English Validation of the Short Form of the Physical Self-Inventory (PSI-S)

Alexandre J. S. Morin*, Institute for Positive Psychology and Education, Australian Catholic University, Australia

Christophe Maïano*, Cyberpsychology Laboratory, Department of Psychoeducation and Psychology, Université du Québec en Outaouais (UQO), Canada

Rhiannon Lee White, Institute for Positive Psychology and Education, Australian Catholic University, Australia

Katherine B. Owen, Institute for Positive Psychology and Education, Australian Catholic University, Australia

Danielle Tracey, Western Sydney University, Australia

Nicolas Mascret, Aix Marseille Univ, CNRS, ISM, Inst Movement Sci, Marseille, France

Chris Lonsdale, Institute for Positive Psychology and Education, Australian Catholic University, Australia

* 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.

Acknowledgements

Preparation of this article was supported in part by grants from the Australian Research Council (DP140101559 and DP130104659).

This is the final pre-publication version of a manuscript accepted for publication:

Morin A. J. S., Maïano, C., White, R.L., Owen, K., Tracey, D., Mascret, N., & Lonsdale, C. (2016). English Validation of the Short Form of the Physical Self-Inventory (PSI-S). *Psychology of Sport & Exercise*, *27*, 180-194. doi: 10.1016/j.psychsport.2016.08.016

Corresponding author:
Alexandre J.S. Morin, Institute for Positive Psychology and Education,
Australian Catholic University, Strathfield Campus,
Locked Bag 2002, Strathfield, NSW 2135, Australia
E-mail: Alexandre.Morin@acu.edu.au

Highlights

- We assess the psychometric properties of a revised Physical Self Inventory-Short (PSI-S).
- This assessment relies on Exploratory Structural Equation Modeling (ESEM).
- Results support the psychometric properties of the English and French versions.
- Ratings were longitudinally invariant across 7-8 months.
- We found no differential item functioning (age, gender, BMI and physical activity).

Abstract

The study examined the psychometric properties of scores on the English version of the short Physical Self-Inventory (PSI-S) among 1,368 English-Speaking and 224 French-Speaking adolescents. Participants completed the original PSI-S and a revised version including positively-worded reformulations of the original negatively-worded items. The results supported the a priori factor structure, composite reliability, and test-retest reliability (7-8 months) of scores on the revised PSI-S, and its equivalence with the French version. Compared to confirmatory factor analyses, exploratory structural equation modeling resulted in a superior solution, and more cleanly differentiated factors. PSI-S responses were longitudinally invariant across 7-8 months, and presented no evidence of differential item functioning as a function of age, gender, body mass index (BMI) and physical activity. The results showed latent mean differences as a function of gender, BMI (linear and curvilinear effects), and physical activity, supporting the criterion-related validity of the PSI-S factors.

Key words: Physical self-concept, physical self-inventory, short form, exploratory structural equation modeling, ESEM, English.

Fox and Corbin (1989; Fox, 2000) developed a multidimensional and hierarchical model of the physical self-concept inspired by Shavelson, Hubner, and Stanton's (1976) representation of the self-concept. Global self-worth (GSW), which represents the positive or negative way people feel about themselves as a whole, occupies the upper level of this representation. Global physical self-worth (PSW), which refers to people's general feelings of satisfaction regarding their physical self, occupies the intermediate (domain) level of this hierarchical model. Finally, the four dimensions of sport competence (SC: ability to learn and to perform in sports), physical condition (PC: physical endurance and fitness), physical attractiveness (PA: positive body image) and physical strength (PS: strength and muscularity) occupy the lowest (subdomain) level of this hierarchy.

