Journal of Sports Sciences, March 2008; 26(5): 499–510

Routledge
Taylor & Francis Group

The Preference for and Tolerance of the Intensity of Exercise Questionnaire: A psychometric evaluation among college women

PANTELEIMON EKKEKAKIS[1], JENNIFER THOME[2], STEVEN J. PETRUZZELLO[3], & ERIC E. HALL[4]

[1]*Department of Kinesiology, Iowa State University,* [2]*Student Counselling Services, Illinois State University,* [3]*Department of Kinesiology and Community Health, University of Illinois at Urbana-Champaign and* [4]*Department of Health and Human Performance, Elon University, USA*

(Accepted 10 August 2007)

Abstract

Individuals differ in the intensity of exercise that they prefer and the intensity that they can tolerate. We examined the psychometric properties of the Preference for and Tolerance of the Intensity of Exercise Questionnaire (PRETIE-Q; Ekkekakis, Hall, & Petruzzello, 2005: *Journal of Sport and Exercise Psychology, 27*, 350–374). A sample of 601 college women (mean age 20 years) completed the PRETIE-Q and the Godin Leisure-Time Exercise Questionnaire. Both the Preference and the Tolerance scale were internally consistent (alpha coefficients of 0.89 and 0.86, respectively), with all items making positive contributions. A confirmatory factor analysis showed that model fit was reasonable. Both Preference and Tolerance were related to the frequency of participation in strenuous exercise and the total leisure-time exercise score. The PRETIE-Q appears to be an internally consistent and structurally valid measure, with a broad range of possible applications in exercise science.

Keywords: *Arousability, sensory modulation, affect, psychometrics, personality*

Introduction

Exercise intensity is one component of the exercise dose that is of great interest to exercise and public health professionals, not only because it is "the most important exercise prescription variable to maintain a cardiovascular training response" (American College of Sports Medicine, 2006, p. 161) and is "associated with an increased risk of orthopaedic injury [and] cardiovascular incidence" (American College of Sports Medicine, 2006, p. 147), but also because it is inversely related to adherence (Cox, Burke, Gorely, Beilin, & Puddey, 2003; Lee *et al.*, 1996; Perri *et al.*, 2002). One possible explanation for the negative intensity–adherence relationship, which is relevant to the present investigation, is that, as the intensity of exercise increases, the pleasure that participants experience is typically reduced, particularly at intensities that exceed the ventilatory or lactate threshold (Ekkekakis, Hall, & Petruzzello, 2005a; Ekkekakis & Petruzzello, 1999). In turn, reduced pleasure could lower the intrinsic motivation for future participation.

Although researchers have shown that a relationship between exercise intensity and affective responses (such as pleasure-displeasure) exists, authors have questioned whether this relationship can be meaningfully described by a single, universal dose–response curve (Ekkekakis & Petruzzello, 1999; Ekkekakis *et al.*, 2005a; Rejeski, 1994). This argument is based on evidence that individuals differ greatly in the intensity of exercise that they prefer (i.e. that which would elicit an optimal affective response) and the intensity they can tolerate (i.e. that which would still permit the maintenance of a pleasant affective state). For example, with respect to preference for exercise intensity, Lind and colleagues (Lind, Joens-Matre, & Ekkekakis, 2005) found that, on average, middle-aged women selected an intensity corresponding to their ventilatory threshold but individuals ranged from 60% of the oxygen uptake associated with the ventilatory threshold to 160%. Similarly, with respect to tolerance, Gulati *et al.* (2005) reported that, in their cohort of 5721 asymptomatic women, the average peak exercise

Correspondence: P. Ekkekakis, Department of Kinesiology, Iowa State University, 253 Barbara E. Forker Building, Ames, IA 50011, USA.
E-mail: ekkekaki@iastate.edu

ISSN 0264-0414 print/ISSN 1466-447X online © 2008 Taylor & Francis
DOI: 10.1080/02640410701624523

capacity determined from a symptom-limited treadmill test was 8.0 metabolic equivalents, but individual values ranged from a low of 1.4 to a high of 20.0 (17.5% to 250.0% of average, or 20.0% to 150.0% after taking age into account).

Besides the role played by age, physiological and health-related factors, it is assumed that the variation in the intensities of exercise that different individuals prefer or can tolerate is partly due to psychological traits (presumably with a genetic basis) and partly due to life experiences, social and cultural influences, and associated patterns of situational appraisals (e.g. self-presentation or acquired fear). Of these two possibilities, until now research attention has focused mainly on the latter. For example, the construct of self-efficacy, which is central in social-cognitive theory, has been shown to account for significant portions of inter-individual variation in affective responses and perceptions of exertion (e.g. McAuley, Talbot, & Martinez, 1999; Pender, Bar-Or, Wilk, & Mitchell, 2002; Rudolph & McAuley, 1996). On the other hand, there has been limited research attention to the role played by psychological traits and dispositional factors. This is surprising given the fact that psychological theories of personality and temperament include several variables of apparent relevance, namely factors hypothesised to influence arousability and sensory modulation (e.g. extraversion – introversion, sensation-seeking, strength of the nervous system, perceptual augmenting – reducing).

This relative inattention to the role of psychological traits is not unique to the exercise literature. It has also been observed in the pain literature (Phillips & Gatchel, 2000), leading one to suspect that this commonality (both exercise and pain being primarily somatic stimuli) is not coincidental. Specifically, it has been suggested that the items in the standard self-report questionnaires that have been developed for the assessment of these theoretically relevant traits focus almost entirely on responses to exteroceptive stimuli (e.g. visual, auditory, tactile) and the manifestations of these traits in social behaviour (e.g. sociability), to the exclusion of stimulation that arises from within the body, such as during vigorous exercise (Ekkekakis, Hall, & Petruzzello, 2005b). However, it is clear that interoceptive and exteroceptive sensory systems have very distinct anatomical and physiological characteristics (Craig, 2002) and, consequently, extrapolations from one domain to the other would be imprudent.

Thus, based on the proposal that there is a need for an exercise-specific measure of individual differences in preference for and tolerance of exercise intensity, researchers recently developed the Preference for and Tolerance of the Intensity of Exercise Questionnaire (PRETIE-Q; Ekkekakis *et al.*, 2005b). Preference for exercise intensity was defined as a "predisposition to select a particular level of exercise intensity when given the opportunity (e.g. when engaging in self-selected or unsupervised exercise)", whereas tolerance of exercise intensity was defined as "a trait that influences one's ability to continue exercising at an imposed level of intensity beyond the point at which the activity becomes uncomfortable or unpleasant" (Ekkekakis *et al.*, 2005b, p. 354). It is crucial to clarify that intensity-preference and intensity-tolerance are not considered the sole (or the most important) determinants of intensity preference or tolerance. Other factors, including physical (e.g. age, fitness, health status), experiential (e.g. learned coping skills, exercise history), and situational (e.g. self-efficacy, social physique anxiety, perceived social evaluation) factors, should be expected to play important roles as well. The conceptualisation of intensity-preference and intensity-tolerance as traits (presumably with a biological and genetic basis) was based on a series of arguments, including the fact that large and temporally consistent inter-individual variability in preference for and tolerance of exercise intensity, beyond what can be accounted for by key somatometric and physiological differences, is observed in humans, as well as in other species.

