

## Reliability, Validity, and Measurement Invariance of the German Version of the Physical Activity Enjoyment Scale

Darko Jekauc,<sup>1</sup> PhD, Manuel Voelkle,<sup>2</sup> PhD, Matthias O. Wagner,<sup>1</sup> PhD, Nadine Mewes,<sup>1</sup> PhD, and Alexander Woll,<sup>1</sup> PhD

<sup>1</sup>Department for Sport Science, University of Konstanz, and <sup>2</sup>Max Planck Institute for Human Development Berlin

All correspondence concerning this article should be addressed to Dr. Darko Jekauc, Department for Sport Science, University of Konstanz, Box 30, 78457 Konstanz, Germany. E-mail: [darko.jekauc@uni-konstanz.de](mailto:darko.jekauc@uni-konstanz.de)

**Objective** The purpose of this work is to examine reliability, factorial- and criterion-related validity, and measurement invariance across age and gender of the Physical Activity Enjoyment Scale for children and adolescents in the German-speaking population. **Methods** Confirmatory factor analysis was applied to questionnaire responses obtained from a cross-sectional sample (Study 1) of 250 girls and 254 boys and a longitudinal sample (Study 2) of 109 boys and 87 girls aged 9 to 17 years. **Results** Results indicated sufficient test-retest reliability (ICC = 0.76), internal consistency ( $\alpha = 0.89$ ), and criterion-related validity ( $r = 0.42$  with a physical activity diary;  $r = 0.16$  with accelerometry data). Confirmatory factor analyses partially supported the factorial validity and invariance. **Conclusions** The German version of the Physical Activity Enjoyment Scale is sufficiently reliable and valid to be used for German-speaking children and adolescents. Further research examining the invariance over a longer period is warranted.

**Key words** confirmatory factor analysis; measurement invariance; PACES; physical activity; reliability; validity.

Physical activity (PA) has positive effects on physical and mental health in both clinical and nonclinical populations (WHO, 2010). These findings have led to the proposal that exercise could serve as a supplement to traditional forms of therapy (Martinsen, 2008). Two recent meta-analyses showed that PA interventions were also effective in preventing and reducing the symptoms of anxiety and depression in children and adolescents (Ahn & Fedewa, 2011; Larun, Nordheim, Ekeland, Hagen, & Heian, 2006). However, to realize the benefits of PA, two issues should be considered (Wankel, 1993). First, how can patient adherence to exercise be enhanced? Second, how can the effects of exercise and PA on mental health be explained? Theory and empirical studies support the idea that PA enjoyment is relevant for both issues.

Regarding the issue of adherence to PA and exercise, most major exercise motivation theories such as achievement goal theory (Nicholls, 1989), competence motivation

theory (Harter, 1981), and the sport commitment model (Scanlan, Carpenter, Lobel, & Simons, 1993) include enjoyment of PA as an important component, which is significantly correlated with PA maintenance in children and adolescents (Sallis, Prochaska, & Taylor, 2000; van der Horst, Paw, Twisk, & van Mechelen, 2007). For instance, in a sample of 1,504 children and adolescents in grades 4 to 12, enjoyment of physical education was strongly and consistently associated with the child PA index (Sallis, Prochaska, Taylor, Hill, & Geraci, 1999). In another study, Di Lorenzo, Stucky-Ropp, Vander Wal, and Gotham (1998) found that among several psychological and environmental predictors of PA, only enjoyment had a consistent effect on PA in fifth- and sixth-grade boys and girls. In addition, PA enjoyment was found to be the key mediator of the effects of several predictors, such as social support, personal investments, attractive alternatives, social constraints, and perceived competence (Weiss, Kimmel,

& Smith, 2001), as well as a mediator of the effects of a school-based intervention (Dishman et al., 2005).

Contributing to an explanation of the effects of exercise on mental health, Wankel (1993) argued that enjoyment is an essential mediator of psychological benefits of exercise and that offering enjoyable bouts of exercise might foster positive affective states (e.g., agility). Several empirical studies provide evidence for an association between enjoyment and psychological responses to exercise. For instance, Motl, Berger, and Leuschen (2000) found that PA enjoyment mediated positive effects of rock-climbing on mood, and Raedeke (2007) showed that enjoyment was related with increases in positive affect, but not related with changes in negative affect. In a therapeutic recreation setting, Dattilo, Kleiber, and Williams (1998) suggested that experiencing enjoyment during leisure activities leads to an improvement in functional abilities. Correspondingly, the results of an empirical study largely supported the importance of the role of enjoyment in therapeutic recreation practice (Hutchinson, LeBlanc, & Booth, 2006).

Both the association between PA enjoyment and exercise adherence on the one hand and the association between exercise and mental health on the other reflect the importance of enjoyment in ensuring the benefits of PA. However, to improve research on the effects of enjoyment in children and adolescents, properties of the measuring instruments must be examined more closely. In this endeavor, reliability, validity, and measurement invariance are the most important aspects. As the importance of reliability and validity for assessing the quality of a self-report instrument is well recognized, measurement invariance as a measurement property is increasingly being discussed (Vandenberg, 2002). To compare the results of a self-reporting instrument across two or more groups or across two or more time points, the invariance of measurement must be established (Meredith, 1993). Otherwise, detected differences between groups or time points may actually be due to the measurement instrument and not to the construct in question. To test measurement invariance, five steps have been proposed (Vandenberg & Lance, 2000): equivalence of structure, equivalence of factor loadings, equivalence of measurement intercepts, equivalence of structural covariance, and equivalence of item errors (uniquenesses).

Most studies that have examined the relationship between enjoyment and PA used single-item measures or scales that have not been adequately validated (Kendzierski & De Carlo, 1991; Moore, et al., 2009; Motl, et al., 2000). To provide a better measurement of enjoyment of PA, Kendziersky and DeCarlo (1991) designed the Physical Activity Enjoyment Scale (PACES) for which satisfying

internal consistency and test retest reliability have been demonstrated (Crocker, Bouffard, Gessaroli, 1995; Kendzierski & DeCarlo, 1991). However, the results of a confirmatory factor analysis (CFA) did not support the unidimensional structure of the PACES (Crocker, Bouffard, Gessaroli, 1995). Motl et al. (2000) speculated that the deviation from the unidimensional structure was a methodological artefact caused by the shared variance of positively and negatively worded items and showed that the factor model with correlated uniqueness among positively worded items provided a better fit in adolescent girls. On the other hand, Moore et al. (2009) identified a better model fit for the one factor model with correlated uniqueness among negatively worded items and speculated that differential response patterns for boys and girls could be responsible for these findings.

