



Affective, but hardly effective: a reply to Gauvin and Rejeski (2001)[☆]

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, IL, USA

Received 4 October 2002; received in revised form 25 April 2003; accepted 19 May 2003

Abstract

Background and purpose: This paper is a reply to Gauvin and Rejeski's rebuttal (Psychol. Sport Exerc. 2 (2001) 73) of a previously published conceptual and methodological critique (Psychol. Sport Exerc. 2 (2001) 1) of the Exercise-induced Feeling Inventory (EFI; J. Sport Exerc. Psychol. 15 (1993) 403).

Methods: Our responses focus on (a) issues regarding scientific debates, (b) the necessity of psychometric scrutiny, (c) the ongoing search for a definition of "exercise-induced feeling states," (d) concerns regarding the underrepresentation of the intended domain of content of the EFI and its appropriate uses, (e) the implications of inductive and deductive scale development, (f) several methodological issues, (g) the value of a circumplex model for exercise psychology research, and (h) the compatibility of categorical and dimensional models of affect.

Results and conclusions: We maintain that the most important issues raised in the original critique of the EFI, such as the definition, the demarcation, and the structure of its intended domain of content, were not addressed in Gauvin and Rejeski's rejoinder and remain unclear. Researchers are urged to contemplate the theoretical bases and to scrutinize the psychometric data of the available measures before making their selection.

© 2003 Elsevier Ltd. All rights reserved.

Keywords: Affect; Feeling states; Scale development; Inductive and deductive; Circumplex

[☆] PII of original article S1469-0292(01)00008-5

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

Any criticism or challenge to consider changing our methods strikes of course at all our ego defenses (Platt, 1964, p. 349).

In 2001, as part of a four-paper series on the measurement of affect in the context of exercise, we published a conceptual and methodological critique (Ekkekakis & Petruzzello, 2001) of the Exercise-induced Feeling Inventory (EFI; Gauvin & Rejeski, 1993). Gauvin and Rejeski (2001) then published a rebuttal. It had been the hope of the editor-in-chief of *Psychology of Sport and Exercise* that the anticipated “constructive and healthy debate” could “only be good for our field” (Biddle, 2000, p. 69). Herein, we respond to Gauvin and Rejeski’s rebuttal.

We recognize that critiques of scientific works are bound to elicit what Chamberlin (1897) called the “partiality of paternalism” (p. 841). Thus, a certain dose of emotionality in Gauvin and Rejeski’s (2001) response and the conclusion that they “entertain optimism regarding the continued utility (of the EFI) and are confident about its psychometric properties” (p. 86) were not particularly surprising. The response was surprising, however, in that it provided no clarifications to the many confusing aspects of the original publication of the EFI. In this paper, we address (a) issues regarding scientific debates, (b) the necessity of psychometric scrutiny, (c) the definition of “exercise-induced feeling states,” (d) the extent to which the EFI is representative of its intended content domain, (e) the implications of inductive and deductive scale development strategies, (f) several methodological issues, (g) the value of the circumplex model for studying affective responses to exercise, and (h) the compatibility of categorical and dimensional models of affect within integrative hierarchical structures.

On debate and “acrimony”

In one of the most incisive parts of his brief autobiography, written for his children in 1876 under the title “Recollections of the development of my mind and character,” Darwin (1899/1959) wrote: “I cannot remember a single first-formed hypothesis which had not after a time to be given up or greatly modified” (p. 83). Such is the nature of scientific ideas, particularly novel ones, that their review and revision is essentially inevitable. Accordingly, the *Standards for Educational and Psychological Testing* of the American Educational Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurement in Education (NCME, 1999; hereafter referred to as the *Standards*) state that it is the *responsibility* of test developers to monitor the research literature, to review, and, when appropriate, to revise their measures. Several authors have published specific guidelines for such test revisions (e.g., Nelson, 2000; Smith & McCarthy, 1995). It could be argued that it is also the responsibility of other researchers to raise issues of theoretical or methodological concern regarding published measures. As Schutz (1993) emphasized, “It is my view that no test is sacred—just because an instrument . . . has been around for years and used in dozens of published studies does not necessarily mean it is a valid and reliable instrument. In other fields, even the most venerable of tests are constantly being reevaluated and questioned” (p. 128).

Publishing a critique of a published work, however, is not a simple decision and numerous factors must be considered, which go beyond the theoretical and technical aspects of the cri-

tique. Before we started developing the series on the “affect measurement conundrum in exercise psychology,” we admittedly spent a long time contemplating Hyman’s (1995) advice:

Despite your indignation or sincere desire to set the record straight, more often than not the best course is to do nothing. If, indeed, the target article has erred in some way, subsequent research and developments in the field will often suffice to remedy the error. Both the reply and subsequent rejoinder by the original author will consume much of the author’s and your time and resources. Ask yourself whether you and your opponent might better contribute to science and humanity by devoting your time to other matters (p. 179).

Why, then, did we choose to undertake this, by all accounts, risky endeavor? While attending an international conference in the late 1990s, one of us (PE) was struck by the exuberance reflected in the references made by presenters to the new, exercise-specific measures of affect. Commenting on the EFI and the Subjective Exercise Experiences Scale (SEES; McAuley & Courneya, 1994), one speaker exclaimed that “we finally have measures that are theory-based!” It was then that the decision was made to attempt an in-depth analysis of the theoretical and methodological issues surrounding the measurement of affect in exercise psychology, as we realized that “research and developments in the field,” contrary to Hyman’s assurance, were not moving toward “remedying the error,” but in the opposite direction.