Fox and Corbin (1989) created the Physical Self-Perception Profile (PSPP) in order to measure the various dimensions of Shavelson et al.'s model. Since then, scores on the PSPP have shown validity evidence across English-speaking samples of adults (e.g., Hagger, Asç1, & Lindwall, 2004; Page, Ashford, Fox, & Biddle, 1993), and cross-cultural validity evidence in other languages among samples from countries such as Portugal (Fonseca & Fox, 2002), Spain (Atzienga, Balaguer, Moreno, & Fox, 2004), Sweden (Hagger et al., 2004), Belgium, the Netherlands (Van de Vliet et al., 2002), and Turkey (Hagger et al., 2004; Marsh, Asçl, & Marco, 2002). Still, over the years, a variety of concerns about diverse characteristics of the PSPP have been expressed. First, the PSPP relies on items taken from the Rosenberg Self-Esteem Inventory (RSEI; Rosenberg, 1965) to assess GSW, an instrument including a mixture of negatively- and positively-worded items known to be associated with methodological artifacts interfering with the assessment of the latent constructs (Marsh, Scalas, & Nagengast, 2010). Second, the PSPP relies on a structured alternative response scale (i.e., a 4-point scale involving paired forced-choices[1]), which has often been criticized as confusing and associated with important methodological effects also interfering with the ratings (Eiser, Eiser, & Haversmans, 1995; Marsh, Bar-Eli, Zach, & Richards, 2006; Marsh et al., 2002; Wichstrøm, 1995). Third, some

[1] PSPP items encompass two opposite statements (e.g., "some people feel that they are not very good when it comes to playing sports" BUT "others feel that they are really good at just about every sport"). One statement reflects a high level of self-concept (the second one in this example) while the other one reflects a low level of self-concept. Respondents are first asked to select the statement which best describes them. Then, they are asked to indicate whether this statement is "really true" or "sort of true" for them, resulting in a 4-point rating scale.

have questioned the appropriateness of the PSPP for children and adolescents who might have more difficulties discerning between PSPP subdomains due to a combination of abstract item formulation, complex response scale, and their more limited cognitive and linguistic skills relative to adults (Biddle, Page, Ashford, Jennings, Brooke, & Fox, 1993; Marsh, Richards, Johnson, Roche, & Tremayne, 1994). These criticisms led to the development of a revised version of the PSPP, specifically designed for North-American children and adolescents (Eklund, Whitehead, & Welk, 1997; Whitehead, 1995). This revised version has since been cross-validated in other languages among a variety of non-English speaking populations (Aşç1, Eklund, Whitehead, Kirazci, & Koca, 2005; Bernardo & Matos, 2003; Hagger, Ashford, & Stambulova, 1997; Moreno, Cervellò, Vear, & Ruiz, 2007). Still, this revised version continues to rely on the potentially problematic structured alternative format answer scale.

In their review of available measures of the physical self-concept, Marsh and Cheng (2012) identify the Physical Self-Description Questionnaire (PSDQ; Marsh et al., 1994) as a strong alternative to the PSPP for the assessment of physical self-conceptions across a wide variety of age groups and languages (Marsh et al., 2002, 2006). Arguably one of the most commonly studied measure of the physical self-concept, the PSDQ also appears free of the complications associated with the PSPP. Still, the length of the PSDQ (70 items, versus 30 for the PSPP) is impractical for large-scale epidemiologic or longitudinal studies. In these contexts, shorter instruments help to maximize the amount of information collected without overtaxing participants' cognitive resources. Even though a shorter 40-item PSDQ-S has been developed (Maïano, Morin, & Mascret, 2015; Marsh, Martin, & Jackson, 2010), this "*short*" version remains long for studies aiming to assess a wide variety of constructs.