Despite the growing support for the validity and internal consistency of the PACES, to date, the test retest reliability and composite reliability for the age-group of children and adolescents remains unknown. Furthermore, Moore et al. (2009) and Dunton, Tscherne, and Rodriguez (2009) found contradicting results with respect to gender invariance. Because a lack of gender invariance would imply a major limitation of the PACES, additional research on this topic is necessary. In addition, there is no evidence for its factorial validity and invariance across age-groups and time in non-English speaking populations. Especially in longitudinal studies that analyse predictive and mediation effects of enjoyment of PA, it is essential to establish measurement invariance across time. Finally, there is no evidence for predictive validity of the PACES for accelerometer-measured PA. Only one study with 168 girls aged 13 years examined the association between the PACES and accelerometer-measured PA and found a low significant correlation of 0.21 (Davidson, Werder, Trost, Baker, & Birch, 2007).

The purposes of the present study were to (a) determine the test retest reliability of the PACES for children and adolescents, (b1) investigate the factorial validity for the German-speaking population, (b2) examine the differential response patterns for boys and girls, (c) test the measurement invariance across time, gender, and age-groups, and (d) validate the PACES by using accelerometer-based and subjective PA measures.

## Method

To address these research questions, two separate studies have been conducted.

## Study I

### Participants

Participants were 250 girls and 254 boys aged between 9 and 17 years ( $M = 13.9$ ;  $SD = 2.2$ ) from the MoMo Study (Motorik-Modul), which is based on a representative sample of children and adolescents in Germany (Woll, Kurth, Opper, Worth, & Bös, 2011). The sample of the MoMo Study is a subsample of the German Health Interview and Examination Survey for Children and Adolescents (KiGGS) conducted by the Robert Koch Institute in Berlin (Kurth et al., 2008).

### Procedure

Participants were enrolled using a three-step process. First, a systematic sample of 167 primary sampling units was selected from an inventory of German communities stratified according to the classification system that measures the level of urbanization and the geographic distribution (Kurth et al., 2008). Second, an age-stratified sample of randomly selected children and adolescents was drawn from the official registers of local residents for the KiGGS Study with a total of 17,641 participants aged 0–17 years (Kurth et al., 2008). Third, 7,866 participants aged between 4 and 17 years in the KiGGS sample were randomly assigned to be included in the MoMo Study. Teams of three testers from a total of 23 testers collected data at each location during 2–3 testing days, and each subject was involved for approximately 1 hour. Testers were trained in motor test data collection and questionnaire administration. All participants and their guardians gave written informed consent before study participation. Analyses were performed on data collected from September 2009 until December 2010.

### Measurement of Enjoyment

The PACES was originally developed by Kendzierski and DeCarlo (1991) for measuring positive effects associated with involvement in physical activity among college students. The original PACES consists of 18 statements on a scale between two bipolar adjectives (e.g., enjoy-hate, bored-interested, pleasant-unpleasant) with seven response categories. Motl et al. (2000) revised the PACES for its use in adolescents. The revised version consists of 16 items beginning with “When I am physically active...” Motl et al. (2000) shortened the answer categories to a 5-point continuum (1 = “disagree a lot” to 5 = “agree a lot”). The scale showed acceptable internal consistency with Cronbach’s  $\alpha = 0.87$  (Moore et al., 2009).

For the present study, a qualified staff member (native speaker) translated the revised PACES from English into German. A second person, working without reference to

the original instrument, translated the revised PACES from German back into English. The comparison of the second version of the revised PACES with the original revealed four wording differences that were subsequently resolved by the translators. Finally, the German version of the PACES was completed by five 7th-grade students, who were asked to evaluate the comprehensibility of the translation using two items. The first item asked for comprehensibility of the translated PACES on a four-point continuum (easy to understand – impossible to understand). The second item asked which aspects of the PACES caused comprehension difficulties, using an open response format. None of the five pupils reported comprehension difficulties. These pupils did not take part in the main study. The German and English version of the PACES contain the same item and scale formatting.

## Study II

To replicate the results of the reliability and validity analyses of Study I and to additionally assess test retest reliability, longitudinal invariance, and predictive validity of the PACES, the Study II was conducted.

### Participants

For this study, 109 boys and 87 girls aged between 9 and 17 years ( $M = 12.8$ ;  $SD = 1.6$ ) were recruited from a comprehensive secondary school in Konstanz, Germany, with all three traditional types of the tripartite German secondary school system: Hauptschule ( $n = 28$ ), Realschule ( $n = 63$ ), and Gymnasium ( $n = 105$ ). All participants and their guardians provided informed consent.

### Procedure

Participants completed the PACES and the MoMo Physical Activity Questionnaire for adolescents (MoMo PAQ) two times at an interval of 7 days. Between the two measurement occasions, participants wore an accelerometer and completed the Previous Day Physical Activity Recall (PDPAR; Weston, Petrosa, & Pate, 1997) on a daily basis. The study was conducted from April to July 2009 on school days.

### Measurement

*Enjoyment.* Enjoyment was measured with the German version of the PACES described earlier.

*Accelerometer.* Assessments of PA were obtained using the Actigraph GT1M accelerometer (Pensacola, FL, USA). The Actigraph is a biaxial accelerometer designed to detect vertical and horizontal accelerations ranging in magnitude from 0.05 to 2.00 G’s with a frequency response of 0.25–2.50 Hz. The filtered acceleration signal is digitized,

rectified, and integrated over a user-specified time interval. At the end of each interval, the summed value or “activity count” is stored in memory, and the integrator is reset. To minimize error among individual estimates, a 10-s interval was used. This accelerometer can be used to discriminate between light, moderate, and vigorous levels of PA (Puyau, Adolph, Vohra, Zakeri, & Butte, 2004).

Movement counts were converted to average minutes per day spent in resting or light [ $<3$  metabolic equivalents (METs)], moderate (3–6 METs), vigorous (6–9 METs), and very vigorous ( $>9$  METs) PA. The minutes per day spent in each level of moderate, vigorous, and very vigorous PA were combined into one variable. The duration of time during which the device was worn was estimated using an algorithm proposed by Troiano et al. (2008) in which the time threshold was determined at 60 min and the activity count threshold at 50 counts per min. The measurement day was accepted as valid when the device was worn for  $\geq 10$  h. To be accepted for analysis, each case needed to have at least five valid days. In accordance with this procedure, 139 participants showed valid measurements on this device.

The accelerometer was attached securely to the right hip by an elastic waist belt. Participants were asked to wear the device during waking hours on seven consecutive days and instructed to remove the accelerometer during swimming and bathing. The Actigraph has been shown to be a valid and reliable tool for assessing PA in children and adolescents (De Vries, Bakker, Hopman-Rock, Hirasings, & van Mechelen, 2006; Freedson, Pober, & Janz, 2005).

**PA questionnaire.** The MoMo-PAQ was used to assess self-reported habitual PA in children and adolescents (Jekauc, Wagner, Kahlert, & Woll, 2012). The MoMo-PAQ consists of 28 items and measures frequency, duration, intensity, and type of PA in four domains: daily PA, PA at school, PA in organized sports clubs, and PA outside of organized sports clubs. Data obtained with the MoMo-PAQ are sufficiently reliable (test retest reliability = .68) and significantly correlate ( $r = 0.29$ ) with data obtained using accelerometry (Jekauc, Wagner, Kahlert, & Woll, 2012).