Several disconcerting signs had started to appear in the literature. The published psychometric evaluations of both the EFI and the SEES were exclusively concerned with whether the various goodness of fit indices reached or exceeded established standards, and none contained a critical discussion of the theoretical bases of the measures or the methodological steps that were followed in their development (Lox & Rudolph, 1994; Markland, Emberton, & Tallon, 1997; Vlachopoulos, Biddle, & Fox, 1996). Likewise, the selection of either the EFI or the SEES in published applied studies was presented as a *de facto* choice, as if there was no need to substantiate this decision by theoretical reasoning or psychometric evidence. Furthermore, as the popularity of the new measures grew, there were increasing signs of uncertainty and confusion regarding the meaning of “exercise-induced feeling states” and “subjective exercise experiences.” In one case, the authors justified the use of the EFI not by stating why the EFI was selected, but rather by explaining why certain other popular measures, namely the Profile of Mood States (POMS; McNair, Lorr, & Droppleman, 1971), the positive and negative affect schedule (PANAS; Watson, Clark, & Tellegen, 1988), the Feeling Scale (FS; Hardy & Rejeski, 1989), and the SEES (McAuley & Courneya, 1994), were not (Karageorghis, Vlachopoulos, & Terry, 2000). Other authors stated that their intended construct of interest was exercise-induced “feeling states,” but, instead of using the EFI, chose to employ the SEES, presumably based on the assumption that the constructs of “exercise-induced feeling states” and “subjective exercise experiences” are identical and, thus, interchangeable (Blanchard, Rogers, Spence, & Courneya, 2001; Lox & Treasure, 2000; Spence & Blanchard, 2001). Others have characterized both the EFI and the SEES as measures of “exercise-induced affect,” but, based on empirical results, have pointed out that the two measures provide non-redundant information and, therefore, the two (Szabo, Mesko, Caputo, & Gill, 1998), along with other measures (Szabo & Bak, 1999), should be used in combination. Also assuming that the EFI and the SEES tap identical or largely overlapping constructs, other researchers have advocated a merger of the two measures (Lox, Jackson,

Tuholski, Wesley, & Treasure, 2000). Surprisingly, although McAuley and Courneya (1994) had initially drawn a distinction between the content domains of the EFI and the SEES, noting that the constructs assessed by the two measures occupy different levels along a “hierarchy of psychological responses to exercise participation,” with the scales of the EFI representing “further underlying structural aspects of the more general psychological responses assessed by the SEES” (p. 173), McAuley and coworkers recently characterized the SEES as a measure of “feeling states” (Gerome et al., 2002). This adds to the confusion by seemingly contradicting the earlier position that the content domains of the EFI (i.e., “exercise-induced feeling states”) and the SEES (i.e., “subjective exercise experiences”) are conceptually distinct.

In sport psychology, Jones and Hanton (2001), citing Gauvin and Rejeski (1993), adopted the term “feeling states,” but used it to describe a different domain of content, whose structure consists of a “positive” and “negative” component. Furthermore, according to Jones and Hanton, “the use of feeling states embraces athletes’ perceptions of physical states” (p. 387), although Gauvin and Rejeski (1993) had chosen to exclude “physical symptoms” (p. 407) from the content domain of the EFI. These diverse and occasionally conflicting views have arisen because, in the absence of specific information on what “exercise-induced feeling states” and “subjective exercise experiences” are and what they are not, researchers have resorted to devising their own, often divergent, interpretations of these terms. We believe that continuing down this path can only perpetuate the confusion and exacerbate the problems that the new measures were developed to solve.

It had always been our belief that embarking on an exploration of the affective space, its nature, its limits, and its structure, regardless of whether one’s domain of interest was exercise or something else, was akin to attempting to reinvent the proverbial wheel. Not taking advantage of available maps and compasses while undertaking such an exploration is difficult to understand, since so much has already been discovered about the affective domain in the last century; one need only refer to the literature. In developing our critique, our approach consisted simply of comparing the goings-on in exercise psychology with principles and procedures from affective psychology and psychometrics that had been around, in some cases, for over a century and, in most cases, for several decades. To underscore the fact that virtually none of the ideas we put forth was novel or controversial and, thus, to illustrate that our disagreements were not merely a matter of subjective preference, we invested considerable effort in documenting each point.

It was, therefore, puzzling to see that Gauvin and Rejeski’s response was focused not on clarifying the issues we had raised and possibly laying the groundwork for a revision of the EFI, but rather on (a) questioning our own preferred measurement approach, namely the affect circumplex (Ekkekakis & Petruzzello, 2002a) for about 1/8 of their paper, (b) commenting on a small sample of methodological issues for another 1/8, and (c) for the remainder of their paper, reviewing the studies in which the EFI had been used and treating this review as evidence of its merit. As stated earlier, on this issue, we agree with the view expressed by Schutz (1993), that “just because an instrument . . . has been around for years and used in dozens of published studies does not necessarily mean it is a valid and reliable instrument” (p. 128). Both Gauvin (e.g., Gauvin & Spence, 1998) and Rejeski (e.g., Rejeski, Hardy, & Shaw, 1991) must certainly agree with this point, since, like us (Ekkekakis, Hall, & Petruzzello, 1999), they have criticized the use of the state anxiety scale of the State Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, &

Lushene, 1970) in the context of exercise, disagreeing with authors who consider the extensive use of this measure as adequate evidence of its reliability and validity (O'Connor, Raglin, & Martinsen, 2000; Raglin, 1997). For this reason, the studies in which the EFI has been used will not be discussed in this paper, just as they were not discussed in our initial critique.

Gauvin and Rejeski (2001) charged that we exhibited “ardor in dissecting the EFI” (p. 75) and that we formulated our statements in an “acrimonious manner” (p. 75). However, they identified no specific examples, other than a few statements of fact, such as that the development of exercise-specific measures grew out of “escalating dissatisfaction” with the traditional measures (e.g., the POMS). This, of course, is hardly “acrimonious,” since Gauvin and Rejeski’s (1993) own rationale for the development of the EFI was based on a critique of the traditional measures (see pp. 403–404). In fact, Gauvin and Spence (1998) used language almost identical to ours when they stated that the exercise-specific instruments “emerged as a result of an effort to overcome lacunae in the measurement of exercise-induced psychological states” (p. 331). Furthermore, the accusations of “ardor” and “acrimony” are groundless for several important reasons. First, our paper had undergone peer review. Given its critical nature, it is reasonable to infer that, had there been any signs of acrimony, our critique would not have been accepted for publication. Second, although we have opted not to use the EFI in our research for the reasons we discussed, we consistently use the FS, a measure co-developed by Rejeski (Hardy & Rejeski, 1989; Kenney, Rejeski, & Messier, 1987). Third, in most of our research, we cite the insightful analysis by Gauvin and Brawley (1993) on the relative strengths and weaknesses of categorical and dimensional models of affect for research in the context of exercise and their suggestions to exercise psychology researcher to consider the circumplex model. Fourth, as previously noted, we have published research examining problems associated with the use of the state anxiety scale of the STAI in the context of exercise (Ekkekakis et al., 1999). This research was not only inspired by a similar earlier study by Rejeski et al. (1991), but also strongly defended the position of these authors against criticisms that had been expressed by others in the interim (e.g., Raglin, 1997). Fifth, Rejeski’s (1981, 1985) pioneering social–psychophysiological model of perceived exertion is at the core of our more recent work on the mechanisms underlying the generation of affect in the context of exercise (Ekkekakis, 2003; Ekkekakis & Petruzzello, 2002b; Hall, Ekkekakis, & Petruzzello, 2002). In short, our work has benefitted considerably from Gauvin and Rejeski’s contributions and we consistently give credit where credit is due.

