Beyond this point, however, lie some controversial issues that have been the topic of a large and continuously expanding literature in affective psychology and psychometrics.

The question at the center of this controversy deals with the nature of the relationship between positive and negative affect and, more specifically, with whether positive and negative affect are independent dimensions or the two opposite poles of a single bipolar dimension. This is an important dilemma and, clearly, a conceptual position in support of one or the other view can have significant implications for the development and construct validation of a dimensional measure of affect. Given the importance of the topic, the intensity of the controversy that surrounds it, the size of the relevant literature, and its practical implications for the development of the SEES (and every measure of affect for that matter), it is necessary to trace its history and take a critical look into the most prominent 'sticky points'.

In the introduction of the SEES paper, McAuley and Courneya (1994) noted that "from a conceptual perspective, we concur with the broader social psychological literature that suggests emotional or affective responses vary along two (positive and negative)… dimensions" (p. 165). Although this statement points to a position in favor of the notion of the independence between positive and negative affect, McAuley and Courneya did not elaborate further and did not explain the reasons that led them to adopt this position. This creates some ambiguity that is exacerbated by several statements and methodological decisions that seem to alternate between the two opposite views, the one that favors the independence of positive and negative affect and the one that favors bipolarity.

First, McAuley and Courneya (1994) made several statements that seem to favor the model proposed by Tellegen, Watson, and associates (Tellegen, 1985; Watson & Tellegen, 1985; Zevon & Tellegen, 1982). According to this model, the structure of affect is characterized by two dimensions, namely 'Positive Affect' and 'Negative Affect', which are theorized to be independent and, more specifically, *orthogonal*. On the other hand, there were also several statements by McAuley and Courneya that seem to support a *bipolar* conceptualization of health, including 'general psychological health'. According to this conceptualization, which goes against the traditional notion of a disease continuum, health is viewed as a bipolar construct, with disease on one end and well-being on the other (also see McAuley, 1994).

Second, the ambiguity extends to McAuley and Courneya's (1994) review and critique of previous theory and research. For example, they criticized the notion of the "independence of positive and negative affect" (p. 173) which formed the conceptual basis of the PANAS (Watson et al., 1988), but they also criticized the notion of bipolarity, which formed the conceptual basis of the Feeling Scale (FS; Hardy & Rejeski, 1989), stating that "the presumption of affect as bipolar and therefore unidimensional (i.e. positive and negative affect as opposite ends of the same continuum) is troublesome from both conceptual and theoretical perspectives" (p. 165). In support of the former position, they cited the work of Green, Goldman, and Salovey (1993) and in support of the latter they cited the work of Watson et al. (1988). These authors have, however, criticized each other, and hold opposing views on the issue of independence versus bipolarity (also see Green & Salovey, 1999; Green, Salovey, & Truax, 1999; Tellegen, Watson, & Clark, 1999a, 1999b; Watson & Clark, 1997; Watson & Tellegen, 1999; Watson, Wiese, Vaidya, & Tellegen, 1999, for recent follow-ups).

Third, McAuley and Courneya (1994) used the terms 'dimensions' and 'poles' interchangeably throughout their paper. For example, they noted that "the [Positive Well-Being] and [Psychological Distress] *dimensions* [of the SEES] represent the positive and negative *poles* of overall psychological health" (p. 167, emphasis added) and that their data "show support for the conceptual perspective that psychological responses to exercise assume a multi-*dimensional* structure anchored by positive and negative *poles*" (p. 172, emphasis added).

Fourth, the conceptual ambiguity appears to have also affected the methodological steps that were followed in the development of the SEES. Specifically, in the factor analyses conducted to refine the content and test the structural validity of the scale, McAuley and Courneya (1994) used an orthogonal (varimax) rotation in their exploratory factor analysis. This points to the assumption that the resultant factors are unrelated, but later allowed the latent factors to correlate freely in a confirmatory factor analysis.

In sum, despite the statement that "emotional or affective responses vary along two (positive and negative)… dimensions" (p. 165), McAuley and Courneya's (1994) theoretical position on the issue of the independence versus bipolarity of positive and negative affect is obfuscated by several apparently conflicting statements and methodological decisions. Ultimately, however, the subscales of the SEES that were theorized to tap positive affect (i.e. Positive Well-Being) and negative affect (i.e. Psychological Distress) were developed as separate entities. Previous reviewers and users of the SEES seem to have accepted this position as a given, in spite of the aforementioned controversy that surrounds this issue. With the benefit of hindsight and the numerous insightful papers that have appeared on this topic since the publication of the SEES (Carroll, Yik, Russell, & Feldman-Barrett, 1999; Feldman-Barrett & Russell, 1998, 1999; Green & Salovey,

1999; Green et al., 1999; Russell & Carroll, 1999a, 1999b; Russell & Feldman-Barrett, 1999; Tellegen et al., 1999a,b; Watson & Clark, 1997; Watson & Tellegen, 1999; Watson et al., 1999), it might be useful to revisit the question of independence versus bipolarity.

What a difference a name can make

As McAuley and Courneya (1994) noted, the idea that 'positive affect' and 'negative affect' represent independent dimensions has been very popular in social psychology. In the 1980s, in particular, the popularity of this notion was fueled following the work of Tellegen, Watson, and associates (Tellegen, 1985; Watson & Tellegen, 1985; Zevon & Tellegen, 1982). These researchers presented a dimensional model of affect in which two dimensions, labeled 'Positive Affect' (PA) and 'Negative Affect' (NA), theorized to be orthogonal and bipolar, were considered primary. The model evolved from fairly extensive and rigorous structural analyses and its empirical basis is generally regarded as valid. What has generated abysmal confusion, however, are the labels that were given to the two dimensions. Although most people consider the adjective 'happy' to be a prime example of 'positive affect' and the adjective 'sad' to be a prime example of 'negative affect' (Shaver, Schwartz, Kirson, & O'Connor, 1987), this was not the case in Watson and Tellegen's (1985) model. The items 'happy' and 'sad' were located on the opposite poles of another bipolar dimension labeled 'Pleasantness–Unpleasantness' (see Fig. 1, panel 'a'). As confusing and counterintuitive as it may seem, 'Positive Affect' does not necessarily imply pleasure and 'Negative Affect' does not necessarily imply displeasure. Despite the fact that the terms 'Positive Affect' and 'Negative Affect' imply unipolarity, the dimensions were bipolar (having 'high' and 'low' poles). The PA and NA dimensions were hybrids of affective *valence* (referred to as the 'Pleasantness–Unpleasantness' dimension in Tellegen and Watson's terminology) and *activation* (referred to as the 'Strong Engagement–Disengagement' dimension in Tellegen and Watson's terminology). Thus, PA includes pleasant states at its high activation pole (e.g. enthusiastic, excited, etc.), but unpleasant states at its low activation pole (e.g. drowsy, sluggish, etc.). Conversely, NA includes unpleasant affective states at its high activation pole (e.g. distressed, fearful, etc.), but pleasant states at its low activation pole (e.g. relaxed, calm, etc).

The PANAS, the measure of PA and NA developed by Watson and his associates (1988), only assesses the high-activation poles of the theoretical PA and NA dimensions, in essence providing an operationalization of only one half of the affective space described by the original PA–NA model (compare panels 'a' and 'b' in Fig. 1). As explained by Watson and Clark (1997), by including only the high-activation poles of the dimensions, they were able to keep the correlation between the PA and NA scales closer to zero (as posited by the PA–NA model). When bipolar scales were used, which included both the high-activation and the low-activation poles of the two dimensions, the correlation between the bipolar versions of the PA and NA scales deviated from zero. This is not an uncommon finding in the psychometric literature that deals with the structure of affect and possible reasons for this phenomenon will be discussed shortly.

