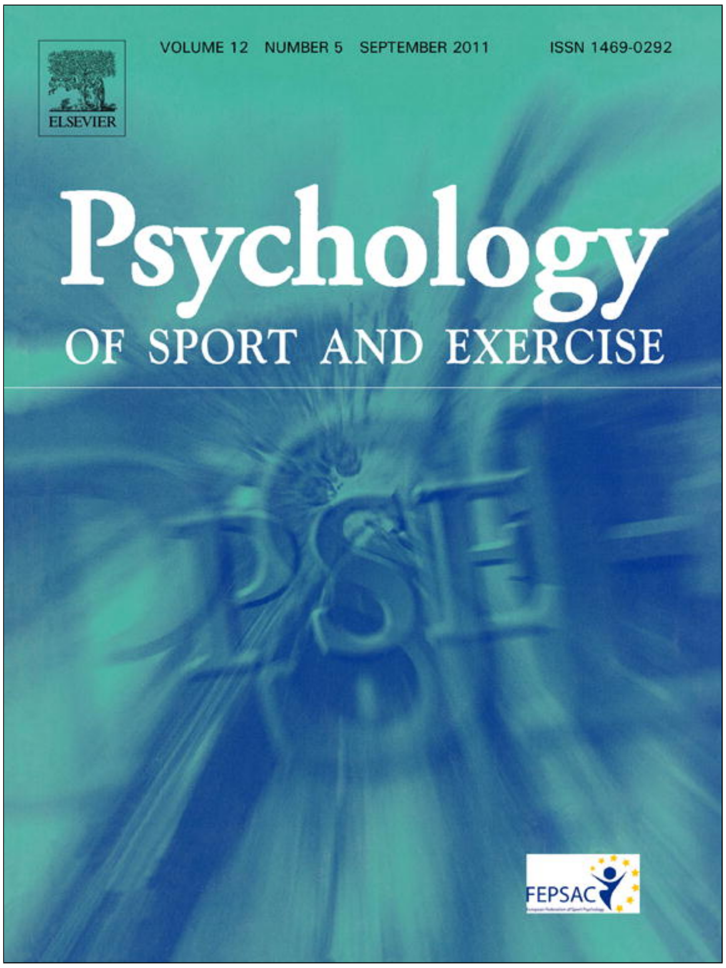
VOLUME 12 NUMBER 5 SEPTEMBER 2011
ISSN 1469-0292
ELSEVIER
Psychology
OF SPORT AND EXERCISE
FEPSAC

Psychology of Sport and Exercise 12 (2011) 540–554

Contents lists available at ScienceDirect

Psychology of Sport and Exercise

journal homepage: www.elsevier.com/locate/psychsport

Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM)

Alexandre J.S. Morin [a,b,*,1], Christophe Maïano [c,1]

[a] Department of Psychology, University of Sherbrooke, Canada
[b] Educational Excellence and Equity (E3) Research Program, Center for Educational Research, University of Western Sydney, Australia
[c] Institute of Movement Sciences Etienne-Jules Marey (UMR 6233), CNRS–University of Aix-Marseille II, Marseille, France

ARTICLE INFO

Article history:
Received 2 February 2011
Received in revised form
7 April 2011
Accepted 10 April 2011
Available online 21 April 2011

Keywords:
Physical self-concept
Physical self-inventory
Short form
Exploratory structural equation modeling
ESEM
Convergent validity
Reliability
Measurement invariance

ABSTRACT

Objectives: In a review of physical self-concept instruments Marsh and Cheng (in press) noted that the short version (18 item) of the physical self-inventory (PSI-S) represents an important contribution to applied research but that further research was needed to investigate the robustness of its psychometric properties in new and diversified samples and to investigate the reasons for the elevated correlations observed between the six PSI-S subscales.
Design and Method: A sample of sample 2029 French adolescents completed the PSI-S and their answers were analyzed with exploratory structural equation modeling (ESEM).
Results: The results show that the PSI-S ESEM measurement model is robust and fully invariant across subgroups of students formed on the basis of gender, weight categories, age categories, and ethnicity. The results also confirm the convergent validity and reliability of the PSI-S subscales. Most importantly, the ESEM model results in importantly deflated latent factor correlations and suggest that the previously reported inflated correlations may have been due to the fact that traditional confirmatory factor analytic (CFA) models arbitrarily constrain all cross-loadings to zero. In addition, the ESEM model reveals that the negatively worded items from the PSI-S may be suboptimal, a result that was not obvious from the CFA results.
Conclusion: The obtained results clearly confirm the robustness of the psychometric properties of the PSI-S and the usefulness of ESEM for more detailed analyses of measurement scale psychometric properties. Reformulations for the negatively worded items are proposed and directions for future studies of the PSI-S are noted.

In their classic review of self-concept research Shavelson, Hubner, and Stanton (1976) represented self-concept as a pyramid, with global self-esteem at the apex and more specific constructs at the next-lower level, such as the academic self, the social self and the physical self. Specificity increases downward with the most situation-specific self-perceptions at the base. With the recognition of the multi-dimensionality of the self-concept (Marsh, 1997), came more refined conceptualizations and studies of its sub-components (Fox, 2000). Following Sonstroem's (1976, 1978) work, Fox and Corbin (1989) developed a multidimensional and hierarchical model of the physical self-concept of particular interest to sport psychologists. In this model, the upper level is occupied by a generic construct representing global self-worth (GSW). GSW refers to the positive or negative way people feel about themselves as a whole, which is also often called global self-esteem (e.g. Brown, Dutton, & Cook, 2001). The next level (the domain level) is occupied by a global construct representing physical self-worth (PSW; general feelings of happiness, satisfaction and pride in the physical self). Finally, the lower level (the subdomain level) is occupied by four constructs: 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.).

From this model, Fox and Corbin (1989) developed the *Physical Self-Perception Profile* (PSPP) and validated it among North

* Corresponding author at: University of Sherbrooke, Department of Psychology, 2500 boulevard de l'Université, Sherbrooke, QC J1K 2R1, Canada.
E-mail address: alexandre.morin@usherbrooke.ca (A.J.S. Morin).

[1] Both contributed equally to the preparation of this paper, the order of appearance of the first and second authors (A.J.S.M and C.M.) was determined at random: both should be considered first authors.

1469-0292/$ – see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.psychsport.2011.04.003

American college students.[2] This model and instrument were successfully cross-validated in adults English-speaking samples (Hagger, Asç1, & Lindwall, 2004; Page, Ashford, Fox, & Biddle, 1993; Sonstroem, Speliotis, & Fava, 1992) and cross-culturally adapted and validated in non-English European countries, such as Belgium and the Netherlands (Van de Vliet et al., 2002), Portugal (Fonseca & Fox, 2002), Spain (Atzienga, Balaguer, Moreno, & Fox, 2004), Sweden (Hagger et al., 2004), and Turkey (Hagger et al., 2004; Marsh et al., 2002).

Nevertheless, concerns have been expressed about PSPP, particularly about its non-standard structured alternative format (i.e. paired forced-choice with a 4-point answer scale) which has been found to be confusing for respondents and associated with substantial method effects (Eiser, Eiser, & Havermans, 1995; Marsh et al., 1994, 2002, 2006; Wichstrøm, 1995; Wylie, 1989). In addition, the PSPP assess GSW with items from the Rosenberg Self-Esteem Inventory (RSEI; Rosenberg, 1965), which is also associated with substantial method effects (Marsh, Scalas, & Nagengast, 2010; Tomás & Oliver, 1999). Finally, and importantly, many have expressed concerns regarding the appropriateness of this instrument for children and adolescents (Biddle et al., 1993; Marsh et al., 1994). Indeed, because youths' cognitive abilities are more limited than those of adults, it might be harder for them to distinguish their own physical self-evaluations across a variety of specific subdomains and to fully comprehend the items' abstract formulations (which are made worse by the non-standard answering scale). Fortunately, some of these concerns were addressed (i.e. sport competence items were replaced and age-appropriate terminology was used) with the development of a version of the PSPP for North American children and adolescents (Eklund, Whitehead, & Welk, 1997; Whitehead, 1995). This instrument has since been similarly validated for youths from non-English-speaking European countries (e.g. Aşç1, Eklund, Whitehead, Kirazci, & Koca, 2005; Bernardo & Matos, 2003; Hagger, Ashford, & Stambulova, 1997; Moreno, Cervelló, Vear, & Ruiz, 2007). However, this instrument still relies on a structured alternative format answer scale.

In France, Ninot, Delignières, and Fortes (2000) developed, for adults, the *physical self-inventory* (PSI). The PSI is based on the PSPP and provides a promising way of circumventing the problems typically associated with the PSPP: (i) the original response format was replaced by a 6-point Likert scale (1: *not at all*, 2: *very little*, 3: *some*, 4: *enough*, 5: *a lot*, 6: *entirely*); (ii) GSW was assessed with 5 items from the school version of the Coppersmith's (1967, 1984) Self-Esteem Inventory, rather than with items from the RSEI; (iii) following initial analyses, the items from the original PSW scale were replaced with five items taken from Marsh and O'Neill's (1984) Self-Description Questionnaire-III. Maïano et al. (2008) adapted the PSI for use with adolescents and developed short form of this instrument (PSI-S; 18 items, with 3 items per dimension). The factor validity and reliability of this instrument was tested with a sample 1018 French adolescents (541 boys and 477 girls), aged between 11 and 16 years. Maïano et al. (2008) conducted a series of confirmatory factor analyses (CFA) to test the original six-factor measurement model. Results from analyses performed in two independent subsamples provided support for the: (i) factorial validity of the measurement model of the PSI-S; (ii) invariance of the PSI-S intercepts across gender; and (iii) a lack of latent mean invariance, showing that girls presented a lower level on most PSI-S dimensions (GSW, PSW, SC, PA and PS), confirming the results from previous studies conducted with similar instruments (e.g. Aşç1, 2002; Hagger, Biddle, & Wang, 2005; Marsh et al., 2006; Marsh, Hau, Sung, & Yu, 2007). Subsequent analyses also confirmed that the PSI-S was characterized by: (i) satisfactory internal consistency coefficients ranging from .73 to .75; (ii) acceptable test–retest correlations, ranging from .74 to .84, and (iii) elevated latent factor correlations that still provided evidence of discriminant validity (r = .50–.91; M = .71; SD = .12). With the sole exception of a subsequent study in which Maïano, Bégarie, Morin, and Ninot (2009) validated another adaptation of the PSI for use among adolescents with intellectual disability (n = 362) and replicated the results from their original study, no other attempt was made to replicate theses results on new samples of "normal" adolescents. This is worrying since it is a known fact that a single study is insufficient to reach clear conclusions regarding the psychometric properties of an instrument. This is especially true given the fact that Maïano et al. (2008) developed the PSI-S from the 25-item adult version of the PSI in order to obtain a reasonable fit from an initially suboptimal measurement model and never really cross-validated it on a new independent sample of adolescents. Moreover, the methodological limitations mentioned by Maïano et al. (2008) remain unresolved and stress the need for additional cross-validation efforts.

First, Maïano et al.'s (2008) study was based on a sample of normal-weight adolescents. It is thus uncertain whether the observed psychometric properties could generalize to youth with different weight statuses. However, current research evidence reveal that overweight and obesity represent a highly prevalent phenomenon in multiple countries around the world (e.g. Lissau et al., 2004) with prevalence rates sometimes reaching over 30% for overweight and 15% for obesity in some subpopulations. As overweight adolescents (or very skinny ones for that matter) present a higher risk of being discriminated against on the basis of their weight, the resulting stigmatization may strongly influence their individual self-concepts, particularly in the physical domain and sub-domains (e.g. Puhl & Latner, 2007; Wardle & Cooke, 2005). Thus, when overweight and obese adolescents are compared to normal-weight peers they tend to present significantly different relations to their bodies and lower level of GSW and physical self-perceptions (e.g. French, Story, & Perry, 1995; Griffiths, Parsons, & Hill, 2010; Hau, Sung, Yu, Marsh, & Lau, 2005; Marsh et al., 2007; Sung, Yu, So, Lam, & Hau, 2005). However, the validity of this conclusion relies on the often untested assumption that the measurement model used to assess physical self-concept is invariant across weight categories; whereas it is highly possible that overweight or obesity may completely modify the way the physical self-concept is organized. To our knowledge, this assumption was only verified once among a sample of Chinese children (Hau et al., 2005), using the Chinese version of the *Physical Self-Description Questionnaire* (PSDQ), and never amongst Western populations or using PSPP-based instruments.