The two factors of the PRETIE-Q, consisting of eight items each, have been shown to be internally consistent, temporally stable (over 3- and 4-month intervals), and structurally valid. Both the Preference and the Tolerance factors were also found to be related cross-sectionally to the intensity of habitual physical activity. Furthermore, although both the Preference and Tolerance scales were significantly related to affective responses to a 15-min bout of running at the intensity of the ventilatory threshold, only the Tolerance scale was significantly related to affective responses when the intensity was even higher (10% of aerobic capacity above the ventilatory threshold), thus providing evidence of discriminant validity. Similarly, the Preference – but not the Tolerance – scale significantly predicted the intensity of physical activity (percentage of oxygen uptake at the ventilatory threshold) that a sample of middle-aged women selected (Ekkekakis, Lind, & Joens-Matre, 2006). Finally, the Tolerance scale and, to a lesser extent, the Preference scale predicted the period of time two samples of individuals persevered during incremental treadmill tests to volitional exhaustion beyond the point at which they reached their ventilatory threshold (Ekkekakis, Lind, Hall, & Petruzzello, 2007).

This early evidence suggests that the PRETIE-Q could be useful as a research tool for capturing part of the inter-individual variability in the preference for and tolerance of exercise intensity. Based on the principle that psychometric evaluation is a continuous process and that ongoing validation with independent samples is essential, our aim in this study was to examine the psychometric properties of the PRETIE-Q in a large sample of young women. Specifically, we provide information on internal consistency, structural validity (by means of a confirmatory factor analysis), and construct validity (by means of cross-sectional correlations with self-reported physical activity habits). Moreover, given the large size of the sample (much larger than the samples used in the initial psychometric evaluation of the PRETIE-Q), we provide norms for college-age women.

Methods

Participants

The participants were 601 female undergraduate and postgraduate students at a large university in the Midwestern United States. They were approached at the university fitness centre, campus sorority organisations, and classes in educational psychology, psychology, and kinesiology. Before participating, they read and signed an informed consent form that had been approved by the university's Institutional Review Board. The participants had a mean age of 20.1 years ($s = 1.4$; range 17–29).

The participants represented the following ethnic or racial groups: 77.0% Caucasian, 8.0% Asian or Asian-American, 6.0% African-American, 5.7% Latinas, 0.3% Native American, 2.0% biracial, and 1.0% "other". The ethnic/racial composition was representative of the campus student population. According to enrolment statistics, the student body consisted of 77.0% Caucasian, 6.1% African-American, 5.9% Latino/a, 10.7% Asian-American, and 0.3% Native American. With the exception of a lower proportion of African-Americans, the sample was also representative of the racial composition of the local community, according to Census Bureau statistics. Specifically, the 89,363 women in the local community were 80.8% Caucasian, 12.4% African-American, 6.4% Asian or Asian-American, 0.6% Native American, and 1.6% "other".

The mean physical characteristics of the participants in the sample were as follows: height 1.66 m ($s = 0.07$), body mass 60.6 kg ($s = 9.0$), body mass index (BMI) 21.99 $kg \cdot m^{-2}$ ($s = 2.94$). Of all participants, 88% were classified as being of normal weight (BMI < 25 $kg \cdot m^{-2}$), 10% as overweight (BMI ≥ 25 $kg \cdot m^{-2}$), and 2% as obese (BMI ≥ 30 $kg \cdot m^{-2}$). These figures approximate those based on a sample of women from another large university in the Midwestern United States (Huang *et al.*, 2003). In that sample, 79.7–85.6% of the female students were deemed to be of normal weight, 11.3–16.1% as overweight, and 3.1–4.2% as obese. On the other hand, in the United States overall, only 48% of women aged 20–39 years are of normal weight, whereas 23% are overweight and 29% are obese (Ogden *et al.*, 2006).

Finally, more than half (59.6%) of the participants reported being physically active at least three days per week. Of those who were active on three or more days, 20.3% were involved in intramural sports and 6.8% were involved in college sports. In the United States overall, only 15.0% of college women are physically inactive, the lowest percentage among 23 countries examined by Haase and colleagues (Haase, Steptoe, Sallis, & Wardle, 2004). According to data from the 2001 Behavioural Risk Factor Surveillance System, 49.8% of women in the United States aged 18–29 years report being moderately or vigorously active on a regular basis (Macera *et al.*, 2005).

Of the 601 women who responded, 572 had complete responses on the preference-related items of the PRETIE-Q, 580 had complete responses on the tolerance-related items, and 557 had complete responses on all PRETIE-Q items. The 44 women with incomplete PRETIE-Q data and the 557 with complete data did not differ significantly in terms of age, height, body mass, body mass index, or the likelihood of being physically active on a regular basis.

Measures

The PRETIE-Q was used to assess the variables of preference for and tolerance of exercise intensity. The eight-item Preference scale contains four items that tap preference for high intensity (e.g. "I would rather have a short, intense workout than a long, low-intensity workout") and four that tap preference for low intensity (e.g. "When I exercise, I usually prefer a slow, steady pace"). Similarly, the eight-item Tolerance scale contains four items that tap high exercise tolerance (e.g. "I always push through muscle soreness and fatigue when working out") and four that tap low exercise tolerance (e.g. "During exercise, if my muscles begin to burn excessively or if I find myself breathing very hard, it is time for me to ease off"). The complete questionnaire is available in Ekkekakis *et al.* (2005b). Each item is accompanied by a 5-point response scale (1 = "I totally disagree"; 2 = "I disagree"; 3 = "Neither agree or disagree"; 4 = "I agree"; 5 = "I strongly agree"). Items indicative of preference for low intensity (Items 2, 4, 8, 12) and items indicative of low tolerance (Items 1, 3, 9, 13) are reversed-scored. Ekkekakis *et al.* (2005b) reported alpha coefficients of internal consistency ranging from 0.81 to 0.85 for the Preference scale and from 0.82 to 0.87 for the Tolerance scale. Furthermore, test–retest reliability coefficients of 0.67 and 0.80 for the Preference scale and 0.85 and 0.72 for the Tolerance scale were reported after 3- and 4-month intervals, respectively.

