Psychology of Sport and Exercise 9 (2008) 830–847

Psychology
OF SPORT AND EXERCISE

www.elsevier.com/locate/psychsport

A short and very short form of the physical self-inventory for adolescents: Development and factor validity

Christophe Maïano[a,*], Alexandre J.S. Morin[b], Grégory Ninot[c], Johana Monthuy-Blanc[c], Yannick Stephan[d], Jean-François Florent[e], Philippe Vallée[f]

[a] UMR 6233, CNRS, Movement Sciences Institut, Etienne-Jules Marey, University of Aix-Marseille II, 163 Avenue de Luminy, CP 910, 13288 Marseille Cedex 9, France
[b] Department of Psychology, University of Sherbrooke, Sherbrooke, Québec, Canada
[c] EA 4206, Addictive, Performance and Health Behaviors, University of Montpellier I, Montpellier, France
[d] JE 2494, Psychology of Physical Activity, University of South-Paris 11, Orsay, France
[e] Benjamin Morel Technical High School, Dunkerque, France
[f] André Malraux Junior High School, Cagnes Sur Mer, France

Received 25 May 2007; received in revised form 26 October 2007; accepted 31 October 2007
Available online 29 January 2008

Abstract

Objectives: The Physical Self-Inventory (PSI)—a French adaptation of the Fox and Corbin's [1989. The Physical Self-Perception Profile: Development and preliminary validation. *Journal of Sport and Exercise Psychology*, *11*, 408–430] Physical Self-Perception Profile—was originally developed for use with adults and no study has systematically verified its psychometric properties in adolescent populations. Additionally, this instrument remains too long to be efficiently completed in combination with multiple other instruments within extensive longitudinal or idiographic studies. The purpose of the present investigation was thus threefold: (a) testing the factor validity and reliability of the original PSI in a sample of adolescents; (b) developing and testing the factor validity and reliability of a *very short* (i.e., two items per scale) form of the PSI in a sample of adolescents; and (c) testing the equivalence of the factor pattern, structural parameters, latent mean structure, and criterion-related validity of both forms of the PSI.
Design: Structural equation modeling approach.
Method: Two samples participated in this series of studies. In Study 1, a sample of 1018 adolescents completed the adult PSI (25 items) and was randomly split in two sub-samples. In Study 2, a new sample of 320 adolescents completed a *very short* form of the PSI (PSI-VSF). Factorial validity and gender and multigroup invariance of these instruments (PSI, PSI-VSF) were tested using confirmatory factorial analysis (CFA) and structural equation modeling (SEM).
Results: In Study 1, CFA and SEM analyses provided evidence for the factor validity and reliability of a *short* (PSI-SF: 18 items) and *very short* (PSI-VSF: 12 items) form of the PSI for adolescents. In Study 2, CFAs and SEMs supported the equivalence of the factor pattern, structural parameters, latent mean structure, and criterion-related validity of both forms of the PSI (i.e., PSI-SF, PSI-VSF).

---

[*] Corresponding author. Tel.: +33 491 759 653; fax: +33 491 170 415.
E-mail address: christophe.maiano@univmed.fr (C. Maïano).

1469-0292/$ - see front matter
doi:10.1016/j.psychsport.2007.10.003

Conclusion: The present results thus provide preliminary evidence regarding the reliability and validity of a *short* and a *very short* form of the PSI for French adolescents.

Keywords: Global self-esteem; Physical self-concept; Multidimensionality; Questionnaire; French; Gender invariance

Introduction

Self-concept is a multidimensional construct encompassing many characteristics, competencies, and roles possessed or played by individuals (Marsh, 1997). As such, self-concept includes many sub-selves or domains, such as the academic self, the family self, the emotional self, the social self, and the physical self (Marsh, 1997). Shavelson, Hubner, and Stanton (1976) hypothesized that these various domains are organized hierarchically: the self-concept being like a pyramid, with global self-concept (GSC)[1] at the apex and general constructs at the next lower level (Shavelson et al., 1976). Specificity increases downward, with the most situation-specific self-perceptions at the base (Shavelson et al., 1976).

With the recognition of the multidimensionality of the self-concept has come more detailed study of its various sub-components (Fox, 2000). Thus, in the late 1980s, following the work of Sonstroem (1976, 1978), Fox and Corbin (1989) developed a multidimensional and hierarchical model of the physical self-concept following Shavelson et al. (1976) structure. In this model, the upper level is occupied by a generic construct representing GSC. The GSC is then subdivided into specific domains (physical, academic, social, etc.), which themselves may be subdivided into subdomains (Marsh, 1997). In the case of Fox and Corbin (1989) model, the domain level is occupied by a construct representing the generic physical self-concept domain (i.e., general feelings of happiness, satisfaction, and pride in the physical self). The subdomain level is then occupied by four more specific dimensions of physical self-perceptions: sport competence (SC: athletic ability, ability to learn sports, etc.), physical condition (PC: stamina, fitness, etc.), physical attractiveness (PA: physical attractiveness, ability to maintain an attractive body over time, etc.), and physical strength (PS: perceived strength, muscle development, etc.). This multidimensional and hierarchical conception is represented in Fig. 1.

To measure and validate this representation, two prominent instruments were successively developed: the *Physical Self-Perception Profile* (PSPP) of Fox and Corbin (1989) and the *Physical Self-Description Questionnaire* (PSDQ) of Marsh and Redmayne (1994). Both instruments are based on a multidimensional and hierarchical conception of the self-concept in the physical domain. Differences between these instruments are numerous and include the number of measured subdomains (the PSDQ has 11 subscales, while the PSPP has 6) and the item-response format (the PSDQ relies on Likert-type scales, while the PSPP relies on a structured alternative format).[2] For the present investigation, only the characteristics of the PSPP of Fox and Corbin (1989) will be presented in details.

The PSPP comprises 30 items (i.e., 6 for each dimension of the physical self-concept). Additionally, 6 items from the Rosenberg Self-Esteem Inventory (RSEI; Rosenberg, 1965) are often used in conjunction with the PSPP to assess GSC. The PSPP utilizes a structured alternative format (i.e., paired forced choice with a 4-point answer scale) providing a possible range of 1–4 for each item. The RSEI rely on a 4-point Likert scale, also providing a possible range of 1–4. The internal consistency coefficients for the original instrument ranged from $\alpha = .81$ to .92 for the six scales (including the RSEI), and the test–retest stability correlations coefficients ranged from $r = .74$ to .92 over a 16-day period and between $r = .81$ to .88 over a 23-day period. The content and factor validity of this instrument was originally verified in a large sample of North American college students (Fox & Corbin, 1989). Further studies conducted in college and adult samples from other countries primarily focused on English-speaking populations (Hagger, Asçı, & Lindwall, 2004; Page, Ashford, Fox, & Biddle, 1993; Sonstroem, Speliotis, & Fava, 1992). Fortunately, this model was also replicated in

---

[1] In the present investigation, GSC and global self-esteem are considered as synonyms since their distinction has not yet been clearly established (Byrne, 1996).

[2] For more details on the psychometric differences between the PSDQ and PSPP in various samples differing in age, culture or language, see Marsh et al. (1994, 2002, 2006).

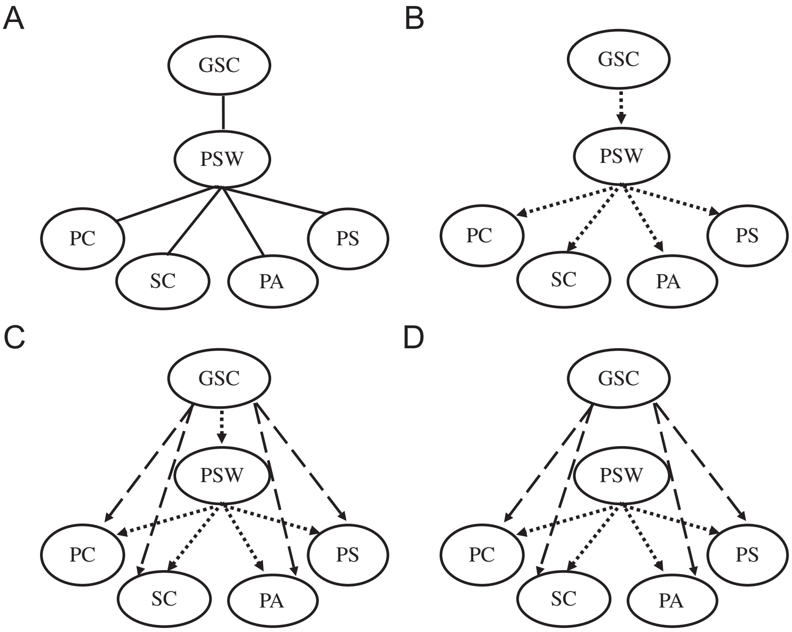

Fig. 1. Multidimensional and hierarchical conceptualization of the physical self-concept of Fox and Corbin (1989) and alternative models proposed by Sonstroem et al. (1994). *Notes*: GSC: global self-concept; PSW: physical self-worth; PC: physical condition; SC: sport competence; PA: physical attractiveness; PS: physical strength; A: multidimensional and hierarchical conceptualization of the physical self-concept; B: structural equation model suggested by Fox and Corbin (1989), showing relations between the four physical self-perceptions subdomain factors, the physical self-worth factor, and the global self-concept factor; C: first alternative model with direct effects of global self-concept factor on each of the four physical self-perception subdomain factors proposed by Sonstroem et al. (1994); D: second alternative model with effect of global self-concept on physical self-worth factor constrained to zero proposed by Sonstroem et al. (1994).

non-English-speaking countries, such as Belgium and the Netherlands (Van de Vliet et al., 2002), Israel (Marsh, Bar-Eli, Zach, & Richards, 2006), Portugal (Fonseca & Fox, 2002), Spain (Atzienga, Balaguer, Moreno, & Fox, 2004), Sweden (Hagger et al., 2004), and Turkey (Hagger et al., 2004; Marsh, Asçl, & Marco, 2002).