Similarly, although Maïano et al. (2008) did confirm the gender-based invariance of the PSI-S, these results also need to be replicated. In addition, they did not examine the measurement invariance of the PSI-S across age categories (i.e. early and late adolescence) and ethnicity. Adolescence is a period of multiple social and physical transformations in which youths implicitly and explicitly learn about themselves psychologically and physically and these transformations exert a determining impact on how they perceive themselves and even on how they organize their self-perceptions (Cicchetti & Rogosch, 2002; Eccles et al., 1993;

[2] Marsh and Redmayne's (1994) also developed and validated a multidimensional and hierarchical instrument: the *Physical Self-Description Questionnaire* (PSDQ). For details on the psychometric properties of this instrument and differences with the PSPP in various samples differing in age, culture or language, see Marsh et al. (Marsh, Aşç1, & Marco, 2002; Marsh, Bar-Eli, Zach, & Richards, 2006; Marsh, Richards, Johnson, Roche, & Tremayne, 1994). In addition, for a recent review of the various instruments that may be used to assess the physical self-concept, see Marsh and Cheng (in press).

Steinberg & Morris, 2001). Whereas early adolescence is generally characterized by efforts to cope with these multiples changes, late adolescence is generally a period were these changes are progressively integrated into a new self definition. Indeed, results from cross-sectional and longitudinal studies show that early adolescents, when compared to middle or late adolescents, tend to present higher level of GSW (e.g. Hagger et al., 2005; Robins, Trzesniewski, Tracy, Gosling, & Potter, 2002), PSW (e.g. Hagger et al., 2005; Marsh, 1998; Marsh et al., 2007) and more specific physical self-perceptions (e.g. Hagger et al., 2005; Marsh, 1998; Marsh et al., 2007). However, in addition to these mean-levels changes in physical self-perceptions, it is also possible that these biopsychosocial changes may directly modify the way physical self-perceptions are organized over adolescence, thus precluding valid mean-level comparison due to a change in the construct itself.

Gender and ethnicity are also known to influence how these changes, particularly pubertal maturation, are integrated into adolescents' self-concept. Indeed, puberty often results in body fat accumulation in girls, an often undesired change, whereas for boys it usually results in muscle increase and the emergence of other culturally valued attributes (Alsaker, 1995; Angold & Worthman, 1993; Stice & Bearman, 2001). Furthermore, some studies revealed that the potentially deleterious effects of early puberty could be limited to, or stronger for, girls of Caucasian European/North American origin (Halpern, Udry, Campbell, & Suchindran, 1999; Morin, Maïano, Marsh, Janosz, & Nagengast, 2011; Siegel, Yancey, Aneshensel, & Schuler, 1999), suggesting that social factors may moderate these relations. For instance, whereas the European/North American culture emphasizes lean "prepubertal" looks for girls, African-American/Black, Hispanic/Latin or Arabic/Maghreb cultures put less emphasis on leanness and more on the fuller forms emerging with puberty (Siegel et al., 1999; Stice & Bearman, 2001). These observations suggest that ethnicity, age and gender may each exert an influence not only on adolescents' average levels on the physical self-concept domains and subdomains, but also that these variables may influence the way physical self-concept is defined and organized in these specific subgroups. This possibility suggest that physical self-concept measures may not be fully invariant across these grouping variables, thus precluding valid mean-level comparisons between them.

Third, the convergent validity of the PSI-S was never evaluated among normal adolescents. This clearly brings into question the construct validity of this instrument that, if it correctly measures what it is supposed to measure, should be significantly related to constructs know to be related to physical self-concept, such as physical self-image congruence, social physique anxiety, body image avoidance, fear of negative appearance evaluation, and disturbed eating attitudes and behaviors (e.g. Crocker et al., 2001, 2003; Hagger & Stevenson, 2010; Hagger et al., 2010; Lau, Cheung, & Ransdell, 2008; Marsh, 1996; Marsh et al., 2007; Monsma & Malina, 2004; Monsma, Malina, & Feltz, 2006).

Fourth, the correlations reported by Maïano et al. (2008) between the PSI-S subscales are very elevated (i.e. >.50 and sometimes even >.90). This issue has broad generalizability to physical self-concept research as similarly elevated correlations are apparently the norm in studies of other PSPP-based self-concept instruments (e.g. Atzienga et al., 2004; Fox & Corbin, 1989; Hagger et al., 2004, 2005; Maïano, Bégarie, et al., 2009; Marsh et al., 1994, 2002, 2006), Maïano et al. (2008, p. 844; also see Marsh & Cheng, in press) noted that 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*". Marsh et al. (1994, 2002, 2006) reached a similar conclusion and suggested that these inflated correlations could potentially be related to the structured alternative response scale used in the PSPP, which was replaced by a more conventional Likert scale for the PSI-S. These authors also noted that such a result may be observed with short scales that attempt to cover a broad range of dimensions with few items, as it is the case for the PSI-S.

More recently however, Marsh et al. (2009; Marsh, Liem, et al., in press; Marsh, Lüdtke, et al., 2010; Marsh, Nagengast, et al., in press) suggested that the independent cluster model (ICM) inherent in CFA – in which each item loads on a single factor – could be too restrictive for many multidimensional constructs, a conclusion that was already noted by McCrae, Zonderman, Costa, Bond, and Paunonen (1996) in personality research. Marsh et al. (2009) noted that many instruments have a well replicated structure based on Exploratory Factor Analyses (EFA) but that this structure sometimes fails to be replicated with CFA, and that many widely used psychometric instruments do not even reach minimal standards of fit with CFA. They suggest that these observations could be related to the unrealistic assumption of zero cross-loadings. Furthermore, even when the CFA model fits well in the first place (see Marsh, Liem, et al., in press; Marsh, Nagengast, et al., in press), they note that in all CFA applications, "*factor correlations will be at least somewhat inflated unless all non-target loadings are close to zero. This results in multicollinearity and undermines discriminant validity in relation to predicting other outcomes*" (Marsh, Lüdtke, et al., 2010, p. 486). Interestingly, recent simulations studies apparently confirm that EFA models are better at recovering true population latent correlations and that CFA-based latent correlations can be severely inflated by the presence of few small cross-loadings erroneously fixed to 0 (Asparouhov & Muthén, 2009; Marsh et al., under review). Unfortunately, EFA methods have been outshined by the methodological advances associated with CFA and structural equation models (SEM; tests of measurement invariance, fit indices, growth models, etc.) and by the erroneous semantically-based assumption that EFA was "*exploratory*" and that "*confirmatory*" methods were better in studies based on a priori hypotheses regarding factor structure, when in fact EFA only differs from CFA by the fact that all cross-loadings are freely estimated. Clearly, EFA models are better suited then CFA models to data-driven studies were no a priori hypotheses exist to guide the selection of the optimal measurement model; however, they are also perfectly well suited to theory-driven investigations, providing a stronger test that the items will relate to factor in the a priori hypothesized manner – imposing no ICM constraints on the model. Recently, Asparouhov and Muthén (2009; also see Dolan, Oort, Stoel, & Wicherts, 2009) developed Exploratory Structural Equation Modeling (ESEM), an integration of EFA within the CFA/SEM framework. ESEM thus makes it possible to conduct all types of psychometric tests typically reversed to CFA within an EFA measurement models and thus represent a valuable tool to explore the reasons for the elevated factor correlations of the PSI-S and other PSPP-based instruments. However, it should be noted that the free estimation of all cross-loadings does create an additional complexity for EFA models: rotational indeterminacy. Rotational indeterminacy means that EFA models based on different rotation procedures may converge on different solutions that will all have equivalent implications for the covariance structure and thus the same fit to the data. As noted by Sass and Schmitt (2010, p. 99) "*In many circumstances, different rotation criteria yield nearly identical rotated pattern loading matrices*" resulting in the identification of the same factors based on the same patterns of main loadings and cross-loadings. However, this will not always be the case and thus, estimation of EFA models should generally include an exploration of different rotation procedures.

In a recent review of the various instruments that may be used to assess the physical self-concept, Marsh and Cheng (in press) included a single non-English instrument, the PSI, and noted that the short form of the PSI "*make a potentially important contribution to applied research. However, further research is needed to more fully evaluate the robustness of support for construct validity and application in non-French-speaking settings*." In the present study, we address the first of these objectives by verifying the robustness of the psychometric properties of the PSI-S across multiple subgroups of French adolescents. More specifically, this cross-validation study will examine the factor validity and reliability of the PSI-S among a new independent sample of adolescents, as well as various subsamples defined according to gender, age (early and late adolescents), ethnicity, and weight categories (underweight, normal weight, and overweight). In addition, these verifications will be done with ESEM and compared to CFA results as an attempt to investigate the reasons for the elevated correlations noted between the PSI subscales in the context of previous studies. As a secondary objective, this study will also investigate the convergent validity of the PSI-S subscales with measures of social physique anxiety, behavioral manifestations of body image disturbances, fear of negative appearance evaluation, disturbed eating attitudes and behaviors and physical self-image congruence.

Method

Sample and procedure

A sample of 2029 adolescents (aged 11–18 years; M = 14.66 years, SD = 1.89), was recruited from 21 middle and high schools located in southern France. Participants were schooled in regular classes and, as such (according to French education policies) a priori presented no motor, intellectual, or sensory disability. The questionnaires were administered to participants in quiet conditions, during physical education classes of up to 30 students, and participants gave written informed consent.

This sample comprise: (i) 946 boys (46.6%) and 1083 girls (53.4%); (ii) 944 (46.5%) early (aged 11–14) and 1085 (53.5%) late adolescents (aged 15–18); (iii) 1499 (73.9%) participants whose parents are of European origin and 530 (26.1%) whose parents are from other origins (mostly from North Africa; there was an insufficient number of students from other descents to create finer distinctions); and (iv) 223 (11.0%) could be considered as underweight, 1588 (78.3%), as having a normal weight and 218 (10.7%) as overweight (including 32 obese that were too few to be considered separately). Weight categories were determined on the basis of participants body mass index [BMI = weight/height × height (self-reported); Cole, 1979] and of gender-specific cut-off scores provided by Cole, Bellizzi, Flegal, and Dietz (2000) and Cole, Flegal, Nicholls, and Jackson (2007).

Measures

Demographics

Participants were asked to self-reported their height, weight, gender, age and the birth country of their parents (as a proxy for ethnicity). These variables were used to define the groups used in the tests of measurement invariance.

Physical self-concept

Participants' physical self-concept was assessed with the previously described short form of the PSI-S (Maïano et al., 2008). Negatively worded items were recoded beforehand. The items are reported in the Appendix.

Disturbed eating attitudes and behaviors

The French version of the Eating Attitudes Test-26 (EAT-26: Garner, Olmstead, Bohr, & Garfinkel, 1982; Leichner, Steiger, Puentes-Newman, Perreault, & Gottheil, 1994) was used to assess participants' disturbed eating attitudes and behaviors. Participants were asked to indicate whether each item applied to them on a six-point scale ranging from always to never. The French EAT-26 (Leichner et al., 1994) was validated in samples of adolescent and adults, clinical (i.e. ED) and nonclinical females. Results provided support for the original three factor structure, for the internal consistency of the extracted factors (ranging from .54 to .86) and confirmed the criterion-related validity of this version (by the comparison of ED and nonclinical participants). This instrument includes three subscales: (i) Dieting (EAT-D; 13 items); (ii) Bulimia and food preoccupation (EAT-BFP; 6 items); and (iii) Oral control (EAT-OC; 7 items).

Social physique anxiety

The French version of the Social Physique Anxiety Scale (SPAS: Hart, Leary, & Rejeski, 1989; Maïano, Morin, Eklund, et al., 2010) was used to assess participants' social physique anxiety (i.e. the degree to which they become anxious to the real or perceived evaluation of their physique by others). The seven items of the French version were rated on a five-point Likert scale ranging from not at all (1) to extremely (5) and form a single global scale. The French version of the SPAS (Maïano, Morin, Eklund, et al., 2010) was validated on a mixed (males and females) sample of adolescents in a series of six studies (n = 1563). The results confirmed that the psychometric properties of the French version were adequate and similar to those from the original version. Those results gave support to the proposed factor model across two independent samples, and found acceptable internal consistency (ranging from .81 to .87 across studies and subsamples) and test–retest coefficients ($r = .78$). The results also supported the convergent validity of this instrument with measures of social anxiety, self-esteem, fear of negative evaluation, and body image disturbance.