Among the various instruments covered in their review, Marsh and Cheng (2012) identified the Physical Self-Inventory (PSI) as one of the shortest measures of the physical self-concept available to date, and one showing the greatest promise. Using the PSPP as a basis, the PSI (Ninot, Delignières, & Fortes, 2000) was developed with the objective of overcoming some of these limitations. First, the original PSPP 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*). Second, the PSW and GSW scales were respectively replaced by five items from the Self-Description Questionnaire-III (Marsh & O'Neill, 1984) and by 5 items from Coopersmith's (1967) Self-Esteem Inventory. Third, starting from the slightly longer 25-item PSI version developed in French by Ninot et al. (2000), Maïano et al. (2008) developed an 18-item short form of this instrument (PSI-S; 3 items per dimension), specifically adapted for French-speaking adolescents. Among a sample of 1,018 French adolescents (aged 11 to 16) and relying on confirmatory factor analyses (CFA), Maïano et al. (2008) provided evidence of reliability for PSI-S ratings, and demonstrated that these ratings followed the expected factor structure and remained invariant across gender. Their results further showed meaningful latent mean differences, revealing lower levels on most PSI-S dimensions (GSW, PSW, SC, PA and PS) among girls relative to boys, a result that supported conclusions from prior research conducted in the physical self-concept area (Aşç1, 2002; Hagger, Biddle, & Wang, 2005; Marsh et al., 2006; Marsh, Hau, Sung, & Yu, 2007). More recently, Maïano et al. (2015) also demonstrated the convergent validity of PSI-S ratings with matching subscales from the PDSQ-S. Keeping in mind that a key criticism of the original PSPP is related to its complexity for children and adolescents, it is particularly interesting to note that a version of the PSI has since been successfully adapted to samples of youth with intellectual disabilities (Maïano, Bégarie, Morin, & Ninot, 2009; Maïano, Morin, Bégarie, & Ninot, 2011).

The PSI-S is one of the shortest multidimensional measures of physical self-perceptions available to date. As such, it is not surprising that it was the only non-English language instrument highlighted by Marsh and Cheng's (2012) review as representing a significant contribution to applied research on physical self-concept measurement. However, Marsh and Cheng (2012) noted two critical limitations of the PSI-S related to: (i) the high correlations observed among PSI-S subscales, and (ii) the need to develop PSI-S versions applicable to a greater variety of linguistic groups, including English-speaking populations. We address these two limitations in sequence.

Factor Correlations, Cross-Loadings, and Exploratory Structural Equation Modeling

The first limitation noted by Marsh and Cheng (2012) relates to the observation that the factor correlations reported by Maïano et al. (2008) among the PSI-S factors (i.e., r = .50 to .91) are high enough to call into question the independence of the factors. This limitation, however, is not restricted to the PSI-S, as it is also shared by most PSPP-based self-concept measures for which factor correlations reaching or higher than .80 are often reported (e.g., Atzienga et al., 2004; Fox & Corbin, 1989; Hagger et al., 2004, 2005; Marsh & Cheng, 2012; Marsh et al., 1994, 2006). Initial interpretations attributed these high factor correlations to the PSPP idiosyncratic response scale (Marsh et al., 1994, 2006). Yet, the PSI-S relies on a traditional Likert-type format, and recent research suggests that structured alternative responses may perform better than previously anticipated when analyzed with proper measurement models (Arens & Morin, 2016).

More precisely, the independent cluster model (ICM) inherent in CFA measurement models, which forces cross-loadings between items and non-target factors to be exactly zero, has recently been called into question for the assessment of constructs closely related to one another conceptually (Marsh, Morin, Parker, & Kaur, 2014; Morin, Marsh, & Nagengast, 2013). In fact, the idea that psychometric ratings (i.e., the items) tend to include more than one source of true score variance represents one of the core tenets of classical test theory (e.g., Nunnally & Bernstein, 1994). Thus, whenever multiple conceptually-related constructs are represented within a single measurement model, ratings on the items can be expected to present at least some level of true score association with non-target constructs (i.e., cross-loadings; Morin, Arens, & Marsh, 2016). ICM restrictions force these cross-loadings to be exactly zero. As a result, the only way through which these cross-loadings can be expressed is through the inflation of the CFA factor correlations. A variety of statistical simulation studies, as well as studies of simulated data, have supported this interpretation (for a recent review, see Asparouhov, Muthén, & Morin, 2015). These studies have revealed that the free estimation of all cross-loadings results in more exact estimates of the true factor correlations when cross-loadings are present in the population model, but still provide unbiased estimates of factor correlations when population models correspond to ICM assumptions. Because the true meaning of psychometric constructs stems from the manner in which they relate to other constructs, this means that ICM-CFA models carry the risk of inducing fundamental biases in terms of construct definition.