Given that the PA and NA dimensions were, in fact, orthogonal to each other, the ambiguity of the terms sparked an outbreak of papers where the research by Tellegen and Watson is erroneously cited as supporting the notion that positive affective states (i.e. pleasure, not PA) are orthogonal to negative affective states (i.e. displeasure, not NA). Larsen and Diener (1992) list several examples from the literature where the 'positive affect–negative affect' nomenclature is used in reference to affective states that are not parts of Tellegen and Watson's orthogonal PA

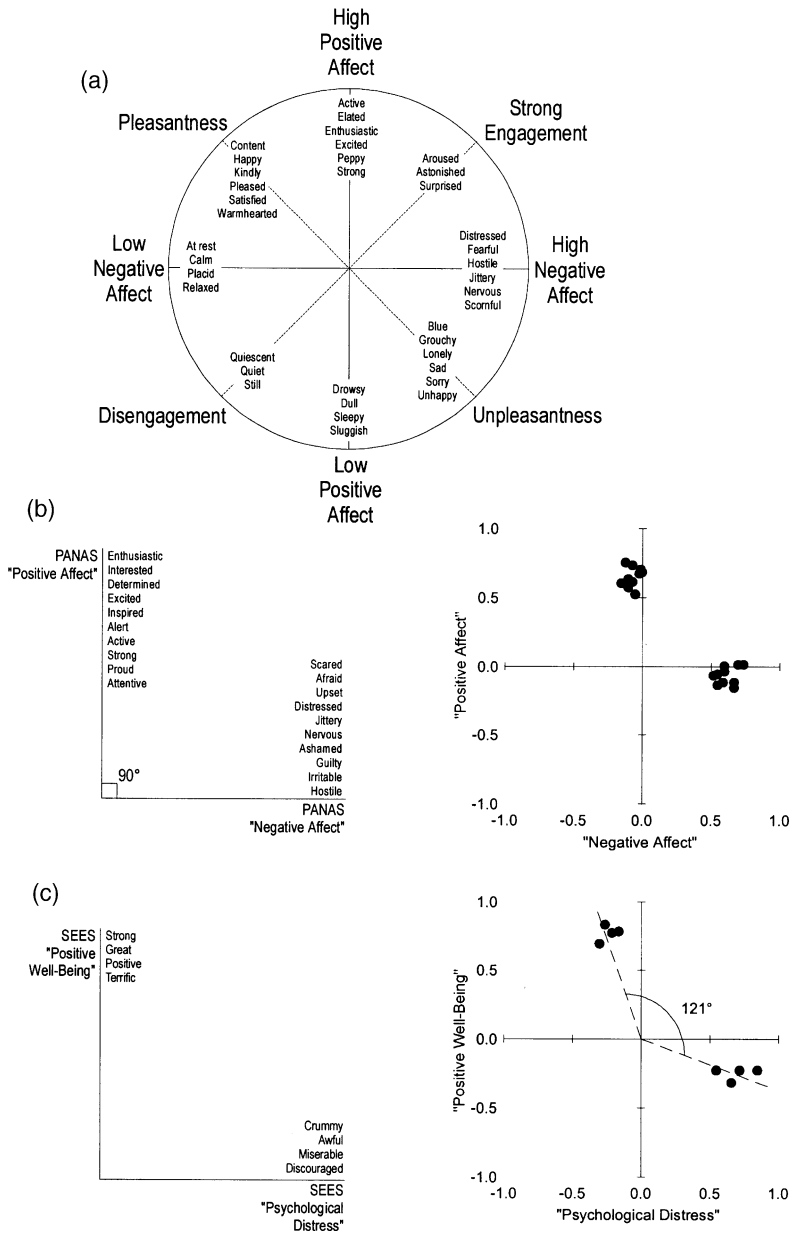

Fig. 1. Panel a: Tellegen and Watson's positive affect — negative affect model. Panel b: The hypothesized structure of the PANAS (left) and its factor plot (right). Panel c: The hypothesized structure of the SEES (left) and its factor plot (right).

and NA dimensions, but rather map onto either end of the bipolar 'Pleasantness–Unpleasantness' dimension.

It is important to emphasize that, since 1985, Watson and Tellegen had cautioned that not all positively valenced states are representatives of the PA dimension and not all negatively valenced states are representatives of the NA dimension. Therefore, the thesis of orthogonality (independence) does not apply to all positively and negatively valenced states. States such as 'happy' and 'sad' are at opposite ends of a bipolar dimension. Only states such as 'enthusiastic' (i.e. pleasant high-activation states) and 'distressed' (unpleasant high-activation states) are located on orthogonal axes. More recently, Watson and Tellegen (1999) reissued the same caveat:

> The terms 'positive affect' and 'negative affect' have been used inconsistently by different writers. In early studies of self-rated affect, researchers tended to use the terms indiscriminately to refer to any positively and negatively valenced feeling states (for a discussion, see Watson and Tellegen, 1985). This produced widespread confusion in the literature, because… different types of mood descriptors actually show substantially different intercorrelations… Markers of positive affect and negative affect should consistently show weak negative correlations, whereas terms reflecting pleasantness and unpleasantness should tend to be strongly negatively correlated (and, hence, define a single bipolar dimension). In recent years, researchers have increasingly ignored these conceptual/terminological distinctions and have reverted to using the terms 'positive affect' and 'negative affect' indiscriminately… Now, we again face the extremely confusing situation that researchers may report low, moderate — even strong — negative correlations between measures of positive and negative affect because of substantial differences in the descriptors used to create the scales. The literature is so confused at this point that the terms 'positive affect' and 'negative affect' perhaps should indeed be used only as inclusive terms referring to any positive and negative feeling states (pp. 602–603).

If one ignores the fundamental "conceptual/terminological distinctions" noted by Tellegen and Watson, the statement "Positive Affect and Negative Affect are independent" (notice the capital letters in the labels, referring to Tellegen and Watson's bipolar dimensions) can easily be misconstrued as being equivalent to the statement "positive and negative affect are independent" (notice the lower case letters, referring to generic pleasure and displeasure, respectively). This was enough to open the floodgates of confusion. Researchers in social psychology have engaged in large-scale efforts to find substantive explanations for why someone would feel both happy and sad at the same time (Cacioppo & Berntson, 1994; Cacioppo, Gardner, & Berntson, 1997; Goldstein & Strube, 1994). In all such efforts, the work of Tellegen and Watson is consistently, albeit erroneously, cited as evidence in support of the concept of independence between 'positive affect' and 'negative affect'.

Several researchers have repeatedly cautioned about the serious ramifications of the widespread confusion (Carroll et al., 1999; Egloff, 1998; Feldman-Barrett & Russell, 1998; Feldman-Barrett & Russell, 1999; Green & Salovey, 1999; Green et al. 1993, 1999; Larsen & Diener, 1992; Mossholder, Kemery, Harris, Armenakis, & McGrath, 1994; Nemanick & Munz, 1994; Russell & Carroll, 1999a, 1999b). Feldman-Barrett and Russell (1998) pleaded for a change in the PA and NA labels: "Tellegen and his colleagues could bring much clarity to this area of research, especially to all the research they have inspired, by renaming PA something like 'Surgency' or

'Pleasant Activated Affect' and renaming NA something like 'Upset' or 'Distress' or 'Unpleasant Activated Affect'" (p. 980). As Feldman-Barrett and Russell put it, "when the literal meaning of a term must be denied in explaining its meaning, it may be time for a new term… Names are arbitrary, but what mischief can they create!" (p. 980). Responding to this plea, Tellegen and Watson (Tellegen et al., 1999a,b; Watson & Tellegen, 1999; Watson et al., 1999) recently renamed the dimensions of PA and NA from 'Positive Affect' and 'Negative Affect' to 'Positive Activation' and 'Negative Activation', respectively.

If pleasure and displeasure are polar opposites, where is the evidence?

Larsen and Diener (1992) noted that "virtually nobody has found that positive and negative affect [pleasure and displeasure, not PA and NA] are completely independent, except in those cases where they are made to be independent because a two-factor solution with a varimax rotation is requested in a factor analysis program" (p. 50). Yet, at the same time, measures of positively and negatively valenced affective states have seldom been shown to have near-perfect negative correlations and, consequently, strong evidence of bipolarity has been elusive.