In a series of three studies, Ninot, Delignières, and Fortes (2000), developed and validated a French adaptation of the PSPP–RSEI combination: the *Physical Self-Inventory* (PSI). In the first study, different pre-test versions of the French adaptation of the PSPP were tested in a sample of 127 participants (55.1% males). The first pre-test involved a literal translation of the PSPP items and relied on the original structured alternative format. Some of these pre-test participants expressed difficulty in understanding items addressing feelings of pride in themselves and in their physical capabilities simultaneously. To address this issue while preserving the original meaning of the PSPP, a second pre-test version was developed using the committee technique. This second pre-test revealed that the problem directly involved the structured alternative response format of the original PSPP. This is consistent with the results from other studies of the PSPP (Marsh et al., 2002, 2006; Marsh, Richards, Johnson, Roche, & Tremayne, 1994). In accordance with the literature on this topic (Wichstrom, 1995), Ninot et al. (2000) produced a third pre-test version of the instrument in which the original response format was replaced by a six-point Likert scale. Following these pre-tests, the structure of the resulting PSI was verified using the principal component analyses (PCA) and internal consistency analyses. Results were good overall but the internal consistency coefficients remained unsatisfactory ($\alpha < .60$) for the physical self-worth (PSW) scale. Consequently, the authors decided to remove the items from the original PSW scale and to replace them with five items taken from the Self-Description Questionnaire-III of Marsh and O'Neill (1984). In addition, to assess GSC, five items taken from the *Self-Esteem Inventory, school version* of Coppersmith (1967, 1984) were also incorporated.

In the second study, the resulting 30-item version (5 items per subscale, including GSC) was administered to 168 participants (44.0% males). The results from the PCA were generally satisfactory but suggested significant cross-factor loadings involving five of the items. These problematic items were removed and the analysis was

Table 1
Items of the physical self-inventory (Ninot, Delignières, & Fortes, 2000)

1. J'ai une bonne opinion de moi-même (*I have a good opinion of myself*)	GSC	SF	VSF
2. Globalement, je suis satisfait de mes capacités physiques (*Globally, I'm proud of what I can do physically*)	PSW	SF	VSF
3. Je ne peux pas courir longtemps sans m'arrêter (*I can't run for a long time without stopping*)	PC		
4. Je trouve la plupart des sports faciles (*I find that all sports are easy for me*)	SC		
5. Je n'aime pas beaucoup mon apparence physique (*I don't like very much the appearance of my body*)	PA	SF	
6. Je pense être plus fort que la moyenne (*I'm physically stronger than most people*)	PS	SF	VSF
7. Il y a des tas de choses en moi que j'aimerais changer (*There are many things in myself that I would change*)	GSC	SF	
8. Je suis content de ce que je suis et de ce que je peux faire physiquement (*I'm happy with who I am and what I can do physically*)	PSW	SF	VSF
9. Je serais bon dans une épreuve d'endurance (*I would be good at aerobic exercise*)	PC	SF	VSF
10. Je trouve que je suis bon dans tous les sports (*I find that I'm good in all sports*)	SC	SF	
11. J'ai un corps agréable à regarder (*I have a nice body to look at*)	PA	SF	VSF
12. Je serais bon dans une épreuve de force (*I would be good at exercise that depends on strength*)	PS	SF	VSF
13. Je regrette souvent ce que j'ai fait (*I often regret what I have done*)	GSC		
14. Je suis confiant vis-à-vis de ma valeur physique (*I'm confident about my physical self-worth*)	PSW	SF	
15. Je pense pouvoir courir longtemps sans être fatigué (*I think I could run for a long time without tiring*)	PC	SF	VSF
16. Je me débrouille bien dans tous les sports (*I can find a way out of difficulties in all sports*)	SC	SF	VSF
17. Personne ne me trouve beau (*Nobody finds me good-looking*)	PA	SF	VSF
18. Face à des situations demandant de la force, je suis le premier à proposer mes services (*Faced with a situation requiring physical strength, I'm the first to offer help*)	PS	SF	
19. J'ai souvent honte de moi (*I'm often ashamed of myself*)	GSC		
20. En général, je suis fier de mes possibilités physiques (*Globally, I'm proud of my physical abilities*)	PSW		
21. Je pourrais courir 5 km sans m'arrêter (*I could run five kilometers without stopping*)	PC	SF	
22. Je réussis bien en sport (*I do well in sports*)	SC	SF	VSF
23. Je voudrais rester comme je suis (*I would like to stay as I am*)	GSC	SF	VSF
24. Je suis bien avec mon corps (*I feel comfortable in my body*)	PSW		
25. Je ne suis pas très bon dans les activités d'endurance telles que le vélo ou la course (*I'm really not good at endurance activities, such as cycling or running*)	PC		

Notes: GSC: global self-concept; PSW: physical self-worth; PC: physical condition; SC: sport competence; PA: physical attractiveness; PS: physical strength; SF: items retained in the short form; VSF: items retained in the very short form.

replicated, supporting the proposed structure. Then, a confirmatory factor analysis (CFA) realized on this new version (i.e., GSC, PSW, and PC: 5 items per scale; SC: 4 items; PA and PS: 3 items per scale) provided acceptable goodness of fit indices (GFI: .948; AGFI: .937; IFI: .941; RMSEA: .17). The obtained internal consistency coefficients were also satisfactory for all scales ($\alpha = .76$–.90).

Finally, in the third study the temporal stability of the 25-item PSI (items are reported in Table 1) was estimated on 50 participants (50.0% males) retested over a 1-month period. The results revealed that the responses were highly consistent over time (i.e., $r = .90$–.96). Recent studies confirmed these satisfactory psychometric properties within adult French samples (Masse, Jung, & Pfister, 2001; Stephan, Bilard, Ninot, & Delignières, 2003; Stephan & Maïano, 2007).

The present studies

The PSI therefore appears to provide a viable French alternative to the PSPP and a complementary instrument to the French version of the PSDQ (Guérin, Marsh, & Famose, 2004) for sport and exercise psychology researchers interested in assessing adults' physical self-perceptions. However, the reliability, validity, and appropriateness of this instrument for younger French populations remain unknown and could not be taken for granted given the results obtained from similar verifications involving the PSPP. In fact, whereas some researchers provided satisfactory psychometric properties for the PSDQ in adolescent samples

(Guérin et al., 2004; Marsh et al., 1994); other researchers examining the psychometric properties of the original PSPP in children and adolescents populations expressed concerns regarding its applicability to younger populations (Biddle et al., 1993; Marsh et al., 1994). Indeed, their results suggested that (a) children's and adolescents' cognitive abilities are more limited than those of adults, which makes it harder for them to clearly distinguish their own physical self-evaluations across a variety of highly specific subdomains (i.e., numerous cross-loadings among factors); (b) the items used in the adult's PSPP often rely on abstract formulations that can be hard to understand for younger subjects; and (c) young populations often have difficulties understanding the PSPP structure-alternative response scale. To address some of these concerns, the PSPP was modified according to these observations. These modified versions were then successfully validated in children and adolescents of English-speaking samples (CY-PSPP; Eklund, Whitehead, & Welk, 1997; Whitehead, 1995).[3] The PSPP for younger populations has also been similarly validated cross-culturally (Aşçı, Eklund, Whitehead, Kirazci, & Koca, 2005; Bernardo & Matos, 2003; Hagger, Ashford, & Stambulova, 1997; Hagger, Biddle, Chow, Stambulova, & Kavussanu, 2003; Hagger, Biddle, & Wang, 2005; Moreno, Cervelló, Vear, & Ruiz, 2007).

Accordingly, a recent attempt to directly use the adult PSI in a longitudinal sample of French adolescents indicated that this instrument did indeed present some confusing items and was unwieldy due to its length (approximately 8–10 min), especially when used in combination with multiple other instruments and in the context of multiple, idiographic measurements (Maïano, Ninot, Morin, & Bilard, 2007). In order to address this concern, especially in the context of in-depth idiographic studies (i.e., repeated measures assessment), the present study's goal is to adapt the PSI to French adolescent populations and to reduce its length by half, in order to produce a *very short* form of the PSI comprising a total of 12 items (two items per factor).

More precisely, the purpose of the present study was threefold: (a) testing the factor validity and reliability of the original PSI in a sample of adolescents; (b) developing and testing the factor validity and reliability of a *very short* form of the PSI (PSI-VSF) in adolescents; and (c) testing the equivalence of the factor pattern, structural parameters, latent mean structure, and criterion-related validity of both forms of the PSI. These studies were performed following recent recommendations for the development of *short* form tests (Marsh, Ellis, Parada, Richards, & Heubeck, 2005; Smith, McCarthy, & Anderson, 2000). These suggestions point to the fact that test developers should start with a strong long form of the instrument and show within independent or cross-validation samples that: (a) the *short form* retains content coverage of each factor; (b) the *short form* is adequately reliable (i.e., maintains reliability estimates of at least .80); (c) factor structure of the *short form* provides goodness of fit indexes that meets acceptable standards; (d) the *short form* retains the factor structure (i.e., multidimensionality) of the *long form*; (e) the *short form* preserves the hierarchical structure of the *long form*; (f) the *short form* has adequate overlapping factor structure (i.e., multidimensionality and hierarchy) with the *long form*; (g) the mean structure of latent factors (i.e., reproduced item and latent variable means) are equivalent between the *short* and the *long form*; and (h) criterion-related validity is invariant between the *short* and the *long form*. To assess the criterion-related validity of the resulting PSI, gender was chosen as the external criterion, because it represents one of the most studied variables in the physical self-concept literature. In fact, numerous studies on adolescent samples have previously uncovered gender-based differences favoring boys in global and physical self-concepts (Aşçı, 2002; Hagger et al., 2005; Klomsten, Shaalvik, & Espnes, 2004; Maïano et al., 2006).