The Godin Leisure-Time Exercise Questionnaire (Godin & Shephard, 1985) was used to assess leisure-time exercise habits. It includes three questions on how many times, during a typical 7-day period, the respondent participates in: (a) "strenuous exercise (heart beats rapidly)", such as running, jogging, hockey, football, soccer, squash, basketball, and cross-country skiing; (b) "moderate exercise (not exhausting)", such as fast walking, baseball, tennis, easy cycling, volleyball, badminton, and easy swimming; and (c) "mild exercise (minimal effort)", such as yoga, archery, fishing from river bank, bowling, horseshoes, and golf. Weekly frequencies are multiplied by 9, 5, and 3 for strenuous, moderate, and mild exercise, respectively, to calculate a composite score. The questionnaire also includes one item inquiring how often ("often", "sometimes", or "never/rarely"), during a typical 7-day period, the respondent engages in any regular activity "long enough to work up a sweat". Godin and Shephard (1985) reported 2-week test–retest reliability coefficients of 0.48 for mild, 0.46 for moderate, and 0.94 for strenuous exercise. Strenuous exercise ($r = 0.38$ and $r = 0.21$), the total score ($r = 0.24$ and $r = 0.13$), and the frequency of sweating ($r = 0.26$ and $r = 0.21$) correlated significantly with maximal aerobic capacity and percentage of body fat (respectively).

Results

Mean scores, distributional characteristics and normative data

On a possible range from 8 to 40, the mean score was 24.22 ($s = 6.47$) for Preference and 23.04 ($s = 5.96$) for Tolerance. For Preference, the median was 24 and the mode was 24. For Tolerance, the median was 23 and the mode was 19.

Both Preference and Tolerance were normally distributed. For Preference, skewness was −0.082 and kurtosis was −0.296. For Tolerance, skewness was 0.114 and kurtosis was −0.298. For both scores, responses covered the entire possible range, from 8 to 40. The norms (percentile rankings) for Preference and Tolerance scores are shown in Table I.

Item means and standard deviations, as well as estimates of skewness and kurtosis, are shown in Table II. The skewness of all items was within ± 0.600 (on average, −0.031 for Preference and 0.093 for Tolerance), whereas kurtosis was within ± 1.00 (on average, −0.640 for Preference and −0.811 for Tolerance). The negative kurtosis values for all items indicate distributions that were somewhat platykurtic (flat), although by a modest amount.

Internal consistency

Cronbach's alpha coefficient of internal consistency was 0.89 for the Preference scale and 0.86 for the Tolerance scale. As shown in Table III, each item made a positive contribution to the internal consistency of its scale, since the deletion of any item would result in a decrease in the value of the coefficient for the scale to which it belongs.

The mean inter-item correlation was 0.502 (from 0.373 to 0.709) for Preference and 0.441 (from 0.309 to 0.596) for Tolerance. The mean corrected item-total correlation was 0.664 (from 0.580 to 0.769) for Preference and 0.611 for Tolerance

Table I. Percentile rankings (%) of Preference (Pref) and Tolerance (Tol) scores for college-age women ($N = 557$).

%	Pref	Tol	%	Pref	Tol	%	Pref	Tol	%	Pref	Tol	%	Pref	Tol
1	9.0	9.8	21	18.0	18.0	41	23.0	22.0	61	26.0	25.0	81	30.0	29.0
2	9.0	12.0	22	19.0	18.0	42	23.0	22.0	62	26.0	25.0	82	31.0	29.0
3	11.2	12.0	23	19.0	18.0	43	23.0	22.0	63	27.0	25.0	83	31.0	29.0
4	13.0	13.0	24	19.5	19.0	44	23.0	22.0	64	27.0	25.0	84	31.0	29.0
5	14.0	13.0	25	20.0	19.0	45	23.0	22.0	65	27.0	25.0	85	31.0	29.0
6	14.0	14.0	26	20.0	19.0	46	24.0	22.0	66	27.0	25.0	86	31.0	30.0
7	15.0	14.0	27	20.0	19.0	47	24.0	22.0	67	27.0	26.0	87	32.0	30.0
8	15.0	15.0	28	20.0	19.0	48	24.0	23.0	68	27.0	26.0	88	32.0	30.0
9	16.0	15.0	29	21.0	19.0	49	24.0	23.0	69	28.0	26.0	89	32.0	31.0
10	16.0	16.0	30	21.0	19.0	50	24.0	23.0	70	28.0	26.0	90	32.0	31.0
11	16.0	16.0	31	21.0	20.0	51	24.0	23.0	71	28.0	26.0	91	33.0	31.0
12	17.0	16.0	32	21.0	20.0	52	25.0	23.0	72	28.0	27.0	92	33.0	32.0
13	17.0	16.5	33	21.0	20.0	53	25.0	23.0	73	29.0	27.0	93	33.9	32.0
14	17.0	17.0	34	22.0	20.0	54	25.0	24.0	74	29.0	27.0	94	34.0	33.0
15	17.0	17.0	35	22.0	20.0	55	25.0	24.0	75	29.0	27.0	95	34.0	33.0
16	18.0	17.0	36	22.0	20.2	56	25.0	24.0	76	29.0	27.0	96	35.0	33.0
17	18.0	17.0	37	22.0	21.0	57	25.0	24.0	77	29.0	28.0	97	36.8	34.0
18	18.0	18.0	38	22.0	21.0	58	26.0	24.0	78	30.0	28.0	98	37.0	35.0
19	18.0	18.0	39	22.5	21.0	59	26.0	24.0	79	30.0	28.0	99	39.0	38.0
20	18.0	18.0	40	23.0	21.0	60	26.0	24.6	80	30.0	28.0			

Table II. Descriptive and distributional statistics for the 16 PRETIE-Q items.

PRETIE-Q item	$\bar{x}$	s	Skewness	Kurtosis
Preference				
2. I would rather work out at low-intensity levels for a long duration rather than at high-intensity levels for a short duration	2.82	1.11	0.12	−0.74
4. I'd rather go slow during my workout, even if that means taking more time	2.99	1.10	0.02	−0.80
6. I would rather have a short, intense workout than a long, low-intensity workout	2.96	1.05	0.10	−0.64
8. When I exercise, I usually prefer a slow, steady pace	3.06	1.08	−0.05	−0.87
10. Exercising at a low intensity does not appeal to me at all	2.67	1.11	0.37	−0.65
12. While exercising, I prefer activities that are slow-paced and do not require much exertion	3.58	0.98	−0.60	−0.05
14. The faster and harder the workout, the more pleasant I feel	3.20	1.03	−0.28	−0.53
16. Low-intensity exercise is boring	2.93	1.13	0.09	−0.84
Tolerance				
1. Feeling tired during exercise is my signal to slow down or stop	2.81	1.07	0.16	−0.91
3. During exercise, if my muscles begin to burn excessively or if I find myself breathing very hard, it is time for me to ease off	2.48	1.10	0.44	−0.79
5. While exercising, I try to keep going even after I feel exhausted	3.10	1.03	−0.22	−0.85
7. I block out the feeling of fatigue when exercising	2.90	1.00	−0.02	−0.71
9. I'd rather slow down or stop when a workout starts to get too tough	3.09	1.08	0.02	−0.95
11. Fatigue is the last thing that affects when I stop a workout; I have a goal and stop only when I reach it	2.87	1.10	0.00	−0.89
13. When my muscles start burning during exercise, I usually ease off some	2.80	1.01	0.45	−0.68
15. I always push through muscle soreness and fatigue when working out	2.98	0.98	−0.09	−0.71

Table III. Item analysis for the 16 PRETIE-Q items.