Finally, it should be noted that, since the publication of our series began, additional cautionary notes and critiques of the exercise-specific measures of affect have appeared (Biddle & Mutrie, 2001; Buckworth & Dishman, 2002; Lox, Ekkekakis, & Lochbaum, 2001; Lox et al., 2000). To the extent that our papers are responsible for raising awareness, not only for the problems surrounding the EFI, but also for the measurement of affect in the context of exercise in general, then the series is starting to accomplish its mission.

On doing away with psychometric scrutiny

On a couple of occasions, Gauvin and Rejeski invoked excerpts from a paper by Landy (1986), leading to perhaps some of the most confusing parts of Gauvin and Rejeski's response. At one point, they cited the following:

Researchers are not really interested in the properties of tests. Instead, they are interested in the attributes of people who take such tests. Thus, validation processes are not so much directed toward the integrity of tests as they are directed toward inferences that can be made about the attributes of the people who have produced those test scores (Landy, 1986, p. 1186; cited in Gauvin & Rejeski, 2001, p. 74).

On the basis of this excerpt, Gauvin and Rejeski (2001) noted that our critique of the EFI “exemplifies the pitfall of conferring primacy to the properties of tests over knowledge advancement” (p. 74). We are uncertain how we have done this but, more importantly, we are deeply troubled by the implication that investigators should not be concerned with the properties of the tests they use in their research. How is one to make valid and reliable inferences about the people who take a test if the test itself is not valid and reliable?

Landy (1986) himself makes no such argument. He is arguing for a unitarian approach to validity, which challenges the traditional typology of content, construct, and criterion-related validity. In this view, validation is treated as regular hypothesis testing and, as such, it can be accomplished by a variety of designs and types of analysis that do not fall in distinct categories. In the excerpt cited by Gauvin and Rejeski, Landy makes the rather uncontroversial claim that, ultimately, researchers are interested in learning about the people, not about the tests themselves. Yet, by adhering to the traditional “trinitarian” view of validity, many psychologists had been primed to think of validation as collecting pieces of evidence that fall neatly under the three traditional categories. Landy likens them to stamp collectors: “there are only three spaces to be filled—the content space, the construct space, and the criterion-related space” (p. 1184).

As intriguing as Landy's (1986) views are, we do not see their relevance to our critique of the EFI. Examining the properties of tests may not have “primacy over knowledge development,” but there is hardly any question that, without conceptually and psychometrically sound tests, the process of knowledge development can easily be derailed. It is certainly not yet time to discard the psychometric textbooks, as the news of the death of psychometrics was highly exaggerated!

Consider the following. Gauvin and Spence (1998, p. 334) cited Kerlinger (1979), who noted that “measurement can be the Achilles' heel of behavioral research” (p. 141). Here is the noteworthy continuation of that excerpt:

Too often investigations are carefully planned and executed with too little attention paid to the measurement of the variables of the research ... All fields of human effort have their share of mythology and nonsense. Measurement is unfortunately particularly burdened with both. Negative attitudes toward psychological measurement are part of the cause. But ignorance and misunderstanding are probably a greater part. One of the core difficulties seems to be that the basic purpose of measurement in science is misunderstood. It is believed that psychologists, for example, measure people, and that you cannot measure people. But scientists

do not really measure people, nor do they say they do (or they should not say they do). They measure variables, just as in experiments they do not manipulate people but variables. . . To believe that psychologists measure people, then, is naive. To criticize measurement because it does not measure people is also naive. Indeed, the whole measurement procedure, even of relatively “simple” variables, is highly indirect, complex, and often difficult. . . Like anything else, tests and measures must be used with caution and discretion. . . We must always be very careful to ascertain the reliability and validity of our measures (pp. 141–142).

On the elusive definition of “exercise-induced feeling states”: the search continues

One of the primary issues of concern raised in our critique of the EFI was the absence of a definition of the construct of “feeling states” and the absence of a clear specification of the intended domain of content of the scale. The fallout from this problem is becoming evident, as researchers appear unable to provide conceptual rationales for selecting the EFI or distinguishing the content domain of the EFI from those of other measures.

Given the importance of this problem, we were hopeful that Gauvin and Rejeski would provide a definition of “feeling states,” but they did not. Instead, they claimed that we failed to “grasp the primary intent underlying the development of the EFI” (p. 74), arguing that “an examination of the original paper (Gauvin & Rejeski, 1993) would reveal that the EFI is not a measure of exercise-specific affect. Such a claim was never enunciated” (p. 74). Contrary to these claims, however, Gauvin and Spence (1998) listed “exercise-induced feelings” under the rubric of “variables adopted in the study of exercise-related *affect*” (p. 327, italics added) and the EFI under the rubric of “measures of *affective* states” (p. 331, italics added). Likewise, Rejeski, Reboussin, Dunn, King, and Sallis (1999) stated that the concepts measured by the EFI (i.e., Revitalization, Positive Engagement, Tranquility, and Exhaustion) “appear to be the primary forms of *affect* that are directly influenced by physical activity” (p. 98, italics added).

According to existing guidelines, the definition and demarcation of the content domain is crucial for the development of a measure. This is even more so in the case of a construct, such as “exercise-induced feeling states,” which is unique in the psychological literature and, as such, represents unfamiliar territory. We had noted that the only account of what was meant by “feeling states” that we were able to find was one provided by Gauvin and Spence (1998) several years after the publication of the EFI. According to this description, the term “feeling states” refers “to those human experiences that include bodily reactions, cognitive appraisals, actual or potential instrumental responses, or some combination thereof” (p. 326). Furthermore, Gauvin and Spence noted that exercise can “elicit feeling states including but not limited to Positive Engagement, Revitalization, Physical Exhaustion, and Tranquility” (p. 327). Our comments were that (a) the definition of feeling states was too broad, as it subsumed under the rubric of “feeling states” three extremely large domains of psychological research (perception, cognition, and behavior), while mysteriously excluding affect, and (b) if “exercise-induced feeling states” include more constructs than just the four tapped by the EFI, some explanation must be provided for why others were excluded. To these concerns, Gauvin and Rejeski (2001) replied that “when Gauvin and Spence (1998) defined feeling states, they were not referring to the EFI at all”

(p. 77). We find this perplexing, as we cannot understand why Gauvin and Spence's (1998) definition of "feeling states" would be different from Gauvin and Rejeski's (1993) definition. This means that the definition of feeling states that was used by Gauvin and Rejeski in the development of the EFI remains elusive. Notably, Gauvin and Rejeski (2001) seem to agree with us that Gauvin and Spence's (1998) definition, which follows Averill's (1994) is, in fact, too broad and vague.