Method

Sample and procedure

Study 1: A sample of 1018 adolescents, composed of 541 boys (53.14%, $M = 12.88$ years, s.d. = 1.52) and 477 girls (46.86%, $M = 13.06$ years, s.d. = 1.53) aged between 11 and 16 years and attending regular physical

---

[3] In these studies, the PSPP was modified using age-appropriate terminology. In addition, the sport and athletic competence scale of Harter (1982) was used in place of the original sport competence scale. They found that the structured alternative format did not pose problem. Indeed, Welk, Corbin, Dowell, and Harris (1997) compared two rating scales: a four-point structured alternative format and a four-point Likert scale. The structured alternative format produced cleaner results.

education classes, was recruited from five middle and high schools located in southern and northern France. This overall sample was first stratified according to both gender (boys and girls) and age (11–16 years). Then, each participant with similar characteristics (e.g., 11 year-old boys) was pooled in the same category. Finally, in each category the participants were randomly assigned into two sub-samples using a numerical table. The first sub-sample (i.e., sub-sample A: $n = 509$, $M = 12.97$ years, s.d. = 1.56) was composed of 271 boys (53.24%, $M = 12.88$ years, s.d. = 1.56) and 238 girls (46.76%, $M = 13.08$ years, s.d. = 1.56). The second sub-sample (i.e., sub-sample B: $n = 509$, $M = 12.95$ years, s.d. = 1.52) comprised 270 boys (53.04%, $M = 12.85$ years, s.d. = 1.51) and 239 girls (46.96%, $M = 13.05$ years, s.d. = 1.52). As intended, the two groups were nearly identical with respect to the gender and mean age [$t(1016) = -.23$, $p = .82$]. The adult version of the PSI was administered to all participants in quiet classroom conditions, during physical education classes of up to 30 students. In addition, 132 adolescents (64 girls and 68 boys) from sub-sample B were retested after 2 weeks.

Study 2: A cross-validation sample (i.e., sample 2) of 320 adolescents ($M = 12.81$ years, s.d. = 1.57), aged between 11 and 16 years and attending physical education classes, was recruited from two middle and high schools located in southern and northern France. The overall sample included 160 boys (50%, $M = 12.82$ years, s.d. = 1.57) and 160 girls (50%, $M = 12.88$ years, s.d. = 1.68). The PSI-VSF developed in study 1 was administered to all participants in quiet classroom conditions, during physical education classes of up to 30 students. As intended, the samples from studies 1 and 2 were nearly identical with respect to gender and age [$t(1336) = 1.09$, $p = .27$].

Data analysis

In the two studies, analyses were conducted in several stages. Because of the significant multivariate non-normality of the data (normalized Mardia's coefficients values for skewness and kurtosis: 134.571 and 41.316 for sub-sample A; 68.224 and 28.682 for sub-sample B, 17.329 and 8.456 for sample 2, respectively), they were performed using bootstrapped maximum likelihood estimation with AMOS 4.0 (Arbuckle & Wothke, 1999). Thus, all fit indices provided in these studies were based on Bollen–Stine bootstrap p-value and bootstrap adjusted chi-square (χ^2) and goodness-of-fit indexes (Fouladi, 2000; Yuan & Hayashi, 2003).

Study 1: In the first stage, a CFA was conducted to verify the factor structure of the adult PSI in sub-sample A ($n = 509$). This CFA model a priori hypothesized that: (a) answers to the PSI could be explained by six factors; (b) each item would have a non-zero loading on the PSI factor it was designed to measure, and zero loadings on all other factors; (c) the six factors would be correlated; and (d) measurement error terms would be uncorrelated.

Given findings of inadequate fit for the initially hypothesized model, modifications to the CFA model were performed based on analyses of items' (a) intercorrelations, (b) factor loadings, (c) square multiple correlations, (d) standard error, (e) t-values, and (f) modification indices. The CFA was then reproduced to determine whether the modification resulted in an improved fit. This modification process was continued until a reasonable fit was obtained for the model, as indicated by the fit indices. Finally, structural equation modeling (SEM) was conducted to test the hypothesized hierarchical relationships among the PSI factors (Fig. 1). As a complement, in order to question the role of the PSW domain in the self-concept hierarchy, we also tested the adequacy of two alternative nested hierarchical models to the model proposed by Fox and Corbin (1989), as suggested by Sonstroem, Harlow, and Josephs (1994) and Hagger et al. (2005).[4] In the first alternative model, direct effects of GSC on the four subdomains—mediated by the PSW domain—were estimated as free parameters. In the second alternative model, the direct effects of GSC on the four subdomains were freed and the effect of GSC on PSW was constrained to zero (i.e., unmediated model).

In the second stage, CFA and SEM were conducted on sub-sample B ($n = 509$) to cross-validate the multidimensional and hierarchical factor structure of the PSI version generated in sub-sample A (i.e., the 18-item short form: PSI-SF). Finally, the temporal stability of the adolescent version of the PSI was verified using test–retest reliability correlation for scale scores uncorrected for measurement errors on 132 adolescents from sub-sample B, who were retested over a 2-week period.

---

[4] The authors wish to thank an anonymous reviewer for this suggestion.

In the third stage, sub-sample B was used to develop the 12-item PSI-VSF for adolescents by selecting two appropriate items from each of the six PSI-SF subscales. Following the recommendations of Marsh et al. (2005), the following criteria were used in the selection of items: (1) items that best measured the intended construct, as inferred on the basis of corrected item-total correlation and the size of standardized factor loading in CFA; (2) items that had minimal cross-loadings, as evidenced by AMOS modification indices; (3) items that had a minimal correlated uniquenesses, particularly with other items in the same scale; (4) items that were very seldom left blank by respondents; and (5) positive subjective evaluations of the content of each item in order to maintain the breadth of content of the original construct. CFA and SEM analyses were then used to verify whether the previously obtained multidimensional and hierarchical structure remained stable for the PSI-VSF. Finally, the temporal stability of the PSI-VSF was also estimated using the test–retest reliability correlation for scale scores uncorrected for measurement errors on the data of 132 adolescents from sub-sample B who were retested over a 2-week period.

Study 2: In the first stage, CFA and SEM were applied to sample 2 ($n = 320$) to cross-validate the determined multidimensional and hierarchical structure of the PSI-VSF for adolescents. The invariance of both models (i.e., CFA and SEM) across gender was also verified in the sequential order recommended by Byrne (2001).

In the second stage, multigroups CFAs and SEMs were conducted to test the invariance (based on the PSI-VSF across samples) of the (a) obtained multidimensional (CFA) and hierarchical (SEM) structure; (b) latent mean structures (i.e., intercepts and means of the latent factors; and (c) gender effects. To this end, the invariance of the results based on answers provided to the same PSI-VSF items was verified by comparing sample 2 (which only completed the PSI-VSF) with sub-samples A and B (which completed the PSI-SF). The multiple group: (a) CFAs were conducted in the sequential order recommended by Marsh et al. (2005); (b) SEMs were performed in the sequential order recommended by Hagger et al. (2005); (c) SEM gender effect-invariance tests across samples where conducted in the sequential order recommended by Marsh et al. (2005); (d) equivalence tests of latent mean structures across samples in the hypothesized models were performed in the sequential order recommended by Byrne, Shavelson, and Muthén (1989). In order to perform the multigroup SEM gender effect-invariance tests in the PSI-VSF, a single indicator representing gender (1 = boys; 2 = girls) was built and added to the best model, which was evaluated in the multigroup CFAs performed in this stage. Thus, significant negative path coefficients from the single indicator (gender) to a latent subdomain indicate that the reproduced mean of girls in that subdomain is lower than that of the boys.

In the third and final stage, multigroup CFAs and SEMs were conducted to test the invariance of the (a) obtained multidimensional and hierarchical structure; (b) latent mean structures (i.e., intercept and means of the latent factors; and (c) gender effects based on the PSI-VSF (completed by sample 2) and the PSI-SF (completed by sub-samples A and B). These multigroup analyses were performed by allowing sub-samples A and B to have a different number of observed variables per factor. The invariance tests were conducted in the sequential order recommended respectively by Marsh et al. (2005) for CFA and SEM gender effect-invariance tests across samples, Hagger et al. (2005) for SEMs, and Byrne et al. (1989) for the analyses of equivalence of latent mean structures across samples in the hypothesized models. These multigroup SEMs gender-invariance tests were completed in a manner similar to the one used in the second step.