When a sample of affective adjectives is subjected to factor analysis and a two-factor solution is requested followed by a varimax rotation, the emergent solution more often than not yields separate factors for positively and negatively valenced adjectives (e.g. Diener & Emmons, 1984). This has been demonstrated in numerous studies, including some in the context of exercise (e.g. Choi & Salmon, 1995; Choi, Van Horn, Picker, & Roberts, 1993; Morris & Salmon, 1994). Interestingly, the distinction between positively and negatively valenced adjectives prevails even when the items have been specifically selected to reflect some other important theoretical distinction, such as that between state and trait anxiety or between anxiety and depression or between anxiety, depression, and hostility (e.g. Bernstein & Eveland, 1982; Gotlib & Meyer, 1986; Mook, Kleijn, & van der Ploeg, 1991; Mook, van der Ploeg, & Kleijn, 1992; Spielberger, Vagg, Barker, Donham, & Westberry, 1980; Vagg, Spielberger, & O'Hearn, 1980; van der Ploeg, 1989). Furthermore, between positively and negatively worded items considered to tap the same construct, the positively worded items consistently yield higher average scores compared to the negatively worded ones (e.g. Watson, 1988b).

A common interpretation of these findings, of course, is that they constitute evidence for the independence between positive and negative affect. Several authors, however, have pointed out that, to a certain extent, these findings might have been influenced by several measurement-related factors. These include random measurement error, acquiescent response sets, the use of improper response scales, item-intensity specificity (differential sensitivity of items for low vs high levels of the measured construct), social desirability (affecting negatively worded items to a larger extent), and the use of inappropriate methods of structural analysis (Carroll et al., 1999; Chen, Dai, Spector, & Jex, 1997; Feldman-Barrett, 1996; Fisher, 1997; Fisher, Heise, Bohrnstedt, & Lucke, 1985; Green et al. 1993, 1999; Lorr, 1989; Russell & Carroll, 1999a; Schmitt & Stults, 1985; van Schuur & Kiers, 1994; van Schuur & Kruijtbosch, 1995; Warr, Barter, & Brownbridge, 1983). We will now examine some of these issues in more detail and later return to them to discuss how they may have influenced the results of the exploratory factor analysis that determined the structure of the SEES.

A first issue, which we have already mentioned, refers to the role played by the location or sector of affective space from which items are sampled. Depending on the position of the items

on the dimensions of affective valence and activation, the resultant structure could be unidimensional and bipolar, two-dimensional and orthogonal, or anywhere in between (Carroll et al., 1999; Watson, 1988a; Watson & Tellegen, 1985, 1999; also see Fig. 1, panel 'a'). Items that reflect generic pleasure and displeasure, such as happy and sad, are likely to form a single bipolar factor. On the other hand, items that reflect high-activation pleasant (e.g. excited) and high-activation unpleasant affect (e.g. distressed) are likely to form independent and, more specifically, orthogonal factors. The same is true for items that reflect low-activation pleasant (e.g. calm) and low-activation unpleasant affect (e.g. sluggish).

Second, some authors have provided evidence that independence might emerge as the time frame of the response increases and the temporal relationship to an emotional stimulus loosens (Diener & Emmons, 1984). This is because, in such cases, the respondent is asked to reflect upon periods of time that might, in fact, have included both pleasant and unpleasant affective experiences. On the contrary, at any single point in time (i.e. as is the case with the acute exercise paradigm, in which respondents are typically asked how they feel 'right now'), pleasure and displeasure exhibit substantial negative correlations. As one example, Diener and Emmons (1984) reported that for within-subject analyses of daily reports (having eliminated between-subject variability), correlations between positive and negative affect were −0.10 over a three-week period, −0.31 over the course of a single day, −0.57 for momentary ratings, and −0.85 during moments during which the participants were experiencing some form of affective excitation.

Third, independence might emerge when respondents find themselves in an affectively neutral, mundane, or ambiguous situation. On the other hand, bipolarity emerges when, as in the findings of Diener and Emmons (1984) above, respondents are experiencing an affective episode, be it positive or negative. After examining frequency counts of positive and negative affect ratings made over the course of six weeks, Diener and Iran-Nejad (1986) concluded that "people do not simultaneously experience both positive and negative affect at intense levels" (p. 1036). A similar finding was reported by Folkman and Lazarus (1985). They found that the correlation between positive affect (*exhilarated*, *pleased*, *happy*, *relieved*) and negative affect (*angry*, *sad*, *disappointed*, *guilty*, *disgusted*) became increasingly more negative as a situation turned from highly ambiguous to highly unambiguous. Specifically, the correlation was 0.08 before an academic examination, −0.25 after the examination but before the grades were announced, and −0.50 after the grades were announced. Zautra and coworkers (Potter, Zautra, & Reich, 2000; Zautra, Potter, & Reich, 1997; Zautra, Reich, Davis, Potter, & Nicolson, 2000) have recently presented additional empirical evidence for this effect.

Fourth, a series of studies have examined the effect of random and systematic measurement error on the magnitude of the negative correlation between positive and negative affect (Bentler, 1969; Feldman-Barrett & Russell, 1998; Green & Salovey, 1999; Green et al. 1993, 1999; Lorr, McNair, & Fisher, 1982; Lorr, Shi, & Youniss, 1989; Lorr & Wunderlich, 1988; Meddis, 1972; Russell, 1979). According to classical test theory (Green et al., 1993; Muchinsky, 1996; Nunnally & Bernstein, 1994), random measurement error will have an attenuating effect on the magnitude of correlations (i.e. will bias correlation coefficients toward zero) and systematic measurement error may even produce correlations that have the incorrect sign. In the aforementioned studies, estimating and removing the effects of measurement error was consistently shown to produce stronger negative correlations between positive and negative affect compared to correlations based on raw data. For example, in one case reported by Green et al. (1993), the correlation

between raw scores on happiness and sadness was −0.25, but after controlling for the attenuating effect of random measurement error (unreliability) the correlation was raised to −0.85 and, after controlling for the effect of non-random error due to the response format used (adjective check list), the correlation was raised to −0.84. In light of similar findings, researchers who had initially interpreted their results in terms of unipolar factors (e.g. McNair & Lorr, 1964; Thayer, 1967) later came to support bipolarity (Lorr, 1989; Thayer 1978, 1986).

Fifth, increasing attention is being directed toward the role played by response formats in conjunction with correlation-based models of statistical analysis, including factor analysis, in masking bipolarity. Meddis (1972) was one of the first to demonstrate that the type of response format affected the composition of factors in factor analysis. He showed that a commonly used asymmetric response format (definitely feel, slightly feel, cannot decide, do not feel) had an attenuating effect on the correlations between positive and negative affect items and, as a result, led to the formation of independent positive and negative affect factors. In contrast, the use of a symmetric response format (definitely feel, slightly feel, do not feel, definitely do not feel) for the same items produced bipolar factors. These findings were later replicated and extended by Svensson (1977). Diener and Iran-Nejad (1986) were instrumental in shedding some much needed light into the causes of this problem. They showed that the relationship between positive and negative affect was not linear negative, but rather formed an L-shape. For example, when one is very *happy* (e.g. six on a six-point rating scale) he or she is not *sad* (i.e. 0 on the rating scale). When the same person is moderately *happy* (e.g. three on the rating scale), he or she is not moderately *sad* (i.e. three), but rather, again, not *sad* (i.e. 0). In other words, bipolarity should not be taken to necessarily imply a negative linear relationship between positive and negative affect, as has been traditionally assumed, but rather a relationship of mutual exclusion. Diener and Iran-Nejad found that, although the linear component of the relationship was significant, so was the quadratic component. Therefore, although L-shaped data may produce significant and moderately strong Pearson product-moment correlation coefficients, the non-linear form of the relationship makes this kind of analysis inappropriate and one that cannot produce a perfect −1 correlation, traditionally presumed to be the 'true test' of bipolarity. The implications of this phenomenon were explored further by Russell and Carroll (1999a) and Schimmack (2001). Russell and Carroll provided evidence that the maximum theoretic correlation between responses to two items that are on the opposite ends of a bipolar dimension, if assessed via unipolar response scales (i.e. ranging from *slightly* to *extremely* or even from *not at all* to *very much*), is not −1, but rather −0.467 (see the paper for the mathematical derivation). For two items to have a perfect negative correlation, the information that they provide should be completely redundant and this is clearly impossible with unipolar response scales. More importantly, however, due to the non-linear form of the relationship, the correlation coefficient and other statistics based on correlations, including factor analysis, are rendered inappropriate methods for investigating bipolarity.