To close this section, Gauvin and Rejeski (2001) are right when they say that "the mere juxtaposition of statements from our various publications further confuses matters and does not advance the debate on what is still a highly contentious issue" (p. 75). However, the confusion stems from the contradictions in these statements. Although authors are certainly allowed to change their positions, it is not constructive to do so without providing a clear rationale.

On the representativeness of the content domain

According to the *Standards* (AERA, APA, & NCME, 1999), test developers should give careful consideration to whether "a test measures less or more than its proposed construct" (p. 10). In particular, "construct underrepresentation refers to the degree to which a test fails to capture important aspects of the construct" (p. 10). In our critique of the EFI, we commented extensively on the possibility of construct underrepresentation as a result of the item selection and content validation procedures that were followed. To recap, the item pool of the EFI was reduced based on (a) the opinions of three experts regarding the presumed relevance of items to exercise (from over 500 items to 145), (b) the confirmation of item relevance to exercise by a sample of students (all 145 retained), (c) the formation of 15 clusters based on subjective criteria, and (d) the progressive elimination of 11 clusters, also based on subjective and mostly undisclosed criteria. Given the fact that at no point during these procedures were diverse samples of exercisers queried in an open-ended fashion about their responses to exercise, we expressed the concern that there may be aspects of "exercise-induced feelings" that may not be tapped by the remaining items and scales. In particular, we pointed out that the final form of the EFI contained no scales that were unequivocally negative. This seemed problematic inasmuch as research has shown that such responses do occur, particularly *during* exercise, such as increases in tension (e.g., Ekkekakis et al., 1999).

Gauvin and Rejeski (2001) responded that interpreting increases in tension as negative was an "inferential leap" (p. 79). Furthermore, they argued that, even if participants experience an increase in tension during a maximal graded exercise test as negative, this, in essence, falls outside the scope of the EFI, as Gauvin and Rejeski's "interest in developing the EFI was to provide a measure that would be beneficial for research that has more of a public health focus" (p. 79). They concluded by asserting that "public health interventionists rarely, if ever, encourage individuals to participate in maximal intensity exercise" (p. 79).

Contrary to the assumption that negative affective responses are limited to maximal exercise, several studies have shown declines in affective valence with considerably less strenuous intensities. Hardy and Rejeski (1989), for example, showed that ratings of affective valence (assessed by the FS) declined as exercise intensity was increased from just 30% to 60% (and, finally, to 90%) of estimated maximal aerobic capacity. Van Landuyt, Ekkekakis, Hall, and Petruzzello

(2000) showed that 41.3% of a sample of young and healthy participants reported declines in affective valence (also assessed by the FS) during a moderate-intensity (60% of estimated maximal aerobic capacity, 141 beats·min⁻¹) bout of stationary cycling. Bixby, Spalding and Hatfield (2001) showed a similar decline in affective valence during a bout of stationary cycling, when the intensity corresponded to the ventilatory threshold (158 beats·min⁻¹). In fact, several other studies have also demonstrated negative affective responses during or immediately after sub-maximal bouts of exercise (see Ekkekakis & Petruzzello, 1999, for a review). It should be emphasized that these results have come from studies of young, healthy, and mostly physically active participants. Therefore, it is reasonable to speculate that studies involving older, obese, or otherwise physically limited individuals may yield even more negative patterns. Notably, Gauvin and Spence (1996) stated that “there is no doubt that acute vigorous exercise constitutes a very profound stimulus accompanied by many physical symptoms that can be perceived as *distressing*, instrumental, pleasant, or *unpleasant*” (p. S60, italics added).

In our view, perhaps the most important development that arose from Gauvin and Rejeski's response on this issue is the acknowledgment that the EFI was not intended for use with high-intensity and maximal exercise stimuli and, presumably, any mode or dose of exercise that can elicit unpleasant responses. This is an important caveat, which should have been issued from the beginning. According to the *Standards* (AERA, APA, & NCME, 1999), the recommended uses of a test must be specified and, “where particular misuses of a test can be reasonably anticipated, cautions against such misuses must be specified” (p. 68). Recall that Gauvin and Rejeski (1993) had initially asserted that the components of the EFI reflect the “phenomenology of people involved in exercise in the real world” (p. 408) and Rejeski et al. (1999) stated that these components represent the “primary forms of affect that are directly influenced by physical activity” (p. 98). Researchers should be cautioned that, according to Gauvin and Rejeski's revised position, the intended uses of the EFI exclude high intensity and maximal exercise.

Finally, we feel compelled to reiterate that construct underrepresentation might be a problem when the participants involved are not young, healthy, and physically active. The *Standards* state that “the population for whom the test is intended . . . should be documented” (p. 69). As we had pointed out, although Gauvin and Rejeski (1993) had noted that the “universe of content” of the EFI “is consistent with the phenomenology of men and women who exercise with some regularity” (p. 419), the EFI has been used in several studies that involved participants selected to be sedentary (i.e., that do not “exercise with some regularity”; see Ekkekakis & Petruzzello, 2001, for references). Researchers who undertake such studies should keep in mind that, according to the *Standards*, “if a test is used in a way that has not been validated, it is incumbent on the user to justify the new use, collecting new evidence if necessary” (p. 18).

On inductive and deductive scale development

One of our criticisms of the EFI centered on the fact that, although the focus on Positive Engagement, Revitalization, Tranquility, and Physical Exhaustion was presented as driven by an “*a priori* conceptual framework” (Gauvin & Rejeski, 1993, p. 411), pointing to a deductive approach to scale development, the methods that were followed during the item selection and content validation phases were incompatible with a deductive approach. We further noted that

the item selection procedure did not follow established guidelines, as it consisted of the atypical process of forming clusters of items based on subjective and mostly undisclosed criteria and then gradually pruning the item pool from 15 to 4 clusters, again with insufficient information about this process to allow its replication by independent researchers.