Preference				Tolerance			
Item #	Corrected item-total correlation	Squared multiple correlation	Alpha if item deleted	Item #	Corrected item-total correlation	Squared multiple correlation	Alpha if item deleted
2	0.682	0.579	0.874	1	0.605	0.380	0.846
4	0.735	0.606	0.869	3	0.562	0.371	0.851
6	0.650	0.493	0.878	5	0.571	0.374	0.850
8	0.769	0.624	0.866	7	0.671	0.504	0.839
10	0.659	0.480	0.877	9	0.603	0.384	0.847
12	0.580	0.389	0.884	11	0.627	0.459	0.844
14	0.602	0.409	0.882	13	0.573	0.384	0.850
16	0.637	0.454	0.879	15	0.675	0.515	0.839
$\bar{x}$	0.664	0.504		$\bar{x}$	0.611	0.421	

(from 0.562 to 0.675). The mean squared multiple correlation was 0.504 (from 0.389 to 0.624) for Preference and 0.421 (from 0.371 to 0.515) for Tolerance.

Confirmatory factor analysis

A confirmatory factor analysis was conducted using Amos 6.0 (Arbuckle, 2005) with the maximum likelihood method of estimation. The correlation among the Preference and Tolerance latent factors was not constrained but was left to be estimated. The following fit indices were considered.

First, the χ^2 index represents the ratio of the sample covariance matrix to the hypothesised covariance matrix (Bollen, 1989). Under the assumption of multivariate normality, this index follows the χ^2 distribution and is thus amenable to testing the null hypothesis that all residuals from the model are zero (i.e. that the model fits perfectly in the population). However, it has become evident that theoretical models at best approximate estimated population values and, thus, particularly in large samples (necessary to obtain stable parameter estimates), the null hypothesis of zero discrepancy will always be rejected. Therefore, alternative fit indices have been proposed that are not based on hypothesis-testing, but rather indicate the degree of model misspecification.

Second, the goodness-of-fit index (GFI) represents a measure of "absolute fit" and assesses the extent to which the hypothesised model can reproduce the sample variance/covariance matrix (Jöreskog & Sörbom, 1993). It can be interpreted in a manner analogous to the coefficient of determination (R^2) in multiple regression analyses. For the maximum likelihood model of estimation and samples sizes > 250, a level of 0.90 or higher is considered indicative of adequate fit (Hu & Bentler, 1995). In contrast, according to Bentler and Bonett (1980), "models with overall fit indices of less than .9 can usually be improved substantially" (p. 600).

Third, the comparative fit index (CFI) is a type of incremental fit index, measuring the improvement in fit offered by the hypothesised model over a baseline model, usually the "null" or independence model (Bentler, 1990). For the maximum likelihood model of estimation and samples sizes > 250, a level of 0.90 or higher is considered indicative of adequate fit (Hu & Bentler, 1995).

Fourth, the root mean square error of approximation (RMSEA) is a measure of "badness of fit" or, more specifically, a "measure of discrepancy per degree of freedom" (Browne & Cudeck, 1992, p. 238). A value of 0.05 or lower is considered indicative of "close fit" and one of 0.08 is considered indicative of "reasonable fit" (Browne & Cudeck, 1992).

It should be noted that, based on an extensive Monte-Carlo analysis, Hu and Bentler (1998, 1999) have suggested that adopting more stringent cut-off values than those cited above (i.e. 0.95 for CFI and 0.06 for RMSEA) would reduce the chance of committing Type I and Type II errors. These suggestions have sparked considerable controversy, since the distributions of these goodness-of-fit indices and the factors that affect them remain unknown (Beauducel & Wittmann, 2005; Sivo, Fan, Witta, & Willse, 2006; Yuan, 2005). It has also been argued that the proposed cut-off values "appear to be largely unobtainable in appropriate practice" (Marsh, Hau, & Wen, 2004, p. 326). Given this controversy, the fit of the PRETIE-Q model to the data was evaluated mainly on the basis of whether any misspecifications were attributable to serious problems, such as item cross-loadings, and whether the overall fit was comparable to that attained in the original validation sample.

Three models were examined. Model A was a single-factor model, in which all 16 items represented indicators of the same latent construct. This model was examined because, by not constraining the inter-factor correlation to independence (given that preference and tolerance were conceptualised as oblique factors), it would be possible for a two-factor solution to show better fit than a single-factor solution only as a result of reduced parsimony. Model B was a two-factor model consisting of one eight-item Preference factor and one eight-item Tolerance factor, with no correlated errors. Model C was a two-factor model that included the four correlated errors identified in the original analysis by Ekkekakis *et al.* (2005b).

Regarding Model C, using a *post-hoc* Lagrange multiplier test, Ekkekakis *et al.* (2005b) had noted the presence of correlated errors between the following items: (a) Item 2 ("I would rather work out at low-intensity levels for a long duration than at high-intensity levels for a short duration") and Item 6 ("I would rather have a short, intense workout than a long, low-intensity workout"), with both items referring to comparisons between short and intense versus long and low-intensity exercise; (b) Item 4 ("I'd rather go slow during my workout, even if that means taking more time") and Item 8 ("When I exercise, I usually prefer a slow, steady pace"), with both items using the word "slow"; (c) Item 10 ("Exercising at a low intensity does not appeal to me at all") and Item 16 ("Low-intensity exercise is boring"), with both items referring to boredom or lack of enjoyment associated with low-intensity exercise); (d) Item 3 ("During exercise, if my muscles begin to burn excessively or if I find myself breathing very hard, it is time for me to ease off") and Item 13 ("When my muscles start burning during exercise, I usually ease off some"), with both items referring to "burning" sensations in the muscles. The first three of these items are from the Preference scale and the fourth is from the Tolerance scale. Incorporating these correlated errors in the structural equation model was considered acceptable in the original analysis on the basis of the similarities in the content and/or structure of the items involved.

There were no serious deviations from univariate normality that could jeopardise the accuracy of the maximum likelihood estimates, based on the skewness and kurtosis values presented in Table II. According to a Monte-Carlo study that examined the effects of "slight" (two-thirds of the variables having skewness and kurtosis of about $\pm$ 1.0) and "moderate" non-normality (two-thirds of the variables having skewness of about $\pm$ 1.5 and kurtosis between +3 and +4), departures from normality of such magnitudes resulted in an "almost complete absence of any obvious adverse effect" (Fan & Wang, 1998, p. 730). Mardia's (1970) normalised estimate of multivariate skewness was also computed following the procedure developed by DeCarlo (1997). Although this coefficient, at 17.41, indicated some departure from multivariate normality, its magnitude did not suggest severe non-normality (particularly considering that the index is known to be positively affected by sample size). Nevertheless, in addition to the regular maximum-likelihood method of estimation, we used EQS (Bentler, 1995) to also compute test statistics with the corrections developed by Satorra and Bentler as an alternative to asymptotic methods for the analysis of non-normal data in small to medium-sized samples (Chou, Bentler, & Satorra, 1991; Hu, Bentler, & Kano, 1992; Satorra & Bentler, 1994).