**PA diary.** The PDPAR (Weston, Petrosa, & Pate, 1997) is a short-term, self-report measure of PA designed specifically for children and adolescents (Trost, 2007). A separate table for each day of the week is provided to measure PA of the previous day. Every table is divided into several 60-min time blocks. For each time block, participants can choose one of 38 enumerated activities that are grouped into the following categories: eating,

sleeping/bathing, transportation, work/school, spare time physical activities, and sports. Each activity can be rated according to its intensity (light, moderate, hard, and very hard). The score was calculated starting on the baseline for the 7 consecutive days. PDPAR has been shown to be a reliable and valid measure of PA (Pate et al., 2003; Trost, Ward, McGraw, & Pate, 1999).

## Data Analysis

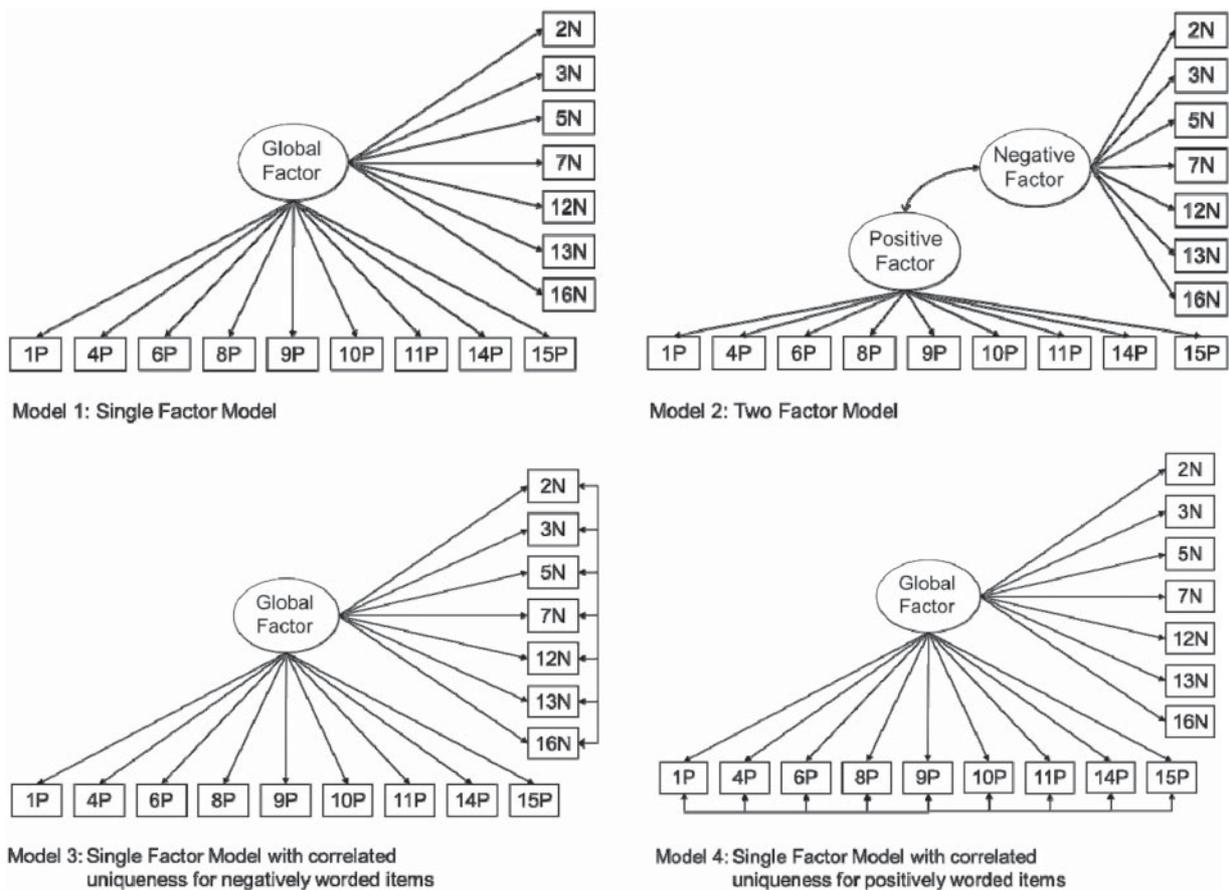
### Reliability

Test retest reliability was assessed by intraclass correlation at a 7-day interval. Internal consistency was estimated by Cronbach’s  $\alpha$  and by composite reliability based on CFA. Because of the assumption of uncorrelated uniqueness among indicators, Cronbach’s  $\alpha$  coefficient underestimates the reliability of the composite score, especially for multidimensional scales (Bollen, 1989). The composite reliability is estimated by the formula of Raykov (1997). All coefficients are presented by gender and age-groups. To create two age-groups of approximately equal size, the samples from both studies were divided into one group of participants aged 9 to 12 years and another group of participants aged 13 to 17 years.

### Factorial Validity

To test the hypothesized one-dimensional structure of CFA, full-information maximum likelihood estimation was performed using AMOS 19 (Arbuckle, 2006). Under the assumption of missing at random or missing completely at random and a multivariate normal distribution, full-information maximum likelihood provides unbiased parameter estimates (Enders & Bandalos, 2001). Even though we cannot be absolutely certain, we have no reason to believe that the assumption of missing at random would be violated. The proportion of missing item responses for each scale ranged from 0.5% to 2.6%. Overall, the proportion of missing items was 1.5% (46 of 3,090 responses).

As described earlier, Motl et al. (2000) assumed that a two-factor solution results primarily from positively and negatively worded items. To test this hypothesis, four models were analysed using the correlated trait, correlated uniqueness framework (Marsh, 1996). The four models are illustrated in Figure 1. The first model assumed a single-factor structure with uncorrelated uniqueness. The second model adopted a two-factor structure where the negatively worded items load on one factor and the positively worded items on the other factor. The third and the fourth model had a single-factor structure with correlated uniqueness among negatively and positively worded items. In accordance with Motl et al. (2000) and Moore et al. (2009),



**Figure 1.** Representation of the four models to test factorial validity.

Model 4 was used to assess factorial validity, measurement invariance, and composite reliabilities.

To assess the appropriateness of each model, several indices of fit were used. The  $\chi^2$  statistic assesses the absolute model fit. For large sample sizes, however, the test is very powerful to detect even minor differences between the observed and model implied covariance matrix, thus rejecting the null hypothesis (good model fit) even in cases where model misspecification is practically negligible (Bollen, 1989). The root mean square error of approximation (RMSEA) describes closeness of fit. Values of the  $RMSEA \leq 0.06$  reflect close and acceptable fit of the model (Hu & Bentler, 1999). The 90% confidence interval (CI) around the RMSEA point estimate should also contain zero to indicate a good fit. The Comparative Fit Index (CFI) tests the relative improvement in fit by comparing the proposed model with a baseline model (Bentler & Bonett, 1980). Values for the CFI around 0.90 are considered acceptable, whereas values around 0.95 indicate a good fit (Bentler & Bonett, 1980; Hu & Bentler, 1999). Because the Models 2, 3, and 4 are not nested, we used Akaike information criterion (AIC) in addition to the absolute fit

indices and model parsimoniousness to determine the best fitting model.