In their response, [Gauvin and Rejeski \(2001\)](#) acknowledged that, in fact, the development of the EFI was not based on a deductive approach. They noted that “we had not initially opted for a dimensional or a categorical approach” (p. 77) and “we did not label Positive Engagement, Revitalization, Physical Exhaustion, and Tranquility as ‘feeling states’ until after the data had been collected and analyzed” (p. 77). This certainly contradicts the original claim of an “*a priori* conceptual framework” ([Gauvin & Rejeski, 1993, p. 411](#)). Such an acknowledgment that no conceptual postulates regarding the content and structure of the domain of “exercise-induced feeling states” had been formulated in advance ([Kerr, 1998](#)) is very important information to those interested in using the EFI. We will simply reiterate one of the cardinal rules of scale development: “Even if there is no available theory to guide the investigators, they must lay out their own conceptual formulations prior to trying to operationalize them” ([DeVellis, 1991, p. 52](#)).

On methodological issues

[Gauvin and Rejeski \(2001\)](#) commented on only a small sample of the methodological issues we raised. Our responses follow.

On the reciprocal relationship between parsimony and goodness of fit

In our critique of the EFI, we questioned the decision to split an apparently homogenous factor into two (Positive Engagement and Revitalization) and then attempt to establish the superiority of this solution through confirmatory factor analysis (FA). We argued that, because of the reciprocal relationship between parsimony and goodness of fit, the solution that contains the additional factor has a high likelihood of yielding a better fit, especially if the correlation between factors is not constrained to some degree of independence. We illustrated this point with a simple computational exercise, using a matrix of factor loadings from a principal components analysis (PCA) reported by [Gauvin and Rejeski](#) and working backwards to reconstruct the item intercorrelation matrix. We then used these data to run a series of confirmatory factor analyses to illustrate our point. [Gauvin and Rejeski \(2001\)](#) commented that “the reader is left to wonder why [Ekkekakis and Petruzzello’s](#) derived analyses were performed in the first place, particularly since they did not state what specific question they were attempting to answer through this demonstration” and concluded that the analysis was “nonsensical” (p. 78). They also claimed that “in a last ironic twist, if [Ekkekakis and Petruzzello](#) had read our original paper on the EFI more carefully, they would have concluded that this was the wrong data set to analyze in the first place” (p. 78).

As a first observation, the structure of a measure should remain invariant across the situations in which it is likely to be employed. Therefore, whether the data were collected before or after exercise should make no difference in the structure of the EFI, as long as the usage of

the scale at both times is deemed acceptable by its developers. If the structure was expected to vary, then a comparison between pre-exercise and during- or post-exercise responses would be improper, necessitating the issuance of an appropriate warning to prospective users.

Second, the fact that our secondary analysis used the matrix from the PCA was not due to a failure to distinguish the “right” from the so-called “wrong” data set. This matrix was selected because it allowed us to estimate the item intercorrelation matrix using standard formulas, since the item intercorrelation matrix that was used for the confirmatory FA was not reported.

Third, as was clearly explained in our paper (Ekkekakis & Petruzzello, 2001, pp. 18–20), the purpose of our computational exercise was to show, through a series of confirmatory factor analyses, that (a) one can split the homogeneous factor either the way that Gauvin and Rejeski did or some other way and still obtain solutions that have a better fit compared to the three-factor solution, and (b) if one splits up the homogeneous factor even further into three rather than two factors (i.e., five rather than four factors in total), then the fit can be improved even more. Our goal was to underscore that finding better fit by splitting a homogenous factor into more factors is essentially trivial. Absent of a strong theoretical basis, such a finding cannot be construed as an adequate rationale for maintaining the four-factor over the three-factor solution.

On convergent and discriminant validity

Commenting on our discussion of the convergent and discriminant validity data of the EFI, Gauvin and Rejeski (2001) stated that “Ekkekakis and Petruzzello apparently use different criteria at different times but do not explain why. We are perplexed by their methods of analysis and interpretation” (p. 78). Both our methods and our points were plainly explained. We had noted that, given the absence of theoretical or empirical evidence for the separability of the Positive Engagement and Revitalization factors, we were not surprised by the fact that these two factors had similar correlations with external correlates, namely the Positive Affect (PA) Scale of the PANAS (Watson et al., 1988) and the Energy Scale of the activation deactivation adjective check list (AD ACL; Thayer, 1989). If these scales represented substantively different constructs, this should have been reflected in their relationships with these external correlates. Instead, the correlations of Positive Engagement and Revitalization with these two scales were 0.69 and 0.58 with PA and 0.28 and 0.47 with Energy, respectively. Based on tests of the differences between the z -transformed values of these coefficients (for a single-sample case, with $N = 115$), the former coefficients are not significantly different whereas the latter are only marginally different at 0.05 (Cohen, 1988). We maintain that these data do not constitute compelling evidence of discriminant validity.

We had also pointed out that the Tranquility Scale of the EFI had a stronger correlation with the conceptually unrelated PA Scale of the PANAS (i.e., 0.40) than with the conceptually related Calmness Scale of the AD ACL (i.e., 0.30). The rationale for this observation was that both Calmness and Tranquility are pleasant low-activation states (e.g., see Rejeski et al., 1999, p. 99, Fig. 1) and should, therefore, be highly interrelated (approaching unity). Instead, they had a correlation of 0.30. On the other hand, PA is a state of pleasant high activation and should, therefore, have a near-zero correlation with Tranquility (i.e., separated by 90° ; see Rejeski et al., 1999). Instead, they had a correlation of 0.40. Again, we maintain that these data do not constitute convincing evidence of convergent and discriminant validity.

On error of measurement and correlation coefficients

Also on the issue of the separability of the Positive Engagement and Revitalization Scales, we had cited the opinion of [Nunnally and Bernstein \(1994\)](#) that a correlation of 0.50 “should make you consider the option” and a correlation of 0.70 “would be a very strong reason” (p. 501) to merge two scales into one. The correlation between Positive Engagement and Revitalization reported by [Gauvin and Rejeski](#) was 0.68, leading us to believe that, absent of a theoretical argument to the contrary, there was no statistical reason to separate these two scales. As an additional argument, we pointed out that, once the error of measurement was taken into account by “correcting” the correlation coefficient for unreliability, the two scales appeared almost perfectly redundant, with a correlation of 0.90.

[Gauvin and Rejeski](#)’s response was that (a) only test–retest, and not internal consistency coefficients, can be used in the formula for correction of attenuation due to unreliability, and (b) that “these authors chose a level of reliability that was optimal for their point of discussion” (p. 79). Neither point is accurate.