Fear of negative appearance evaluation

The French version of the Fear of Negative Appearance Evaluation Scale (FNAES: Lundgren, Anderson, & Thompson, 2004; Maïano, Morin, Monthuy-Blanc, & Garbarino, 2010) was used to evaluate participants' fear of having their physical appearance negatively evaluated by others. The five items of the French version were rated on a five-point Likert scale ranging from not at all (1) to extremely (5) and form a single global scale. The French FNAES (Maïano, Morin, Monthuy-Blanc, et al., 2010) was validated on a mixed (males and females) sample of adolescents in a series of three studies (n = 684). The results confirmed that the psychometric properties of the French version were adequate and similar to those from the original version. Those results gave support to the proposed single factor model in two independent samples, and found acceptable internal consistency (of .83 in both samples) and test–retest coefficients ($r = .77$). The results also supported the convergent validity of this instrument with measures of social physique anxiety, self-esteem, fear of negative evaluation, and disturbed eating attitudes and behaviors.

Physical self-image congruence

Participants' physical self-image congruence (or satisfaction with the physical appearance of their body) was assessed with the Silhouette Matching Task (SMT: Marsh & Roche, 1996; Stunkard, Sorenson, & Schulsinger, 1983). The SMT assess the discrepancy between participants actual and ideal physical appearance by asking them to self match to a series of nine body silhouette images that vary from very thin to very fat (different series are presented for boys and girls) in relation to actual and ideal body image ratings. The discrepancies between the two ratings yield information regarding the level of satisfaction participants have with their physical self-image, coded so that higher scores reflect higher congruence.

Behavioral manifestations of body image disturbances

The French version of the Body Image Avoidance Questionnaire (BIAQ: Rosen, Srebnik, Saltzberg, & Wendt, 1991; Maïano, Morin, Monthuy-Blanc, & Garbarino, 2009) was used to assess the behavioral manifestations of body image disturbances through situations that may provoke body image concerns (i.e. wearing tight-fitting clothes, social activities, physical intimacy, weighing, exercising, and eating with others). The 19 items of the BIAQ are rated on a six-point Likert scale ranging from never (0) to always (5). The French version of the BIAQ (Maïano, Morin, et al., 2009) was validated on a mixed (males and females) sample of adolescents in a series of four studies ($n = 945$). The results confirmed that the psychometric properties of the French version were adequate and similar to those from the original version. The results supported the proposed four factor model across two independent samples, and revealed acceptable internal consistency coefficients, ranging from .73 to .90. The test–retest correlations were also satisfactory (ranging from .78 to .83) and the results supported the convergent validity of the BIAQ with measures of disturbed eating attitudes and behaviors and self-esteem. The BIAQ includes four subscales: (i) Clothing (BIAQ-C; 9 items); (ii) Social activities (BIAQ-SA; 4 items); (iii) Eating restraint (BIAQ-ER; 3 items); (iv) Grooming and Weighing (BIAQ-GW; 3 items).

Data analysis

All analyses in the present investigation were conducted with Mplus 6.0 (Muthén & Muthén, 2010), using the robust maximum likelihood estimator (MLR). This estimator provides standard errors and tests of fit that are robust in relation to non-normality and the use of ordered-categorical variables involving at least five response categories (e.g., Beauducel & Herzberg, 2006; Dolan, 1994; Lei, 2009; Lubke & Muthén, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2010). The full-information MLR estimator was used to correct for the small amounts of missing data present at the item level of the PSI-S (.79–5.22%; $M_{missing} = 2.18\%$; $SD_{missing} = 1.11\%$; see Enders, 2010; Graham, 2009).

In the first stage, the a priori factor structure of the PSI-S was tested in the total sample with CFA and ESEM models. In the CFA models, it was hypothesized that: (i) answers to the PSI-S would be explained by six correlated factors (previously defined); (ii) each item would have a non-zero loading on the factor it was designed to measure, and zero loadings on all other factors; and (iii) error terms would be uncorrelated. The a priori ESEM model was estimated following Marsh et al. (2009; Marsh, Liem, et al., in press; Marsh, Lüdtke, et al., 2010; Marsh, Nagengast, et al., in press) recommendations with an oblique geomin rotation and an epsilon value of .5, and hypothesized that the answers to the PSI-S items would be explained by six correlated factors. The ESEM approach differs from CFA in that all factor loadings are estimated, subject to constraints necessary for identification (for further details of the ESEM approach and identification issues, see Asparouhov & Muthén, 2009; Marsh, Lüdtke, et al., 2010; Marsh et al., 2009).

In the second stage, the measurement invariance of the PSI-S was tested in the following sequence (Meredith, 1993; also see Marsh, Lüdtke, et al., 2010; Marsh et al., 2009): (i) configural invariance (all loadings, intercepts and uniquenesses are freely estimated, the latent variances are constrained to 1 and the latent means are constrained to 0); (ii) loadings invariance (metric invariance: constraining the loadings to invariance allows for the free estimation of the factors variances in all but one group); (iii) intercepts invariance (strong invariance: constraining the intercepts to invariance allows for the free estimation of the factors means in all but one group); (iv) uniquenesses invariance (strict invariance); (v) variance/covariance invariance; and (vi) latent means invariance. In each sequence of invariance the preceding model served as reference.

Assessment of fit for the models was based on multiple indicators: the chi-square statistic (χ^2), the comparative fit index (CFI), the Tucker–Lewis index (TLI), the root mean square error of approximation (RMSEA), the 90% confidence interval of the RMSEA, and the Standardized Root Mean Square Residual (SRMR) Values greater than .90 for CFI and TLI are considered to be indicative of adequate model fit, although values greater than .95 are preferable (Byrne, 2005; Hu & Bentler, 1999). Values smaller than .08 or .06 for the RMSEA and smaller than .10 and .08 for the SRMR support respectively acceptable and good model fit (Hu & Bentler, 1999; Vandenberg & Lance, 2000). Concerning the RMSEA 90% CI, values less than .05 for the lower bound (left side) and less than .08 for the upper bounds (right side) or containing 0 for the lower bound and less .05 for the upper bounds (right side) provide respectively acceptable and good model fit (MacCallum, Browne, & Sugawara, 1996). Since chi-square tests of model fit are known to be overly sensitive to sample size and to minor deviations from multivariate normality, it is typical for applied CFA research to focus on sample size independent indices (Marsh, Balla, & Hau, 1996; Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004), particularly the CFI, TLI and RMSEA. However, as there are still very few applications of ESEM, and none that investigated the adequacy of these fit indices and proposed cut-off scores, their relevance to ESEM is not clear (Marsh, Lüdtke, et al., 2010; Marsh et al., 2009). Thus, we emphasize that these proposed cut-off scores should be considered as rough guidelines. In addressing this issue, Marsh et al. (2009; Marsh, Lüdtke, et al., 2010) suggest that fit indices that correct for parsimony (TLI and RMSEA) may be particularly important in ESEM given the large number of estimated parameters (e.g. in ESEM, the number of factor loadings is the product of the number of items times the number of factors).

Tests of measurement invariance were evaluated by the examination of χ^2 difference tests[3]. However, recent studies suggest complementing this information with changes in CFIs and RMSEAs (Chen, 2007; Cheung & Rensvold, 2002; Vandenberg & Lance, 2000). Indeed, these studies suggest that those additional indices tend to be more trustworthy than chi-square difference tests that present the same limitations as the chi-square. Here, chi-square differences tests are reported but changes in fit indices will be more closely inspected. A ΔCFI of .01 or less and a ΔRMSEA of .015 or less between a more restricted model and the preceding one indicate that the invariance hypothesis should not be rejected. It should also be noted that for indices incorporating a penalty for lack of parsimony such as the TLI and RMSEA, it is possible for a more restrictive model to result in better fit than a less restricted model; thus changes in TLI should also be inspected (Marsh, Hau, et al., 2005).

[3] As this study relied on MLR, the scaling correction composite needed to be taken into account in the calculation of chi-square differences tests. These tests were computed as minus two times the difference in the log likelihood of the nested models and are interpreted as chi square with degrees of freedom equal to the difference in free parameters between both models. The resulting difference then needs to be divided by its scaling correction composite, *cd*, where: (i) $cd = (p0 \times co \times p1 \times c1)/(p0 \times p1)$; (ii) *p0* and *p1* are the number of free parameters in the nested and comparison models; and (iii) *c0* and *c1* are the scaling correction factors for the nested and comparison models (Muthén & Muthén, 2010; Satorra & Bentler, 1999). We worked from model log likelihoods for greater precision as these statistical indices are less affected by rounding in Mplus.

Given the aforementioned parsimony issue in ESEM, Marsh et al. (2009; Marsh, Lüdtke, et al., 2010) noted that changes in fit indices correcting for parsimony might be particularly important. Indeed, ESEM models differ so much in degrees of freedom that relying on indices that do not adjust for parsimony may amplify the risk of capitalizing on chance (Marsh, Lüdtke, et al., 2010; Marsh et al., 2009).

Results

Factor validity, discriminant validity, cross-validation of the PSI-S, and reliability

The goodness-of-fit statistics and factor loadings-uniquenesses of the a priori CFA measurement model are respectively displayed in Tables 1 and 2. The results showed that the a priori six-factor CFA model provided an adequate degree of fit to the data according to the RMSEA (≤.06) and SRMR (≤.08), but a suboptimal degree of fit to the data according to the CFI and TLI which remain slightly under .95; although these indices remain in the acceptable range (≤.90). Examination of the standardized parameter estimates from this CFA model reveals that most loadings (with the exception of item GSW2) are substantial. However, the latent variables correlations are quite elevated ($r = .52–.93$; M = .69; SD = .14). These results are consistent with those from previous studies of the PSI-S (Maïano et al., 2008) and of other physical self-concepts instruments (e.g. Atzienga et al., 2004; Fox & Corbin, 1989; Hagger et al., 2004; Marsh et al., 2002, 2006; also see Marsh & Cheng, in press). Bagozzi and Kimmel (1995) propose to evaluate the discriminant validity of highly correlated factors through the calculation of the 95% confidence intervals (CI) for the standardized factor correlations (±1.96 times the standard error of the correlation). When the upper bound of these CI exceeds one it indicates that the correlated factors could in fact be taken to represent a single underlying construct. Here, many of these CI have an upper bound that is very close to one, although none exceeded one. These results suggest the inadequacy of the six-factor a priori CFA solution for the PSI-S. However, this a priori CFA model was compared to a series of post-hoc alternative models in which highly correlated factors were combined by pairs. None of these models provided a better fit to the data than the a priori model and none of these alternatives models resulted in deflated correlations between the remaining factors, and thus they are not reported.

In contrast with the results from the CFA model, the results revealed that the a priori 6-factor ESEM models (Table 1) provided a satisfactory degree a fit to the data with: (i) significant χ^2 values, (ii) CFI and TLI ≥ .95; (iii) RMSEA ≤ .06; (iv) SRMR ≤ .08. The standardized parameter estimates from this ESEM model are reported in Table 2. Most of the estimated loadings of the items in their a priori factors were also substantial (≥.30; M = .53, SD = .18), with few exceptions (Items GSW2, PSW2, PA2 and PS1). In addition, most cross-loadings remained small (≤.30; M = .08, SD = .10), with again few exceptions related to items GSW1, PSW2, and PA2.

These observations suggest that item GSW2 ("*There are many things in me that I would change*"), as well as item PS1 ("*I'm physically stronger than most people*") apparently do not represent optimal indicators of their respective constructs. In the case of item GSW2, this might be due to its negative formulation, as well as to its imprecise referent ("things in me" can still be physical). For future studies based on the PSI-S, it would be interesting to verify this hypothesis by the addition of a positively worded version of this item including a clearer "global" referent such as "*Globalement, je m'accepte tel que je suis/Overall, I am satisfied with being the way I am*". It is interesting to note that Maïano et al. (2008) did not retain

Table 1
Goodness-of-fit statistics of the confirmatory factor analytic and exploratory structural equation models.