As shown in Table IV, Models A and B showed clearly unacceptable fit. On the other hand, Model C, which allowed for four correlated errors, showed "reasonable" fit (Browne & Cudeck, 1992), with a goodness-of-fit index of 0.89, a comparative fit index of 0.91, and a root mean square error of approximation of 0.08 (note that the comparative fit index based on the Satorra-Bentler correction did not differ until the third decimal). However, it should be noted that these indices were lower than those reported by Ekkekakis *et al.* (2005b) based on a much smaller ($N = 184$) sample of both male and female students (GFI = 0.92, CFI = 0.97, RMSEA = 0.04). Model C is represented graphically in Figure 1. As can be seen in this figure, the standardised factor loadings were acceptably high (from 0.65 to 0.81 for Preference and from 0.56 to 0.76 for Tolerance). The correlation between the latent factors of Preference and Tolerance was 0.59. Although this was higher than the values (0.42 and 0.45) reported by Ekkekakis *et al.* (2005b), the clearly unacceptable fit of the single-factor solution (Model A) suggests that a two-factor model fit the data much better than a single-factor model.

The modification indices were examined to determine whether the same correlated errors identified in Ekkekakis and colleagues' (2005b) analysis also had the largest modification indices in the present analysis. Of the four correlated errors, three (those associated with Items 2–6, 10–16, and 3–13) appeared again in the top three positions,

Table IV. Degrees of freedom (d.f.), χ^2, Satorra-Bentler scaled χ^2 (SB χ^2), goodness-of-fit index (GFI), comparative fit index (CFI), EQS "robust" comparative fit index (CFI*), and root mean square error of approximation (RMSEA) with associated 90% confidence intervals for the three covariance structure models tested.

	d.f.	χ^2	SB χ^2	GFI	CFI	CFI*	RMSEA
Model A	104	1529.4	1204	0.63	0.67	0.68	0.16 (0.15–0.16)
Model B	103	697.4	555.3	0.84	0.86	0.87	0.10 (0.09–0.11)
Model C	99	492.7	403.2	0.89	0.91	0.91	0.08 (0.07–0.09)

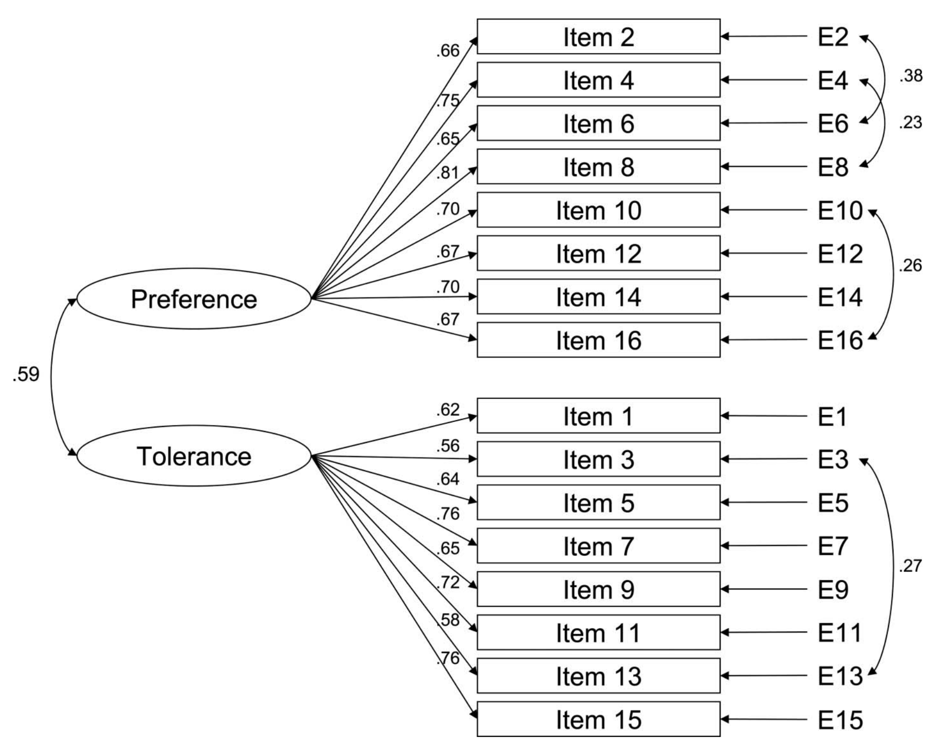

Figure 1. Results of the confirmatory factor analysis of the PRETIE-Q (Model C). The numbers represent standardised coefficients.

whereas one (associated with Items 4–8) was at a lower position. The correlated errors between Items 2–6, 10–16, and 3–13 had modification indices of 64.04, 58.86, and 33.23 respectively, whereas the modification index for the errors of Items 4–8 was 12.37. The fourth largest modification index (26.00) referred to the correlated errors of Items 2–4. However, a new analysis without correlated errors between Items 4–8 but with correlated errors between Items 2–4 resulted in identical GFI, CFI, and RMSEA indices as Model C (despite the somewhat lower χ^2, consistent with the larger modification index). Therefore, no additional re-specifications of the model were deemed necessary. As suggested by MacCallum and colleagues (MacCallum, Roznowski, & Necowitz, 1992), modifications based on smaller discrepancies between the model and the data often tend to be sample-specific. A series of χ^2 difference tests was conducted, allowing only one of the four pairs of correlated errors at a time (i.e. with one degree of freedom), following the order suggested by the modifications indices. All were significant ($P < 0.001$), suggesting that each parameter helped to improve the fit of the model to a significant degree.

Since the goodness-of-fit indices for Model C suggested that the fit was only marginally "acceptable" by conventional guidelines, the modification indices were also examined for any fundamental structural problems, such as cross-loading items. No such parameter appeared among the top 15 suggested modifications. Furthermore, an examination of the matrix of residual item covariances did not single out any items as being outstanding contributors to model misspecification. Of the 136 residuals, the one item that appeared most frequently (5 times) in the list of the 20 largest residuals was Item 14 ("The faster and harder the workout, the more pleasant I feel"). However, upon closer inspection, its average residual was only 0.077, not much higher than the average residual across all items of 0.053. Overall, it appears that the two-factor model with the four correlated errors fit the data "reasonably" well, without any fundamental structural problems (i.e. item cross-loadings) and with no single badly fitting item.

Correlations with physical activity

The following types of association of the Preference and Tolerance scores with the variables in the Godin Leisure-Time Exercise Questionnaire were examined. First, we computed correlations with the number of times per typical week the women reported participating in "strenuous", "moderate", and "mild" exercise, as well as with the total score derived from the questionnaire. We hypothesised that the Preference and Tolerance scales of the PRETIE-Q would correlate with the tendency to engage in "strenuous" (i.e. more intense) types of exercise (e.g. running, jogging, hockey, football, soccer, squash, basketball, cross-country skiing). Second, we examined whether the women who reported working out "long enough to work up a sweat" – "often", "sometimes", and "never/rarely" – differed in terms of their Preference and Tolerance scores. We hypothesised that those participants who reported working up a sweat "often" would tend to have higher Preference and Tolerance scores than those who did not "work up a sweat" as frequently in a typical week. Third, we identified the women whose most frequent type of exercise in a typical week was "strenuous", "moderate", or "mild" and examined whether their Preference and Tolerance scores differed. We hypothesised that those who indicated a clear preference for "strenuous" activities (i.e. participated in such activities more often than less intense activities per typical week) would report higher Preference and Tolerance scores.