Sixth, the appropriateness of the factor analytic model itself as a means of analyzing data that represent a single bipolar dimension has come under some additional scrutiny. Sjöberg, Svensson, and Persson (1979) had initially pointed out that different items represent the underlying trait to a different extent. For example, not all items assumed to express displeasure reflect an equal degree of unhappiness or sadness. Although factor analysis assumes that values on the observed variables (item scores) are linearly related to values on the underlying trait, this will hold true for only those items that reflect the extreme end of the latent trait. In contrast, values on intermedi-

ate items will relate to the latent trait in a curvilinear (quadratic) fashion. The implications of this earlier observation were articulated more recently by van Schuur and Kiers (1994) and van Schuur and Kruijtbosch (1995). These authors provided both mathematical proof and empirical evidence that, because of this problem, when factor analysis is used to examine the structure of data that represent a single bipolar dimension, factor analysis produces not one bipolar, but rather two unipolar factors. The appropriate analytic model, according to van Schuur and Kiers and van Schuur and Kruijtbosch, is the unidimensional unfolding model, a distance model that takes into account not only the position of the subjects, but also the position of the items along the latent trait.

Recapitulation

As noted earlier, McAuley and Courneya (1994) stated that one of the assumptions on which the development of the SEES was based was that "emotional or affective responses vary along two (positive and negative)… dimensions" (p. 165) and, eventually, the Positive Well-Being and Psychological Distress scales were developed as independent components. Although McAuley and Courneya were correct in pointing out that the idea that positive and negative affect are independent dimensions rather than polar opposites is widely popular in social psychology, we have reviewed evidence that this belief may be the result of a misunderstanding, exacerbated by the long disregard of a number of important factors that may influence the results of structural analyses. *The important point is that, by all accounts, the general statement that positive and negative affect are independent dimensions is incorrect*. In this review, we examined evidence that: (a) the nature of the items (their location along the dimensions of pleasure–displeasure and low-high activation); (b) the time frame of the responses (immediate versus long-term); (c) the context of the assessment (affectively charged versus mundane); (d) the (un)reliability of measurement; (e) the format of the response scales (balance, unipolarity); and (f) the method of structural analysis are all factors that have been shown to influence the results of structural analyses. Regardless of the impact of these factors, however, there is consensus that, when the immediate responses to an affectively charged stimulus are assessed, pleasure and displeasure are polar opposites, not independent dimensions. In an extensive study of the relationship between positive and negative responses to affective slides, the vast majority of participants (n=239) "consistently showed a reciprocal relation in their positivity and negativity scores", whereas very few (n=8) "showed a consistently uncoupled relation between the valent systems" (Ito, Cacioppo, & Lang, 1998, p. 876).

Because the scales that are presumed to tap positive affect (i.e. PWB) and negative affect (i.e. PD) are two of the three scales of the SEES, the relationship between positive and negative affect is a critical issue that impacts the scale as a whole. As we noted, despite the statement that "emotional or affective responses vary along two (positive and negative)… dimensions" (p. 165), McAuley and Courneya's (1994) conceptual position on this issue was obfuscated by several other, apparently conflicting, statements. Furthermore, an examination of the methodological steps that were followed in the development of the SEES does not reveal a systematic effort to form independent positive and negative affect components and, therefore, the formation of independent PWB and PD scales appears to be the product of induction, not deduction. The problem with taking an inductive approach in this situation is that, if one does not take into consideration the points that we reviewed here, it is almost impossible to escape the inadvertent formation of independent positive and negative affect factors. Inattention to the elements that we identified will

consistently sway the results in one direction, namely the formation of independent factors. As Russell and Carroll (1999a) put it, "finding a substantial negative correlation (and, thus, a bipolar factor) is an uphill struggle, finding a weaker correlation (and, thus, two independent factors) is a downhill run" (p. 18). Not taking a proactive stance by anticipating the pitfalls that have been highlighted in the literature is almost tantamount to ensuring that an exploratory factor analysis will produce independent factors. According to Russell and Carroll:

> In an uncountable number of studies, affect items were incidentally administered to the participants, and those items were then submitted to exploratory factor analysis with varimax rotation. The result — one factor labeled *positive affect* and another labeled *negative affect* — is so commonplace nowadays as to warrant little more than passing mention… This method cannot be decisive (p. 16).

In this review, we intentionally covered material that was available long before the publication of the SEES, as well as more recently published material. This was done in the name of fairness. Virtually every factor that can reduce the likelihood of finding bipolarity and, instead, increase the likelihood of independent factors had been discussed in the literature before the development of the SEES. In that sense, it could be argued that the measure would have benefited substantially from a thorough and systematic examination of the relevant theoretical and methodological literature. It is also true, however, that many of these issues have come into much sharper focus since then as a result of the increased attention toward affect in general psychology. Either way, there is sufficient evidence to suggest that the independence of positive and negative affect can no longer be regarded as a given simply as a function of the popularity of this idea in social psychology.

Methodological implementation of conceptual postulates

Item selection and content validation

Content validation should reflect the scale developers' fundamental theoretical postulates regarding the content domain and the structure of the constructs of interest (Carmines & Zeller, 1979; Clark & Watson, 1995; Crocker & Algina, 1986; DeVellis, 1991; Haynes, Richard, & Kubany, 1995; Loevinger, 1957). An analysis of the introduction of the SEES publication reveals two such postulates. First, McAuley and Courneya (1994) posited that the subjective experiences associated with exercise vary along at least two dimensions: a *positive* and a *negative* dimension and, perhaps, an additional dimension associated with physical exertion. Second, McAuley and Courneya sought to develop a measure of "subjective experiences that are unique to the exercise domain" (p. 165).

Based on the fact that the development of the SEES relied on *two* main conceptual premises, one would expect that the item selection and content validation criteria that were used would address both. This was not the case, however. Only the latter of these two premises was actually used as an item selection and content validation criterion. Specifically, the purpose of item selection was to identify and exclude those items that were not "likely to be influenced, either positively

or negatively, by exercise participation" (McAuley & Courneya, 1994, p. 166). Thus, the former conceptual premise (i.e. identifying those items that would best reflect a positive and an independent negative dimension, as well as a dimension of physical exertion) was not taken into account. The distinctions between the positive, the negative, and the physical exertion dimensions were left to emerge from a subsequent exploratory factor analysis. Therefore, the item selection methodology did not fully reflect the nature and structure of the content domain of interest as that was initially described.

As we have already discussed, when adjectives denoting affect are sampled indiscriminately (i.e. without taking into consideration their location along the dimensions of affective valence and activation) and are later subjected to some form of factor analysis with a varimax rotation, the primary distinction that will emerge will almost invariably be that between the adjectives denoting positive and those denoting negative affective valence. As we discussed earlier, however, numerous investigations have shown that the formation of separate positive and negative factors might be due to a host of inadvertently introduced measurement and analytic biases rather than due to the actual independence between positive and negative affect. Therefore, to avoid misleading findings, it is important that researchers be aware of these biases and attempt to prevent their occurrence or apply appropriate corrections to offset their effects.

Exploratory factor analysis

The initial item pool of 367 items was reduced to 46 items, for which there was at least 86% agreement between expert judges regarding their presumed 'relevance' to the context of exercise. The remaining 46 items were then administered to a sample of undergraduate students enrolled in physical activity classes. The students were asked to indicate whether, in their experience, exercise participation leads to increases, decreases, or no change in the affective state described by each item. Their responses were then subjected to a principal axis factor analysis followed by a varimax rotation. The theoretical and practical problems associated with these judgements of 'relevance' and 'irrelevance' of items to exercise have already been discussed in a previous section and these criticisms will not been repeated here.