Model fit: Assessment of fit for the models was based on multiple indicators: the χ^2 statistics, the Comparative Fit Index (CFI), the Tucker–Lewis Index (TLI), and the root mean square error of approximation (RMSEA). Values greater than .90 for CFI and TLI are considered to be indicative of adequate model fit (Byrne, 2005; Hu & Bentler, 1999), although values greater than .95 are preferable. Values smaller than .08 or .06 for the RMSEA support, respectively, acceptable and good model fit (Hu & Bentler, 1999; Vandenberg & Lance, 2000). The factor loadings, square multiple correlations, standard errors, and *t*-values were inspected for appropriate sign and/or magnitude. Critical values for the tests of gender and multigroup invariance in CFA models and SEMs were evaluated by examination of several criterions: χ^2 difference tests (with α set to be .01 to control for error inflation associated with multiple comparisons), CFI changes (Cheung & Rensvold, 2002); and RMSEA fluctuations (Marsh et al., 2005). A CFI difference of −.01 or less between a baseline model and the resulting model indicates that the invariance hypothesis should not be rejected. Alternately, when the difference lies between −.01 and −.02, the researcher should suspect that differences may exist (Vandenberg & Lance, 2000). Non-invariance of a model can be identified when the change in CFI exceeded −.02 (Vandenberg & Lance, 2000). Regarding RMSEA differences, Marsh et al. (2005) indicated that very small differences do represent substantively unimportant variations in model fit. Finally, reliability was computed from the model's standardized parameters estimates, using the formula (Bagozzi & Kimmel, 1995) $\rho = (\Sigma\lambda i)^2/([\Sigma\lambda i]^2 + \Sigma\delta ii)$ where λi are the factor loadings and δii the error variances.

Results

Study 1: factor validity and reliability of the adult PSI and PSI-VSF

Stage 1: CFA models for sub-sample A are displayed in Table 2. The 25-item CFA model exhibited significant bootstrapped χ^2-values, and CFI and TLI under .90. Moreover, the value of the RMSEA indicator exceeded the .08 criterion. Examination of items revealed that some of them: (a) had low factor loadings and square multiple correlations; (b) were highly correlated and cross-loaded; and (c) presented elevated modifications indices. Thus, as illustrated in Table 3, the most problematic items (i.e., 3, 4, 13, 19, 20, 24, and 25) were removed from the adult version of the PSI. After deleting those seven items, a PSI-SF of 18 items adapted for adolescents was obtained.

CFA and SEM for this PSI-SF showed significant bootstrapped χ^2-values (Table 2). As displayed in Table 2, CFI and TLI fit to index values of this version exceeded .90 and RMSEA values were equal to .07. All loadings and uniquenesses in this CFA model were significant and substantial (Table 3). Structural parameter estimates among the latent variables for the SEM analyses were all large and significant (Table 3).

Results from the SEM evaluation of alternative models[5] are displayed in Table 2. Goodness-of-fit statistics for the first of these alternative models were acceptable (i.e., CFI-TLI > .90 and RMSEA equal to .07) and overlapping (i.e., non significant $\Delta\chi^2$-values; ΔCFI < .01; unsubstantial RMSEA variations). The second alternative model exhibited inadequate goodness-of-fit statistics (i.e., CFI-TLI < .90 and RMSEA > .08) and overlapping (i.e., non significant $\Delta\chi^2$-values; ΔCFI > .01; substantial RMSEA variations). Importantly, examination of the *t*-values provided by AMOS showed that three of the four structural paths between GSC and the four subdomains were non-significant. According to these results, and following the recommendations of Sonstroem et al. (1994) and Hagger et al. (2005), the direct effects of GSC on the four subdomains should be restricted to zero. Moreover, modification indices provided by AMOS were also used to identify parameters in the second alternative model that if freed would result in a statistically significant improvement in the model goodness-of-fit. The modification indices indicated that freeing the relationship between PSW and GSC would restore model goodness-of-fit indices to acceptable levels. Moreover, in this second alternative model, the factor pattern coefficients for the relationships from PSW and the four subdomains were higher (i.e., from .670 to .935, Mdn $\lambda = .732$) than the effect of GSC on the subdomains (i.e., from .002 to .369, Mdn $\lambda = .068$). As suggested by Sonstroem et al. (1994) and Hagger et al. (2005), direct effects of GSC on the subdomain factors are redundant and provide further support for the role of PSW in the proposed hierarchical structure.

Descriptive statistics for the PSI-SF adapted for adolescents in sub-sample A are provided in Table 4. In this version, all scales presented modest internal consistency coefficients (i.e., ranging from .74 to .75) (Table 4). Latent variable intercorrelations are provided in Table 5. All of these correlations are quite elevated (i.e., > .50) and statistically significant. In this stage, all latent factor correlations provide evidence for the discriminant validity of the PSI, according to the criteria of Bagozzi and Kimmel (1995). These criteria state that the validity of two distinct constructs is supported when the result of the multiplication of the standard error of the factor correlation by 1.96 is less than unity. They revealed, as found by Fox and Corbin (1989), that the (a) relationship between GSC and PSW was significantly stronger than any of the relationships between GSC and subscales; (b) all of the subscales were significantly and positively correlated with the PSW domain and exhibited stronger significant relationships with PSW than with GSC.

Stage 2: CFA and SEM for the PSI-SF in sub-sample B showed significant bootstrapped χ^2-values (Table 2). As displayed in Table 2, CFI and TLI fit index values of this version exceeded .90 and RMSEA values were equal to or smaller than .08. All loadings and uniquenesses in this CFA model were significant and

---

[5] Unreported alternative models were also tested with sub-sample B and sample 2. These analyses yielded similar findings and are available on request from the first author.

Table 2
Goodness-of-fit statistics of PSI, PSI-SF, and PSI-VSF models[a]

Study and sample	No. of items	Model	Description	χ^2(B–S)	df	CFI	TLI	RMSEA	$\Delta\chi^2$	Δdf	ΔCFI	Δ RMSEA
Study 1—sub-sample A	25	CFA		322.191*	260	.859	.837	.085	–	–		
($n = 509$)	18	CFA		151.401*	120	.935	.917	.072	–	–	–	–
		SEM, hierarchical model	Model A	164.265*	131	.921	.908	.076	–	–	–	–
		SEM, hierarchical model	Model B	159.682*	127	.931	.917	.072	−4.583	4	−.010	−.004
		SEM, hierarchical model	Model C	159.328*	128	.874	.850	.097	−4.937	3	−.047	.021
Study 1—sub-sample B	18	CFA		150.168*	120	.943	.927	.067	–	–	–	–
($n = 509$)		SEM, hierarchical model		163.021*	131	.918	.904	.077	–	–	–	–
Study 1—sub-sample B	12	CFA		47.491*	39	.972	.953	.065	–	–	–	–
($n = 509$)		SEM, hierarchical model		60.558*	50	.933	.911	.077	–	–	–	–
Study 2—sample 2	12	CFA		43.040*	39	.983	.972	.043	–	–	–	–
($n = 320$)		CFA, gender-invariance tests	A—no invariance	88.653*	78	.974	.957	.051	–	–	–	–
			B—λs invariant	93.648*	84	.971	.954	.052	4.995	6	−.003	−.001
			C—λs, δs invariant	105.463*	94	.967	.954	.052	16.810	16	−.007	−.001
			D—λs, δs, ξs invariant	110.074*	100	.965	.954	.052	21.421	22	−.009	−.001
			E—λs, δs, ξs, φs invariant	128.413*	115	.964	.955	.055	39.760	37	−.010	.004
		SEM, hierarchical model		56.873*	50	.940	.921	.073	–	–	–	–
		SEM, gender-invariance tests	A—no invariance	113.940*	100	.913	.885	.082	–	–	–	–
			B—λs invariant	119.901*	106	.913	.892	.079	5.961	6	.000	−.003
			C—λs, ξs/ζs invariant	126.461*	112	.904	.887	.082	12.521	18	−.009	.000
			D—λs, ξs/ζs, γs invariant	129.441*	116	.903	.888	.082	15.501	34	−.010	.000
Study 2—sub-sample A/sub-sample B/sample 2	12/12	Multiple-group CFA	A—no invariance	138.243*	117	.973	.954	.062	–	–	–	–
			B—λs invariant	149.340*	129	.970	.955	.061	11.097	12	−.003	−.002
			C—λs, ξs invariant	159.949*	141	.969	.967	.061	21.706	24	−.004	−.002
			D—λs, φs invariant	181.700*	159	.968	.958	.059	43.457	42	−.005	−.003
			E—λs, φs, ξs invariant	196.603*	171	.966	.957	.059	58.360	54	−.007	−.003
			F—λs, φs, δs invariant	214.840*	183	.964	.957	.061	76.597	78	−.009	−.002
			G—λs, φs, ξs, δs invariant	229.138*	195	.963	.957	.059	90.895	66	−.010	−.003
		Multiple-group SEM	A—no invariance	178.441*	150	.936	.916	.083	–	–	–	–
			B—λs invariant	190.307*	162	.934	.919	.081	11.866	12	−.002	−.002
			C—λs, ξs/ζs invariant	203.374*	174	.929	.919	.081	24.933	24	−.007	−.002
			D—λs, ξs/ζs, γs invariant	209.490*	182	.928	.922	.080	31.049	32	−.008	−.003