#### Measurement Invariance

Measurement invariance across age-groups, with respect to gender, for the PACES was examined by testing and comparing five nested models (Model A to Model E) using multiple group analysis (Byrne, 2004; Vandenberg & Lance, 2000). Each successive model included the previous model restrictions plus additional constraints (Wu, Li, & Zumbo, 2007). Model A tested the equivalence of the structure, Model B the equivalence of factor loadings, Model C the equivalence of measurement intercepts, Model D the invariance of structural covariances, and Model E the invariance of item uniquenesses and correlations between uniquenesses across time, gender, and age-groups (Vandenberg & Lance, 2000). The successive, nested models were tested by  $\chi^2$  difference tests. Because the  $\chi^2$  difference test is sensitive to sample size, Cheung and Rensvold (2002) recommend using the difference of CFI ( $\Delta CFI$ ). A value of  $\Delta CFI \leq 0.01$  indicates that the null hypothesis of invariance should not be rejected (Cheung & Rensvold, 2002).

### Predictive Validity

To test the predictive validity of the PACES, different measures of PA were used. First, correlations of the PACES (measured at the baseline) with accelerometry during 7 consecutive days were computed. Second, the PACES was correlated with a 7-day aggregate of PDPAR. Finally, the PACES was related to the MoMo-PAQ (measured 7 days after the baseline).

## Results

### Study I

#### Descriptive Statistics and Reliability

Scale means, confidence intervals, standard deviations, Cronbach  $\alpha$ , and composite reliabilities are shown in Table I. Cronbach's  $\alpha$  ranged from 0.92 to 0.93, indicating good internal consistency. The composite reliabilities of the two-factor model with correlated uniqueness among positively worded items (Model 4) were on average (overall:  $cr=0.96$ ) slightly higher than the average Cronbach's  $\alpha$  reflecting the presence of a method artefact because of positively and negatively worded items.

#### Factorial Validity

Results are provided in Table II. The CFA indicated that Model 1 provided a poor fit to the data ( $\chi^2_{104}=710.3$ ; CFI=0.85; RMSEA=0.108 [90% CI=0.100 0.115]; AIC=806.3). Model 2, which represents the two-factor model, indicated a better fit ( $\chi^2_{103}=459.2$ ; CFI=0.91; RMSEA=0.083 [90% CI=0.075 0.091]; AIC=557.2) than Model 1. Model 3, which allowed for correlated uniqueness among negatively worded items, indicated an acceptable model fit ( $\chi^2_{83}=288.1$ ; CFI=0.95; RMSEA=0.070 [90% CI=0.061 0.079]; AIC=426.1). In terms of the AIC, it was superior to both Model 1 and 2. The one-factor model with correlated uniqueness among the positively worded items (Model 4) also showed an acceptable model fit ( $\chi^2_{68}=285.5$ ; CFI=0.95; RMSEA=0.080 [90% CI=0.070 0.089]; AIC=453.5). As apparent from the AIC, it was superior to Model 1 and 2. All parameter estimates (factor loadings, intercepts, variances, and covariances) in all four models were significantly different from zero. Factor loadings for Model 4 are presented in Table III.

#### Measurement Invariance

The analysis of measurement invariance was conducted with Model 4 according to Moore et al. (2009) and Motl et al. (2000). Results are reported in Table IV. The invariance was tested across age-groups and gender. The  $\chi^2$  difference tests were significant for the difference between

Model B and C and between Model D and E for the age-groups. However, the CFI decreased substantially only between Model D and E. This means that the item uniquenesses (variances and covariances of the "error" terms) differ significantly between younger and older participants. For gender, the differences between Model A and B and between Model D and E were significant according to the  $\chi^2$  difference test. Considering the changes of CFI, however, a noteworthy impairment of the model fit was only detected between Model D and E, suggesting that the assumption of invariance of item uniquenesses and correlations between uniquenesses for gender may be unreasonable.

### Study II

#### Descriptive Statistics and Reliability

Descriptive statistics presented in Table V indicate comparable means and standard deviations for gender and age-groups for those in Study I (compare Table I). Cronbach's  $\alpha$  ranged between 0.89 and 0.91. The composite reliabilities for Model 4 (overall:  $cr=0.94$ ) were on average slightly lower than those in Study I. The intraclass correlations ranged from 0.73 to 0.84, indicating that the scale was stable over time.

#### Factorial Validity

The results of CFA, which are provided in Table VI, indicated a poor fit of Model 1 ( $\chi^2_{104}=456.6$ ; CFI=0.74; RMSEA=0.132 [90% CI=0.120 0.144]; AIC=552.5). Assuming a two-factor structure, Model 2 produced an improvement of the model fit ( $\chi^2_{103}=300.8$ ; CFI=0.853; RMSEA=0.099 [90% CI=0.086 0.112]; AIC=398.8). Model 3 ( $\chi^2_{83}=257.9$ ; CFI=0.870; RMSEA=0.104 [90% CI=0.090 0.118]; AIC=395.9) showed a comparable model fit with Model 2. In contrast to the sample of Study I, Model 4 ( $\chi^2_{68}=141.3$ ; CFI=0.95; RMSEA=0.074 [90% CI=0.057 0.092]; AIC=309.3) exhibited a substantially lower AIC value than Model 3, indicating a better model fit. However, the superiority of Model 4 over Model 3 was only shown for boys, but not for girls (see Table VI). With the exception of four covariances (Items 1 6, 1 10, 1 11, and 1 15), all parameter estimates (factor loadings, intercepts, variances, and covariances) significantly differed from zero. Factor loadings for Model 4 are presented in Table III.

#### Measurement Invariance

The results of the analysis of multigroup and longitudinal invariance are presented in Table VII. The invariance analysis with the  $\chi^2$  difference test across age-groups yielded significant differences between Models A and B, B and C,

**Table I. Descriptive Statistics and Reliability of the PACES in Study I**

	N	M (SD)	95% CI	$\alpha$	cr
Overall	504	66.1 (9.7)	(65.3 67.0)	0.92	0.96
Boys	254	66.4 (9.8)	(65.2 67.6)	0.92	0.95
Girls	250	65.9 (9.5)	(64.6 67.1)	0.92	0.95
Age $\leq 12$ years	246	67.0 (9.8)	(65.7 68.2)	0.92	0.95
Age $\geq 13$ years	258	65.4 (9.5)	(64.2 66.6)	0.93	0.97

Note. N = sample size; M = mean; SD = standard deviation; cr = composite reliability.