Based on the experience that has been accumulated from the long debate over the issue of the independence versus bipolarity of positive and negative affect, a number of observations can be made regarding this factor analysis. First, no attempt was made to sample items from specific sectors of the affective space. In other words, the location of the items on the dimensions of affective valence and activation was not taken into account. As has been argued by both those authors who support bipolarity (e.g. Carroll et al., 1999) and those in favor of independence (e.g. Watson, 1988a; Watson & Tellegen, 1985, 1999), the indiscriminate sampling of items is one of the primary reasons for confusion and conflicting findings as this practice can result in factors that are bipolar, unipolar, or anywhere in between. In the case of the SEES, the factors appear to be somewhere in-between. This is illustrated in Fig. 1, panels 'b' and 'c', where the structure of the SEES is compared to that of the PANAS. In the development of the PANAS, the items that formed the PA and NA scales were specifically chosen to be on orthogonal vectors (90° separation) and this resulted in PA–NA correlations that are consistently near zero. This was done by selecting primarily items that reflect high levels of activation for both the PA and the NA scales (e.g. enthusiastic, excited, active, scared, nervous, etc.), consistent with the results of Tellegen and

his associates' earlier structural analyses (Tellegen, 1985; Watson & Tellegen, 1985; Zevon & Tellegen, 1982; also see Fig. 1 panel 'a'). In the development of the SEES, the content of the items in terms of their valence and activation value was not taken into account in selecting them. As can be seen from the factor plot in panel 'c', the PWB and PD items fall between Tellegen and Watson's PA–NA dimensions and tend to approximate the bipolar Pleasantness–Unpleasantness dimension (compare to Fig. 1, panel 'a'). Not surprisingly, the PWB–PD correlation was −0.52, far from the near-zero values typically found between the PA and NA scales of the PANAS.

Second, in McAuley and Courneya's (1994) factor analysis, each of the hypothesized factors was represented by an unequal number of items. This imbalance has been identified as one of the major sources of bias in factor analyses leading to separate positive and negative factors. According to Russell (1979), "to the extent that the sample of emotion words studied underrepresents one end of a bipolar continuum, bipolar factors are less likely to emerge" (p. 347). McAuley and Courneya did not report the exact balance of items in their factor analysis, but judging from the unbalanced numbers of positive (15) and negative (4) items that satisfied the criteria for retention, it is reasonable to assume that the factors were unequally represented in the first place.

Third, as explained earlier, it is well known that long time frames of responses diminish the negative correlation between positive and negative affect (Diener & Emmons, 1984; Diener & Iran-Nejad, 1986) and facilitate the formation of independent factors. According to Russell (1989), "if the instructions ask subjects how they felt over an extended period of time (such as over a week), they may describe several, perhaps opposite, emotional experiences" (p. 347). From McAuley and Courneya's (1994) description, it appears that the responses that were used in the factor analysis reflected the respondents' overall experiences with exercise. As such, consistent with Russell's observation, they may very well have included several, perhaps both positive and negative, experiences.

Fourth, the response scale that accompanied the items used in the factor analysis was rather unique (a bipolar scale, asking respondents to indicate whether, in their experience, the affective state represented by each item increases or decreases with exercise), so its role in influencing the results is difficult to contemplate. However, the issue raised by van Schuur and Kiers (1994) and van Schuur and Kruijtbosch (1995) regarding the appropriateness of the factor analytic model for examining bipolar data remains. It is also noteworthy that the format of the response scale that was used in the factor analytic study (i.e. a bipolar scale asking whether each state increases or decreases with exercise) is different from the format of the response scale that accompanies the items in the final version of the SEES (i.e. a unipolar scale of intensity, ranging from 'not at all' to 'very much so'). The possible effects of this discrepancy are open to investigation but, given previous reports in the literature, it is possible that the response format may have been partly responsible for the outcome of the factor analysis.

Fifth, one must question the decision to use only an orthogonal rotation, especially given the substantial (r=−0.52) correlation reported between the PWB and PD scales. To repeat the excerpt from Larsen and Diener (1992), "virtually nobody has found that positive and negative affect are completely independent, except in those cases where they are made to be independent because a two-factor solution with a varimax rotation is requested in a factor analysis program" (p. 50). An examination of the factor plot of the PWB and PD factors in Fig. 1, panel 'c' indicates that there is a visible deviation from orthogonality.

In sum, in conjunction with the item selection procedures, the factor analysis did not avoid

several elements which, based on the findings of extensive previous research, may have decreased the likelihood of bipolarity and, instead, may have increased the likelihood of independent positive affect and negative affect factors. A couple of additional ambiguities in McAuley and Courneya's (1994) factor analysis are noteworthy but of lesser theoretical and practical significance. First, it is unclear whether the factor loadings reported (p. 167) are based on the analysis of only the 12 items shown (i.e. the ones that were eventually retained) or on the analysis of all 46 items that were initially factor-analyzed or, finally, on the 23 items that satisfied the criteria for retention. This is confusing because the solution presented by McAuley and Courneya accounts for 63.6% of the variance (by calculating and summing the communalities), whereas the percentages of variance explained by each factor as reported by McAuley and Courneya add up to 65.6% of the variance. Second, no information is provided regarding the criteria that were used to retain three factors. The omitted information makes it difficult to evaluate and replicate the factor analysis.

Confirmatory factor analysis

Like many other reports on structural equation modeling (see Hoyle & Panter, 1995; Raykov, Tomer, & Nesselroade, 1991), McAuley and Courneya (1994) provided little information on the confirmatory factor analysis of the SEES. No data were reported on the viability of the assumption of multivariate normality, no alternative models were considered and compared, and the sources of bad fit were not discussed. More importantly, however, although the latent variables were apparently allowed to correlate freely, the correlations were not reported. Reporting these correlations would have provided some information regarding the relationship between positive and negative affect as reflected in the SEES. Because latent variables are assumed to be error-free, the correlation between positive and negative affect factors in structural equation models has been compared to the correlation derived from raw data in order to get an estimate of the attenuating effect of random measurement error on the observed correlation between positive and negative affect (e.g. Feldman-Barrett & Russell, 1998; Green et al., 1993).

Concluding comments

From a purely technical-methodological standpoint, the procedures that were followed in the development of the SEES are, with perhaps a few minor exceptions, in line with established guidelines and accepted conventions. It could be argued that, had this been a measure in a truly novel and unexplored area, with no previous theory and empirical research to serve as a guide, these procedures would have been a reasonable first exploratory step. However, the domain of affect is not a novel and unexplored area. On the contrary, over a century of accumulated experience makes this one of the richest areas of psychological investigation. Although controversy and confusion abound, this experience can only be seen as an invaluable resource. As our review demonstrated, even before the 1990s, research on the structure of affect had made considerable progress, uncovering the fundamentals and alerting of impending obstacles.

We agree with McAuley and Courneya (1994) on an important point, namely that a dimensional measure would be preferable to a categorical measure for the majority of the studies conducted at the present stage of knowledge development. Beyond this point, however, the promise of the

SEES as a dimensional measure was not fully realized. The primary reason for this is that its conceptual basis did not incorporate the knowledge on the structure of the affective domain that previous research had uncovered. Instead, the development of the SEES followed a clearly inductive approach with no apparent guidance by theoretical postulates (the only exception being the dubious process of selecting 'exercise-relevant' items). The exact nature and the limits of the content domain of the scale were unclear and the concept of 'subjective experiences unique to exercise' appears to have been formulated without having fully contemplated the theoretical and practical problems that it raises. Furthermore, a major part of the present analysis focused on the important issue of the relationship between positive and negative affect. As it became evident from the review of an extensive literature, this has been the subject of a long and often heated debate in affective psychology and psychometrics. Although the answers to several questions remain elusive, our review was aimed to demonstrate that there is emerging consensus on a number of points, including two that are of particular relevance to the SEES. First, there is agreement that positive and negative affect are not necessarily independent dimensions and, if by those terms one refers to generic pleasure and displeasure, these are, in fact, polar opposites. Second, there is agreement that a number of factors in the data collection and analytical methodology can substantially influence the relationship between positive and negative affect, swaying the results away from bipolarity and toward independence. As our examination of McAuley and Courneya's (1994) exploratory factor analysis showed, it is possible, on the basis of previous research, that several of the methodological and analytical choices involved might have contributed to the emergence of PWB and PD as separate factors. Nevertheless, it is also interesting that, as a result of not taking into consideration the location of the items along the dimensions of affective valence and activation, these factors had a substantial negative correlation ($r=-0.52$).