		Multiple-group latent mean	A—baseline	149.340*	129	.994	.990	.061	–	–	–	–
			B—FINT invariant	161.170*	141	.993	.989	.064	11.83	12	−.001	.003
			C—LT, FINT invariant	172.664*	152	.992	.988	.068	23.324	23	−.002	.007
		Multiple-group gender effect-invariance tests	A—λs, φs, ξs invariant; gender effect free	217.239*	189	.957	.947	.062	–	–	–	–
			B—λs, φs, ξs invariant; gender effect invariant	229.232*	201	.947	.937	.067	11.99	12	−.010	.005
Study 2—sub-sample A/sub-sample B/sample 2	18/12	Multiple-group CFA	A—No invariance	343.792*	279	.945	.926	.071	–	–	–	–
			B—λs invariant	360.532*	297	.942	.927	.069	16.74	18	−.003	−.002
			C—λs, ξs invariant	370.309*	309	.941	.928	.069	26.517	30	−.004	−.002
			D—λs, φs invariant	392.187*	327	.939	.931	.068	48.395	48	−.006	−.003
			E—λs, φs, ξs invariant	407.104*	339	.937	.930	.068	63.312	60	−.008	−.003
			F—λs, φs, δs invariant	425.234*	349	.935	.930	.069	81.442	82	−.010	−.002
			G—λs, φs, ξs, δs invariant	439.608*	361	.934	.930	.068	95.816	64	−.011	−.003
		Multiple-group SEM	A—no invariance	384.049*	312	.922	.907	.080	–	–	–	–
			B—λs invariant	400.947*	330	.920	.909	.078	16.898	18	−.002	−.002
			C—λs, ξs/ζs invariant	414.707*	342	.916	.909	.078	30.658	30	−.006	−.002
			D—λs, ξs/ζs, γs invariant	424.677*	354	.912	.905	.080	40.628	42	−.010	.000
		Multiple-group latent mean	A—baseline	23.818*	291	.988	.983	.071	–	–	–	–
			B—FINT invariant	35.843*	315	.987	.983	.071	5.968	24	−.001	.000
			C—LT. FINT invariant	69.958*	327	.986	.982	.071	12.101	36	−.002	.000
		Multiple-group gender effect-invariance tests	A—λs, φs, ξs invariant; gender effect free	442.329*	369	.939	.931	.066	–	–	–	–
			B—λs, φs, ξs invariant; gender effect invariant	454.131*	381	.929	.921	.071	11.80	12	−.010	−.005

Notes: CFA: measurement confirmatory factor analytic model; SEM: hypothesized hierarchical multidimensional structural equation model of physical self-concept; χ^2(B–S): Bollen–Stine chi-square; df: degrees of freedom; CFI: comparative fit index; TLI: Tucker–Lewis index; RMSEA: root mean square error of approximation; Model A: structural equation model suggested by Fox and Corbin (1989); Model B: assuming model A, with direct effects of global self-concept factor on each of the four physical self-perception subdomain factors; Model C: assuming model B, with effect of global self-concept on physical self-worth factor constrained to zero; λ: factor loading; ξ: factor variance; φ: factor covariance; δ: uniquenesses; ζ: factor error terms; γ: structural relations among latent constructs; FINT: factor intercepts; LT: latent means; $\Delta\chi^2$: change in goodness-of-fit χ^2 relative to baseline model; Δ df: change in degrees of freedom relative to the baseline model; ΔCFI: change in comparative fit index relative to the baseline model; Δ RMSEA: change in root mean square error of approximation relative to the baseline model.

[a] Bootstrapped goodness of fit indexes are reported in this table because of the significant multivariate non-normality within these data; $*p<.01$.

Table 3
CFA's factor loadings-uniquenesses and SEM's standardized structures loadings[a]

Factor	Item no.	Study 1—sub sample A		Study 1—sub sample B		Study 2—sample 2
		PSI[b]	PSI-SF[c]	PSI-SF[c]	PSI-VSF[d]	PSI-VSF[d]
		$\lambda(\delta)$	$\lambda(\delta)$	$\lambda(\delta)$	$\lambda(\delta)$	$\lambda(\delta)$
GSC	1	.637(.41)†	.658(.43)†	.712(.51)†	.694(.38)†	.630(.30)†
	7	.380(.14)	.351(.12)	.381(.14)	–	–
	13	.186(.03)	–	–	–	–
	19	.344(.12)	–	–	–	–
	23	.686(.47)	.649(.42)	.617(.38)	.597(.26)	.554(.21)
PSW	2	.782(.61)†	.784(.61)†	.773(.60)†	.781(.51)†	.686(.37)†
	8	.789(.62)	.783(.61)	.783(.61)	.789(.52)	.794(.53)
	14	.817(.67)	.746(.56)	.771(.60)	–	–
	20	.736(.54)	–	–	–	–
	24	.598(.36)	–	–	–	–
PC	3	.588(.35)	–	–	–	–
	9	.806(.65)†	.775(.60)†	.809(.66)†	.836(.60)†	.842(.61)†
	15	.882(.78)	.913(.83)	.943(.89)	.908(.72)	.860(.64)
	21	.793(.63)	.793(.63)	.774(.60)	–	–
	25	.436(.19)	–	–	–	–
SC	4	.621(.39)	–	–	–	–
	10	.865(.75)†	.863(.74)†	.853(.73)†	–	–
	16	.830(.69)	.830(.69)	.861(.74)	.845(.66)	.797(.54)
	22	.889(.79)	.887(.79)	.871(.76)	.873(.61)†	.859(.64)†
PA	5	.443(.20)†	.443(.20)†	.377(.14)†	–	–
	11	.767(.59)	.784(.61)	.763(.58)	.732(.34)†	.596(.25)†
	17	.540(.29)	.520(.27)	.440(.19)	.382(.15)	.380(.14)
PS	6	.796(.63)†	.791(.63)†	.729(.53)†	.738(.44)†	.732(.44)†
	12	.807(.65)	.812(.66)	.787(.62)	.718(.41)	.503(.15)
	18	.737(.54)	.737(.54)	.639(.41)	–	–
Parameter			PSI-SF[c]	PSI-SF[c]	PSI-VSF[d]	PSI-VSF[d]
$\xi\text{GSC} \rightarrow \xi\text{PSW}$		–	.863	.841	.866	.679
$\xi\text{PSW} \rightarrow \xi\text{PC}$		–	.689	.733	.772	.699
$\xi\text{PSW} \rightarrow \xi\text{SC}$		–	.903	.833	.897	.923
$\xi\text{PSW} \rightarrow \xi\text{PA}$		–	.834	.858	.896	.737
$\xi\text{PSW} \rightarrow \xi\text{PS}$		–	.727	.790	.861	.935

Notes: †: item that was set to be 1.0; λ: loadings; δ: uniquenesses; GSC: global self-concept; PSW: physical self-worth; PC: physical condition; SC: sport competence; PA: physical attractiveness; PS: physical strength.

[a] All loadings and uniquenesses are significant ($p<.001$).

[b] 25-item.

[c] 18-item.

[d] 12-item.

substantial (Table 3). Structural parameter estimates among the latent variables for the SEM analyses were all large and significant (Table 3). These results support the multidimensional nature and hierarchical structure of the 18-item PSI-SF adapted for adolescents.

The descriptive statistics of PSI-SF from sub-sample B are displayed in Table 4. All of these scales offered modest internal consistency coefficients (i.e., .73–.75) and were similar to those obtained in sub-sample A (Table 4). The latent variable intercorrelations are presented in Table 5. It can be seen that all correlations are again elevated (i.e., >.50) and statistically significant. Once again, according to the criteria of Bagozzi and Kimmel (1995), these results provide evidence for the discriminant validity of the PSI: the relationship between

Table 4
Descriptive statistics for the PSI-SF and PSI-VSF versions based on responses from the pooled sample

Scale	Study 1		Study 1						Study 2	
	Sub-sample A		Sub-sample B						Sample 2	
	PSI-SF[a]		PSI-SF[a]			PSI-VSF[b]			PSI-VSF[b]	
	ρ	M(SD)	ρ	M(SD)	r_{tt}	ρ	M(SD)	r_{tt}	ρ	M(SD)
GSC	.74	4.10(1.13)	.74	3.89(1.17)	.81*	.72	4.10(1.27)	.79*	.73	4.37(1.18)
PSW	.75	4.39(1.17)	.75	3.99(1.25)	.84*	.71	4.10(1.29)	.81*	.71	4.32(1.22)
PC	.75	3.61(1.56)	.75	3.35(1.56)	.80*	.70	3.42(1.58)	.79*	.70	3.05(1.68)
SC	.75	4.11(1.33)	.75	3.77(1.35)	.83*	.70	3.94(1.37)	.81*	.70	4.12(1.31)
PA	.74	3.91(1.25)	.73	3.62(1.26)	.74*	.72	3.62(1.32)	.74*	.71	3.78(1.20)
PS	.75	3.61(1.50)	.75	3.32(1.39)	.82*	.71	3.28(1.47)	.82*	.76	3.36(1.45)

Notes: Responses to items on all forms of the PSI varied from 1 to 6; GSC: global self-concept; PSW: physical self-worth; PC: physical condition; SC: sport competence; PA: physical attractiveness; PS: physical strength; ρ: composite reliability estimate; M: mean; SD: standard deviations; r_{tt}: test–retest reliability correlations; $^{*}p < .001$.

[a] 18-item.

[b] 12-item.

Table 5
Factor correlations among latent factors according to PSI-SF or PSI-VSF version*

Scale	GSC	PSW	PC	SC	PA	PS
GSC	1.00					
PSW	.912[a] .897[b] .944[c] .702[d]	1.00				
PC	.583 .543 .588 .369	.638 .676 .714 .592	1.00			
SC	.688 .652 .669 .534	.887 .859 .866 .945	.701 .739 .770 .650	1.00		
PA	.825 .903 .998 .661	.835 .846 .865 .831	.550 .625 .681 .572	.717 .678 .721 .685	1.00	
PS	.571 .545 .637 .614	.668 .742 .771 .906	.503 .594 .688 .656	.727 .787 .857 .874	.630 .713 .847 .609	1.00

Notes: GSC: global self-concept; PSW: physical self-worth; PC: physical condition; SC: sport competence; PA: physical attractiveness; PS: physical strength.