Model	Description	χ^2 (df)	CFI	TLI	RMSEA	90% CI	SRMR	CM	$\Delta S\chi^2$ (df)	ΔCFI	ΔTLI	ΔRMSEA
CFA	1-1. CFA 6-factor model.	816.823 (120)*	.948	.934	.053	.050–.057	.038	–	–	–	–	–
ESEM	1-2. ESEM 6-factor model	201.185 (60)*	.990	.973	.034	.029–.039	.012	1-1	583.822 (60)*	+.042	+.039	–.019
Invariance across gender	2-1-Configural invariance	246.501 (120)*	.990	.975	.032	.026–.038	.012	–	–	–	–	–
	2-2-λ Invariant	384.866 (192)*	.985	.976	.031	.027–.036	.029	2-1	138.561 (72)*	–.005	+.001	–.001
	2-3-λ, τs Invariant	439.210 (204)*	.982	.972	.034	.029–.038	.030	2-2	56.109 (12)*	–.003	–.004	+.003
	2-4-λ, τs, δs Invariant	464.534 (222)*	.981	.974	.033	.029–.037	.033	2-3	27.685 (18)	–.001	+.002	–.001
	2-5-λ, τs, δs, ξ/φ Invariant	520.155 (243)*	.978	.973	.034	.030–.038	.041	2-4	54.549 (21)*	–.003	–.001	+.001
	2-6-λ, τs, δs, ξ/φ, ηs Invariant	706.553 (249)*	.964	.956	.043	.039–.046	.082	2-5	200.555 (6)*	–.014	–.017	+.009
Invariance across age categories	3-1-Configural invariance	260.530 (120)*	.990	.974	.034	.028–.040	.013	–	–	–	–	–
	3-2-λ Invariant	332.176 (192)*	.990	.984	.027	.022–.032	.020	3-1	75.37 (72)	.000	+.010	–.007
	3-3-λ, τs Invariant	364.366 (204)*	.988	.982	.028	.023–.032	.021	3-2	41.76 (12)*	–.002	–.002	+.001
	3-4-λ, τs, δs Invariant	419.616 (222)*	.986	.980	.030	.025–.034	.025	3-3	50.09 (18)*	–.002	–.002	+.002
	3-5-λ, τs, δs, ξ/φ Invariant	473.557 (243)*	.983	.979	.031	.026–.035	.040	3-4	52.78 (21)*	–.003	–.001	+.001
	3-6-λ, τs, δs, ξ/φ, ηs Invariant	483.335 (249)*	.983	.979	.030	.026–.034	.040	3-5	9.96 (6)	.000	.000	–.001
Invariance across parental origin	4-1-Configural invariance	298.211 (120)*	.987	.967	.038	.033–.044	.014	–	–	–	–	–
	4-2-λ Invariant	389.053 (192)*	.986	.977	.032	.027–.036	.022	4-1	106.227 (72)*	–.001	+.010	–.006
	4-3-λ, τs Invariant	411.258 (204)*	.985	.977	.032	.027–.036	.022	4-2	21.777 (12)	–.001	.000	.000
	4-4-λ, τs, δs Invariant	455.609 (222)*	.983	.976	.032	.028–.036	.025	4-3	43.774 (18)*	–.002	–.001	.000
	4-5-λ, τs, δs, ξ/φ Invariant	494.567 (243)*	.982	.977	.032	.028–.036	.032	4-4	39.036 (21)*	–.001	+.001	.000
	4-6-λ, τs, δs, ξ/φ, ηs Invariant	501.236 (249)*	.982	.977	.032	.028–.036	.030	4-5	6.996 (6)	.000	.000	.000
Invariance across weight categories	5-1-Configural invariance	579.276 (180)*	.971	.926	.057	.052–.063	.015	–	–	–	–	–
	5-2-λ Invariant	498.311 (324)*	.987	.982	.028	.023–.033	.023	5-1	114.185 (144)	+.016	+.056	–.029
	5-3-λ, τs Invariant	548.816 (348)*	.985	.981	.029	.024–.034	.025	5-2	51.906 (24)*	–.002	–.001	+.001
	5-4-λ, τs, δs Invariant	573.988 (384)*	.985	.984	.027	.022–.032	.026	5-3	29.397 (36)	.000	+.003	–.002
	5-5-λ, τs, δs, ξ/φ Invariant	633.342 (426)*	.985	.984	.027	.022–.031	.035	5-4	54.443 (42)	.000	.000	.000
	5-6-λ, τs, δs, ξ/φ, ηs Invariant	728.537 (438)*	.979	.978	.031	.027–.035	.044	5-5	99.462 (12)*	–.006	–.006	+.004

Note. CFA: confirmatory factor analytic model; ESEM: exploratory structural equation modeling; χ^2: chi square; df: degrees of freedom; CFI: comparative fit index; TLI: Tucker–Lewis index; RMSEA: root mean square error of approximation; 90% CI: 90% confidence interval of the RMSEA; SRMR: standardized root mean square error of approximation; λ: factor loading; τ: intercept; δ: uniquenesses; ξ: factor variance; φ: factor covariance; η: factor means; CM: comparison model; $\Delta S\chi^2$: scaled chi-square difference tests (calculated from models log likelihoods for greater precision); Δdf: change in degrees of freedom; ΔCFI: change in CFI; ΔTLI: change in TLI; ΔRMSEA: change in RMSEA; *p < .01.

Table 2
Standardized parameters estimates from the confirmatory factor analytic and exploratory structural equation models of the PSI-S.

Confirmatory factor analysis								Exploratory structural equation modeling						
Standardized factor loadings and uniquenesses														
Items	GSW (λ)	PSW (λ)	PC (λ)	SC (λ)	PA (λ)	PS (λ)	δ	GSW (λ)	PSW (λ)	PC (λ)	SC (λ)	PA (λ)	PS (λ)	δ
GSW1	.709						.498	**.361**	**.387**	−.072	.003	.083	.113	.491
GSW2	.162						.974	**.234**	−.056	.023	−.009	.057	−.044	.945
GSW3	.676						.543	**.514**	.141	.064	.014	.180	−.024	.497
PSW1		.774					.401	.055	**.704**	.072	.107	.032	.045	.279
PSW2		.800					.360	**.387**	**.269**	.123	.198	−.038	.181	.342
PSW3		.753					.433	.174	**.436**	.108	.066	.222	.058	.408
PC1			.867				.249	.068	.086	**.760**	.078	.020	.052	.219
PC2			.840				.294	.008	.201	**.657**	.095	.021	.045	.303
PC3			.769				.408	.085	−.035	**.606**	.151	.035	.146	.394
SC1				.823			.323	−.060	.138	.157	**.503**	.255	.145	.300
SC2				.845			.286	.183	.039	.107	**.625**	−.012	.152	.269
SC3				.814			.337	−.011	.121	.042	**.719**	.018	.091	.274
PA1					.542		.706	.215	.031	−.009	−.021	**.569**	−.032	.538
PA2					.786		.382	**.451**	.109	.077	.044	**.218**	.128	.466
PA3					.410		.832	−.058	.002	−.031	.023	**.660**	.036	.577
PS1						.483	.767	−.140	.151	.131	.089	.134	**.278**	.749
PS2						.822	.325	.055	.064	.029	.066	.063	**.717**	.316
PS3						.665	.558	.018	.004	.028	.100	−.029	**.628**	.517
Factor correlations (95% confidence intervals)														
Factor		PSW	PC	SC	PA	PS			PSW	PC	SC	PA	PS	
GSW		.93	.52	.59	.92	.52			.51	.23	.22	.37	.23	
		(.89–.96)	(.47–.57)	(.54–.64)	(.87–.96)	(.47–.58)			(.46–.57)	(.18–.28)	(.18–.27)	(.32–.41)	(.18–.29)	
PSW			.73	.81	.81	.70				.37	.43	.35	.37	
			(.70–.77)	(.78–.84)	(.77–.85)	(.66–.74)				(.32–.41)	(.38–.48)	(.29–.40)	(.32–.41)	
PC				.76	.52	.61					.45	.16	.33	
				(.73–.79)	(.47–.57)	(.57–.66)					(.42–.48)	(.12–.20)	(.30–.37)	
SC					.60	.78						.21	.51	
					(.55–.64)	(.74–.82)						(.16–.26)	(.47–.54)	
PA						.54							.20	
						(.59–.60)							(.15–.25)	

Note: λ = standardized factor loading; δ = standardized uniquenesses; GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA = physical attractiveness; PS = physical strength. All correlations are statistically significant (p ≤ .01). Greyscale entries: Target loadings.

item GSW2 in the development of a very short (12-items) form of the PSI. Regarding item PS1, the observed results are hard to explain, but it should be noted that PS1 standardized loading on the PS factor remains very close to the arbitrarily selected .30 level for acceptable loadings. These items were kept in the solution to maintain the local identification of each ESEM factor (requiring three items per factor), keeping in mind that these items only have a low impact on the parameter estimates, most of their variability being treated as uniquenesses.

Regarding items GSW1 ("*I have a good opinion of myself*") and PSW2 ("*I'm happy with what I am and what I can do physically*") they both apparently contribute as much to the definition of GSW as PSW, which is consistent with the general formulation of item GSW1 which might be taken to encompass physical self-worth and with the double-barreled formulation of item PSW2. For item PSW2, we recommend taking out the "*what I am*" part of this item in future studies. Similarly PA2 ("*I have a nice body to look at*") apparently contribute even more to the definition of GSW as of PA, which is consistent with the observation made by previous scholars that physical appearance represent a major component of global self-esteem in modern Western societies (e.g. Fox, 1998; Harter, 1999; Marsh & Redmayne, 1994). However, it is also interesting to note that amongst the items forming the PA subscale, two items are negatively worded (PA1 and PA3), which may also potentially explain why the only positively worded item from this subscale has its main loading on another factor and a very low contribution to the PA factor. Regarding the possibility that negatively worded items may have induced a bias in the measurement model (as it was the case for the only other negatively worded item of the PSI-S: item GSW2), we propose that future studies also include positively worded version of these items such as "PA1: *J'aime beaucoup mon apparence physique/I am really pleased with the appearance of my body*" and "PA3: *Tout le monde me trouve beau(belle)/Everybody thinks that I am good-looking*". The observed pattern of cross-loadings involving GSW, PSW and PA also apparently explains why the CFA correlations between these constructs needs to be so elevated (also see Atzienga et al., 2004; Fox & Corbin, 1989; Hagger et al., 2004; Maïano et al., 2008, 2009; Marsh et al., 2002, 2006; Marsh & Redmayne, 1994) in order to compensate that some of these constructs were in fact defined by shared indicators. We also explored the possibility that problematic results related to negatively worded items may be explained by unmodeled method factors (Marsh & Grayson, 1995; Marsh, Scalas et al., 2010; Tomás & Oliver, 1999). However, alternative models including method factors resulted in a negligible improvement in fit and on substantively identical parameters estimates, showing the same pattern of problematic cross-loadings and low loadings (these results are available upon request from first author).

Most importantly, the ESEM solution also resulted in clearly deflated factor correlations that provide a clear support to the discriminant validity of the extracted factors ($r = .16–.51$; $M = .33$; $SD = .11$). More specifically, the results showed that: (i) the relation between GSW and PSW was significantly stronger than any of the relationships between GSW and the other subscales; (ii) all of the subscales were significantly and positively correlated with the PSW domain and exhibited stronger significant relationships with PSW than with GSW, with the exception of PA for which was similarly related to both PSW and GSW. These results are also consistent with the results from previous studies of the PSI-S (Maïano et al., 2008) and of the physical self-concept more generally (e.g. Marsh & Redmayne, 1994; Marsh et al., 1994, 2002, 2006), as well as with theoretical propositions regarding the hierarchical nature of the physical self-concept (Fox & Corbin, 1989).

Finally, the internal consistency coefficients calculated on the full sample are fully satisfactory for most subscales ($\alpha = .82$ for PSW, .86 for PC, .87 for SC), modest yet acceptable given the reduced length of this instrument (see Streiner, 2003) for other subscales ($\alpha = .65$ for PA and .68 for PS), but unsatisfactory for the GSW subscale ($\alpha = .47$). However, when the problematic GSW2 item is taken out from the GSW subscale, its internal consistency reach the acceptable range ($\alpha = .64$), reinforcing the need to reformulate this item and revalidate this subscale in future studies.

Multiple group measurement invariance tests

To ensure that between-group comparisons based on the PSI-S are meaningful, it needs to be shown that the measurement scales are psychometrically equivalent (i.e. measure the same thing) across different subsamples (e.g. Vandenberg & Lance, 2000). To this end, a series of measurement invariance tests were conducted across gender (models 2-1 to 2-6), age categories (models 3-1 to 3-6), ethnicity (i.e. parental origin; models 4-1 to 4-6) and weight categories (models 5-1 to 5-6).

The results from the measurement invariance tests conducted according to participants gender showed that (i) all of the χ^2 were significant and most of the χ^2 difference tests were; (ii) the CFI, TLI, RMSEA and SRMR indicated adequate model fit at all steps; (iii) the ΔCFI and ΔTLI never showed a decrease superior to .01, except between models 2-5 and 2-6, suggesting the non-invariance of the latent factor means; (iv) the ΔRMSEA never showed an increase superior to .015; (v) the fit indices that control for model parsimony were similar or slightly lower for model 2-5 than at the beginning of the sequence (TLI moved from .975 to .973 and RMSEA moved from .032 to .034), suggesting the complete invariance of the measurement model of the PSI-S, up to the level of the factor variance–covariance matrix. Given the non-invariance of the latent factor means across gender and that gender-related differences are of substantive interest to this study latent means differences across gender were probed. These comparisons showed that when boys latent means are fixed to zero (from model 2-5), girls latent means were significantly ($p \leq .01$) lower by a quarter to half of a standard deviation on all subscales of the PSI-S. More specifically, girls latent means are of: (i) −.474 on GSW; (ii) −.255 on PSW; (iii) −.545 on PC; (iv) −.484 on SC; (v) −.225 on PA and; (vi) −.541 on PS. These results thus show the presence of gender-based mean differences on the PSI-S factors that confirm the results from previous studies (e.g. Aşçı, 2002; Hagger et al., 2005; Marsh et al., 2006, 2007), while demonstrating that these differences are not an artefact of a lack of measurement invariance.