**Table II. CFA Testing Factorial Validity of the PACES in Study I**

Group	Model	$\chi^2$	df	p	CFI	RMSEA	AIC
Overall	Model 1	710.3	104	<.001	0.848	0.108	806.3
Overall	Model 2	459.2	103	<.001	0.911	0.083	557.2
Overall	Model 3	288.1	83	<.001	0.949	0.070	426.1
Overall	Model 4	285.5	68	<.001	0.946	0.080	453.5
Boys	Model 3	208.4	83	<.001	0.939	0.077	346.4
Boys	Model 4	189.1	68	<.001	0.941	0.084	357.1
Girls	Model 3	207.6	83	<.001	0.938	0.078	345.6
Girls	Model 4	191.3	68	<.001	0.938	0.085	359.3

Note. CFA = confirmatory factor analysis;  $\chi^2$  = chi-square statistic; df = degrees of freedom; p = probability value; CFI = comparative fit index; RMSEA: root mean square error of approximation; AIC = Akaike Information Criterion; Model 1 = single factor model; Model 2 = two factor model; Model 3 = single factor model with correlated uniqueness for negatively worded items; Model 4 = single factor model with correlated uniqueness for positively worded items.

**Table III. Lambda Loadings, Standard Errors, and Critical Ratios for Model 4 in Studies I and II**

Item	Study 1				Study 2			
	$\lambda$	SE	CR	p	$\lambda$	SE	CR	p
Item 1	1				1			
Item 2	0.94	0.06	15.24	<.001	0.97	0.11	8.63	<.001
Item 3	0.84	0.07	11.57	<.001	0.60	0.12	5.09	<.001
Item 4	0.83	0.06	13.44	<.001	0.67	0.11	6.27	<.001
Item 5	0.84	0.07	12.72	<.001	0.73	0.10	7.10	<.001
Item 6	0.83	0.07	12.76	<.001	0.74	0.12	6.03	<.001
Item 7	0.87	0.08	11.42	<.001	0.58	0.14	4.16	<.001
Item 8	0.72	0.07	10.32	<.001	0.45	0.13	3.41	<.001
Item 9	0.90	0.06	15.92	<.001	0.91	0.12	7.30	<.001
Item 10	1.01	0.06	15.98	<.001	1.26	0.11	10.98	<.001
Item 11	0.99	0.07	14.34	<.001	0.85	0.10	8.76	<.001
Item 12	0.84	0.06	14.17	<.001	0.72	0.09	7.70	<.001
Item 13	0.52	0.05	10.50	<.001	0.42	0.09	4.85	<.001
Item 14	0.51	0.05	10.13	<.001	0.45	0.08	5.79	<.001
Item 15	0.91	0.07	13.76	<.001	0.76	0.10	7.61	<.001
Item 16	1.03	0.07	14.42	<.001	0.97	0.11	9.02	<.001

Note. SE = standard error of lambda; CR = critical ratio; p = probability value.

**Table IV. Analysis of Invariance of the PACES Across Age and Gender Groups in Study I**

	$\chi^2$	df	p	CFI	RMSEA	$\Delta\chi^2$	$\Delta df$	p
Invariance by age								
Model A	375.2	136	<.001	0.941	0.059			
Model B	393.2	151	<.001	0.940	0.057	18	15	.263
Model C	429.8	167	<.001	0.935	0.056	36.6	16	.002
Model D	430.5	168	<.001	0.935	0.056	0.7	1	.403
Model E	553.7	220	<.001	0.917	0.055	123.2	52	<.001
Invariance by gender								
Model A	380.4	136	<.001	0.940	0.06			
Model B	417.5	151	<.001	0.934	0.059	37.1	15	.001
Model C	434.6	167	<.001	0.934	0.057	17.1	16	.379
Model D	435.2	168	<.001	0.934	0.056	0.6	1	.439
Model E	532.5	220	<.001	0.923	0.053	97.3	52	<.001

Note.  $\chi^2$  = chi-square statistic; df = degrees of freedom; p = probability value; CFI = comparative fit index; RMSEA: root mean square error of approximation;  $\Delta\chi^2$  = chi-square difference;  $\Delta df$  = difference of degrees of freedom; Model A = equivalence of the structure; Model B = Model A + equivalence of factor loadings; Model C = Model B + equivalence of measurement intercepts; Model D = Model C + invariance of structural covariances; Model E = Model D + invariance of item uniquenesses and correlations between uniquenesses.

**Table V. Descriptive Statistics and Reliability of the PACES in Study II**

	N	M (SD)	95% CI	$\alpha$	cr	ICC
Overall	196	66.3 (9.6)	(64.9 67.6)	0.89	0.94	0.76
Boys	109	66.3 (9.6)	(64.5 68.2)	0.89	0.94	0.73
Girls	87	66.2 (9.5)	(64.2 68.2)	0.90	0.93	0.84
Age $\leq 12$ years	125	67.3 (9.0)	(66.2 68.9)	0.89	0.91	0.74
Age $\geq 13$ years	71	64.4 (10.2)	(62.0 66.9)	0.91	0.95	0.77

Note. N = sample size; M = mean; SD = standard deviation; cr = composite reliability; ICC = intraclass correlation.

**Table VI. CFA Testing Factorial Validity of the PACES in Study II**

Group	Model	$\chi^2$	df	p	CFI	RMSEA	AIC
Overall	Model 1	456.6	104	<.001	0.738	0.132	552.6
Overall	Model 2	300.8	103	<.001	0.853	0.099	398.8
Overall	Model 3	257.9	83	<.001	0.870	0.104	395.9
Overall	Model 4	141.3	68	<.001	0.946	0.074	309.3
Boys	Model 3	203.5	83	<.001	0.838	0.116	341.5
Boys	Model 4	113.9	68	<.001	0.938	0.079	281.9
Girls	Model 3	159.0	83	<.001	0.886	0.103	300.0
Girls	Model 4	130.9	68	<.001	0.908	0.104	298.9

Note. CFA = confirmatory factor analysis;  $\chi^2$  = chi-square statistic; df = degrees of freedom; p = probability value; CFI = comparative fit index; RMSEA: root mean square error of approximation; AIC = Akaike Information Criterion; Model 1 = single factor model; Model 2 = two factor model; Model 3 = single factor model with correlated uniqueness for negatively worded items; Model 4 = single factor model with correlated uniqueness for positively worded items.