*All correlations are significant ($p < .01$).

[a] PSI-SF (i.e., 18-item) from sub-sample A.

[b] PSI-SF (i.e., 18-item) from sub-sample B.

[c] PSI-VSF from sub-sample B.

[d] PSI-VSF from sample 2.

GSC and PSW was significant but not stronger than the relationship between GSC and the PA scale (Table 5). Nevertheless, all of the subscales were significantly and positively correlated with the PSW domain (Table 5). They also showed stronger significant relationships with PSW than with GSC, with an exception of the PA scale (Table 5). The results from the test–retest stability correlation coefficients of the PSI-SF adapted for adolescents are reported in Table 2. They were highly satisfactory in all cases.

Stage 3: Following the aforementioned criterion for the selection of items in the construction of shorter scales, six items (5, 7, 10, 14, 18, 21) were selected for deletion from the PSI-SF according to the results from sub-sample B. The remaining 12 items form the PSI-VSF. CFA model and SEM hierarchical model of the PSI-VSF from sub-sample B showed significant bootstrapped χ^2-values (Table 2). As shown in Table 2, CFI and TLI fit index values exceeded .90 and RMSEA values were smaller than .08.[6] Table 3 shows that all loadings and uniquenesses in this CFA model were significant and substantial. The structural parameters among the latent variables (SEM) were all large and significant (Table 3). Thus, the multidimensional and hierarchical nature of the PSI-VSF is supported by the results.

The descriptive statistics for PSI-VSF from sub-sample B are displayed in Table 4. The internal consistency coefficients were modest (i.e., .70–.72) but exceeded the recommended minimum of .70 (Bagozzi & Kimmel, 1995), which can be considered satisfactory given the shortness of the subscales (Streiner, 2003). Latent intercorrelations are presented in Table 5. The results indicate that all correlations are elevated (i.e., $>.50$) and statistically significant. Once again, these results provide evidence for the discriminant validity of the PSI (Bagozzi & Kimmel, 1995): (a) relationships between GSC and PSW were significant but not stronger than the relationships between GSC and the PA; (b) all subscales were significantly and positively correlated with the PSW domain and showed stronger significant relationships with PSW than with GSC, with an exception for PA. As displayed in Table 4, the test–retest reliability correlation coefficients of the PSI-VSF were highly satisfactory in all cases.

Study 2: cross-validation of the PSI-SF and PSI-VSF

Stage 1: As presented in Table 2, the CFA and SEM models of the PSI-VSF showed significant bootstrapped χ^2-values. The CFI and TLI fit index values of this version exceeded .90 and the RMSEA values were smaller than .08 (Table 2). All factor loadings and uniquenesses in this CFA model were significant and important (Table 3). The structural parameter estimates among the latent variables for the SEM analyses were also all large and significant (Table 3). Thus, CFA and SEM results confirmed the multidimensional nature and hierarchical ordering of the PSI-VSF.

The descriptive statistics for the PSI-VSF administered to sample 2 are displayed in Table 4. As shown in this table, the internal consistency coefficients were modest in all cases (i.e., .70–.76). Results from the latent variable intercorrelations are presented in Table 5. Again, all of these correlations are elevated (i.e., $>.50$) and statistically significant. Analyses of the latent variable intercorrelations provide evidence for the discriminant validity of the PSI (Bagozzi & Kimmel, 1995): (a) the relation between GSC and PSW was significantly stronger than any of the relationships between GSC and the other subscales; (b) all of the subscales were significantly and positively correlated with the PSW domain and exhibited stronger significant relationships with PSW than with GSC, with the exception of PA.

The results from the CFA and SEM gender-invariance analyses are presented in Table 2.[7] These results show that: (a) CFI and TLI values exceeded .90, except the TLI values of the SEMs; (b) ΔCFI values did not exceed $-.01$; (c) RMSEA values for CFAs and SEMs were, respectively, equal to .05 and .08 and the ΔRMSEA values remained small and non-substantial; (d) the χ^2 tests were all significant but none of the χ^2

---

[6] Relying on two indicators to estimate each construct creates a locally underidentified latent construct (although the overall model may be overidentified). To address this problem, Little, Lindenberger, and Nesselroade (1999) suggest placing equality constraints on the loadings of both indicators to help locate the construct at the true centroid of the indicators. Little et al. (1999) noted that this procedure results in a decrease in the fit of the measurement models, which should not overly concern researchers if it does not become dramatic. The measurement models were thus estimated twice for PSI-VSF, once without the equality constraints and once including them. Similar results were found in both cases.

[7] Unreported gender-invariance tests were also conducted on sub-samples A and B. These analyses provided similar findings and are available on request from the first author.

difference test was significant. These results suggest that the measurement model and hierarchical structure of the PSI-VSF is fully invariant across gender.

Stage 2: The results from the CFA and SEM multigroup analyses for the between PSI-VSF completed by sub-samples A and B and sample 2 are presented in Table 2. These results show that: (a) the CFI, TLI, and RMSEA values all indicate adequate model fit; (b) ΔCFI values did not exceed .01; (c) the ΔRMSEA values remained small and non-substantial; (d) the χ^2 tests were all significant but none of the $\Delta\chi^2$ tests was significant. These results suggest that the measurement model and hierarchical structure of the PSI-VSF was fully invariant across the three samples. The results from the multigroup invariance analyses of PSI-VSF latent mean structures are presented in Table 2 and parallel to those from the more classical multiple-group invariance tests, which supports the equivalence of the mean structure of the very short form (item intercepts and latent means) across the three samples.

In order to evaluate the criterion-related validity of the PSI-VSF, the invariance of the effect of gender on the latent variables across samples was also estimated in the context of SEM models. In these models, in conformity with the previous analyses, the measurement model of the PSI was also constrained to be invariant across groups (i.e., factor loading, factor variance, and factor covariance invariant). The results from these analyses are presented in Table 2 and show that: (a) the CFI, TLI, and RMSEA values all indicate adequate model fit; (b) ΔCFI values did not exceed .01; (c) the ΔRMSEA values remained small and non-substantial; (d) the χ^2 tests were all significant but none of the $\Delta\chi^2$ tests was significant. These results support the invariance of the gender effect across samples. Specific results from the standardized path coefficients reflecting the effects of gender indicate that boys tended to present higher scores than girls on the GSC ($b = -.16$, $t = -2.72$, $p < .05$, $d = .15$), PSW ($b = -.27$, $t = -4.10$, $p < .05$, $d = .23$), SC ($b = -.37$, $t = -4.94$, $p < .05$, $d = .27$), PA ($b = -.32$, $t = -3.82$, $p < .05$, $d = .21$), and PS subscales ($b = -.40$, $t = -4.50$, $p < .05$, $d = .25$). However, no statistically significant gender difference was found on the PC subscale ($b = .04$, $t = -0.46$, $p > .05$, $d = .03$).

Stage 3: The results from CFA and SEM multiple-group invariance tests comparing the results from the PSI-SF completed by sub-samples A and B with those from the PSI-VSF completed by sample 2 are displayed in the last section of Table 2. These results show that: (a) the CFI, TLI, and RMSEA values all indicate adequate model fit; (b) ΔCFI values did not exceed .01; (c) the ΔRMSEA values remained small and non-substantial; and (d) the χ^2 tests were all significant but none of the χ^2 difference tests was significant. These results suggest that the measurement model and hierarchical structure of the PSI-VSF fully overlapped those from the PSI-SF. The results from the multigroup invariance analyses of PSI-VSF latent mean structures are also presented in the last section of Table 2. They parallel those from the preceding multiple-group invariance tests and support the equivalence of the mean structures of the PSI-SF and PSI-VSF (item intercepts and latent means).

Finally, in order to evaluate the equivalence of the criterion-related validity of the various versions of the PSI (SF and VSF), the invariance of the effect of gender on the latent variables across versions was also estimated in the context of SEM models. In these models, in conformity with the previous analyses, the measurement model of the PSI was also constrained to be invariant across versions (i.e., factor loading, factor variance, and factor covariance invariant). The results from these analyses are presented at the end of Table 2 and show that: (a) the CFI, TLI, and RMSEA values all indicate adequate model fit; (b) ΔCFI values did not exceed .01; (c) the ΔRMSEA values remained small and non-substantial; (d) the χ^2 tests were all significant but none of the χ^2 difference test was significant. These results support the invariance of the gender effect across PSI versions. Specific results from the standardized path coefficients reflecting the effects of gender indicate that boys tended to present higher scores than girls on the GSC ($b = -.23$, $t = -3.91$, $p < .05$, $d = .21$), PSW ($b = -.37$, $t = -5.92$, $p < .05$, $d = .32$), SC ($b = -.41$, $t = -5.65$, $p < .05$, $d = .31$), PA ($b = -.43$, $t = -5.22$, $p < .05$, $d = .29$), and PS subscales ($b = -.48$, $t = -5.47$, $p < .05$, $d = .30$). However, no statistically significant gender difference was found on the PC subscale ($b = .11$, $t = -1.24$, $p > .05$, $d = .07$).