Exploratory factor analysis (EFA) allows for the free estimation of cross-loadings, and has been recently incorporated with CFA and structural equation modeling (SEM) into a broader exploratory structural equation modeling (ESEM) statistical framework (Asparouhov & Muthén, 2009; Morin et al., 2013). ESEM makes it possible to apply to EFA measurement models most types of psychometric tests typically limited to CFA (goodness-of-fit assessment, measurement invariance, estimation of regressions between latent factors, etc.). Through target rotation it is now even possible to rely on a truly “confirmatory” approach in the estimation of EFA/ESEM measurement models (Asparouhov & Muthén, 2009; Browne, 2001). More precisely, target rotation allows the analyst to pre-specify target loadings in a confirmatory manner and to “target” cross-loadings to be as close to zero as possible.

Based on these considerations, Morin and Maïano (2011a) applied ESEM to PSI-S answers provided by a sample of 2,029 French adolescents (aged 11 to 18) to verify whether inappropriate ICM-CFA assumptions may explain the previously reported high factor correlations. Their results first supported the superiority of the ESEM solution, showing that this solution resulted in a higher degree of fit to the data and fully acceptable factor correlations (r = .16 to .51), whereas the ICM-CFA factor correlations remained problematically high (r = .52 to .93). Using the retained ESEM solution, their results further supported the scale score reliability (α =.64 to .87) and convergent validity of PSI-S ratings (with measures of physical self-image congruence, body image avoidance, disturbed eating attitudes and behaviors, fear of negative appearance evaluation, and social physique anxiety).

Finally, Morin and Maïano’s (2011a) ESEM results revealed problems with the three negatively worded items of the PSI-S (two PA items and one GSW item), which could not be controlled by the inclusion of methodological controls for wording effects. These results thus added to the bulk of previous research showing that negatively-worded items tended to be particularly challenging for self-concept measurement (DiStefano, & Motl, 2006; Lindwall, Aşçı, & Hagger, 2011; Marsh, Scalas et al., 2010), particularly for the adaptation of self-concept measures to new cultural or linguistic groups (Aşçı, Fletcher, & Çağlar, 2009; Watkins & Cheung, 1995). For these various reasons, Morin and Maïano (2011a) proposed a positively-worded reformulation of these items, recommending that future research should systematically aim to verify “whether the psychometric properties of the original PSI-S can be preserved, and even improved, with the proposed reformulations of these items” (p. 550).

Further evidence of Measurement Invariance and Group Differences

Using the retained ESEM solution, Morin and Maïano’s (2011a) results also supported the measurement invariance of the PSI-S ratings (using the retained ESEM measurement model) across genders (male versus female), age groups (early versus late adolescence), and weight categories calculated from participants’ body mass index (BMI; underweight, normal weight, overweight, or obese), but revealed some significant latent mean differences across genders and weight categories. More precisely, they found that girls tended to present a lower level than boys on all PSI-S dimensions (GSW, PSW, PC, SC, PA and PS), and that overweight/obese participants tended to present lower GSW, PSW and PC than underweight and normal weight participants, whereas PS scores tended to increase as a function of participants' weight categories. Altogether, these results supported those from prior research on the PSI-S (Maïano et al., 2008) and other physical self-concept instruments (e.g., Aşç1, 2002; Griffiths, Parsons, & Hill 2010; Hagger et al., 2005; Hau, Sung, Yu, Marsh, & Lau, 2005; Marsh et al., 2006, 2007; Sung, Yu, So, Lam, & Hau, 2005). The observed increases in PS levels as a function of participants' weight categories are consistent with the observation that high BMI might reflect muscular structure in addition to body fat (Morin & Maïano, 2011a). This interpretation suggests that future research would do well to take into account participants' levels of involvement in physical activities, as well as possible interactions between physical activity and BMI levels.