Both Preference ($r = 0.299$ and $r = 0.181$, both $P < .001$) and Tolerance ($r = 0.355$ and $r = 0.301$, both $P < 0.001$) were significantly related to the frequency of strenuous exercise and the total score (respectively). On the other hand, Preference ($r = -0.007$ and $r = -0.009$) and Tolerance ($r = 0.112$ and $r = 0.072$) were unrelated to the number of times per typical week that the women participated in moderate or mild exercise (respectively).

Both Preference ($r = 0.210$, $P < 0.001$) and Tolerance ($r = 0.314$, $P < 0.001$) were related to the number of times per week the women reported working out "long enough to work up a sweat". Also, those who reported working out long enough to work up a sweat "often" had the highest Preference (26.37) and Tolerance (25.75) scores, whereas the individuals who responded "never/rarely" had the lowest scores (22.22 and 20.30, respectively). The one-way analysis of variance was significant for both Preference ($F_{2,556} = 17.76$, $P < 0.001$) and Tolerance ($F_{2,565} = 37.90$, $P < 0.001$). All the Bonferroni-corrected inter-group differences were significant (see Figure 2 for significant differences and effect sizes).

A total of 289 women were identified who had complete Preference and Tolerance data and indicated a clear preference for participating in strenuous, moderate, or mild exercise (i.e. the women who reported an equal number of times per typical week for two or more types of exercise were eliminated). Then, the Preference and Tolerance scores of the 48 women who preferred strenuous exercise, the 67 women who preferred moderate exercise, and the 174 women who preferred mild

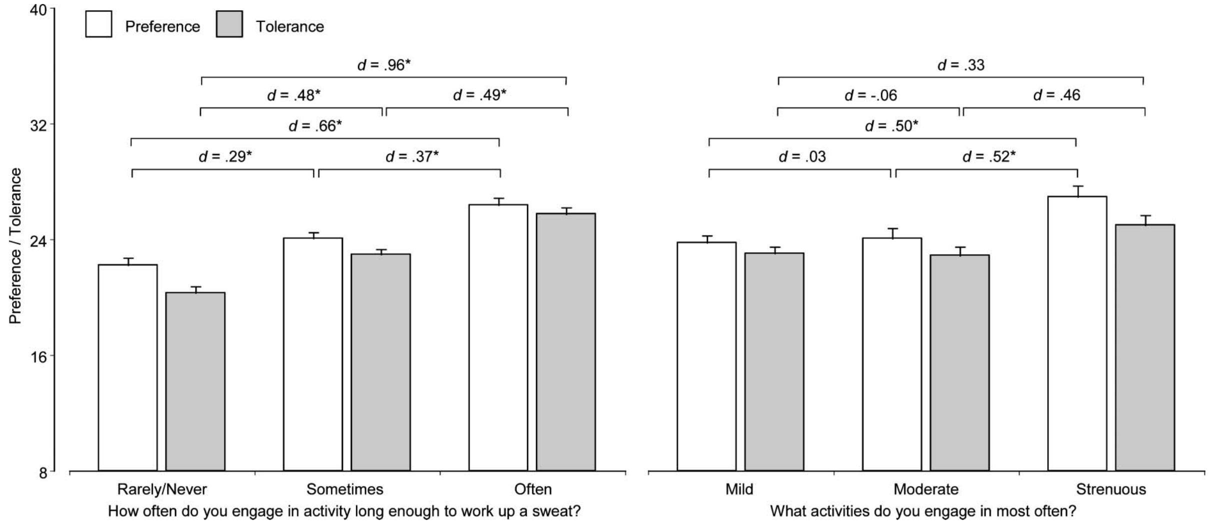

Figure 2. Relationships of Preference and Tolerance with variables from the Godin Leisure-Time Exercise Questionnaire. (Left) Differences in Preference and Tolerance between women who reported exercising "long enough to work up a sweat" "rarely/never", "sometimes", or "often". (Right) Differences in Preference and Tolerance between women who reported a preference for mild, moderate, or strenuous exercise. Bonferroni-corrected statistically significant comparisons are indicated by an asterisk (*). Effect sizes were computed as $d = (\bar{x}_i - \bar{x}_j)/s_{pooled}$.

exercise were compared. The one-way analysis of variance was significant for Preference ($F_{2,286} = 5.00$, $P < 0.01$), but not for Tolerance ($F_{2,286} = 2.66$, $P = 0.07$). *Post-hoc* comparisons showed that the women who preferred strenuous exercise had significantly higher Preference scores (26.85) than those who preferred moderate (23.88) or mild exercise (23.71), whereas those in the latter two groups did not differ (for significant differences and effect sizes, see Figure 2). The pattern for Tolerance was the same (scores of 25.02, 22.76, 23.12, respectively), but the differences were less marked.

Discussion

Individuals differ in the intensity of exercise they prefer and the intensity they can tolerate. The extent of this variability is so large that it is unlikely to be fully accounted for by differences in physiological, somatometric, or health-related variables. In other words, presumably, some part of this variability must lie within the psychological domain. The psychological sources of this variability probably encompass a wide range of factors, including dispositions (e.g. traits related to arousability or somatosensory modulation, with a possible genetic basis), past experiences, and situational influences. The present investigation focused only on the first of these possibilities, namely the dispositional factors of preference for and tolerance of exercise intensity, as these were operationally defined by the PRETIE-Q (Ekkekakis *et al.*, 2005b). The purpose of the present study was to extend the initial psychometric evaluation of the PRETIE-Q by examining its internal consistency, structural validity, and construct validity in an independent sample of college women.

The results of the psychometric analyses reported here largely agree with the initial assessment that the PRETIE-Q is an internally consistent and structurally valid self-report instrument that can continue to be tested in research and clinical contexts (Ekkekakis *et al.*, 2005b). The indices of internal consistency (0.89 for Preference and 0.86 for Tolerance) are very satisfactory for eight-item scales and similar to those reported previously (i.e. 0.81 to 0.85 for Preference and 0.82 to 0.87 for Tolerance). The item analysis showed that all items related well with others in their respective scales and all contributed positively to internal consistency.