Table VII. Analysis of Invariance of the PACES Across Age, Gender, and Time Groups in Study II

Invariance by	$\chi^2$	df	p	CFI	RMSEA	$\Delta\chi^2$	$\Delta df$	p
Age								
Model A	280.3	136	<.001	0.901	0.074			
Model B	321.1	151	<.001	0.893	0.076	40.8	15	<.001
Model C	355.5	167	<.001	0.884	0.076	34.4	16	.005
Model D	355.5	168	<.001	0.881	0.076	0.0	1	0.999
Model E	517.8	220	<.001	0.795	0.084	162.3	52	<.001
Gender								
Model A	244.8	136	<.001	0.923	0.064			
Model B	269.3	151	<.001	0.916	0.064	24.5	15	.057
Model C	282.7	167	<.001	0.918	0.060	13.4	16	.643
Model D	283.0	168	<.001	0.918	0.059	0.3	1	.584
Model E	371.6	220	<.001	0.892	0.060	88.6	52	.001
Time								
Model A	321.7	136	<.001	0.944	0.059			
Model B	369.8	151	<.001	0.935	0.061	48.1	15	<.001
Model C	384.2	167	<.001	0.935	0.058	14.4	16	.569
Model D	385.1	168	<.001	0.932	0.058	0.9	1	.343
Model E	532.0	220	<.001	0.907	0.060	146.9	52	<.001

Note.  $\chi^2$  = chi-square statistic; df = degrees of freedom; p = probability value; CFI = comparative fit index; RMSEA: root mean square error of approximation;  $\Delta\chi^2$  = chi-square difference;  $\Delta df$  = difference of degrees of freedom; Model A = equivalence of the structure; Model B = Model A + equivalence of factor loadings; Model C = Model B + equivalence of measurement intercepts; Model D = Model C + invariance of structural covariances; Model E = Model D + invariance of item uniquenesses and correlations between uniquenesses.

as well as D and E. However, the first two differences indicate only a slight impairment of CFI and RMSEA, whereas the difference between Model D and E represents a substantial decline of model fit in terms of the CFI. Similar to the results of the invariance tests for gender described earlier, this indicates a deviation of the postulated assumption of measurement error (uniqueness) invariance. Only one of the five  $\chi^2$  difference tests across gender groups was significant (Model D E). The decrease of CFI by 0.026 indicates that the measurement error (uniqueness) structure differs between boys and girls. The results of the invariance across time showed two significant  $\chi^2$  differences between Models A and B and between Models D and E. Although CFI and RMSEA indicated only a slight difference between Model A and B, the difference between Model D and E represents a considerable deviation, which means that the measurement error (uniqueness) structure differs between older and younger participants.

### Predictive Validity

To assess the predictive validity of the PACES, its effects on different measures of PA (PDPAR, accelerometer, MoMo-PAQ) were computed. Measured at baseline, the

PACES was significantly correlated with the PDPAR ( $r=0.42$ ;  $p<.001$ ) and accelerometer data ( $r=0.16$ ;  $p=.025$ ) measured during the subsequent 7 days, and MoMo-PAQ ( $r=0.26$ ;  $p<.001$ ) measured 7 days after the baseline. Even after controlling for gender and age, the effects of the PACES remained significant.

### Discussion

Previous research suggests that PA enjoyment may enhance adherence to exercise and the effects of exercise on mental health. However, especially for children and adolescents, measurement properties of questionnaires to measure PA enjoyment have not been adequately established. The main purpose of the present work was to examine (a) reliability, (b1) factorial validity, (b2) differential response patterns for boys and girls, (c) measurement invariance, and (d) criterion-related validity of the PACES for children and adolescents in the German-speaking population.

The results of both Study I and Study II indicate an appropriate internal consistency of the German version of the revised PACES for children and adolescents. Cronbach's  $\alpha$  for Study I and for Study II are comparable with the results of other studies with children and adolescents (Davidson, Werder, Trost, Baker, & Birch, 2007; Moore et al., 2009). Composite reliabilities for Study I and Study II were slightly higher than the corresponding alpha coefficients, suggesting a method effect of positively and negatively worded items. The overall intraclass correlation with 1 week's distance between measurements in Study II indicates a satisfactory test retest reliability of the scale with the reliability being slightly higher for girls than boys.

Regarding factorial validity, our results support the suspicion that independent of the actual construct more similarly worded items are more closely related (a so-called method effect because of positively vs. negatively worded items). However, the two studies revealed an incoherent picture. The data of Study I do not provide evidence for a better fit of Model 4 (single factor model with correlated uniqueness among positively worded items). In this case, the more parsimonious Model 3 (single-factor model with correlated uniqueness among negatively worded items) should be preferred (see Figure 1). However, the results of Study II support the findings of Motl et al. (2000) and Dunton, Tscherne, and Rodriguez (2009) that Model 4 has a superior fit to Model 3. Therefore, we are inconclusive about the preference for the model.

The findings of both our studies are only partially consistent for boys and girls. In Study I, the fit of Model 4 is not better than the fit of Model 3 neither for boys nor girls.

In Study II, Model 4 is only superior for boys, but not for girls. These findings contradict the hypothesis that Model 3 is more appropriate for boys than Model 4. The assumption of differential response patterns for negatively and positively worded items in boys and girls could not be confirmed.

To compare the results of the PACES across two or more groups or across two or more time points, invariance of measurements must be established (Meredith, 1993). Motl et al. (2000) and Moore et al. (2009) found significant deviations from the assumption of measurement invariance. In our study, similar deviations were found for the assumption of invariant uniquenesses and correlations between uniquenesses, which were consistent across gender, age, and time. Overall, these findings suggest that the method effect related to the positive and negative wording of items could impair the invariance assumption. These findings raise the question of why children and adolescents might respond differently to positively and negatively worded items of the PACES. There are two possible explanations. First, according to Tourangeau, Rips, and Rasinski (2000) there are four cognitive components involved in responding to a questionnaire: comprehension, retrieval, judgement, and response. A person faced with a question has to comprehend the meaning of the question, to retrieve from memory the relevant information, to assess the relevant information, and to give a response according to this assessment in a scale. Looking at it from this perspective, it might be expected that children and adolescents have difficulties in consistently answering positively and negatively worded questions containing emotional content (Tourangeau, Rips, & Rasinski, 2000). Second, the PACES was constructed in tradition of the bipolar model of affect structure (Frijda, 1986; Green, 1988; Russell, 1980; Russell & Barrett, 1999; Smith & Ellsworth, 1985) in which positive and negative affect can be seen as two extreme, opposite endpoints of a continuum. However, considerable evidence was found for the unipolar model, which implies a (partial) independence of positive and negative affect (Cacioppo & Berntson, 1994; Davidson, 1998; Diener & Emmons, 1985; Watson & Tellegen, 1985; Watson & Tellegen, 1999). Assuming that a unipolar model holds for the PACES, the effect of positive and negative worded items would not be interpreted as a methodological artefact, but as an effect of two partially independent affects. Further studies are required to test these assumptions.

Finally, the results of this study support the predictive validity of the translated version of the PACES. Measured at baseline, the PACES was significantly correlated with one self-reported measure of habitual PA (MoMo-PAQ), one

short-term, self-reported measure of PA (PDPAR), and with one accelerometer-measured indicator of PA 7 days after the baseline. The correlation of the PACES with accelerometer data was low, and the correlations with self-report measures of PA were moderate. These results are comparable with the findings of other studies (Davidson, Werder, Trost, Baker, & Birch, 2007; Moore et al., 2009). In Study II, the highest longitudinal correlation was observed with PDPAR ( $r=0.42$ ), a short-term, self-report measure of PA, which was significantly correlated with accelerometer data ( $r=0.54$ ) and the MoMo-PAQ ( $r=0.51$ ). The results of this study suggest that the PACES is better suited to predict short term, self-reported measures of PA than habitual and accelerometer-measured indicators of PA. In addition, higher correlations with self-reported measures of PA could be explained by a method effect as the PACES is a self-report measure. It is assumed that enjoyment has a short-term mechanism of action on PA, which explains why the PACES can better predict short-term PA than habitual PA.