Although it is well-documented that physical activity involvement tends to be associated with more positive physical self-perceptions (Babic, Morgan, Plotnikoff, Lonsdale, White, & Lubans, 2014; Bowker, 2006; Findlay & Bowker, 2007; Schmalz & Davison, 2006), similar associations have never been tested with the PSI-S. Furthermore, although prior research suggested that physical self-perceptions may decrease as a function of age (Hagger et al., 2005; Marsh et al., 2007), Morin and Maïano (2011a) did not observe mean-level differences across samples of early (11-14 years) and late (15-18 years) adolescents. However, Morin and Maïano's (2011a) study was limited by the reliance on a rough categorization of BMI and age into a limited number of subgroups, whereas the categorization of continuous variables is known to result in substantial decreases in terms of statistical power (MacCallum, Zhang, Preacher, & Rucker, 2002; Marsh, Nagengast, & Morin, 2013).

Furthermore, the possibility that associations between physical self-conceptions and participants' age, BMI, and physical activity levels may differ across genders was never systematically investigated. This is particularly surprising given that physical standards are known to differ as a function of gender-differentiated sociocultural norms and socialization processes (Klomsten, Shaalvik, & Espnes, 2004; Smith, Noll, & Bryant, 1999; McCabe & Ricciardelli, 2003; Tiggeman, 2003). These differences may be particularly salient in adolescence due to puberty, which often results in body fat accumulation in girls, an undesired change according to the thin-ideal Caucasian beauty standards but a desired one among cultures valuing “fuller” forms. In contrast, for boys adolescence usually involves muscular increase and the emergence of other valued attributes (Morin, Maïano, Marsh, Janosz, & Nagengast, 2011; Siegel, Yancey, Aneshensel, & Schuler, 1999; Stice & Bearman, 2001).

English Adaptation of the PSI-S: The Present Study

The second limitation noted by Marsh and Cheng (2012) is related to the need to develop a more diversified set of linguistic versions of the PSI-S, which was only available in French, and particularly an English version of this instrument. Clearly, the lack of an English version is a serious limitation to the PSI-S, and a severe impediment to its more widespread use in research. When adapting psychometric measures to new cultural or linguistic groups, one of the key challenges is to develop a measure that preserves the psychometric properties of the original version when adapted to the new linguistic or cultural context (i.e., measurement invariance: Millsap, 2011). In order to answer the need for an English version of the PSI-S, the present study was designed to verify whether Morin and Maïano’s (2011a) results regarding the superiority of an ESEM representation of the PSI-S ratings could be replicated using an English language adaptation of the PSI-S. We also extended this previous research by contrasting two different versions of the PSI-S, the original version (including the initial pool of 18 items) and the revised version (in which the three reversed-keyed items have been replaced by their positively-worded reformulations). In addition to testing the stability of PSI-S responses over time in order to obtain an estimate of the test-retest reliability of the English PSI-S version, we also assessed the measurement invariance of ratings on the retained PSI-S (original or revised) version across samples of English- and French- speaking participants. We then systematically tested the measurement invariance and latent mean differences of ratings on the retained English language PSI-S version across gender. Finally, we extended these tests to an assessment of differential item functioning (DIF) and possible latent mean differences on the PSI-S factors as a function of age, BMI, and physical activity, and assessed the extent to which these results generalize across genders. This last set of analyses aimed to extend Morin and Maïano’s (2011a) results by testing the relations between physical self-concept levels and participants’ age, BMI, and physical activity without relying on a suboptimal categorization of continuous variables (age, BMI, and physical activity involvement).

Method

Samples and Procedures

The main sample used in this study included 1,368 English-speaking Australian adolescents (606 girls and 762 boys) aged between 12 and 14 years old (M_{age} = 12.94; SD_{age} = .53). We recruited these participants from 14 government-funded secondary schools located in the greater Western Sydney area (for additional details see Lonsdale et al., 2016). Participants completed the English version of the PSI-S in the first school term (February-April) of the 2014 school year when they were in Grade 8. From those, 1,159 (519 girls and 640 boys) completed a retest questionnaire during Term 4 of the same year (September-December, 2014), 7-8 months after the initial testing (84.7% retention). Preliminary analyses (MANOVAs) showed that participants who completed the second measurement points did not significantly differ from the other participants in terms of gender, age, or BMI ($p \geq .05$), but tended to be slightly less physically active ($eta^2 = .018$).