The results from the measurement invariance tests conducted according to participants age and ethnicity/parental origin were highly similar and showed that (i) all of the χ^2 were significant and most of the χ^2 difference tests were also significant; (ii) the CFI, TLI, RMSEA and SRMR indicated adequate model fit at all steps; (iii) the ΔCFI and ΔTLI never showed a decrease superior to .01; (iv) the ΔRMSEA never showed an increase superior to .015; (v) the fit indices that control for model parsimony were similar or improved at the end of the sequence than at the beginning (for age categories: TLI moved from .974 to .979 and RMSEA moved from .034 to .030; for parental origin: TLI moved from .967 to .977 and RMSEA moved from .038 to .032). These results clearly confirm the complete measurement invariance of the PSI-S across age categories and parental origin.

Finally, the measurement invariance tests conducted according to participants' weight categories showed that (i) all χ^2 were significant and two of the χ^2 difference tests were also significant (when constraining the intercepts and the latent means to equality); (ii) the CFI, TLI, RMSEA and SRMR indicated adequate model fit at all steps; (iii) the ΔCFI and ΔTLI never showed a decrease superior to .01 (once again, although the CFI should be monotonous with model complexity, here the CFI exhibit a non monotonous trend between models 5-1 and 5-2 due to the reliance on the MLR estimator were the scaling correction factor was under 1 for model 5-1 and over 1 for model 5-2, the corresponding ML values for the CFI are .990 and .986); (iv) the ΔRMSEA never showed an increase superior to .015; (v) the fit indices that control for model parsimony improved at the end of the sequence compared to their values at the beginning (TLI: .926–.978; RMSEA: .057–.031). These results clearly confirm the complete measurement invariance of the PSI-S across weight categories, at least up to the invariance of the variance–covariance matrix. However, a closer examination of the results reveal that the observed deterioration in fit indices (CFI, TLI, RMSEA) at the last step when the latent means were constrained to equality are the largest observed in the study, save for the single time the changes reached the recommended cut-off values. In addition, the observed change in model chi-square at this step is not only significant, but also very large. Given these observations and the fact that weight-related differences are of substantive interest to this study latent means differences across weight categories were still probed. These comparisons (from model 5-5) showed that: (i) normal-weight participants latent means on the PS factor (.346, $p \leq .01$) were significantly higher than those from underweight participants (latent means are fixed to zero); (ii) overweight participants latent means on the PS factor (.605, $p \leq .01$) were significantly higher while their latent means on the GSW (−.705, $p \leq .01$), PSW (−.405, $p \leq .01$) and PC (−.221 $p \leq .05$) factors were lower than those from underweight participants (latent means are fixed to zero); (iii) overweight participants latent means on the PS factor (.259, $p \leq .01$) were significantly higher while their latent means on the GSW (−.586, $p \leq .01$), PSW (−.300, $p \leq .01$) and PC (−.302 $p \leq .01$) factors were lower than those from normal-weight participants (latent means are fixed to zero). Thus, these results confirm the presence of weight-related mean differences on the PSI-S factors that confirm the results from previous studies (e.g. French et al., 1995; Griffiths et al., 2010; Hau et al., 2005; Marsh et al., 2007; Sung et al., 2005), while showing that these differences are also not an artefact of a lack of measurement invariance.

A note on rotational indeterminacy

Due to rotational indeterminacy, ESEM models based on different rotation procedures may converge on different solutions that have equivalent implications for the covariance structure and thus the same fit to the data. Although we report a solution based on a geomin rotation with an epsilon value of .5 for consistency with previous ESEM applications (Marsh, Liem, et al., in press; Marsh, Lüdtke, et al., 2010; Marsh et al., 2009; Marsh, Nagengast, et al., in press), we also explored solutions based on alternative rotation procedures following Sass and Schmitt (2010; Schmitt & Sass, 2011) recommendations. The obtained latent variable correlations and latent mean comparisons across gender and weight categories (no significant latent mean differences were found across age and ethnicity) are reported in the Appendix. As shown in the Appendix, these alternative models all resulted in substantially deflated latent factor correlations when compared to the CFA solution, and on the same general pattern of latent factor correlations. In addition, all of these alternative models converged on substantively similar factor loadings estimate showing the same pattern of problematic cross-loadings and low main loadings. Similarly, latent means comparisons resulted in highly similar results across the different rotated solutions. The few exceptions to this pattern concern geomin rotation based on Mplus default epsilon value, and: (i) the lower level of PC observed in overweight participants when compared to underweight participants which became only marginally significant with the Crawford-Ferguson (CF) Facparsim ($p = .074$) and Parsimax ($p = .057$) rotations; (ii) the higher level of PS observed in overweight participants when compared to normal-weight participants, which became only marginally significant with the target rotation ($p = .079$); (iii) the PA level of overweight participants became significantly lower than the level observed in normal-weight and underweight participants with target rotation (however, this difference was marginally significant with the other rotations with $p \leq .10$); (iv) gender differences on PSW and PA became non significant with CF-Quartimin, which tend to maximally limit variable complexity (i.e. cross-loadings) and may have resulted in biased results for the PSW and PA factors that are both defined by items presenting elevated cross-loadings (see Sass & Schmitt, 2010; Schmitt & Sass, 2011). Most of these differences are pretty minor, limited to a small set of rotations and concern generally smaller differences. A closer examination of these results mostly showed highly similar solutions across rotations, which the exception of the solution based on Geomin rotation (with Mplus default epsilon value). This rotation place more weight on reducing variable complexity (i.e. by striving to obtain at least one zero cross-loadings per variable) and was previously found by Sass and Schmitt (2010; Schmitt & Sass, 2011) to yield unstable solutions in situations characterized by substantial variable complexity, as in the present study. In the present context, this rotation resulted in a less well defined PSW construct that merged more with GSW. In summary, rotational indeterminacy does not seem to play a major role in the present results. However, the choice of the optimal rotation criteria remains an open question and different rotational procedures should generally be explored in ESEM studies. More details on alternative rotations are available elsewhere (Asparouhov & Muthén, 2009; Bernaards & Jennrich, 2005; Browne, 2001; Jennrich, 2007; Sass & Schmitt, 2010; Schmitt & Sass, 2011).

Convergent validity

As a final step, we assessed the convergent validity of the PSI-S in relation to measures of disturbed eating attitudes and behaviors (EAT-D, EAT-BFP, EAT-OC), social physique anxiety (SPAS), fear of negative appearance evaluation (FNAES), physical self-image congruence (SMT), and behavioral manifestations of body image disturbances (BIAQ-C, BIAQ-SA, BIAQ-ER, BIAQ-GW). The latent variables correlations are reported in Table 3 (measurement models for the convergent measures were specified as CFA factors in accordance with the results from their validation studies). These results reveal that GSW, PSW and subdomain-specific physical self-perceptions are negatively and modestly to moderately related to (i) disturbed eating attitudes and behaviors: dieting (EAT-D, except for the PS subscale) and bulimia/food preoccupation (EAT-BFP; except for the PC, SC and PS subscales), (ii) social physique anxiety (SPAS), (iii) fear of negative appearance evaluation (FNAES; except for the SC and PS subscales), and (iv) behavioral manifestations of body image disturbances: clothing (BIAQ-C; except for the PC and PS subscales), eating restraint (BIAQ-ER; except for the PC and SC subscales) and social activities (BIAQ-SA; except for the PC and SC subscales). Additionally, most of the PSI-S subscales (except PS) are positively and moderately related to physical self-image congruence (SMT). Interestingly, among the subdomain subscales of the PSI-S, PA showed the highest correlations with the convergent measures, confirming the preeminent role of physical appearance in various manifestations of body image disturbances (e.g. Crocker et al., 2001, 2003; Monsma & Malina, 2004; Monsma et al., 2006). This could be explained by the choice of convergent measures all related to body images disturbances whereas the PS, SC and PC subscales are mostly related to physical capacities/abilities and would most likely be more strongly correlated to measures related to sport practice and physical condition (e.g. Sonstroem & Morgan, 1989). In addition, most of the observed correlations were stronger for the GSW and PSW scales of the PSI-S than for the subdomain-specific subscales, showing that body image disturbances do not depend on highly specific physical self-perceptions as much as on more global self-perceptions (e.g. Hagger & Stevenson, 2010; Hagger et al., 2010; Lau et al., 2008; Marsh et al., 2007; Marsh & Roche, 1996). These results are in line with results previously reported for other physical self-concept instruments, such as the various versions of the PSPP and the PDSQ (e.g. Crocker et al., 2001, 2003; Lau et al., 2008; Marsh, 1996; Marsh et al., 2007; Marsh & Roche, 1996; Monsma & Malina, 2004; Monsma et al., 2006) which clearly support the convergent and construct validity of the PSI-S.

Table 3
Latent variables correlations between the PSI-S subscales and convergent validity measures.

Scales	GSW	PSW	PC	SC	PA	PS
EAT-D	−.580**	−.363**	−.167**	−.128**	−.306**	−.035
EAT-BFP	−.317**	−.303**	−.042	−.033	−.162**	.057
EAT-OC	−.059	−.040	.097**	−.037	.034	−.059
SPAS	−.643**	−.364**	−.174**	−.150**	−.369**	−.153**
FNAES	−.535**	−.212**	−.142**	−.056	−.270**	−.071
SMT	.455**	.300**	.277**	.190**	.286**	.043
BIAQ-C	−.263**	−.313**	−.030	−.074*	−.342**	.075
BIAQ-SA	−.160**	−.193**	.043	.038	−.145**	.106**
BIAQ-ER	−.312**	−.165**	.036	.014	−.186**	.135**
BIAQ-GW	−.112	.017	−.129**	−.092*	−.028	−.035

Note: $^{*}p \leq .05$; $^{**}p \leq .01$; GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA = physical attractiveness; PS = physical strength; EAT = eating attitude test; EAT-D = dieting subscale of the EAT; EAT-BFP = bulimia and food preoccupation subscale of the EAT; EAT-OC = oral control subscale of the EAT; SPAS = social physique anxiety scale; FNAES = fear of negative appearance evaluation scale; SMT = silhouette matching task (physical self-image congruence); BIAQ = body image avoidance questionnaire; BIAQ-C = clothing subscale of the BIAQ; BIAQ-SA = social activities subscale of the BIAQ; BIAQ-ER = eating restraint subscale of the BIAQ; BIAQ-GW = grooming and weighing subscale of the BIAQ.

Discussion

Following Marsh and Cheng (in press) recommendations, the first objective of the present study was to evaluate the robustness of the psychometric properties of the PSI-S across multiple subgroups of French adolescents. More specifically, this cross-validation study first verified the factor validity and reliability of the PSI-S in a large sample of French adolescents. The results from the CFA conducted on this sample perfectly replicated Maïano et al. (2008) results and confirmed that the a priori measurement model of the PSI-S provides an adequate, yet suboptimal (with CFIs and TLIs close to but slightly under .95), fit to the data. However, the PSI-S latent factor correlations are again very elevated (i.e. .52–.93), as in previous studies of the PSI-S (Maïano et al., 2008) and self-concept instruments based on the PSPP (Atzienga et al., 2004; Fox & Corbin, 1989; Hagger et al., 2004; Maïano, Bégarie, et al., 2009; Marsh et al., 1994, 2002, 2006). Although we obtained evidence of discriminant validity for most of the PSI-S subscales (Bagozzi & Kimmel, 1995), these elevated correlations still suggest that the CFA-based discriminant validity of the PSI-S may be suboptimal.