To illustrate the degree to which measurement factors can influence correlations and, thus, partly conceal the nature of the relationships between the underlying constructs, it is interesting to examine what the correlation between PWB and PD would be after taking into account only one such factor, namely random measurement error (unreliability). This can be done by using the standard formula for correction of attenuation due to unreliability (Muchinsky, 1996; Nunnally & Bernstein, 1994) in conjunction with the alpha coefficients of internal consistency of the two scales. In interpreting the results of these calculations, readers should be aware that the formula "really only estimates how high the correlation would be if the two variables were made perfectly reliable rather than a true correction (these estimates can exceed 1!)" (Nunnally & Bernstein, 1994, p. 241). Data were obtained from four studies where all the necessary information was reported (Lox & Rudolph, 1994; Markland et al., 1997; McAuley & Courneya, 1994; Rudolph & Kim, 1996). The direction and magnitude of the correlations reported in these studies are consistent with those from other studies where the alpha coefficients were not reported (e.g. McAuley, Shaffer, & Rudolph, 1995). The results are shown in Table 1. The correlations are also shown in terms of the angle of separation between the hypothetical PWB and PD vectors (the angle of separation is given by the inverse cosine of the value of the correlation coefficient). This angle should be 90° if the scales are orthogonal ($r=0.00$) or 180° if the scales are perfect polar opposites ($r=-1.00$). The first observation is that PWB and PD raw scores consistently show substantial negative correlations. The second observation is that, after taking into account the estimated effects of random measurement error, the negative relationship is raised to a magnitude that makes bipolarity a viable possibility.

Table 1
Correlations between PWB and PD and corresponding angles of separation, uncorrected and corrected for attenuation due to unreliability (random measurement error)

Study	Correlation	Angle of separation (°)	PWB alpha	PD alpha	Corrected correlation	Corrected angle of separation (°)
McAuley and Courneya (1994)	−0.52	121	0.86	0.85	−0.61	128
Lox and Rudolph (1994)	−0.48 to −0.58	119–125	0.86	0.84	−0.57 to −0.68	125–133
Rudolph and Kim (1996)	−0.56 to −0.62	124–128	0.89	0.87	−0.64 to −0.70	130–134
Markland et al. (1997)	−0.36 to −0.74	111–138	0.79	0.71	−0.48 to −0.99	119–172

McAuley and Courneya's interpretation of the negative correlation found between PWB and PD was the following:

> …it should be expected that the PWB and PD dimensions would be inversely and moderately correlated. We are, after all, attempting to measure subjective responses to a particular stimulus (exercise) about which there certainly can be both positive and negative feelings. Developing such measures to assess responses to a particular domain is likely to produce related factors (p. 172).

Of course, if there were 'both positive and negative feelings', the relationship between the factors would have been positive, not negative, or, as was originally hypothesized, the two factors would be independent. What the negative correlation between PWB and PD indicates instead is that people's affective responses to acute exercise tend to be *either positive or negative*. Furthermore, as was explained in our critique of the conceptual underpinnings of the SEES, the relationship between the PWB and PD scores is not necessarily something that 'should be expected'. Instead, as research has demonstrated, this relationship is not static and, to a large extent, can be strengthened or weakened by certain methodological and data-analytic choices.

McAuley and Rudolph (1995) commented that the SEES "may not be the definitive solution to measuring psychological well-being responses resulting from exercise and physical activity stimuli", but that it "should prompt researchers to give greater consideration to this issue" (p. 90). We echo this sentiment and hope that the clarifications we offered in the present analysis will provide the impetus for further progress in this important area of research.

References

Bass, B. M. (1974). The substance and the shadow. *American Psychologist, 29*, 870–886.
Batson, C. D., Shaw, L. L., & Oleson, K. C. (1992). Differentiating affect, mood, and emotion: Toward functionally

based conceptual distinctions. In M. S. Clark (Ed.), *Review of personality and social psychology* (Vol. 13, pp. 294–326). Newbury Park, CA: Sage.

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

Bernstein, I. H., & Eveland, D. C. (1982). State vs trait anxiety: A case study in confirmatory factor analysis. *Personality and Individual Differences, 3*, 361–372.

Blanchard, C. M., Rogers, W. M., Spence, J. C., & Courneya, K. S. (2001). Feeling state responses to acute exercise of high and low intensity. *Journal of Science and Medicine in Sport, 4*, 30–38.

Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401–423.

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review, 1*, 3–25.

Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment.* Newbury Park, CA: Sage.

Carroll, J. M., Yik, M. S. M., Russell, J. A., & Feldman-Barrett, L. (1999). On the psychometric principles of affect. *Review of General Psychology, 3*, 14–22.

Chen, P. Y., Dai, T., Spector, P. E., & Jex, S. M. (1997). Relation between negative affectivity and positive affectivity: Effects of judged desirability of scale items and respondents' social desirability. *Journal of Personality Assessment, 69*, 183–198.

Choi, P. Y. L., & Salmon, P. (1995). Symptom changes across the menstrual cycle in competitive sportswomen, exercisers and sedentary women. *British Journal of Clinical Psychology, 34*, 447–460.

Choi, P. Y. L., Van Horn, J. D., Picker, D. E., & Roberts, H. I. (1993). Mood changes in women after an aerobics class: A preliminary study. *Health Care for Women International, 14*, 167–177.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*, 309–319.

Clore, G. L., Ortony, A., & Foss, M. A. (1987). The psychological foundations of the affective lexicon. *Journal of Personality and Social Psychology, 53*, 751–766.

Cole, D. A. (1987). Utility of confirmatory factor analysis in test validation research. *Journal of Consulting and Clinical Psychology, 55*, 584–594.

Crocker, L., & Algina, J, (1986). *Introduction to classical and modern test theory.* Fort Worth, TX: Harcourt Brace Jovanovich.

DeVellis, R. (1991). *Scale development: Theory and applications.* Newbury Park, CA: Sage.

Diener, E., & Emmons, R. A. (1984). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47*, 1105–1117.

Diener, E., & Iran-Nejad, A. (1986). The relationship in experience between various types of affect. *Journal of Personality and Social Psychology, 50*, 1031–1038.

Egloff, B. (1998). The independence of positive and negative affect depends on the affect measure. *Personality and Individual Differences, 25*, 1101–1109.

Ekkekakis, P., Hall, E. E., & Petruzzello, S. J. (1999). Measuring state anxiety in the context of acute exercise using the State Anxiety Inventory: An attempt to resolve the brouhaha. *Journal of Sport & Exercise Psychology, 21*, 205–229.

Ekkekakis, P., & Petruzzello, S. J. (1999). Acute aerobic exercise and affect: Current status, problems, and prospects regarding dose-response. *Sports Medicine, 28*, 337–374.

Ekkekakis, P., & Petruzzello, S. J. (2000). Analysis of the affect measurement conundrum in exercise psychology: I. Fundamental issues. *Psychology of Sport and Exercise, 1*, 71–88.

Ekkekakis, P., & Petruzzello, S. J. (2001). Analysis of the affect measurement conundrum in exercise psychology: II. A conceptual and methodological critique of the Exercise-induced Feeling Inventory. *Psychology of Sport and Exercise, 2*, 1–26.