A second sample of 224 French-Speaking adolescents from France (129 girls and 95 boys) aged between 12 and 14 years old (M_{age} = 13.22; SD_{age} = .82) was also included in this study. We recruited these participants from a total of five middle schools ("*Collèges*") located in Southern France, and they completed the French version of the PSI-S either in the last term of the 2010-2011 school year (May-June 2011) or in the first term of the 2011-2012 school year (September-October 2011).

This project met ethical requirements for research with human participants in both countries. The Australian project was approved by the human research ethic committee of the Western Sydney University, the Australian Catholic University, and the NSW Department of Education. The French project was approved by the French Advisory Committee on Information Processing in Material Research in the Field of Health and by the Chief Education Officer of the Académie of Aix-Marseille. Authorization to perform the study was first obtained from schools' principals. Appropriate consent procedures were then followed, with permission obtained from the participants' parents prior to the data collection. All participants volunteered and the confidentiality of their responses was guaranteed.

Measures

Demographic Predictors. All participants were asked to self-report their gender (coded 0 for girls and 1 for boys) and age.

Physical Activity. Australian participants also self-reported the frequency with which they were physically active (enough to be out of breath) outside of the school setting, on a 5-point scale ranging from 1 (*once a month or less*) to 5 (*every day*; M = 3.41; SD = 1.20). This measure was adapted from the WHO Health Behaviour in School-aged Children (HBSC) item (Prochaska, Sallis, & Long, 2001), which has been found to represent a reliable and valid measure of physical activity for Australian 8th grade students (Booth, Okely, Chey, & Bauman, 2001).

Body Mass Index. Research assistants measured Australian participants' height (to the nearest 0.1 cm) using a portable stadiometer (Surgical and Medical Products N^{o}. 26SM, Medtone Education Supplies, Melbourne, Australia), as well as their weight (to the nearest 0.1 kg) using digital scales (UC-321, A&D Company LTD, Tokyo, Japan). For purposes of sample description only, French participants self-reported their height and weight. This information was then used to calculate participants' BMI (Australia: M_{BMI} = 22.10, SD_{BMI} = 4.83; France: M_{BMI} = 19.65, SD_{BMI} = 4.83), as well as their normalized BMI (BMI-Z) using the age- and gender- specific parameters recently proposed by the Center for Disease and Control (Flegal & Cole, 2013). Normalized (BMI-Z) scores were used for purposes of analyses, to ensure that BMI scores maintained the same meaning across genders and age groups. BMI levels were then categorized for descriptive purposes using the revised sex- and age-specific cut-off scores provided by the International Obesity Task Force (IOTF; Cole & Lobstein, 2012): (a) 6.2% of Australian participants and 12.5% of French participants were considered to be underweight; (b) 53.9% of the Australians and 74.1% of the French were considered to have a normal weight; (c) 26.0% of the Australian and 11.2% of the French were considered to be overweight; (d) 13.9% of the Australian and 2.2% of the French were considered to be obese.

PSI-S. English adaptations of the original (Maïano et al., 2008) and revised (Morin & Maïano, 2011a) versions of the PSI-S were developed for the present study, and used in combination with the original French versions. These versions were developed through a classical process of translation and back translation by independent bilingual translators (Gudmundsson, 2009; Hambleton, 2005; Van de Vrijver & Hambleton, 1996). Any discrepancies between the original and back-translated versions were resolved through a discussion between the translators, the back translators, and the first two authors of the present study, who are fluent in both languages. All of these versions included 18 items, rated on a 6-point Likert scale (1= *Not at all* to 6 = *Entirely*), and aimed to assess six a priori subscales (GSW, PSW, PA, PS, PC, SC) with three items each. The original version included three negatively-worded items (one associated with the GSW scale, and 2 with the PA scale), which were replaced by positively-worded reformulations in the revised version. All items are presented in Table S1 of the online supplements and available free of charge to researchers.