First, internal consistency coefficients can be used as estimates of unreliability. Although it is true that before the use of Cronbach’s alpha coefficient had spread, [Guilford \(1954\)](#) had argued that “the type of reliability estimate to use for this purpose has been a much debated subject” (p. 400), since that time, psychometricians have adopted alpha as the appropriate estimate (e.g., [Muchinsky, 1996](#); [Nunnally & Bernstein, 1994](#)).

Second, we certainly did not choose the reliability estimate that “was optimal for our point of discussion.” Rather, we chose the same reliability estimates reported by [Gauvin and Rejeski \(1993, p. 415\)](#) for the data set from which the inter-scale correlations were calculated (i.e., post-exercise; specifically, 0.78 for Revitalization and 0.74 for Positive Engagement). As we noted, in conjunction with the reported 0.68 correlation between Positive Engagement and Revitalization, the correction for attenuation yields 0.90. The important point is that, with or without correction for attenuation, the correlation between Positive Engagement and Revitalization is high enough to challenge the separability of these two scales on statistical grounds. Even without a correction, others have reported a correlation as high as 0.857 ([Lox et al., 2000, p. 85](#)), well above the 0.70 level specified by [Nunnally and Bernstein \(1994\)](#) as a “very strong reason” against the separation.

On principal components and factor analyses

Of the numerous comments we made on the PCA of the EFI, [Gauvin and Rejeski \(2001\)](#) limited their response to one, arguing that they opted to use a PCA rather than a FA because they “were in the initial phase of development of the measure and simply wanted an empirical summary of the data” (p. 79). This assessment does not agree with our reading of their paper. As we noted in our critique (see p. 12), the PCA was not conducted “in the initial phase of development.” In fact, we expressed our puzzlement over the fact that an exploratory analysis was conducted in the first place, since, when it was conducted, there was really nothing left to explore. The PCA was conducted when the final four scales of the EFI had already been formed and the final 12 items that made up these scales had already been selected.

On the circumplex as a “simplistic” model

Although the word “circumplex” appeared only once, in a parenthesis that was not even meant to describe or defend the model but rather to specify the work presented in a paper (Russell, 1980) that Gauvin and Rejeski (1993) had reportedly consulted in the development of the EFI, Gauvin and Rejeski (2001) devoted as much space criticizing the circumplex (pp. 75–77) as they did responding to criticisms about the EFI. Therefore, it was with some amusement that we read that “whether or not a circumplex model, like the one described by Russell and others, is useful in harnessing the psychological outcomes of acute exercise is entirely another matter that should not be gratuitously entangled within a discussion of the value of the EFI. The only possible outcome is confusion” (p. 77). Suffice it to say that we agree.

It appears that the critique of the circumplex was prompted by the fact that, according to Gauvin and Rejeski (2001), “Ekkekakis and Petruzzello would like exercise scientists to adopt a dimensional approach to the study of affect, the circumplex model” (p. 75). First, we should point out that, when Gauvin and Rejeski’s (2001) response was published, our conceptual arguments for the use of the circumplex in exercise psychology had not yet been published (Ekkekakis & Petruzzello, 2002a) and, evidently, Gauvin and Rejeski had not read the already published introductory paper of the series (Ekkekakis & Petruzzello, 2000).

Gauvin and Rejeski (2001) attacked the circumplex model, based on three unsubstantiated claims: (a) the fit of the circumplex to self-reported affective data is poor, (b) separable neural substrates of pleasure and displeasure challenge the notion of a bipolar relationship between these states, which is central to the circumplex model, and (c) assessments of perceived activation are problematic. We believe that (a) these issues are, in fact, unrelated to the present discussion on the EFI and (b) although it might take but a few lines of text to condemn decades of research in an invited article, it takes meticulous analysis to set the record straight and inform those who may not be ready or willing to perform their own, independent evaluation of the literature. Therefore, we have decided that it would be best if we dealt with these criticisms in an appropriately detailed fashion elsewhere. Here, we will only examine the position taken by Gauvin and Rejeski regarding the circumplex before they decided to label this model “outdated” (p. 74) and “oversimplified” (p. 76).

Gauvin and Brawley (1993) were probably the first authors in exercise psychology to draw attention to the circumplex. They noted that, despite the considerable potential of the model and its relevance to exercise, the “circumplex model of affect has been virtually ignored by researchers in exercise psychology” (p. 153). They then recommended that this situation be changed, suggesting that “future research should. . .focus on describing when changes in hedonic tone (what we have called affective valence) and self-perceptions of activation occur throughout exercise stimuli of varying intensities, as well as during the recovery period from exercise” (p. 154). This recommendation was based on the assertion that describing responses to exercise “according to the circumplex model of affect appears to be relevant for two reasons” (p. 155). One reason why “the circumplex model of affect. . .seems particularly well suited” to the study of exercise-induced responses is “because exercise may induce changes in both dimensions of the model” (p. 155). The other reason is that, according to research (see Ekkekakis & Petruzzello, 2002a, for an overview), “the dimensions of hedonic tone and activation appear to account for about 45% of the variance in affective experience. Therefore, the importance of these dimensions

cannot be ignored” (p. 155). Besides these comments, which focused specifically on the circumplex model, as we noted in our initial critique of the EFI (Ekkekakis & Petruzzello, 2001), Gauvin and Brawley (1993) made an excellent case for using dimensional models (such as the circumplex) rather than categorical models for the study of affect in the context of exercise. Expressing a position with which we completely agree (Ekkekakis & Petruzzello, 2000, 2002a), they noted that “because the affective experience that accompanies exercise has not been thoroughly described, a model of affect that has a wider breadth (i.e., a dimensional model) is more likely to capture the essence of exercise-induced affect than a model that, at the outset, limits the focus of investigation to specific emotions” (p. 152). We would like to reemphasize that we consider these views to be remarkably insightful and pioneering, as they were formulated at a time that others in exercise psychology had yet to contemplate the merits of the circumplex model or dimensional models of affect, in general. Gauvin and Spence revisited the circumplex model as recently as 1998, yet offered no indication that the model had, in the interim, turned from “relevant” (Gauvin & Brawley, 1993, p. 155) and one that exercise psychology researchers should adopt to one that represents “a conceptual view of the area that is overly simplistic and outdated” (Gauvin & Rejeski, 2001, p. 74). In fact, in 1999, Rejeski et al. not only offered a description of the model, but also proceeded to speculate as to which octant of the circumplex space the scales of the EFI would occupy. Despite the previous endorsements, however, Gauvin and Rejeski (2001) did a complete about-face with regard to the circumplex model. Gauvin and Brawley’s (1993) views were now portrayed as almost inconsequential, describing nothing more than an “alternative” model for the study of exercise and affect (p. 77). The reasons for this dramatic change remain unclear.