In line with previous research on the English version of the PACES (Motl et al., 2000; Moore et al., 2009), the measuring instrument suffers from method effects because of negatively and positively worded items. One possible strategy for avoiding the method effect could be to use either only positively or only negatively worded items. For instance, Dishman et al. (2005) and Paxton et al. (2008) used only the negatively worded items of the PACES. However, there are no examinations to compare the two shortened versions with positively vs. negatively worded items. Although beyond the scope of this work, the psychometric analyses of Study II indicate higher test retest correlations, internal consistencies, and correlations with measures of PA of the shortened version with positively worded items. Nonetheless, further research is needed to examine the psychometric properties, especially discriminant and convergent validity, of these two shortened versions of the PACES. It is possible that they represent different affective processes, which to some extent independently mediate the effects of PA on mental health. A further step of validation of the PACES could be to test the mediation hypothesis in a clinical setting. Furthermore, longitudinal invariance should be tested for a longer period (e.g., several months; Paxton et al., 2008), as enjoyment was found to be a significant mediator (Dishman et al., 2005; Motl et al., 2000; Raedeke, 2007) of PA interventions, which usually take at least several months.

Both studies have a number of limitations. First, our considerations are based on two studies with limited sample size. In particular, the results of test retest

correlations and measurement invariance across time should be interpreted with caution because these results are based on a sample size of only 196 participants. Second, the present work does not include a sample of English-speaking participants to directly compare the results of the German and the English version of the PACES and to confirm the measurement invariance over different cultures. Third, the tests on longitudinal invariance refer to a relatively short period (1 week). Nevertheless, to our knowledge, the present study is the first longitudinal study that facilitates the examination of the test retest reliability and factorial validity of the PACES over time. Additionally, it provides first results on reliability, validity, and measurement invariance, using a representative sample of German adolescents.

## Conclusion

Adequate measurement properties in terms of the reliability, validity, and measurement invariance of the PACES are a prerequisite for its use in scientific and clinical studies. The results of this work showed that the German version of the PACES is sufficiently reliable and its reliability is comparable with that of the English version. In agreement with previous research, factorial validity and invariance of measurement can only partially be confirmed because of a method effect because of negatively and positively worded items. Thus, in future studies, it seems reasonable to use exclusively positively or negatively worded items. However, further research is needed to explore such an approach. Although overall predictive validity of the PACES was good, self-reported and short-term measures of PA were better predicted by the PACES than by accelerometer-measured PA and measures of habitual PA.

## Acknowledgments

This work was supported by a project grant from the German Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research). We would like to thank Dr. Daniela Kahlert and Prof. Dr. Ralf Brandt for providing the accelerometer for Study II, and all of the children and adolescents who participated in both studies.

## Funding

This work was supported by the German Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research) (grant number: 01ER0810).

*Conflicts of interest:* None declared.

## References

- Ahn, S., & Fedewa, A. L. (2011). A meta-analysis of the relationship between children's physical activity and mental health. *Journal of Pediatric Psychology, 36*, 385–397. doi: 10.1093/jpepsy/jsq107.
- Arbuckle, J. L. (2006). *Amos (Version 7.0) [Computer Program]*. Chicago: SPSS.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 8*, 588–606.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.
- Byrne, B. M. (2004). Testing for multigroup invariance using AMOS graphics: A road less traveled. *Structural Equation Modeling, 11*, 272–300.
- Cacioppo, J. T., & Berntson, G. G. (1994). Relationship between attitudes and evaluative space: A critical review, with emphasis on the separability of positive and negative substrates. *Psychological Bulletin, 115*, 401–423.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluation goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9*, 233–255.
- Crocker, P. R. E., Bouffard, M., & Gessaroli, M. E. (1995). Measuring enjoyment in youth settings: A confirmatory factor analysis of the physical activity enjoyment scale. *Journal of Sport & Exercise Psychology, 17*, 200–205.
- Dattilo, J., Kleiber, D., & Williams, R. (1998). Self-determination and enjoyment enhancement: a psychologically-based service delivery model for therapeutic recreation. *Therapeutic Recreation Journal, 32*, 258–271.
- Davidson, K. K., Werder, J. L., Trost, S. G., Baker, B. L., & Birch, L. L. (2007). Why are early maturing girls less active? Links between pubertal development, psychological well-being, and physical activity among girls at ages 11 and 13. *Social Science & Medicine, 64*, 2391–2404.
- Davidson, R. J. (1998). Affective style and affective disorders: Perspectives from affective neuroscience. *Cognition and Emotion, 12*, 307–330.
- De Vries, S. I., Bakker, I., Hopman-Rock, M., Hirasings, R. A., & van Mechelen, W. (2006). Review of prediction models to estimate activity-related energy expenditure in children and adolescents. *Journal of Clinical Epidemiology, 59*, 670–680.
- Diener, E., & Emmons, R. A. (1985). The independence of positive and negative affect. *Journal of Personality and Social Psychology, 47*, 1105–1117.