Analyses

All analyses were conducted using the Mplus 7.4 (Muthén & Muthén, 2016) robust maximum likelihood (MLR) estimator, which is robust to non-normality. These analyses are also robust to students' nesting within schools as they were implemented in conjunction with the Mplus design-based correction of standard errors (Asparouhov, 2005). Full information maximum likelihood (FIML; Enders, 2010) was used to handle the few missing responses present at the item level (Australia Time 1: .07% to 2.34% of missing data per item; M = 1.00 %; Australia retest/Time 2: .09 % to 2.59% of missing data per item; M = 1.02 %; France: .45 % to 4.91% of missing data per item; M = 2.08%).

The a priori factor structure of ratings on the original and revised PSI-S was tested using CFA and ESEM models. In the CFA models, answers to the PSI-S were used to estimate six correlated factors, with no cross-loadings allowed between items and non-target factors. The a priori ESEM model was estimated using confirmatory target rotation (Asparouhov & Muthén, 2009; Browne, 2001; Morin et al., 2016) in which it was hypothesized that answers to the PSI-S items would be explained by six correlated factors and all cross-loadings were targeted to be as close to zero as possible. We first estimated the alternative measurement models based on the original and revised PSI-version separately at both time points. We then estimated longitudinal models including both measurement points, and thus two separate sets of six CFA or ESEM factors. All factors were allowed to correlate within and across time-points. No cross-loadings were allowed in CFA models, whereas cross-loadings were freely estimated but targeted to be as close to zero as possible within time-points in ESEM. Longitudinal models included a priori correlated uniquenesses between matching indicators utilized at the different time-points to reflect the fact that the unique variance of these indicators emerges, in part, from shared sources of influences over time (Marsh, Abdujabbar et al., 2013).

These longitudinal models were used to estimate the longitudinal invariance and test-retest reliability of the PSI-S ratings over a period of 7-8 months. These tests of longitudinal measurement invariance were performed in the following sequence, applied to both CFA and ESEM solutions estimated for the original and revised PSI-S (Millsap, 2011; Guay, Morin, Litalien, Valois, & Vallerand, 2015): (i) configural invariance (the same measurement model is estimated at both time-points with no additional equality constraint save those required for identification purposes); (ii) weak invariance (invariance of the factor loadings); (iii) strong invariance (invariance of the factor loadings and item intercepts); (iv) strict invariance (invariance of the factor loadings, item intercepts, and items uniquenesses); (v) variance/covariance invariance (invariance of the factor loadings, item intercepts, items uniquenesses, and latent variances and covariances); and (vi) latent means invariance (invariance of the factor loadings, item intercepts, items uniquenesses, latent variances and covariances, and latent means). Composite reliability coefficients were computed from these most invariant longitudinal models standardized parameter estimates, using McDonald's (1970) $\omega = (\Sigma|\lambda i|)^2 / ([\Sigma|\lambda i|]^2 + \Sigma\delta ii)$ where λi are the factor loadings and δii, the error variances.

Starting with the most appropriate solution (ESEM or CFA, for the original or revised PSI-S), the remaining analyses focus on Time 1 responses. Given the complexity of the following models, it was not possible to rely on longitudinal models. In the second stage, we test the measurement invariance of the PSI-S ratings across the samples who completed the English (Australia) and French (France) versions of the instrument to ascertain the linguistic equivalence of the PSI-S ratings. These tests of measurement invariance were conducted in the same sequence as the previous models.