On the compatibility of categorical and dimensional models

Gauvin and Rejeski (2001) argued that our interpretation of the relative strengths and weaknesses of categorical and dimensional models was “simply incorrect” (p. 76) and that, in “contemporary views of affect” that are not “constrained by the assumptions being promoted by Ekkekakis and Petruzzello” (p. 76), categorical and dimensional models are compatible.

This statement raises an issue that had already been addressed at length in the introductory paper of our series (Ekkekakis & Petruzzello, 2000, pp. 80–81). As we had emphasized there and later reiterated in the paper on the circumplex (Ekkekakis & Petruzzello, 2002a, pp. 36–37 and 46–47), it is clear that both categorical and dimensional models have their place. Specifically, a year before Gauvin and Rejeski’s (2001) reminder that categorical and dimensional models are compatible within a hierarchical structure, we had noted the following: “acknowledging that both the categorical and the dimensional perspective have relative advantages and limitations, several authors have presented hierarchical models in which a dimensional representation accounts for the commonalities between affective states at a macro-level and categorical models account for the specificity at a micro-level” (Ekkekakis & Petruzzello, 2000, p. 80). We also cited Watson and Clark’s (1997) opinion that categorical and dimensional perspectives “are not incompatible or mutually exclusive; rather, they essentially reflect different levels of a single, integrated hierarchical structure” (p. 269).

The question, then, is which one is preferable for use in the context of exercise at the present stage of knowledge development and, primarily, in descriptive studies, which continue to make up the vast majority of studies in this field. On this issue, we have stated repeatedly that we concur with [Gauvin and Brawley's \(1993\)](#) assessment that “a model of affect that has a wider breadth (i.e., a dimensional model) is more likely to capture the essence of exercise-induced affect than a model that, at the outset, limits the focus of investigation to specific emotions (i.e., a categorical model)” (p. 152).

Conclusion

Perhaps the most important issues discussed in our original critique of the EFI were not addressed in [Gauvin and Rejeski's \(2001\)](#) response. All indications point to the fact that there remains a lot of confusion regarding what “exercise-induced feeling states” are and what they are not. We are still uncertain what the boundaries of the domain of content are and whether the EFI offers an encompassing representation of this domain. We still do not know why the EFI was developed as a categorical measure. We are still troubled by the notion of a universal phenomenology of exercise that can allegedly be represented by four distinct states. We do not understand how, despite the claim that “the stimulus properties of exercise can elicit feeling states including *but not limited to* Positive Engagement, Revitalization, Physical Exhaustion, and Tranquility” ([Gauvin & Spence, 1998, p. 327](#), italics added), the EFI, a measure of the domain of “exercise-induced feeling states,” taps only these four and yet its “exclusive use...can be defended” ([Gauvin & Rejeski, 2001, p. 75](#)). We also remain unaware of any evidence that the content and/or structure of the affective domain in the context of exercise differ from those of global affect, making exercise-specific instruments necessary. We have yet to see evidence that a measure such as the EFI can unveil secrets of exercise-induced affect that general theory-based measures, such as the AD ACL ([Thayer, 1989](#)), cannot. We also do not understand how it is acceptable to use a measure that has been specifically tailored to be responsive to only one type of experimental condition (i.e., exercise) in another, clearly incompatible, type of condition (e.g., quiet rest; see [Gauvin & Rejeski, 2001, p. 74](#)) and then to compare the responses from the two conditions. We do not know what the difference between Positive Engagement and Revitalization is and we have never seen two scales that are correlated as strongly as these two are (e.g., 0.857, as reported by [Lox et al., 2000](#)) be treated as separate entities. [Gauvin and Rejeski \(2001\)](#) argued that they “have responded to (our) specific arguments” (p. 85) although they do not find these to be “pertinent issues for debate” (p. 79). Unfortunately, they have not responded to most of our arguments. Moreover, the issues we raised are not only pertinent, but crucial. As we did throughout our series on the measurement of affect in the context of exercise, we will repeat the call to researchers to contemplate the theoretical bases and to scrutinize the psychometric data of the available measures before making their selection.

References

- American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.
- Averill, J. R. (1994). I feel, therefore I am—I think. In P. Ekman, & R. J. Davidson (Eds.), *The nature of emotion: fundamental questions* (pp. 379–385). New York: Oxford University Press.
- Biddle, S. (2000). Editorial: theory, measurement and debate. *Psychology of Sport and Exercise*, 1, 69.
- Biddle, S. J. H., & Mutrie, N. (2001). *Psychology of physical activity: determinants, well-being, and interventions*. London: Routledge.
- Bixby, W. R., Spalding, T. W., & Hatfield, B. D. (2001). Temporal dynamics and dimensional specificity of the affective response to exercise of varying intensity: differing pathways to a common outcome. *Journal of Sport and Exercise Psychology*, 23, 171–190.
- 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.
- Buckworth, J., & Dishman, R. K. (2002). *Exercise psychology*. Champaign, IL: Human Kinetics.
- Chamberlin, T. C. (1897). The method of multiple working hypotheses. *Journal of Geology*, 5, 837–848.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
- Darwin, C. (1959). Autobiography. In F. Darwin (Ed.), *The life and letters of Charles Darwin, including an autobiographical chapter*, Vol. 1. (pp. 25–86). New York: Basic Books.
- DeVellis, R. (1991). *Scale development: theory and applications*. Newbury Park, CA: Sage.
- Ekkekakis, P. (2003). Pleasure and displeasure from the body: perspectives from exercise. *Cognition and Emotion*, 17, 213–239.
- 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 and 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.
- Ekkekakis, P., & Petruzzello, S. J. (2002a). Analysis of the affect measurement conundrum in exercise psychology: IV. A conceptual case for the affect circumplex. *Psychology of Sport and Exercise*, 3, 35–63.
- Ekkekakis, P., & Petruzzello, S. J. (2002b). Biofeedback in exercise psychology. In B. Blumenstein, M. Bar-Eli, & G. (Eds.), *Brain and body in sport and exercise: biofeedback application in performance enhancement* (pp. 77–100). Chichester, UK: John Wiley & Sons.
- Gauvin, L., & Brawley, L. R. (1993). Alternative psychological models and methodologies for the study of exercise and affect. In P. Seraganian (Ed.), *Exercise psychology: the influence of physical exercise on psychological processes* (pp. 146–171). New York: John Wiley & Sons.
- Gauvin, L., & Rejeski, W. J. (1993). The Exercise-induced Feeling Inventory: development and initial validation. *Journal of Sport and Exercise Psychology*, 15, 403–423.
- Gauvin, L., & Rejeski, W. J. (2001). Disentangling substance from rhetoric: a rebuttal to Ekkekakis and Petruzzello (2001). *Psychology of Sport and Exercise*, 2, 73–88.
- Gauvin, L., & Spence, J. C. (1996). Physical activity and psychosocial well-being: knowledge base, current issues, and caveats. *Nutrition Reviews*, 54, S53–S65.
- Gauvin, L., & Spence, J. C. (1998). Measurement of exercise-induced changes in feeling states, affect, mood, and emotions. In J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 325–336). Morgantown, WV: Fitness Information Technology.