Following Marsh et al. (2009, Marsh, Lüdtke, et al., 2010) observation that the independent cluster model inherent in CFAs could simply be too restrictive for many multidimensional constructs, the measurement model of the PSI-S was re-estimated with ESEM. The estimated ESEM model resulted in an optimal fit to the data, modest to acceptable internal consistency coefficients (with the exception of the GSW factor due to problems associated with item GSW2), and most importantly, drastically deflated latent factor correlations (i.e. .16–.51) providing a clear support to the discriminant validity of the PSI-S subscales. However, the ESEM solution revealed the presence of many cross-loadings, some of which were quite important and may even contribute to modify the interpretations/definition of some subscales (i.e. GSW, PSW, and PA). In most cases, these important cross-loadings are consistent with the observation that, at least in Western societies, physical appearance (and its specific indicators) plays a determining role in how adolescents define their GSW and PSW, and that PSW also plays a determining role in defining GSW (e.g. Fox & Corbin, 1989; Harter, 1999; Marsh & Redmayne, 1994; Sonstroem et al., 1992). These cross-loadings, which do make sense theoretically, explain why the latent factor correlations between the PSI-S subscales were so inflated in the CFA solution in which these cross-loadings were arbitrarily constrained to be zero (also see Asparouhov & Muthén, 2009; Marsh et al., under review). In addition to showing that the GSW, PSW and PA construct shared some common indicators, the ESEM solution also revealed the presence of small main loadings. Overall, these ESEM results suggest that the negatively worded items included in the current version of the PSI-S may not perform as well as once thought, a result which is consistent with what was often observed with other instruments incorporating negatively worded items (Marsh, Scalas, et al., 2010; Motl, Conroy, & Horan, 2001; Tomás & Oliver, 1999). Interestingly, most of these observations made in ESEM (cross-loadings, small main loadings) were not apparent in the CFA solution, with the sole exception of the low main loading of item GSW2 on the GSW factor. This could be related to the fact that in ESEM all cross-loadings are simultaneously estimated whereas the modifications indices commonly used for the post-hoc adjustment of CFA models only consider a single cross-loading at a time, making ESEM a one step process and thus limiting potential capitalization on chance. It should be noted that these results have broad generalizability for the assessment of physical self-perceptions in that severely inflated latent factor correlations are quite prevalent in many physical self-concepts instruments (for a review, see Marsh & Cheng, in press). The present results, in conformity with previous reports regarding the preeminent role of physical appearance in global self-evaluations, especially in adolescence, suggest that these inflated latent correlations may reflect the arbitrary nature of the CFA independent cluster model.

Although the observed internal consistency coefficients may in some case appear lower than conventional rule of thumbs (i.e. .80; see Marsh, Ellis, Parada, Richards, & Heubeck, 2005), they may in most cases (with the exception of GSW) simply reflect the limited number of items used to measure each dimension of the PSI-S (i.e. three items). Indeed, Streiner (2003) noted that internal consistency coefficients are strongly and positively affected by the number of items in a scale and that, consequently, acceptability levels must be adjusted in the context of short measurement scales.

Following the observation that the convergent validity of the PSI-S was never evaluated among normal adolescents, a second objective of the present study was to verify the convergent validity of the PSI-S with measures of constructs theoretically related to the physical self-concept, such as social physique anxiety, body image avoidance, fear of negative appearance evaluation, disturbed eating attitudes and behaviors and physical self-image congruence. These analyses confirmed the convergent validity of the PSI-S and add additional support to its overall construct validity, showing that the PSI-S seems to correctly assess what it is supposed to assess. Thus, the presence of a few elevated cross-loadings could have brought into question the meaning of the subscales. However, the observed correlations between these subscales and convergent measures support their convergent validity, and in turn, their construct validity.

In order to further probe the robustness of these results, multiple group measurement invariance tests were performed across subgroups of students commonly used in group-based comparisons, based on the assumption that physical self-concept may be organized and defined differently in these various subgroups. First, the results confirm the complete measurement and latent means invariance across groups formed on the basis of parental origin (as a proxy for ethnicity) and age categories. This result is encouraging and confirms that the PSI-S psychometric properties are robust across the main ethnic subgroups present in France, as well as across early and late adolescence. Although previous studies suggest that average levels of physical self-concept may differ according to adolescents' ethnicity (e.g. Halpern et al., 1999; Morin et al., in press; Siegel et al., 1999; Twenge & Crocker, 2002) and age (Marsh, 1998; Marsh et al., 2007), these studies also show that these effects differ according to specific ethnic groups (e.g. African Americans, Hispanics, Asians, Arabic; Twenge & Crocker, 2002) and may only emerge in the context of a three way interaction involving gender and age (or pubertal development; see Morin et al., 2011; Siegel et al., 1999). Unfortunately, these distinctions could not be made in the present study due to the low number of participants from specific non-European subgroups. Interestingly, a lack of measurement invariance or a lack or control for measurement errors could also explain the mean-level differences found in the previous studies and not replicated here. Although the present results also suggest a possible loss of precision in group-based comparisons due to the selection of so few items in the creation of the PSI-S, the results based on gender and weight-related comparisons suggest that this is not the case. Overall, and most importantly, the observed results strongly confirm that the PSI-S measurement model (loadings, intercepts, uniquenesses) is fully invariant across ethnic and age-related subgroups and thus could provide unbiased estimates of group-based comparisons.

Second, the present results also confirm the complete measurement (loadings, intercepts, uniquenesses) invariance across groups formed on the basis of gender and weight categories. In addition, these results are in conformity with those from previous studies in showing that girls tend to present lower average levels of most dimensions of the physical self-concept (Aşçı, 2002; Hagger et al., 2005; Maïano et al., 2008; Marsh et al., 2006, 2007) and that overweight adolescents present a lower level of GSW, PSW, and PC than normal-weight and underweight adolescents (e.g. French et al., 1995; Griffiths et al., 2010; Hau et al., 2005; Marsh et al., 2007; Sung et al., 2005). Interestingly, these results also revealed that perceptions of physical strength tended to increase as a linear function of weight (underweight < normal weight < overweight). Although this may seem strange at first, it is consistent with the fact that weight categories were created on the basis of adolescents' BMI, and not on the basis of their percentage of body fat. Thus, it is possible that at least some members of the "overweight" group were included due to a strong muscular or bone structure, rather than due to an elevated proportion of body fat – thus inducing a bias in the mean-level comparisons. Once again, these results strongly confirm the robustness of the PSI-S psychometric properties across gender and weight-related subgroups, and show that mean-level comparisons between these groups adjusted for measurement errors and based on confirmed invariance assumptions are consistent with current knowledge, thus providing further support to the convergent validity of the PSI-S subscales.

Some limitations must be taken into account. First, the internal consistency of the PSI-S remains modest in some cases (PA, PS and especially GSW), potentially due to the brevity of the scales, and to problems associated with specific items identified in the ESEM solution (e.g. GSW2). The sensitivity of the PSI-S in the context of between-group comparisons should also be further probed in the context of future studies, although its measurement invariance was quite clearly demonstrated. However, the observed problems remain minor and should not preclude the use of this instrument for the assessment of the physical self-concept of French adolescents. However, the present study showed that some of the items included in the current version of the PSI-S may be suboptimal and thus we encourage future users of the PSI-S to also incorporate the reformulations we proposed, in addition to the current items, as a way to compare both versions. Clearly, further studies are needed to evaluate whether the psychometric properties of the original PSI-S can be preserved, and even improved, with the proposed reformulations of these items. Second, 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 this instrument should be limited to normally achieving French adolescents, although an alternative version also exists for intellectually disabled adolescents and present comparable psychometric properties (Maïano, Bégarie, et al., 2009). Clearly, the next step in evaluating the robustness and generalizability of the PSI-S, as pointed out by Marsh and Cheng (in press), would be to verify its applicability and validity with non-French-speaking populations. Third, these adolescents were tested in the context of their physical education classes, a context which may have heightened the salience of their physical self-concept. It thus remains unknown whether the present results, especially regarding the discriminant validity of the subscales, could be replicated in other contexts. Fourth, although we found evidence of the convergent and criterion-related validity of the PSI-S, additional tests remain to be conducted in relation to (i) other physical self-concept instruments (e.g. PSDQ) within a multitrait–multimethod framework, and (ii) multiple external criterions (i.e. fitness examinations, percentage of body fat, grades in physical education classes, etc.). Fifth, the reliance on a cross-sectional sample also precludes the verification of the developmental stability of the PSI-S for adolescents. Although Maïano et al. (2008) study showed that the PSI-S presented acceptable levels of two-week test–retest reliability, 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, in verifying the measurement invariance of the PSI-S across age and weight, we arbitrarily categorized these continuous variables in order to conduct complete multiple group tests of measurement invariance, which may have resulted in a loss of precision (MacCallum, Zhang, Preacher, & Rucker, 2002). New approaches were recently developed allowing for tests of measurement invariance of CFA models across levels of a continuous variables (Barendse, Oort, & Garst, 2010; Hildebrandt, Wilhelm, & Robitzsch, 2009), but could not be applied in the present study as these methods have yet to be extended to ESEM. Finally, and although this was not an issue in the present study, ESEM models are submitted to the same issues of rotational indeterminacy as classical EFA models and thus alternative (and equivalent in terms of fit) solutions based on different rotational criterions should be explored before the final model is retained (e.g. Sass & Schmitt, 2010).

Acknowledgements

The authors wish to thank Herbert W. Marsh and Benjamin Nagengast for the ESEM adventure, Pablo Sillon for his invaluable advices at previous stages of this research, every student and staff member from the participating schools as well as every research assistant involved in this project. This manuscript was designed and prepared while A.J.S.M was a visiting scholar at the University of the Mediterranean – Aix-Marseille II.

Appendix. Original items from the PSI-S with English back-translated version in italics

Consignes/directions

Dans ce questionnaire, vous trouverez 18 phrases. Lisez attentivement chacune de ces phrases. Pour chaque phrase, cochez la réponse qui vous ressemble le plus entre: Pas du tout, Très peu, Un peu, Assez, Beaucoup, Tout à fait. **Efforcez-vous de répondre à toutes les phrases**. Sachez qu'aucune réponse n'est juste, elle doit être avant tout personnelle./*This questionnaire comprises 18 sentences. Please read each of them attentively. For each sentence, please check the answer that best describes you: Not at all, Very little, Some, Enough, A lot, Entirely.* ***Try to answer each sentence****. There is no absolute right or wrong answer, a good answer is one more closely reflecting your personal opinion.*

		Pas du tout Not at all	Très peu Very little	Un peu Some	Assez Enough	Beaucoup A lot	Tout à fait Entirely
GSW1	J'ai une bonne opinion de moi-même/*I have a good opinion of myself*	☐	☐	☐	☐	☐	☐
PSW1	Globalement, je suis satisfait(e) de mes capacités physiques/*Globally, I'm proud of what I can do physically*	☐	☐	☐	☐	☐	☐
PA1	Je n'aime pas beaucoup mon apparence physique/*I don't like very much the appearance of my body*	☐	☐	☐	☐	☐	☐
PS1	Je suis physiquement plus fort(e) que les autres/*I'm physically stronger than most people*	☐	☐	☐	☐	☐	☐
GSW2	Il y a des tas de choses en moi que j'aimerais changer/*There are many things in myself that I would change*	☐	☐	☐	☐	☐	☐
PSW2	Je suis content(e) de ce que je suis et de ce que je peux faire physiquement/*I'm happy with what I am and what I can do physically*	☐	☐	☐	☐	☐	☐
PC1	Je serais bon(ne) dans une épreuve d'endurance/*I would be good at physical stamina exercises*	☐	☐	☐	☐	☐	☐
SC1	Je trouve que je suis bon(ne) dans tous les sports/*I find that I'm good in all sports*	☐	☐	☐	☐	☐	☐
PA2	J'ai un corps agréable à regarder/*I have a nice body to look at*	☐	☐	☐	☐	☐	☐
PS2	Je serais bon(ne) dans une épreuve de force/*I would be good at exercises that require strength*	☐	☐	☐	☐	☐	☐

Appendix (*continued*)

		Pas du tout / Not at all	Très peu / Very little	Un peu / Some	Assez / Enough	Beaucoup / A lot	Tout à fait / Entirely
PSW3	Je suis confiant(e) vis-à-vis de ma valeur physique/*I'm confident about my physical self-worth*	☐	☐	☐	☐	☐	☐
PC2	Je pense pouvoir courir longtemps sans être fatigué(e)/*I think I could run for a long time without tiring*	☐	☐	☐	☐	☐	☐
SC2	Je me débrouille bien dans tous les sports/*I can find a way out of difficulties in all sports*	☐	☐	☐	☐	☐	☐
PA3	Personne ne me trouve beau(belle)/*Nobody find me good-looking*	☐	☐	☐	☐	☐	☐
PS3	Face à des situations demandant de la force, je suis le(la) premier(ière) à proposer mes services/*Faced with a situation requiring physical strength, I'm the first to offer assistance*	☐	☐	☐	☐	☐	☐
PC3	Je pourrais courir 5 km sans m'arrêter/*I could run five kilometers without stopping*	☐	☐	☐	☐	☐	☐
SC3	Je réussis bien en sport/*I do well in sports*	☐	☐	☐	☐	☐	☐
GSW3	Je voudrais rester comme je suis/*I would like to stay as I am*	☐	☐	☐	☐	☐	☐

Merci d'avoir répondu à ce questionnaire
Thanks a lot for your answers
Note. Item codes (GSW1 to GSW3) link the items to their parameter estimates in Table 2.