The results of the confirmatory factor analysis were also in general agreement with those reported by Ekkekakis *et al.* (2005b), although overall model fit was lower. As was the case in the initial analysis, the modelling of correlated errors among certain items improved the fit to a reasonable extent. The phenomenon of correlated errors implies that measurement error might not be totally random (as error should be, by definition) but might instead be due to a systematic factor. Common examples of such factors include response sets (Bollen, 1989, p. 232), the response format used (Byrne, 1991, p. 594), or, as in the case of the PRETIE-Q, features such as item content and sentence structure that certain pairs of items have in common but are not shared with other items within the same factor. An examination of the item content of the four items with correlated errors provides a good indication about the causes of this phenomenon, as there are some obvious thematic or syntactical similarities between the items in each pair. Although Ekkekakis *et al.* (2005b) used the minimisation of content overlap as one of the preliminary item selection criteria (see criterion "a" on p. 357), the final selection of items was based on empirical criteria (i.e. item loadings and cross-loadings in an exploratory factor analysis).

The practice of allowing errors to correlate, particularly on the basis of *post-hoc* modification indices, has been criticised as possibly capitalising on chance. However, others have argued that a model that does not allow for correlated errors is "highly restricted" and "rarely appropriate for real data" (Bentler & Chou, 1987, p. 108). Therefore, this can be deemed acceptable but only as long as such decisions are "supported by a strong substantive and/or empirical rationale" (Byrne, 1994, p. 299). As noted, the thematic and syntactical similarities between the items in each pair provided a good empirical rationale in this case (see the discussion of these similarities in the "Confirmatory factor analysis" section). MacCallum and Austin (2000) further noted that, because of the risk of capitalisation on chance, "the modified model must be evaluated by fitting it to an independent sample" (p. 217). The results of the *post-hoc* analysis in this study were generally consistent with those reported by Ekkekakis *et al.* (2005b), providing evidence that the correlated error parameters are stable rather than sample-specific. It is important to point out that the presence of correlated errors *per se* does not have direct negative implications for the structural validity of a measure (other than reducing the parsimony of the model by requiring additional parameters to be modelled). Therefore, this information is not of interest to potential users of the questionnaire but rather to researchers investigating the measurement model of the PRETIE-Q within structural equation models.

One feature of the PRETIE-Q that is of considerable theoretical and practical interest is the relationship between the Preference and Tolerance factors. Ekkekakis *et al.* (2005b) had postulated that Preference and Tolerance are oblique (partially correlated) but distinct constructs. In this study, the two scores were related 0.49, which is at the high end of the range observed in the initial series of psychometric investigations (i.e. between 0.40 and 0.50). Similarly, the correlation between the Preference and Tolerance latent factors in the confirmatory factor analysis, which is theorised to be error-free, was higher in this (0.59) than in the earlier analysis (0.45). Although the two factors are always found to be related, as hypothesised, the magnitude of this relationship certainly allows for a considerable amount of distinct variance within each. In particular, the distinct role of Preference and Tolerance emerges in studies that have specifically operationalised these constructs, such as in the case of self-selection of exercise intensity, which pertains specifically to the construct of intensity-preference (Ekkekakis *et al.*, 2006), and the perseverance during incremental treadmill tests to fatigue, which pertains specifically to the construct of intensity-tolerance (Ekkekakis *et al.*, 2007). In the former case, the Preference scale has been shown to account for 18–19% of the variance in physiologically defined self-selected exercise intensity, beyond the variance accounted for by age, body mass index, and cardiorespiratory fitness (Ekkekakis *et al.*, 2006). In the latter case, the Tolerance scale has been shown to account for 14–20% of the variance in the amount of time individuals persisted during incremental treadmill tests after reaching their ventilatory threshold, beyond the variance accounted for by age, body mass index, physical activity habits, and even cardiorespiratory fitness (Ekkekakis *et al.*, 2007). On the other hand, both Preference and Tolerance relate to broad outcomes, such as physical activity participation in the vigorous or strenuous range, to an almost equal extent, as was shown in the analyses based on the Godin Leisure-Time Exercise Questionnaire reported in this paper.

As the psychometric evaluation of the PRETIE-Q continues, its role in applied settings will also have to be defined. Perhaps one of the main potential areas of application would be exercise prescription. The guidelines of the American College of Sports Medicine (2006) acknowledge that both preference and tolerance need to be taken into account when developing the initial exercise prescription, as well as in adjusting the prescription as the exerciser's fitness improves. In essence, these recommendations could be interpreted as indications of a gradual move towards the formal incorporation of a third guiding principle in exercise prescription, besides the need to provide exercise stimuli that are (a) effective and (b) safe. The third consideration is to provide stimuli that are enjoyable or at least tolerable, thereby increasing the likelihood that intrinsic motivation for future participation can be maintained. As this process evolves, some formalised methods of personalisation will have to be established. This will require the use of a validated battery of measures that can help in tailoring the exercise prescription to each individual and in identifying individuals with a propensity for under-exertion (i.e. choosing exercise doses that are too low to be effective) or over-exertion (i.e. choosing exercise doses that are too high to be safe). Clearly, a lot more groundwork needs to be done before a system like this can be implemented. Essential to this process is the establishment of norms (as was done here for college-age women) and of agreed-upon criteria for what constitutes under- and over-exertion in different populations (upon which validation studies can be based).

Future psychometric investigations of the PRETIE-Q should address some of the limitations of the present study. In particular, the nature of the present sample (i.e. restricted in terms of gender, age, educational level, and socioeconomic status compared with the general population) constitutes an obvious limitation. The appropriate next step would be to perform a similar series of analyses with an equally large sample of males and to examine the factorial invariance of the questionnaire across the sexes. Subsequently, a replication with older and more diverse samples with respect to exercise experiences would be highly desirable. Finally, alternative and stronger methods of validation should be devised that go beyond cross-sectional correlations with self-reported levels of habitual physical activity. These studies will complement the laboratory-based validation work that has already been performed (Ekkekakis *et al*., 2005b, 2006, 2007).

In summary, in this psychometric evaluation with a large, independent sample of college women, the PRETIE-Q was found to possess adequate internal consistency and reasonable structural validity. Furthermore, its factors correlated cross-sectionally with the tendency to engage in vigorous activities such as running, soccer, and basketball. Together with the evidence collected in other validation studies, it appears that the PRETIE-Q could prove useful in a broad range of applications in exercise science.

References

American College of Sports Medicine (2006). *ACSM's guidelines for exercise testing and prescription* (7th edn.). Baltimore, MD: Lippincott, Williams & Wilkins.

Arbuckle, J. L. (2005). *Amos user's guide version 6.0.* Chicago, IL: SPSS.

Beauducel, A., & Wittmann, W. W. (2005). Simulation study on fit indexes in CFA based on data with slightly distorted simple structure. *Structural Equation Modeling*, *12*, 41 – 75.

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin*, *107*, 238 – 246.

Bentler, P. M. (1995). *EQS structural equations program manual.* Encino, CA: Multivariate Software.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin*, *88*, 588 – 606.

Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods and Research*, *16*, 78 – 117.

Bollen, K. A. (1989). *Structural equations with latent variables.* New York: Wiley.

Browne, M. W., & Cudeck, R. (1992). Alternative ways of assessing model fit. *Sociological Methods and Research*, *21*, 230 – 258.