- Gerome, G. J., Marquez, D. X., McAuley, E., Canaklisova, S., Snook, E., & Vickers, M. (2002). Self-efficacy effects on feeling states in women. *International Journal of Behavioral Medicine*, *9*, 139–154.
- Guilford, J. P. (1954). *Psychometric methods*. (2nd ed.). New York: McGraw-Hill.
- Hall, E. E., Ekkekakis, P., & Petruzzello, S. J. (2002). The affective beneficence of vigorous exercise revisited. *British Journal of Health Psychology*, *7*, 47–66.
- Hardy, C. J., & Rejeski, W. J. (1989). Not what, but how one feels: the measurement of affect during exercise. *Journal of Sport and Exercise Psychology*, *11*, 304–317.
- Hyman, R. (1995). How to critique a published article. *Psychological Bulletin*, *118*, 178–182.
- Jones, G., & Hanton, S. (2001). Pre-competitive feeling states and directional anxiety interpretations. *Journal of Sports Sciences*, *19*, 385–395.
- Karageorghis, C. I., Vlachopoulos, S. P., & Terry, P. C. (2000). Latent variable modelling of the relationship between flow and exercise-induced feelings: an intuitive appraisal perspective. *European Physical Education Review*, *6*, 230–248.
- Kenney, E. A., Rejeski, W. J., & Messier, S. P. (1987). Managing exercise distress: the effect of broad spectrum intervention on affect, RPE, and running efficiency. *Canadian Journal of Applied Sport Sciences*, *12*, 97–105.
- Kerlinger, F. N. (1979). *Behavioral research: a conceptual approach*. New York: Holt, Rinehart, and Winston.
- Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*, 196–217.
- Landy, F. J. (1986). Stamp collecting versus science: validation as hypothesis testing. *American Psychologist*, *41*, 1183–1192.
- Lox, C. L., Ekkekakis, P., & Lochbaum, M. R. (2001). Summary of symposium: historical trends and future directions in the measurement of exercise-induced affect. *Journal of Sport and Exercise Psychology*, *23*, S7 (abstract).
- Lox, C. L., Jackson, S., Tuholski, S. W., Wesley, 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 and Exercise Psychology*, *19*, 418–433.
- McAuley, E., & Courneya, K. S. (1994). The Subjective Exercise Experiences Scale (SEES): development and preliminary validation. *Journal of Sport and Exercise Psychology*, *16*, 163–177.
- McNair, D. M., Lorr, M., & Droppleman, L. F. (1971). *Manual for the Profile of Mood States*. San Diego: Educational and Industrial Testing Service.
- Muchinsky, P. M. (1996). The correction for attenuation. *Educational and Psychological Measurement*, *56*, 63–75.
- Nelson, L. D. (2000). Introduction to the special section on methods and implications of revising assessment instruments. *Psychological Assessment*, *12*, 235–236.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. (3rd ed.). New York: McGraw-Hill.
- O'Connor, P. J., Raglin, J. S., & Martinsen, E. W. (2000). Physical activity, anxiety, and anxiety disorders. *International Journal of Sport Psychology*, *31*, 136–155.
- Platt, J. R. (1964). Strong inference. *Science*, *146*, 347–353.
- Raglin, J. S. (1997). Anxiolytic effects of physical activity. In W. P. Morgan (Ed.), *Physical activity and mental health* (pp. 107–126). Washington, DC: Taylor & Francis.
- Rejeski, W. J. (1981). The perception of exertion: a social psychophysiological integration. *Journal of Sport Psychology*, *4*, 305–320.
- Rejeski, W. J. (1985). Perceived exertion: an active or passive process?. *Journal of Sport Psychology*, *7*, 371–378.
- Rejeski, W. J., Hardy, C. J., & Shaw, J. (1991). Psychometric confounds of assessing state anxiety in conjunction with acute bouts of vigorous exercise. *Journal of Sport and Exercise Psychology*, *13*, 65–74.

- Rejeski, W. J., Reboussin, B. A., Dunn, A. L., King, A. C., & Sallis, J. F. (1999). A modified Exercise-induced Feeling Inventory for chronic training and baseline profiles of participants in the Activity Counseling Trial. *Journal of Health Psychology, 4*, 97–108.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology, 39*, 1161–1178.
- Schutz, R. W. (1993). Methodological issues and measurement problems in sport psychology. In S. Serpa, J. Alves, Ferreira, & A. Paulo-Brito (Eds.), *Proceedings of the eighth world congress of sport psychology*. Lisbon, Portugal: International Society of Sport Psychology.
- Smith, G. T., & McCarthy, D. M. (1995). Methodological considerations in the refinement of clinical assessment instruments. *Psychological Assessment, 7*, 300–308.
- Spence, J. C., & Blanchard, C. (2001). Effect of pretesting on feeling states and self-efficacy in acute exercise. *Research Quarterly for Exercise and Sport, 72*, 310–314.
- Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Szabo, A., & Bak, M. (1999). Exercise-induced affect during training and competition in collegiate soccer players. *European Yearbook of Sport Psychology, 3*, 91–104.
- Szabo, A., Mesko, A., Caputo, A., & Gill, E. T. (1998). Examination of exercise-induced feeling states in four modes of exercise. *International Journal of Sport Psychology, 29*, 376–390.
- Thayer, R. E. (1989). *The biopsychology of mood and arousal*. New York: Oxford University Press.
- 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 and Exercise Psychology, 22*, 208–234.
- Vlachopoulos, S., Biddle, S., & Fox, K. (1996). A social-cognitive investigation into the mechanisms of affect generation in children's physical activity. *Journal of Sport and Exercise Psychology, 18*, 174–193.
- 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.