- Di Lorenzo, T. M., Stucky-Ropp, R. C., Vander Wal, J. S., & Gotham, H. J. (1998). Determinants of exercise among children. A longitudinal analysis. *Preventive Medicine, 27*, 470–477.
- Dishman, R. K., Motl, R. W., Saunders, R., Felton, G., Ward, D. S., Dowda, M., & Pate, R. R. (2005). Enjoyment mediates effects of a school-based physical-activity intervention. *Medicine & Science in Sports & Exercise, 37*, 478–487.
- Dunton, G. F., Tscherne, J., & Rodriguez, D. (2009). Factorial validity and gender invariance of the physical activity enjoyment scale (PACES) in older adults. *Research Quarterly for Exercise and Science, 80*, 117–121.
- Enders, C. K., & Bandalos, D. L. (2001). The relative performance of full information maximum likelihood estimation for missing data in structural equation models. *Structural Equation Modeling, 8*, 430–457.
- Freedson, P. S., Pober, D., & Janz, K. F. (2005). Calibration of accelerometer output for children. *Medicine & Science in Sports & Exercise, 37*, 523–530.
- Frijda, N. (1986). *The Emotions*. New York: Cambridge University Press.
- Green, D. P. (1988). On the dimensionality of public sentiment toward partisan and ideological groups. *American Journal of Political Science, 32*, 758–780.
- Harter, S. (1981). The development of competence motivation in the mastery of cognitive and physical skills: Is there a place for joy?. In G. C. Roberts, & D. M. Landers (Eds.), *Psychology of motor behavior and sport* (pp. 3–29). Champaign, IL: Human Kinetics.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55.
- Hutchinson, S., LeBlanc, A., & Booth, R. (2006). More than “just having fun”: reconsidering the role of enjoyment in therapeutic recreation practice. *Therapeutic Recreation Journal, 40*, 220–234.
- Jekauc, D., Wagner, M. O., Kahlert, D., & Woll, A. (2012). *Reliabilität und Validität des MoMo-Aktivitätsfragebogens für Jugendliche (MoMo-AFB)*. [Reliability and Validity of MoMo-Physical-Activity-Questionnaire for Adolescents (MoMo-PAQ)] *Manuscript submitted for publication*.
- Kendzierski, D., & DeCarlo, K. L. (1991). Physical activity enjoyment scale: Two validation studies. *Journal of Sport Exercise Psychology, 13*, 50–64.
- Kurth, B. M., Kamtsiuris, P., Hölling, H., Schlaud, M., Döller, R., Ellert, U., ... Wolf, U. (2008). The challenge of comprehensively mapping children’s health in a nation-wide health survey: Design of the German KiGGS-Study. *BMC Public Health, 8*, 196.
- Larun, L., Nordheim, L. V., Ekland, E., Hagen, K. B., & Heian, F. (2006). Exercise in prevention and treatment of anxiety and depression among children and young people. *Cochrane Database Syst Rev*, CD004691. doi: 10.1002/14651858.CD004691.pub2.
- Marsh, H. W. (1996). Positive and negative global self-esteem: A substantively meaningful distinction or artifacts? *Journal of Personality & Social Psychology, 70*, 810–819.
- Martinsen, E. W. (2008). Physical activity in the prevention and treatment of anxiety and depression. *Nordic Journal of Psychiatry, 62*(47 Suppl), 25–29. doi: 10.1080/08039480802315640.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.
- Moore, J. B., Yin, Z., Hanes, J., Duda, J., Gutin, B., & Barbeau, P. (2009). Measuring enjoyment of physical activity in children: validation of the physical activity enjoyment scale. *Journal of Applied Sport Psychology, 21*(1 Suppl), 116–129.
- Motl, R. W., Berger, B. G., & Leuschen, P. S. (2000). The role of enjoyment in the exercise-mood relationship. *International Journal of Sport Psychology, 31*, 347–363.
- Motl, R. W., Dishman, R. K., Trost, S. G., Saunders, R., Dowda, M., Felton, G., ... Pate, R. R. (2000). Factorial validity and invariance of questionnaires measuring social-cognitive determinants of physical activity among adolescent girls. *Preventive Medicine, 31*, 584–594.
- Nicholls, J. G. (1989). *The competitive ethos and democratic education*. Cambridge, MA: Harvard University Press.
- Pate, R. R., Saunders, R. P., Ward, D.S., Felton, G., Trost, S.G., & Dowda, M. (2003). Evaluation of a community-based intervention to promote physical activity in youth: Lessons from active winners. *American Journal of Health Promotion, 17*, 171–182.
- Paxton, R. J., Nigg, C., Motl, R. W., Yamashita, M., Chung, R., Battista, J., & Chang, J. (2008). Physical activity enjoyment scale short form—Does it fit for children? *Research Quarterly for Exercise and Sport, 79*, 423–427.
- Puyau, M. R., Adolph, A. L., Vohra, F. A., Zakeri, I., & Butte, N. F. (2004). Prediction of activity energy

- expenditure using accelerometers in children. *Medicine and Science in Sports and Exercise*, 36, 1625.
- Raedeke, T. D. (2007). The relationship between enjoyment and affective responses to exercise. *Journal of Applied Sport Psychology*, 19, 105–115.
- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21, 173–184.
- Russell, J. A. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, 39, 1161–1178.
- Russell, J. A., & Barrett, L. F. (1999). Core affect, prototypical emotional episodes, and other things called emotion: Dissecting the elephant. *Journal of Personality and Social Psychology*, 76, 805–819.
- Sallis, J. F., Prochaska, J. J., Taylor, W. C., Hill, J. O., & Geraci, J. C. (1999). Correlates of physical activity in a national sample of girls and boys in grades 4 through 12. *Health Psychology*, 18, 410–415.
- Sallis, J. F., Prochaska, J. J., & Taylor, W. C. (2000). A review of correlates of physical activity of children and adolescents. *Medicine and Science in Sports and Exercise*, 32, 963–975.
- Scanlan, T. K., Carpenter, P. J., Lobel, M., & Simons, J. P. (1993). Sources of enjoyment of youth sport athletes. *Pediatric Exercise Science*, 5, 275–285.
- Smith, C. A., & Ellsworth, P. C. (1985). Patterns of cognitive appraisals in emotions. *Journal of Personality and Social Psychology*, 48, 813–838.
- Tourangeau, R., Rips, L. J., & Rasinski, K. (2000). *The Psychology of Survey Response*. Cambridge: Cambridge University Press.
- Troiano, R. P., Berrigan, D., Dodd, K. W., Mâsse, L. C., Tilert, T., & McDowell, M. (2008). Physical activity in the United States measured by accelerometer. *Medicine & Science in Sports & Exercise*, 40, 181.
- Trost, S. G. (2007). State of the art reviews: Measurement of physical activity in children and adolescents. *American Journal of Lifestyle Medicine*, 1, 299. doi: 10.1177/1559827607301686.
- Trost, S. G., Ward, D. S., McGraw, B., & Pate, R. R. (1999). Validity of the previous day physical activity recall (PDPAR) in fifth-grade children. *Pediatric Exercise Science*, 11, 341–348.
- Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, 5, 139–158. doi: 10.1177/1094428102005002001.
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organisational Research Methods*, 3, 4–70.
- Van Der Horst, K., Paw, M. J., Twisk, J. W., & Van Mechelen, W. (2007). A brief review on correlates of physical activity and sedentariness in youth. *Medicine and Science in Sports and Exercise*, 39, 1241–1250.
- Wankel, L. M. (1993). The importance of enjoyment to adherence and psychological benefits from physical activity. *International Journal of Sport Psychology*, 24, 151–169.
- Watson, D., & Tellegen, A. (1985). Toward a consensual structure of mood. *Psychological Bulletin*, 98, 219–235.
- Watson, D., & Tellegen, A. (1999). Issues in the dimensional structure of affect - Effects of descriptors, measurement error, and response formats: Comment on Russell and Carroll (1999). *Psychological Bulletin*, 125, 601–610.
- Weiss, M. R., Kimmel, L. A., & Smith, A. L. (2001). Determinants of sport commitment among junior tennis players: Enjoyment as a mediating variable. *Pediatric Exercise Science*, 13, 131–144.
- Weston, A. T., Petosa, R., & Pate, R. R. (1997). Validation of an instrument for measurement of physical activity in youth. *Medicine and Science in Sports Exercise*, 29, 138–143.
- Woll, A., Kurth, B. M., Opper, E., Worth, A., & Bös, K. (2011). The 'Motorik-Modul' (MoMo): Physical fitness and physical activity in German children and adolescents. *European Journal of Pediatrics*, 170, 1129–1142.
- World Health Organization (WHO). (2010). *Global Recommendations on Physical Activity for Health*. Geneva: WHO.
- Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research & Evaluation*, 12, 1–26.