Byrne, B. M. (1991). The Maslach Burnout Inventory: Validating factorial structure and invariance across intermediate, secondary, and university educators. *Multivariate Behavioral Research*, *26*, 583 – 605.

Byrne, B. M. (1994). Testing for the factorial validity, replication, and invariance of a measuring instrument: A paradigmatic application based on the Maslach Burnout Inventory. *Multivariate Behavioral Research*, *29*, 289 – 311.

Chou, C.-P., Bentler, P. M., & Satorra, A. (1991). Scaled test statistics and robust standard errors for non-normal data in covariance structure analysis: A Monte Carlo study. *British Journal of Mathematical and Statistical Psychology*, *44*, 347 – 357.

Cox, K. L., Burke, V., Gorely, T. J., Beilin, L. J., & Puddey, I. B. (2003). Controlled comparison of retention and adherence in home- vs. center-initiated exercise interventions in women ages 40 – 65 years: The S.W.E.A.T. study (Sedentary Women Exercise Adherence Trial). *Preventive Medicine*, *36*, 17 – 29.

Craig, A. D. (2002). How do you feel? Interoception: The sense of the physiological condition of the body. *Nature Reviews Neuroscience*, *3*, 655 – 666.

DeCarlo, L. T. (1997). On the meaning and use of kurtosis. *Psychological Methods*, *2*, 292 – 307.

Ekkekakis, P., Hall, E. E., & Petruzzello, S. J. (2005a). Variation and homogeneity in affective responses to physical activity of varying intensities: An alternative perspective on dose – response based on evolutionary considerations. *Journal of Sports Sciences*, *23*, 477 – 500.

Ekkekakis, P., Hall, E. E., & Petruzzello, S. J. (2005b). Some like it vigorous: Measuring individual differences in the preference for and tolerance of exercise intensity. *Journal of Sport and Exercise Psychology*, *27*, 350 – 374.

Ekkekakis, P., Lind, E., Hall, E. E., & Petruzzello, S. J. (2007). Can self-reported tolerance of exercise intensity play a role in exercise testing? *Medicine and Science in Sports and Exercise*, *39*, 1193 – 1199.

Ekkekakis, P., Lind, E., & Joens-Matre, R. R. (2006). Can self-reported preference for exercise intensity predict physiologically defined self-selected exercise intensity? *Research Quarterly for Exercise and Sport*, *77*, 81 – 90.

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

Fan, X., & Wang, L. (1998). Effects of potential confounding factors on fit indices and parameter estimates for true and misspecified SEM models. *Educational and Psychological Measurement*, *58*, 701 – 735.

Godin, G., & Shephard, R. J. (1985). A simple method to assess exercise behavior in the community. *Canadian Journal of Applied Sports Sciences*, *10*, 141 – 146.

Gulati, M., Black, H. R., Shaw, L. J., Arnsdorf, M. F., Bairey Merz, C. N., Lauer, M. S. *et al*. (2005). The prognostic value of a nomogram for exercise capacity in women. *New England Journal of Medicine*, *353*, 468 – 475.

Haase, A., Steptoe, A., Sallis, J. F., & Wardle, J. (2004). Leisure-time physical activity in university students from 23 countries: Associations with health beliefs, risk awareness, and national economic development. *Preventive Medicine*, *39*, 182 – 190.

Hu, L.-T., & Bentler, P. M. (1995). Evaluating model fit. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 76 – 99). Thousand Oaks, CA: Sage.

Hu, L., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, *3*, 424 – 453.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, *6*, 1 – 55.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin*, *112*, 351–362.

Huang, T. T.-K., Harris, K. J., Lee, R. E., Nazir, N., Born, W., & Kaur, H. (2003). Assessing overweight, obesity, diet, and physical activity in college students. *Journal of American College Health*, *52*, 83–86.

Jöreskog, K., & Sörbom, D. (1993). *LISREL 8: Structural equation modeling with the SIMPLIS command language.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Lee, J. Y., Jensen, B. E., Oberman, A., Fletcher, G. F., Fletcher, B. J., & Raczynski, J. M. (1996). Adherence in the Training Levels Comparison Trial. *Medicine and Science in Sports and Exercise*, *28*, 47–52.

Lind, E., Joens-Matre, R. R., & Ekkekakis, P. (2005). What intensity of physical activity do formerly sedentary middle-aged women select? Evidence of a coherent pattern from physiological, perceptual, and affective markers. *Preventive Medicine*, *40*, 407–419.

MacCallum, R. C., & Austin, J. T. (2000). Applications of structural equation modeling in psychological research. *Annual Review of Psychology*, *51*, 201–226.

MacCallum, R. C., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin*, *111*, 490–504.

Macera, C. A., Ham, S. A., Yore, M. M., Jones, D. A., Ainsworth, B. E., Kimsey, C. D. *et al*. (2005). Prevalence of physical activity in the United States: Behavioral Risk Factor Surveillance system, 2001. *Preventing Chronic Disease*, *2* (2), 1–10.

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*, 519–530.

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling*, *11*, 320–341.

McAuley, E., Talbot, H. M., & Martinez, S. (1999). Manipulating self-efficacy in the exercise environment in women: Influences on affective responses. *Health Psychology*, *18*, 288–294.

Ogden, C. L., Carroll, M. D., Curtin, L. R., McDowell, M. A., Tabak, C. J., & Flegal, K. M. (2006). Prevalence of overweight and obesity in the United States, 1999–2004. *Journal of the American Medical Association*, *295*, 1549–1555.

Pender, N. J., Bar-Or, O., Wilk, B., & Mitchell, S. (2002). Self-efficacy and perceived exertion of girls during exercise. *Nursing Research*, *51*, 86–91.

Perri, M. G., Anton, S. D., Durning, P. E., Ketterson, T. U., Sydeman, S. J., Berlant, N. E. *et al*. (2002). Adherence to exercise prescriptions: Effects of prescribing moderate versus higher levels of intensity and frequency. *Health Psychology*, *21*, 452–458.

Phillips, J. M., & Gatchel, R. J. (2000). Extraversion–introversion and chronic pain. In R. J. Gatchel & J. N. Weisberg (Eds.), *Personality characteristics of patients with pain* (pp. 181–202). Washington, DC: American Psychological Association.

Rejeski, W. J. (1994). Dose–response issues from a psychosocial perspective. In C. Bouchard, R. J. Shephard, & T. Stephens (Eds.), *Physical activity, fitness, and health: International proceedings and consensus statement* (pp. 1040–1055). Champaign, IL: Human Kinetics.

Rudolph, D. L., & McAuley, E. (1996). Self-efficacy and perceptions of effort: A reciprocal relationship. *Journal of Sport and Exercise Psychology*, *18*, 216–223.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. Von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Sivo, S. A., Fan, X., Witta, E. L., & Willse, J. T. (2006). The search for "optimal" cutoff properties: Fit index criteria in structural equation modeling. *Journal of Experimental Education*, *74*, 267–288.

Yuan, K.-H. (2005). Fit indexes versus test statistics. *Multivariate Behavioral Research*, *40*, 115–148.