Feldman-Barrett, L. (1996). Hedonic tone, perceived arousal, and item desirability: Three components of self-reported mood. *Cognition and Emotion, 10*, 47–68.

Feldman-Barrett, L., & Russell, J. A. (1998). Independence and bipolarity in the structure of current affect. *Journal of Personality and Social Psychology, 74*, 967–984.

Feldman-Barrett, L., & Russell, J. A. (1999). The structure of current affect: Controversies and emerging consensus. *Current Directions in Psychological Science, 8*, 10–14.

Fisher, G. A. (1997). Theoretical and methodological elaborations of the circumplex model of personality traits and emotions. In R. Plutchik, & H. R. Conte (Eds.), *Circumplex models of personality and emotions* (pp. 245–269). Washington, DC: American Psychological Association.

Fisher, G. A., Heise, D. R., Bohrnstedt, G. W., & Lucke, J. F. (1985). Evidence for extending the circumplex model of personality trait language to self-reported moods. *Journal of Personality and Social Psychology*, *49*, 233–242.

Folkman, S., & Lazarus, R. S. (1985). If it changes, it must be a process: Study of emotion and coping during three stages of a college examination. *Journal of Personality and Social Psychology*, *48*, 150–170.

Frijda, N. H. (1993). Moods, emotion episodes, and emotions. In M. Lewis, & J. M. Haviland, *Handbook of emotions* (pp. 381–403). New York: Guilford.

Gauvin, L., & Brawley, L. R. (1993). Alternative psychological models and methodologies for the study of exercise and affect. In P. Seraganian, *Exercise psychology: The influence of physical exercise on psychological processes* (pp. 146–171). New York: Wiley.

Gauvin, L., & Rejeski, W. J. (1993). The Exercise-induced Feeling Inventory: Development and initial validation. *Journal of Sport & Exercise Psychology*, *15*, 403–423.

Gauvin, L., & Spence, J. C. (1998). Measurement of exercise-induced changes in feeling states, affect, mood, and emotions. In J. L. Duda, *Advances in sport and exercise psychology measurement* (pp. 325–336). Morgantown, WV: Fitness Information Technology.

Goldstein, M. D., & Strube, M. J. (1994). Independence revisited: The relation between positive and negative affect in a naturalistic setting. *Personality and Social Psychology Bulletin*, *20*, 57–64.

Gotlib, I. H., & Meyer, J. P. (1986). Factor analysis of the Multiple Affect Adjective Check List: A separation of positive and negative affect. *Journal of Personality and Social Psychology*, *50*, 1161–1165.

Green, D. P., Goldman, S. L., & Salovey, P. (1993). Measurement error masks bipolarity in affect ratings. *Journal of Personality and Social Psychology*, *64*, 1029–1041.

Green, D. P., & Salovey, P. (1999). In what sense are positive and negative affect independent? A reply to Tellegen, Watson, and Clark. *Psychological Science*, *10*, 304–306.

Green, D. P., Salovey, P., & Truax, K. M. (1999). Static, dynamic, and causative bipolarity of affect. *Journal of Personality and Social Psychology*, *76*, 856–867.

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, *2*, 271–299.

Gross, J. J. (1999). Emotion regulation: Past, present, future. *Cognition and Emotion*, *13*, 551–573.

Hardy, C. J., & Rejeski, W. J. (1989). Not what, but how one feels: The measurement of affect during exercise. *Journal of Sport & Exercise Psychology*, *11*, 304–317.

Haynes, S. N., Richard, D. C. S., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*, 238–247.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle, *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage.

Ito, T. A., Cacioppo, J. T., & Lang, P. J. (1998). Eliciting affect using the International Affective Picture System: Trajectories through evaluative space. *Personality and Social Psychology Bulletin*, *24*, 855–879.

Larsen, R. J., & Diener, E. (1992). Promises and problems with the circumplex model of emotion. In M. S. Clark (Ed.), *Emotion* (pp. 25–29). *Review of personality and social psychology* (Vol. 13). Newbury Park, CA: Sage.

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

Loevinger, J. (1957). Objective tests as instruments of psychological theory. *Psychological Reports*, *3*, 635–694.

Lorr, M. (1989). Models and methods for measurement of mood. In R. Plutchik, & H. Kellerman (Eds.), *The measurement of emotion* (pp. 37–53). *Emotion: Theory, research, and experience*, (Vol. 4). San Diego, CA: Academic Press.

Lorr, M., McNair, D. M., & Fisher, S. (1982). Evidence for bipolar mood states. *Journal of Personality Assessment*, *46*, 432–436.

Lorr, M., Shi, A. Q., & Youniss, R. P. (1989). A bipolar multifactor conception of mood states. *Personality and Individual Differences*, *10*, 155–159.

Lorr, M., & Wunderlich, R. A. (1988). A semantic differential mood scale. *Journal of Clinical Psychology*, *44*, 33–39.

Lox, C. L., Jackson, S., Tuholski, S. W., Wasley, D., & Treasure, D. C. (2000). Revisiting the measurement of exercise-induced feeling states: The Physical Activity Affect Scale (PAAS). *Measurement in Physical Education and Exercise Science*, *4*, 79–95.

Lox, C. L., & Rudolph, D. L. (1994). The Subjective Exercise Experiences Scale (SEES): Factorial validity and effects of acute exercise. *Journal of Social Behavior and Personality, 9*, 837–844.

Lox, C. L., & Treasure, D. C. (2000). Changes in feeling state following aquatic exercise during pregnancy. *Journal of Applied Social Psychology, 30*, 518–527.

Markland, D., Emberton, M., & Tallon, R. (1997). Confirmatory factor analysis of the Subjective Exercise Experiences Scale among children. *Journal of Sport & Exercise Psychology, 19*, 418–433.

McAuley, E. (1994). Physical activity and psychosocial outcomes. In C. Bouchard, R. J. Shephard, & T. Stephens (Eds.), *Physical activity, fitness, and health: International proceedings and consensus statement* (pp. 551–568). Champaign, IL: Human Kinetics.

McAuley, E., & Courneya, K. S. (1994). The Subjective Exercise Experiences Scale (SEES): Development and preliminary validation. *Journal of Sport & Exercise Psychology, 16*, 163–177.

McAuley, E., & Rudolph, D. (1995). Physical activity, aging, and psychological well-being. *Journal of Aging and Physical Activity, 3*, 67–96.

McAuley, E., Shaffer, S. M., & Rudolph, D. (1995). Affective responses to acute exercise in elderly impaired males: The moderating effects of self-efficacy and age. *International Journal of Aging and Human Development, 41*, 13–27.

McNair, D. M., & Lorr, M. (1964). An analysis of mood in neurotics. *Journal of Abnormal and Social Psychology, 69*, 620–627.

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

Meddis, R. (1972). Bipolar factors in mood adjective checklists. *British Journal of Social and Clinical Psychology, 11*, 178–184.

Mook, J., Kleijn, W. C., & van der Ploeg, H. M. (1991). Symptom-positively and negatively worded items in two popular self-report inventories of anxiety and depression. *Psychological Reports, 69*, 551–560.

Mook, J., van der Ploeg, H. M., & Kleijn, W. C. (1992). Symptom-positive and symptom-negative items in the State-Trait Anxiety Inventory: A comparison and replication. *Anxiety, Stress, and Coping, 5*, 113–123.

Morris, M., & Salmon, P. (1994). Qualitative and quantitative effects of running on mood. *Journal of Sports Medicine and Physical Fitness, 34*, 284–291.

Mossholder, K. W., Kemery, E. R., Harris, S. G., Armenakis, A. A., & McGrath, R. (1994). Confounding constructs and levels of constructs in affectivity measurement: An empirical investigation. *Educational and Psychological Measurement, 54*, 336–349.

Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement, 56*, 63–75.

Nemanick, R. C., & Munz, D. C. (1994). Measuring the poles of positive and negative mood using the Positive Affect Negative Affect Schedule and Activation Deactivation Adjective Check List. *Psychological Reports, 74*, 195–199.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory.* (3rd ed.). New York: McGraw-Hill.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotion.* New York: Cambridge University Press.