Latent factor correlations and latent mean comparisons based on alternative rotations

Factor	Correlations (95% confidence intervals)					Latent means comparisons (from models 2-5 and 5-5 in Table 1)			
	PSW	PC	SC	PA	PS	M_{girls} ($M_{boys} = 0$)	M_n ($M_u = 0$)	M_o ($M_u = 0$)	M_o($M_n = 0$)
CF-Facparsim rotation									
GSW	.52 (.47–.58)*	.20 (.15–.25)*	.22 (.17–.27)*	.40 (.36–.44)*	.25 (.19–.30)*	–.49*	–.11	–.70*	–.59*
PSW	1.00	.33 (.28–.38)*	.42 (.37–.48)*	.37 (.32–.42)*	.38 (.34–.43)*	–.32*	–.07	–.34*	–.27*
PC		1.00	.38 (.35–.42)*	.14 (.09–.18)*	.28 (.24–.33)*	–.46*	.08	–.21	–.29*
SC			1.00	.22 (.16–.27)*	.48 (.44–.52)*	–.45*	.10	.12	.02
PA				1.00	.22 (.17–.27)*	–.28*	–.03	–.23	–.19
PS					1.00	–.53*	.36*	.65*	.29*
CF-Parsimax rotation									
GSW	.55 (.49–.61)*	.25 (.19–.30)*	.25 (.19–.30)*	.40 (.35–.45)*	.28 (.22–.34)*	–.49*	–.11	–.70*	–.59*
PSW	1.00	.40 (.34–.46)*	.48 (.41–.55)*	.39 (.34–.45)*	.43 (.37–.49)*	–.30*	–.08	–.37*	–.29*
PC		1.00	.48 (.44–.52)*	.17 (.13–.22)*	.37 (.33–.41)*	–.50*	.08	–.21	–.29*
SC			1.00	.24 (.18–.30)*	.57 (.53–.61)*	–.48*	.11	.11	.00
PA				1.00	.25 (.19–.30)*	–.26*	–.04	–.23	–.19
PS					1.00	–.55*	.35*	.60*	.25*
CF-Equamax rotation									
GSW	.55 (.49–.61)*	.26 (.20–.31)*	.25 (.20–.30)*	.40 (.35–.45)*	.29 (.23–.35)*	–.49*	–.12	–.70*	–.59*
PSW	1.00	.42 (.36–.48)*	.49 (.42–.57)*	.39 (.34–.45)*	.44 (.38–.50)*	–.30*	–.08	–.38*	–.29*
PC		1.00	.50 (.47–.54)*	.18 (.13–.23)*	.39 (.35–.43)*	–.52*	.08	–.21*	–.30*
SC			1.00	.25 (.19–.30)*	.59 (.55–.63)*	–.49*	.11	.11	.00
PA				1.00	.25 (.20–.31)*	–.26*	–.04	–.22	–.19
PS					1.00	–.56*	.34*	.58*	.24*
CF-Varimax rotation									
GSW	.55 (.45–.64)*	.28 (.21–.35)*	.26 (.17–.35)*	.39 (.30–.47)*	.30 (.22–.37)*	–.48*	–.12	–.70*	–.59*
PSW	1.00	.48 (.40–.56)*	.54 (.44–.64)*	.40 (.31–.50)*	.47 (.39–.54)*	–.26*	–.10	–.40*	–.31*
PC		1.00	.60 (.56–.64)*	.22 (.16–.28)*	.48 (.43–.52)*	–.56*	.08	–.21*	–.30*
SC			1.00	.27 (.21–.32)*	.66 (.62–.70)*	–.51*	.11	.09	–.02
PA				1.00	.27 (.20–.33)*	–.24*	–.04	–.22	–.18
PS					1.00	–.57*	.33*	.54*	.21*
CF-Quartimax (direct quartimin) rotation									
GSW	.49 (.21–.77)*	.26 (–.03 to .55)	.23 (–.10 to .55)	.34 (.09–.60)*	.26 (.01–.51)*	–.48*	–.12	–.70*	–.58*
PSW	1.00	.54 (.45–.64)*	.58 (.48–.69)*	.43 (.21–.65)*	.49 (.40–.58)*	–.01	–.13	–.42*	–.29*
PC		1.00	.69 (.66–.73)*	.26 (.17–.35)*	.56 (.52–.61)*	–.59*	.09	–.21*	–.30*
SC			1.00	.30 (.22–.38)*	.72 (.68–.76)*	–.54*	.11	.07	–.05
PA				1.00	.29 (.20–.38)*	–.19	–.05	–.21	–.17
PS					1.00	–.58*	.31*	.49*	.18*
Target rotation[a]									
GSW	.48 (.39–.57)*	.29 (.23–35)*	.29 (.23–.35)*	.34 (.26–.43)*	.29 (.23–.35)*	–.46*	–.11	–.68*	–.57*
PSW	1.00	.55 (.46–.63)*	.61 (.52–.70)*	.50 (.42–.58)*	.52 (.44–.59)*	–.27*	–.11	–.44*	–.32*
PC		1.00	.73 (.70–.77)*	.31 (.26–.36)*	.60 (.55–.64)*	–.60*	.09	–.22*	–.30*
SC			1.00	.36 (.30–.42)*	.76 (.73–.80)*	–.57*	.11	.04	–.07
PA				1.00	.31 (.25–.38)*	–.27*	–.07	–.32*	–.26*
PS					1.00	–.59*	.30*	.46*	.16

(*continued on next page*)

Appendix (*continued*)

Factor	Correlations (95% confidence intervals)					Latent means comparisons (from models 2-5 and 5-5 in Table 1)			
	PSW	PC	SC	PA	PS	M_{girls} ($M_{boys} = 0$)	M_n ($M_u = 0$)	M_o ($M_u = 0$)	M_o($M_n = 0$)
Geomin rotation with default epsilon value									
GSW	.08 (−.28 to .44)	.57 (.51–.64)*	.58 (.47–.69)*	.39 (.31–.47)*	.48 (.40–.56)*	−.57*	−.04	−.51*	−.47*
PSW	1.00	.25 (−.03 to 53)	.36 (.05–.66)*	−.02 (−.27 to .22)	.21 (−.04 to.46)	.12	−.06	.01	.07
PC		1.00	.69 (.65–.73)*	.15 (.07–.23)*	.51 (.45–.57)*	−.57*	.08	−.22*	−.30*
SC			1.00	.16 (.07–.26)*	.70 (.65–.75)*	−.53*	.12	.10	−.02
PA				1.00	.14 (.06–.23)*	−.14	−.06	−.18	−.12
PS					1.00	−.54*	.35*	.64	.29
Geomin rotation with an epsilon value of .5									
GSW	.51 (.46–.57)*	.23 (.18–.28)*	.22 (.18–.27)*	.37 (.32–.41)*	.23 (.18–.29)*	−.47*	−.12	−.71*	−.59
PSW	1.00	.37 (.32–.41)*	.43 (.38–.48)*	.35 (.29–.40)*	.37 (.32–.41)*	−.26*	.11	−.41*	−.30
PC		1.00	.45 (.42–.48)*	.16 (.12–.20)*	.33 (.30–.37)*	−.55*	.08	−.22*	−.30*
SC			1.00	.21 (.16–.26)*	.51 (.47–.54)*	−48*	.10	.10	−.01
PA				1.00	.20 (.15–.25)*	−.23*	−.05	−.22	−.18
PS					1.00	−.54*	.35*	.61*	.26*

Note: *$p \leq .05$; GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA = physical attractiveness; PS = physical strength; M_{girls} = latent means observed in girls; M_{boys} = latent means observed in boys; M_u = latent means observed in underweight participants; M_n = latent means observed in normal-weight participants; M_o = latent means observed in overweight participants; CF = Crawford-Ferguson family of rotations.

[a] Target rotation was specified with the a priori main factor loadings freely estimated (e.g. the loadings of the GSW items on the GSW factor) and the cross-loadings specified with a target value of zero (with the ∼0 function).

References

Alsaker, F. D. (1995). Timing of puberty and reactions to pubertal changes. In M. Rutter (Ed.), *Psychosocial disturbances in young people: Challenges for prevention* (pp. 37–82). New York, NY: Cambridge University.

Angold, A., & Worthman, C. M. (1993). Puberty onset of gender differences in rates of depression: a developmental, epidemiologic and neuroendocrine perspective. *Journal of Affective Disorders, 29*, 145–158.

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 & Exercise, 6*, 33–50.

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling, 16*, 397–438.

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.

Barendse, M. T., Oort, F. J., & Garst, G. J. A. (2010). Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias: a simulation study. *AStA Advances in Statistical Analysis, 94*, 117–127.

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling, 13*, 186–203.

Bernaards, C. A., & Jennrich, R. I. (2005). Gradient projection algorithms and software for arbitrary rotation criteria in factor analysis. *Educational & Psychological Measurement, 65*, 676–696.

Bernardo, R. P. S., & Matos (de), M. G. (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 & Youth, 4*, 93–109.

Brown, J. D., Dutton, K. A., & Cook, K. E. (2001). From the top down: self-esteem and self-evaluation. *Cognition & Emotion, 15*, 615–631.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150.

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

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement. *Structural Equation Modeling, 14*, 464–504.

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

Cicchetti, D., & Rogosch, F. A. (2002). A developmental psychopathology perspective on adolescence. *Journal of Consulting & Clinical Psychology, 70*, 6–20.

Cole, T. J. (1979). A method for assessing age-standardized weight-for-height in children seen cross-sectionally. *Annals of Human Biology, 6*, 249–268.

Cole, T. J., Bellizzi, M., Flegal, K., & Dietz, W. (2000). Overweight and obesity worldwide: international establishing a standard definition for child survey. *British Medical Journal, 320*, 1240–1243.

Cole, T. J., Flegal, K., Nicholls, D., & Jackson, A. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. *British Medical Journal, 335*, 194–197.

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

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

Crocker, P., Kowalski, N., Kowalski, K., Chad, K., Humbert, L., & Forrester, S. (2001). Smoking behaviour and dietary restraint in young adolescent women: the role of physical self-perceptions. *Canadian Journal of Public Health, 92*, 428–432.

Crocker, P., Sabiston, C., Forrestor, S., Kowalski, N., Kowalski, K., & McDonough, M. (2003). Predicting change in physical activity, dietary restraint, and physique anxiety in adolescent girls. Examining covariance in physical self-perceptions. *Canadian Journal of Public Health, 94*, 322–337.

Dolan, C. V. (1994). Factor analysis of variables with 2, 3, 5 and 7 response categories: a comparison of categorical variable estimators using simulated data. *British Journal of Mathematical & Statistical Psychology, 47*, 309–326.

Dolan, C. V., Oort, F. J., Stoel, R. D., & Wicherts, J. M. (2009). Testing measurement invariance in the target rotated multigroup exploratory factor model. *Structural Equation Modeling, 16*, 295–314.

Eccles, J. S., Midgley, C., Buchanan, C. M., Wigfield, A., Reuman, D., & Mac Iver, D. (1993). Development during adolescence: the impact of stage/environment fit. *American Psychologist, 48*, 90–101.

Eiser, C., Eiser, J. R., & Havermans, T. (1995). The measurement of self-esteem: practical and theoretical considerations. *Personality & Individual Differences, 18*, 429–432.

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 & Sport, 68*, 249–256.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.

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.

Fox, K. R. (1998). Advances in the measurement of the physical self. In J. L. Duda (Ed.), *Advances in sport and exercise psychology measurement* (pp. 295–310). Morgantown, WV: Fitness Information Technology.

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 & Exercise Psychology, 11*, 408–430.

French, S. A., Story, M., & Perry, C. L. (1995). Self-esteem and obesity in children and adolescents: literature review. *Obesity Research, 3*, 479–490.

Garner, D. M., Olmstead, M. P., Bohr, Y., & Garfinkel, P. E. (1982). The eating attitude test: psychometric features and clinical correlates. *Psychological Medicine, 12*, 871–878.

Graham, J. W. (2009). Missing data analysis: making it work in the real world. *Annual Review of Psychology, 60*, 549–576.

Griffiths, L. J., Parsons, T. J., & Hill, A. J. (2010). Self-esteem and quality of life in obese children and adolescents: a systematic review. *International Journal of Pediatric Obesity, 5*, 282–304.