Discussion

The first objective of the present study was to verify the factor validity and reliability of the adult PSI (Ninot et al., 2000) in a sample of adolescents. Results from a preliminary CFA performed with sub-sample A suggest that the adult PSI (i.e., 25 items; Ninot et al., 2000) was not suitable for use in adolescent's populations. These results confirm the concerns expressed by other researchers (Biddle et al., 1993; Marsh et al., 1994) regarding the applicability of the PSPP (i.e., differentiation of subdomains; item formulations) to younger populations. In fact, the CFA revealed that some items from the PSI were inappropriate (i.e., cross-loadings, low factor loadings and uniquenesses) for French adolescents. Following the removal of the most problematic items from this version, an 18-item PSI-SF version suited for adolescents was retested. CFAs and SEMs performed with this new version of the PSI (PSI-SF) confirmed, in two randomized sub-samples, that (a) the measurement and hierarchical model of the PSI-SF followed the structure proposed in Fox and Corbin (1989) physical self-concept model and were fully invariant across gender; and that (b) the results of the PSI-SF were highly stable over time (i.e., test–retest reliability).

A second objective of this series of studies was to develop and to test the factor validity and reliability of a PSI-VSF (i.e., 12 items, PSI-VSF), following recent recommendations by Marsh et al. (2005). The fundamental requirement in developing a *short form* of the instrument was to start with a well-established or strong long form. However, this was not the case in the present study given the fact that the full version of the PSI was unsuitable for adolescents. However, despite some limitations that will be addressed later, the PSI-SF appeared to represent a viable alternative in the measurement of French adolescents' physical self-concept. The PSI-VSF was thus developed on the basis of this 18-item version of the PSI (PSI-SF). CFAs and SEMs performed with this *very short* version of the PSI (PSI-VSF) confirmed, in two independent samples, that (a) the measurement and hierarchical model of the PSI-VSF followed the structure proposed in Fox and Corbin (1989) physical self-concept model and were fully invariant across gender; and that (b) the results of the PSI-VSF were highly stable over time (i.e., test–retest reliability). The objective underlying the development of the PSI-VSF was to reduce the time required to assess adolescents' physical self-concepts in the context of extensive longitudinal or idiographic studies. During study 2, participants' completion of the PSI-VSF generally took between 4 and 5 min, half the time required to complete the original 25-items PSI version.

In addition, multiple-groups CFAs and SEMs analyses were also conducted in order to answer the third objective of this series of studies. They demonstrated that the measurement model, mean structure, structural parameters and criterion-related validity (gender effects on the latent variables) of both the PSI-SF and PSI-VSF remained equivalent across samples and versions. The specific results concerning the criterions-related validity of the PSI-SF and PSI-VSF indicate that boys reported significant higher score than girls on most of the evaluated physical self-concept dimensions. These results were also invariant across groups' and versions. This replicates results from other studies in the physical self-concept literature (Aşçı, 2002; Hagger et al., 2005; Klomsten et al., 2004; Maïano et al., 2006) and supports the criterion-related validity of both forms of the instrument.

Nevertheless, additional results indicate that both versions (PSI-SF and PSI-VSF) may not yet be optimal and that replication studies will be needed. Indeed, the latent factor intercorrelations remain elevated (i.e., $>.50$) and the subscale internal consistency coefficients are located near the inferior limit of acceptability (i.e., ranging from .70 to .76). On the one hand, the latent variable intercorrelations from both versions (i.e., *short* and *very short*) do confirm the proposed interrelations between the various PSI sub-dimensions (discriminant validity) empirically and are of similar magnitude to those found in studies evaluating the construct validity of other self-concept instruments such as the PSPP and CY-PSPP (Atzienga et al., 2004; Fox & Corbin, 1989; Hagger et al., 2003, 2004, 2005; Marsh et al., 1994, 2002, 2006; Sonstroem et al., 1992). On the other hand, the strength of those relations also bring into question the real independence of some of the models' sub-dimensions, and by extension their discriminant validity, a finding that has already been observed by Marsh et al. (1994, 2002, 2006) on analyses of the PSPP. According to these authors these very large correlations reflect the fact that the PSPP (and PSI by extension) attempt to cover a broader range of concepts than the PSDQ with a smaller number of dimensions. Hence, the content covered in each of the PSI (which was developed as a French adaptation of the PSPP) factors is more heterogeneous than the more specific dimensions of the PSDQ. Marsh et al. (2002) suggest that these elevated correlations could have been explained by the structured alternative response scale used in the PSPP. As the present study relied on a Likert-type answer scale and also found elevated latent variable intercorrelations, it suggests that this explanation may be erroneous. Regarding the modest internal reliability coefficients observed within the

18-item and 12-item versions (i.e., less than the .80 cutoff point recommended by Marsh et al., 2005), they may simply reflect the limited number of items used to measure each dimension of the PSI (i.e., three to two items per dimension). In fact, Streiner (2003) note that internal consistency coefficients are strongly and positively affected by the number of items in a scale (they increase and decrease as a function of the number of items included in the scale) and that, consequently, acceptability levels must be adjusted in the context of *short* or *very short* measurement scales.

Some limitations of the current studies must be taken into account in the interpretation of the findings. First, according to Marsh et al. (2005), the fundamental requirement in the development of a *short* version of a questionnaire is to start with a strong long form. This requirement was missing from the present studies. However, the acceptable psychometric properties of both the PSI-SF and PSI-VSF suggest that they may represent promising alternatives for the measurement of French adolescents' physical self-concepts pending replication of the current results. Second, the internal consistency and discriminant validity of the PSI-SF and PSI-VSF remain modest, potentially due to the brevity of both scales. For this reason, the use of both instruments cannot still be recommended unless there is a clear need for a *short* or a *very short* scale (i.e., extensive longitudinal or idiographic studies) in a French adolescent sample. Third, the present studies were based on a convenience sample of normally achieving adolescents, which could not be considered as representative of the French adolescent population. Again, this indicates that the use of both instruments should be limited to normally achieving French adolescents. Clearly, before the generalizability of the PSI-VSF to other cultural or linguistic groups (i.e. English-speaking adolescents) can be systematically investigated in additional studies, its cross-cultural or -linguistic use cannot be recommended. Fourth, these adolescents were tested in the context of their physical education classes, a context which may have heightened the salience of their physical self-concepts. It thus remains unknown whether the present results, especially regarding the discriminant validity of the various subscales, could be replicated in other contexts (i.e., regular classroom, youth clubs, sport, leisure associations, etc.).

Fifth, the reliance on a cross-sectional sample also precludes the verification of the developmental stability of the PSI-SF and PSI-VSF for adolescents. Although the present study allowed for the verification of the 2-week test–retest reliability of the instruments, a complete test of their construct validity would involve verifying whether the obtained results followed the same patterns of continuity and change observed in the physical self-concept literature (Marsh, 1996). Sixth, it is currently unknown whether the PSI-VSF represents a practical and useful *short* version of the PSI. Indeed, given the fact that the "long" form of the PSI for adolescents has been reduced to 18 items (PSI-SF), it remains to be verified whether the 12-item version really takes less time to complete, whether it really results in a reduced number of missing data, and whether these and other potential advantages offset the costs of moving from three to two items per dimension. Finally, the criterion-related validity of both instruments was not evaluated with regard to: (a) other physical self-concept instruments (i.e., PSDQ, PSPP, etc.) within a multitrait–multimethod framework, (b) instruments measuring other relevant concepts (i.e., body image, social anxiety, etc.), and (c) multiple external criterions (i.e., fitness examinations, body mass index, grades in physical education classes, etc.). These issues should clearly be addressed in the context of future studies.

In conclusion, it is important that the present study be viewed as a preliminary step in the validation of a *short* and a *very short* form of a physical self-concept instrument that could be used with minimal cost within extensive longitudinal or idiographic studies of adolescents' samples. Regarding the aforementioned limitations of these studies, it would be premature at this time to recommend the use of either instrument (PSI-SF or PSI-VSF), unless the need for a *short* or a *very short* instrument is clear and in that case, the use of the instruments should, for the time being be restricted to samples of normally achieving French adolescents. Despite these limitations, present studies provide promising preliminary evidence regarding the reliability and validity of a *short* and a PSI-VSF for French adolescent.

Acknowledgments

The authors would like to thank two anonymous reviewers for their precious comments on a previous version of this manuscript and Cathy Boissin, Guy Boussion, Françoise Jullien, and Sylvie Massucco for their invaluable assistance in data collection.

References

Arbuckle, J., & Wothke, W. (1999). *AMOS 4.0 user's guide*. Chicago: Smallwaters.

Aşçı, F. H. (2002). An investigation of age and gender differences in physical self-concept among Turkish late adolescents. *Adolescence*, *37*, 365–371.

Aşçı, F. H., Eklund, R. C., Whitehead, J. R., Kirazci, S., & Koca, C. (2005). Use of the CY-PSPP in other cultures: A preliminary investigation of its factorial validity for Turkish children and youth. *Psychology of Sport and Exercise*, *6*, 33–50.

Atzienga, F. L., Balaguer, I., Moreno, Y., & Fox, K. R. (2004). El perfil de autopercepción física: propiedades psicométricas de la versión española y análisis de la estructura jerárquica de las autopercepciones físicas. *Psicothema*, *16*, 461–467.

Bagozzi, R. P., & Kimmel, S. K. (1995). A comparison of leading theories for the prediction of goal directed behaviours. *British Journal of Social Psychology*, *34*, 437–461.