Ortony, A., Clore, G. L., & Foss, M. A. (1987). The referential structure of the affective lexicon. *Cognitive Science, 11*, 341–364.

Parfitt, G., Rose, E. A., & Markland, D. (2000). The effect of prescribed and preferred intensity exercise on psychological affect and the influence of baseline measures of affect. *Journal of Health Psychology, 5*, 231–240.

Potter, P. T., Zautra, A. J., & Reich, J. W. (2000). Stressful events and information processing dispositions moderate the relationship between positive and negative affect: Implications for pain patients. *Annals of Behavioral Medicine, 22*, 191–198.

Raykov, T., Tomer, A., & Nesselroade, J. R. (1991). Reporting structural equation modeling results in Psychology and Aging: Some proposed guidelines. *Psychology and Aging, 6*, 499–503.

Rosenberg, E. L. (1998). Levels of analysis and the organization of affect. *Review of General Psychology, 2*, 247–270.

Rudolph, D. L., & Butki, B. D. (1998). Self-efficacy and affective responses to short bouts of exercise. *Journal of Applied Sport Psychology, 10*, 268–280.

Rudolph, D. L., & Kim, J. G. (1996). Mood responses to recreational sport and exercise in a Korean sample. *Journal of Social Behavior and Personality, 11*, 841–849.

Russell, J. A. (1979). Affective space is bipolar. *Journal of Personality and Social Psychology, 37*, 345–356.

Russell, J. A. (1989). Measures of emotion. In R. Plutchik, & H. Kellerman (Eds.), *The measurement of emotions* (pp. 83–111). *Emotion: Theory, research, and experience* (Vol. 4). San Diego, CA: Academic Press.

Russell, J. A. (1997). How shall an emotion be called? In R. Plutchik, & H. R. Conte (Eds.), *Circumplex models of personality and emotions* (pp. 205–220). Washington, DC: American Psychological Association.

Russell, J. A., & Carroll, J. M. (1999a). On the bipolarity of positive and negative affect. *Psychological Bulletin*, *125*, 3–30.

Russell, J. A., & Carroll, J. M. (1999b). The phoenix of bipolarity: Reply to Watson and Tellegen (1999). *Psychological Bulletin*, *125*, 611–617.

Russell, J. A., & Feldman-Barrett, L. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, *76*, 805–819.

Schimmack, U. (2001). Pleasure, displeasure, and mixed feelings: Are semantic opposites mutually exclusive? *Cognition and Emotion*, *15*, 81–97.

Schmitt, N., & Stults, D. M. (1985). Factors defined by negatively keyed items: The result of careless respondents? *Applied Psychological Measurement*, *9*, 367–373.

Shaver, P., Schwartz, J., Kirson, D., & O'Connor, C. (1987). Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology*, *52*, 1061–1086.

Sjöberg, L., Svensson, E., & Persson, L. O. (1979). The measurement of mood. *Scandinavian Journal of Psychology*, *20*, 1–18.

Spielberger, C. D., Vagg, P. R., Barker, L. R., Donham, G. W., & Westberry, L. G. (1980). The factor structure of the State-Trait Anxiety Inventory. In I. G. Sarason, & C. D. Spielberger (Eds.), *Stress and anxiety* (Vol. 7, pp. 95–109). Washington: Hemisphere.

Stone, A. A. (1995). Measurement of affective response. In S. Cohen, R. C., Kessler, & L. Underwood-Gordon (Eds.), *Measuring stress: A guide for health and social scientists* (pp. 148–171). New York: Oxford University Press.

Svensson, E. (1977). Response format and factor structure in mood adjective check lists. *Scandinavian Journal of Psychology*, *18*, 71–78.

Tellegen, A. (1985). Structures of mood and personality and their relevance to assessing anxiety, with an emphasis on self-report. In A. H. Tuma, & J. D. Maser (Eds.), *Anxiety and anxiety disorders* (pp. 681–706). Hillsdale, NJ: Lawrence Erlbaum.

Tellegen, A., Watson, D., & Clark, L. A. (1999a). On the dimensional and hierarchical structure of affect. *Psychological Science*, *10*, 297–303.

Tellegen, A., Watson, D., & Clark, L. A. (1999b). Further support for a hierarchical model of affect: Reply to Green and Salovey. *Psychological Science*, *10*, 307–309.

Thayer, R. E. (1967). Measurement of activation through self-report. *Psychological Reports*, *20*, 663–678.

Thayer, R. E. (1978). Factor analytic and reliability studies on the Activation-Deactivation Adjective Check List. *Psychological Reports*, *42*, 747–756.

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

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

Vagg, P. R., Spielberger, C. D., & O'Hearn, T. P. (1980). Is the State-Trait Anxiety Inventory multidimensional? *Personality and Individual Differences*, *1*, 207–214.

van der Ploeg, H.M. (1989). A cross-cultural study of the dimensionality of the State-Trait Anxiety Inventory. In: R. Schwarzer, H. M. van der Ploeg, & C. D. Spielberger (Eds.), *Advances in test anxiety research* (Vol. 6, pp. 195–202). Berwyn, PA: Swets North America.

Van Landuyt, L. M., Ekkekakis, P., Hall, E. E., & Petruzzello, S. J. (2000). Throwing the mountains into the lakes: On the perils of nomothetic conceptions of the exercise-affect relationship. *Journal of Sport & Exercise Psychology*, *22*, 208–234.

van Schuur, W. H., & Kiers, H. A. L. (1994). Why factor analysis often is the incorrect model for analyzing bipolar concepts, and what model to use instead. *Applied Psychological Measurement*, *18*, 97–110.

van Schuur, W. H., & Kruijtbosch, M. (1995). Measuring subjective well-being: Unfolding the Bradburn Affect Balance Scale. *Social Indicators Research*, *36*, 49–74.

Wallbott, H. G., & Sherer, K. R. (1989). Assessing emotion by questionnaire. In R. Plutchik, & H. Kellerman (Eds.),

The measurement of emotions (Vol. 4). *Emotion: Theory, research, and experience* (pp. 55–82). San Diego, CA: Academic Press.

Warr, P., Barter, J., & Brownbridge, G. (1983). On the independence of positive and negative affect. *Journal of Personality and Social Psychology*, *44*, 644–651.

Watson, D. (1988a). The vicissitudes of mood measurement: Effects of varying descriptors, time frames, and response formats on measures of positive and negative affect. *Journal of Personality and Social Psychology*, *55*, 128–141.

Watson, D. (1988b). Intraindividual and interindividual analyses of positive and negative affect: Their relation to health complaints, perceived stress, and daily activities. *Journal of Personality and Social Psychology*, *54*, 1020–1030.

Watson, D., & Clark, L. A. (1997). Measurement and mismeasurement of mood: Recurrent and emergent issues. *Journal of Personality Assessment*, *68*, 267–296.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070.

Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, *98*, 219–235.

Watson, D., & Tellegen, A. (1999). Issues in the dimensional structure of affect — effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin*, *125*, 601–610.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology*, *76*, 820–838.

Watt, B. J., & Spinks, W. L. (1997). Dynamics of exercise-induced affect. *Australian Journal of Science and Medicine in Sport*, *29*, 69–74.

Zautra, A. J., Potter, P. T., & Reich, J. W. (1997). The independence of affects is context-dependent: An integrative model for the relationship between positive and negative affect. In K. W. Schaie, & M. P. Lawton (Eds.), *Annual review of gerontology and geriatrics, focus on adult development* (Vol. 7, pp. 75–103). New York: Springer.

Zautra, A. J., Reich, J. W., Davis, M. C., Potter, P. T., & Nicolson, N. A. (2000). The role of stressful events in the relationship between positive and negative affects: Evidence from field and experimental studies. *Journal of Personality*, *68*, 927–951.

Zevon, M. A., & Tellegen, A. (1982). The structure of mood change: An idiographic / nomothetic analysis. *Journal of Personality and Social Psychology*, *43*, 111–122.