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

Hagger, M. S., & Stevenson, A. (2010). Social physique anxiety and physical self-esteem: gender and age effects. *Psychology & Health, 25*, 89–110.

Hagger, M. S., Stevenson, A., Chatzisarantis, N. L. D., Pereira Gaspar, P. M., Leitão Ferreira, P. M., & González Ravé, J. M. (2010). Physical self-concept and social physique anxiety: invariance across culture, gender and age. *Stress & Health, 26*, 304–329.

Halpern, C. T., Udry, J. R., Campbell, B., & Suchindran, C. (1999). Effects of body fat on weight concerns, dating, and sexual activity: a longitudinal analysis of black and white adolescent girls. *Developmental Psychology, 35*, 721–736.

Hart, E. H., Leary, M. R., & Rejeski, W. J. (1989). The measurement of social physique anxiety. *Journal of Sport & Exercise Psychology, 11*, 94–104.

Harter, S. (1999). *The construction of the self: A developmental perspective*. New York, NY: Guilford.

Hau, K. T., Sung, R. Y. T., Yu, C. W., Marsh, H. W., & Lau, P. W. C. (2005). Factorial structure and comparison between obese and nonobese: Chinese children's physical self-concept. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *The new frontiers of self research* (pp. 259–272). Charlotte, NC: Information Age Publishing.

Hildebrandt, A., Wilhelm, O., & Robitzsch, A. (2009). Complementary and competing factor analytic approaches for the investigation of measurement invariance. *Review of Psychology, 16*, 87–102.

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.

Jennrich, R. I. (2007). Rotation methods, algorithms, and standard errors. In R. Cudeck, & R. C. MacCallum (Eds.), *Factor analysis at 100: Historical developments and future directions* (pp. 315–335). Mahwah, NJ: Erlbaum.

Lau, P. W. C., Cheung, M. W. L., & Ransdell, L. B. (2008). A structural equation model of the relationship between body perception and self-esteem: global physical self-concept as the mediator. *Psychology of Sport & Exercise, 9*, 493–509.

Lei, P.-W. (2009). Evaluating estimation methods for ordinal data in structural equation modeling. *Quality & Quantity, 43*, 495–507.

Leichner, P., Steiger, H., Puentes-Neuman, G., Perreault, M., & Gottheil, N. (1994). Validation d'une échelle d'attitudes alimentaires auprès d'une population québécoise francophone. *Revue Canadienne de Psychiatrie, 39*, 49–54.

Lissau, I., Overpeck, M. D., Ruan, W. J., Due, P., Holstein, B. E., & Hediger, M. L. (2004). Body mass index and overweight in adolescents in 13 European countries, Israël, and the United States. *Archives of Pediatrics & Adolescent Medicine, 158*, 27–33.

Lubke, G. H., & Muthén, B. O. (2004). Applying multigroup confirmatory factor models for continuous outcomes to Likert scale data complicates meaningful group comparisons. *Structural Equation Modeling, 11*, 514–534.

Lundgren, J. D., Anderson, D. A., & Thompson, J. K. (2004). Fear of negative appearance evaluation: development and evaluation of a new construct for risk factor work in the field of eating disorders. *Eating Behaviors, 5*, 75–84.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

MacCallum, R. C., Zhang, S., Preacher, K. J., & Rucker, D. D. (2002). On the practice of dichotomization of quantitative variables. *Psychological Methods, 7*, 19–40.

Maïano, C., Bégarie, J., Morin, A. J. S., & Ninot, G. (2009). Assessment of physical self-concept in adolescents with intellectual disability: content and factor validity of the very short form of the physical self-inventory. *Journal of Autism & Developmental Disorders, 39*, 775–787.

Maïano, C., Morin, A. J. S., Eklund, R. C., Monthuy-Blanc, J., Garbarino, J.-M., & Stephan, Y. (2010). Construct validity of the social physique anxiety scale in a French adolescent sample. *Journal of Personality Assessment, 92*, 1–10.

Maïano, C., Morin, A. J. S., Monthuy-Blanc, J., & Garbarino, J.-M. (2009). The body image avoidance questionnaire: assessment of its construct validity in community sample of French adolescents. *International Journal of Behavioral Medicine, 16*, 125–135.

Maïano, C., Morin, A. J. S., Monthuy-Blanc, J., & Garbarino, J.-M. (2010). Construct validity of the fear of negative appearance evaluation scale in a community sample of French adolescents. *European Journal of Psychological Assessment, 26*, 19–27.

Maïano, C., Morin, A. J. S., Ninot, G., Monthuy-Blanc, J., Stephan, Y., Florent, J.-. F., et al. (2008). A short and very short form of the physical self-inventory for adolescents: development and factor validity. *Psychology of Sport & Exercise, 9*, 830–847.

Marsh, H. W. (1996). Construct validity of physical self-description questionnaire responses: relations to external criteria. *Journal of Sport & Exercise Psychology, 18*, 111–131.

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. (1998). Age and gender effects in physical self-concepts for adolescent elite athletes and nonathletes: a multicohort–multioccasion design. *Journal of Sport & Exercise Psychology, 20*, 237–259.

Marsh, H. W., Asçl, F. H., & Marco, I. T. (2002). Multitrait–multimethod analyses of two physical self-concept instruments: a cross-cultural perspective. *Journal of Sport & Exercise Psychology, 24*, 99–119.

Marsh, H. W., Balla, J. R., & Hau, K. T. (1996). An evaluation of incremental fit indices: a clarification of mathematical and empirical processes. In G. A. Marcoulides, & R. E. Schumacker (Eds.), *Advanced structural equation modeling techniques* (pp. 315–353). Hillsdale, NJ: Erlbaum.

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 & Exercise Psychology, 28*, 310–343.

Marsh, H. W., & Cheng, J. H. S. Measuring physical self-concept. In G. Tenenbaum, R. Eklund., & A. Kamata (Eds.), *Handbook of measurement in sport and exercise psychology*. Champaign, IL: Human Kinetics, in press.

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., & Grayson, D. (1995). Latent variable models of multi-trait–multimethod data. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues and application* (pp. 177–198). Thousand Oaks, CA: Sage.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares, & J. McArdle (Eds.), *Psychometrics. A Festschrift to Roderick P. McDonald*. Hillsdale, NJ: Erlbaum.

Marsh, H. W., Hau, K. T., Sung, R. Y. T., & Yu, C. W. (2007). Childhood obesity, gender, actual–ideal body image discrepancies, and physical self-concept in Hong Kong children: cultural differences in the value of moderation. *Developmental Psychology, 43*, 647–662.

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

Marsh, H. W., Liem, G. A. D., Martin, A. J., Morin, A. J. S., & Nagengast, B. Methodological-measurement fruitfulness of exploratory structural equation modeling: new approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, in press.

Marsh, H. W., Lüdtke, O., Muthén, B. O., Asparouhov, T., Morin, A. J. S., Trautwein, U., et al. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22*, 471–491.

Marsh, H. W., Lüdtke, O., Robitzsch, A., Nagengast, B., Morin, A. J. S., & Trautwein, U. Two wrongs do not make a right: camouflaging misfit with item-parcels in CFA models. *Psychological Methods*, under review.

Marsh, H. W., Muthén, B. O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A. J. S., et al. (2009). Exploratory structural equation modeling, integrating CFA and EFA: application to students' evaluations of university teaching. *Structural Equation Modeling, 16*, 439–476.

Marsh, H. W., Nagengast, B., Morin, A. J. S., Parada, R. H., Craven, R. G., & Hamilton, L. R. *Construct validity of the multidimensional structure of bullying and victimization: an application of exploratory structural equation modeling. Journal of Educational Psychology*, in press.

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 & 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 multi-trait–multimethod analysis of relations to existing instruments. *Journal of Sport & Exercise Psychology, 16*, 270–305.

Marsh, H. W., & Roche, L. A. (1996). Predicting self-esteem from perceptions of actual and ideal ratings of body fatness: is there only one ideal supermodel? *Research Quarterly for Exercise & Sport, 6*, 13–23.

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg self-esteem scale: traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22*, 366–381.

McCrae, R. R., Zonderman, A. B., Costa, P. T., Jr., Bond, M. H., & Paunonen, S. (1996). Evaluating the replicability of factors in the revised NEO personality inventory: confirmatory factor analysis versus procrustes rotation. *Journal of Personality & Social Psychology, 70*, 552–566.

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525–543.

Monsma, E. V., & Malina, R. M. (2004). Correlates of eating disorders risk among female figure skates: a profile of adolescent competitors. *Psychology of Sport & Exercise, 5*, 447–460.

Monsma, E. V., Malina, R. M., & Feltz, D. L. (2006). Puberty and physical self-perceptions of competitive female figure skaters: an interdisciplinary approach. *Research Quarterly for Exercise & Sport, 77*, 158–166.

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 & Human Development, 1*, 1–17.

Morin, A. J. S., Maïano, C., Marsh, H. W., Janosz, M., & Nagengast, B. (2011). The longitudinal interplay of adolescents' self-esteem and body image: a conditional autoregressive latent trajectory analysis. *Multivariate Behavioral Research, 46*, 157–201.

Motl, R. W., Conroy, C., & Horan, P. M. (2001). The social physique anxiety scale: an example of the potential consequences of negatively worded items in factorial validity studies. *Journal of Applied Measurement, 1*, 327–345.

Muthén, L. K., & Muthén, B. (2010). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

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. *Personality & Individual Differences, 14*, 585–590.

Puhl, R. M., & Latner, J. D. (2007). Stigma, obesity, and the health of the nation's children. *Psychological Bulletin, 133*, 557–580.

Rhemtulla, M., Brosseau-Liard, P., & Savalei, V. (2010). How many categories is enough to treat data as continuous? A comparison of robust continuous and categorical SEM estimation methods under a range of non-ideal situations. http://www2.psych.ubc.ca/~mijke/files/HowManyCategories.pdf.

Robins, R. W., Trzesniewski, K. H., Tracy, J. L., Gosling, S. D., & Potter, J. (2002). Global self-esteem across the life span. *Psychology & Aging, 17*, 423–434.

Rosen, J., Srebnik, D., Saltzberg, E., & Wendt, S. (1991). Development of a body image avoidance questionnaire. *Psychological Assessment, 3*, 32–37.

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

Sass, D. A., & Schmitt, T. A. (2010). A comparative investigation of rotation criteria within exploratory factor analysis. *Multivariate Behavioral Research, 45*, 73–103.

Satorra, A., & Bentler, P. (1999). *A scaled difference chi-square test statistic for moment structure analysis*. Technical report. Los Angeles: University of California.

Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: implications for factor pattern loadings and inter-factor correlations. *Educational & Psychological Measurement, 71*, 95–113.

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

Siegel, J. M., Yancey, A. K., Aneshensel, C. S., & Schuler, R. (1999). Body image, perceived pubertal timing, and adolescent mental health. *Journal of Adolescent Health, 25*, 155–165.

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

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

Sonstroem, R. J., & Morgan, W. P. (1989). Exercise and self-esteem: rationale and model. *Medicine & Science in Sports & Exercise, 21*, 329–337.

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 & Exercise Psychology, 14*, 207–221.

Steinberg, L., & Morris, A. (2001). Adolescent development. *Annual Review of Psychology, 52*, 83–110.

Stice, E., & Bearman, S. K. (2001). Body-image and eating disturbances prospectively predict increases in depressive symptoms in adolescent girls: a growth curve analysis. *Developmental Psychology, 37*, 597–607.

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

Stunkard, A. J., Sorenson, T., & Schulsinger, F. (1983). Use of the Danish adoption registry for the study of obesity and thinness. In S. Kety (Ed.), *The genetics of neurological and psychiatric disorders* (pp. 115–120). New York: Raven.

Sung, R. Y. T., Yu, C. W., So, R. C. H., Lam, P. K. W., & Hau, K. T. (2005). Self-perception of physical competences in preadolescent overweight Chinese children. *European Journal of Clinical Nutrition, 59*, 101–106.

Tomás, J., & Oliver, A. (1999). Rosenberg's self-esteem scale: two method factors or method effects? *Structural Equation Modeling, 6*, 84–98.

Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: meta-analyses comparing Whites, Blacks, Hispanics, Asians, and American Indians and comment on Gray-Little and Hafdahl (2000). *Psychological Bulletin, 128*, 371–408.

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.

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 & Individual Differences, 32*, 855–863.

Wardle, J., & Cooke, L. (2005). The impact of obesity on psychological well-being. *Best Practice & Research Clinical Endocrinology & Metabolism, 19*, 421–440.

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.

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

Wylie, R. C. (1989). *Measures of self-concept*. Lincoln: University of Nebraska Press.