Bernardo, R. P. S., & Matos, M. G. (de) (2003). Adaptação Portuguesa do Physical Self-Perception Profile for Children and Youth e do Perceived Importance Profile for Children and Youth. *Análise Psicológica*, *21*, 127–144.

Biddle, S., Page, A., Ashford, B., Jennings, D., Brooke, R., & Fox, K. (1993). Assessment of children's physical self-perceptions. *International Journal of Adolescence and Youth*, *4*, 93–109.

Byrne, B. M. (1996). Academic self-concept: Its structure, measurement, and relation to academic achievement. In B. A. Bracken (Ed.), *Handbook of self-concept: Developmental, social, and clinical considerations* (pp. 287–316). New York, NY: Wiley.

Byrne, B. M. (2001). *Structural equation modeling with AMOS*. Rahwah, NJ: Erlbaum.

Byrne, B. M. (2005). Factor analytic models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment*, *85*, 17–32.

Byrne, B. M., Shavelson, R. J., & Muthén, B. (1989). Testing for the equivalence of factor covariance and means structures: The issue of partial measurement invariance. *Psychological Bulletin*, *105*, 456–466.

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255.

Coopersmith, S. (1967). *The antecedents of self-esteem*. San-Francisco: Freeman WH.

Coopersmith, S. (1984). *Inventaire d'estime de soi*. Paris, France: Centre de Psychologie Appliquée.

Eklund, R. C., Whitehead, J. R., & Welk, G. J. (1997). Validity of the children and youth physical self-perception profile: A confirmatory factor analysis. *Research Quarterly for Exercise and Sport*, *68*, 249–256.

Fonseca, A. M., & Fox, K. R. (2002). Como avaliar o modo como as pessoas se percebem fisicamente? Um olhar sobre a versão portuguesa do Physical Self-Perception Profile (PSPP). *Revista Portuguesa de Ciências do Desporto*, *2*, 11–23.

Fouladi, R. T. (2000). Performance of modified test statistics in covariance and correlation structure analysis under conditions of multivariate no normality. *Structural Equation Modeling*, *7*, 356–410.

Fox, K. R. (2000). Self-esteem, self-perception and exercise. *International Journal of Sport Psychology*, *31*, 228–240.

Fox, K. R., & Corbin, C. B. (1989). The Physical Self-Perception Profile: Development and preliminary validation. *Journal of Sport and Exercise Psychology*, *11*, 408–430.

Guérin, F., Marsh, H. W., & Famose, J. P. (2004). Generalizability of the PSDQ and its relationship to physical fitness: The European French connection. *Journal of Sport and Exercise Psychology*, *26*, 19–39.

Hagger, M., Ashford, B., & Stambulova, N. (1997). Physical self-perceptions: A cross-cultural assessment in Russian children. *European Journal of Physical Education*, *2*, 228–245.

Hagger, M. S., Aşçı, F. H., & Lindwall, M. (2004). A cross-cultural evaluation of a multidimensional and hierarchical model of physical self-perceptions in three national samples. *Journal of Applied Social Psychology*, *34*, 1075–1107.

Hagger, M. S., Biddle, S. J. H., Chow, E. W., Stambulova, N., & Kavussanu, M. (2003). Physical self-perceptions in adolescence: Generalizability of a hierarchical multidimensional model across three cultures. *Journal of Cross-Cultural Psychology*, *34*, 611–628.

Hagger, M. S., Biddle, S. J. H., & Wang, C. K. J. (2005). Physical self-concept in adolescence: Generalizability of a multidimensional, hierarchical model across gender and grade. *Educational and Psychological Measurement*, *65*, 297–322.

Harter, S. (1982). The perceived competence scale for children. *Child Development*, *53*, 87–97.

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

Klomsten, A. T., Shaalvik, E. M., & Espnes, G. A. (2004). Physical self-concept and sports: Do gender differences still exist? *Sex Roles*, *50*, 119–127.

Little, T. D., Lindenberger, U., & Nesselroade, J. R. (1999). On selecting indicators for multivariate measurement and modeling with latent variables: When "good" indicators are bad and "bad" indicators are good. *Psychological Methods*, *4*, 192–211.

Maïano, C., Ninot, G., Morin, A. J. S., & Bilard, J. (2007). Long-term effects of athletic competition on basketball skills and self-perceptions of adolescents with conduct disorders. *Adapted Physical Activity Quarterly*, *24*, 178–196.

Maïano, C., Ninot, G., Stephan, Y., Morin, A. J. S., Florent, J. F., & Vallée, P. (2006). Geographic region effects on adolescent physical self: An exploratory study. *International Journal of Psychology*, *41*, 73–84.

Marsh, H. W. (1996). Physical self-description questionnaire: Stability and discriminant validity. *Research Quarterly for Exercise and Sport*, *67*, 249–264.

Marsh, H. W. (1997). The measurement of physical self-concept: A construct validation approach. In K. R. Fox (Ed.), *The physical self* (pp. 27–58). Champaign, IL: Human Kinetics.

Marsh, H. W., Asçı, F. H., & Marco, I. T. (2002). Multi-trait multi-method analyses of two physical self-concept instruments: A cross-cultural perspective. *Journal of Sport and Exercise Psychology*, *24*, 99–119.

Marsh, H. W., Bar-Eli, M., Zach, S., & Richards, G. E. (2006). Construct validation of hebrew versions of three physical self-concept measures: An extended multitrait–multimethod analysis. *Journal of Sport and Exercise Psychology*, *28*, 310–343.

Marsh, H. W., Ellis, L., Parada, R., Richards, G. E., & Heubeck, B. (2005). A short version of the Self-Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment*, *17*, 81–102.

Marsh, H. W., & O'Neill, R. (1984). Self-Description Questionnaire III (SDQ III): The construct validity of multidimensional self-concept ratings by late-adolescents. *Journal of Educational Measurement*, *21*, 153–174.

Marsh, H. W., & Redmayne, R. S. (1994). A multidimensional physical self-concept and its relations to multiple components of physical fitness. *Journal of Sport and Exercise Psychology*, *16*, 43–55.

Marsh, H. W., Richards, G. E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical self-description questionnaire: Psychometric properties and a multitrait–multimethod analysis of relations to existing instruments. *Journal of Sport and Exercise Psychology*, *16*, 270–305.

Masse, C., Jung, J., & Pfister, R. (2001). Agressivité, impulsivité et estime de soi. *Sciences et Techniques des Activités Physiques et Sportives*, *56*, 33–42.

Moreno, J. A., Cervelló, E., Vear, J. A., & Ruiz, L. M. (2007). Physical self-concept of Spanish schoolchildren: Differences by gender, sport practice and levels of sport involvement. *Journal of Education and Human Development*, *1*, 1–17.

Ninot, G., Delignières, D., & Fortes, M. (2000). L'évaluation de l'estime de soi dans le domaine corporel. *Sciences et Techniques des Activités Physiques et Sportives*, *53*, 35–48.

Page, A., Ashford, B., Fox, K., & Biddle, S. (1993). Evidence of cross-cultural validity for the physical self-perception profile. *Personal and Individual Differences*, *14*, 585–590.

Rosenberg, M. (1965). *Society and the adolescent self-image*. New Jersey, NJ: Princeton University.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*, 407–411.

Smith, G., McCarthy, D., & Anderson, K. (2000). On the sins of short-form development. *Psychological Assessment*, *12*, 102–111.

Sonstroem, R. J. (1976). The validity of self-perceptions regarding physical and athletic ability. *Medicine and Science in Sports*, *8*, 126–132.

Sonstroem, R. J. (1978). Physical estimation and attraction scales: Rationale and research. *Medicine and Science in Sports*, *10*, 97–102.

Sonstroem, R. J., Harlow, L. L., & Josephs, L. (1994). Exercise and self-esteem: Validity of model expansion and exercise associations. *Journal of Sport and Exercise Psychology*, *16*, 29–42.

Sonstroem, R. J., Speliotis, E. D., & Fava, J. L. (1992). Perceived physical competence in adults: An examination of the physical self-perception profile. *Journal of Sport and Exercise Psychology*, *14*, 207–221.

Stephan, Y., Bilard, J., Ninot, G., & Delignières, D. (2003). Bodily transition out of elite sport: A one-year study of physical self and global self-esteem among transitional athletes. *International Journal of Sport and Exercise Psychology*, *1*, 192–207.

Stephan, Y., & Maïano, C. (2007). On the social nature of global self-esteem: A replication study. *Journal of Social Psychology*, *147*, 573–575.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment*, *80*, 99–103.

Van de Vliet, P., Knapen, J., Onghena, P., Fox, K., Van Coppenolle, H., David, A., et al. (2002). Assessment of physical self-perceptions in normal Flemish adults versus depressed psychiatric patients. *Personality and Individual Differences*, *32*, 855–863.

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70.

Welk, G. J., Corbin, C. B., Dowell, M. N., & Harris, H. (1997). The validity and reliability of two different versions of the children and youth physical self-perception profile. *Measurement in Physical Education and Exercise Science*, *1*, 163–177.

Whitehead, J. R. (1995). A study of children's physical self-perceptions using an adapted physical self-perception profile questionnaire. *Pediatric Exercise Science*, *7*, 132–151.

Wichstrom, L. (1995). Harter's self-perceptions profile for adolescents: Reliability, validity and evaluation of the question format. *Journal of Personality Assessment*, *65*, 100–116.

Yuan, K.-H., & Hayashi, K. (2003). Bootstrap approach to inference and power analysis based on three statistics for covariance structure models. *British Journal of Mathematical and Statistical Psychology*, *56*, 93–110.