In the third stage, associations between the PSI-S ratings and the predictors (gender, age, BMI, physical activity) were assessed. We first estimated a multi-group measurement model in both gender groups, and proceeded to tests of measurement invariance across genders in the same sequence as used in the previous analyses. Starting from the most invariant of these models, the remaining predictors were included using a multiple indicators multiple causes (MIMIC) approach (e.g., Morin et al., 2013). MIMIC models are essentially regression models in which latent variables are regressed on observed predictors, and can be extended to test for the presence of DIF in relation to the predictors. DIF is a form of measurement non-invariance characterized by direct relations between predictors and item responses over and above the effects of the predictors on the latent factor. Compared to traditional multi-group tests of measurement invariance, MIMIC models are not only more parsimonious, but also consider DIF in relation to multiple continuous (e.g., age, BMI, physical activity) predictors without having to recode them into a smaller number of discrete groups (as would be needed in multi-group tests of invariance). A MIMIC approach also allows for the consideration of curvilinear, and interactive effects, among predictors. We thus also consider curvilinear effect of BMI levels (BMI^2) [2], to acknowledge the fact that optimal BMI levels are located in the average range, rather than at low (underweight) or high (overweight) levels. To facilitate interpretations, age, BMI, and physical activity were standardized prior to the analyses. More precisely, we relied on a hybrid MIMIC multiple-group approach in which a separate MIMIC model is estimated within each gender group, starting from the most invariant measurement model (Marsh, Nagengast et al., 2013).

These models were estimated in the following sequence (Morin et al., 2013): (i) a null effect model in which the paths from the predictors to the latent factors and item responses were constrained to be zero; (ii) a factors-only model in which the paths from the predictors to the latent factors, but not the item responses, were freely estimated; (iii) a saturated model in which the paths from the predictors to the item responses, but not the factors, were freely estimated. In a first step, these models were estimated while allowing all associations to be freely estimated (or equally constrained to be zero) in all gender groups. Then, the retained model was contrasted to an alternative model in which these associations where constrained to be equal across gender. The goodness-of fit of these models was then compared. The observation of an improved level of fit associated with the factors-only and saturated models when compared to the null effects model supports the presence of relations between the predictors and PSI-S ratings, whereas the observation of an improved level of fit associated with the saturated model when compared to the factors-only model supports the presence of DIF.

Given the known oversensitivity of the chi-square test of exact fit to sample size and minor model misspecifications (Marsh, Hau, & Grayson, 2005), we relied on goodness-of-fit indices to assess the fit of the alternative models: the comparative fit index (CFI), the Tucker-Lewis index (TLI), the root

[2] Additional models including curvilinear effects of age or physical activity, as well as models including two- or three-way interactions effects between age, BMI, BMI^2, and physical activity were tested and compared with models in which these interactions or curvilinear effects were constrained to be zero. Goodness-of-fit comparisons, as well as non-significant scaled chi-square difference tests between these models support the lack of interactions or curvilinear effects for these variables.

mean square error of approximation (RMSEA) with its 90% confidence interval. Values greater than .90 and .95 for the CFI and TLI respectively indicate adequate and excellent model fit, while values smaller than .08 or .06 for the RMSEA respectively support acceptable and excellent model fit (e.g., Hu & Bentler, 1999; Marsh, Hau, & Wen, 2004). In the comparison of nested models (e.g., measurement invariance, MIMIC), models differing from one another by less than .01 on the CFI and TLI, or .015 on the RMSEA, can be considered to provide an equivalent level of fit to the data (e.g., Chen, 2007; Cheung & Rensvold, 2002).

Results

Factor Structure of Ratings on the Original and Revised PSI Versions

The results from the time-specific and longitudinal measurement models were nearly identical in terms of goodness-of-fit comparisons conducted between alternative representations and PSI-S versions, as well as in terms of parameter estimates. For the sake of parsimony, we only report the results from the longitudinal measurement models. Focusing on longitudinally invariant parameter estimates helps to maximize the available information (with two sets of observations per participants) in order to achieve maximal stability and generalizability of the parameter estimates. Comparable results from the time-specific measurement models can be found in Table S2 of the online supplements. The goodness-of-fit of the longitudinal ESEM and CFA measurement models of ratings on the original and revised PSI-S versions are reported in Table 1. A first noteworthy observation is that all of the measurement models (CFA and ESEM estimated for the original and revised PSI-S) proved to be fully invariant (loadings, intercepts, uniquenesses, latent variances and covariances, and latent means) across measurement points, with none of the changes in goodness-of-fit indices exceeding the suggested guidelines (ΔCFI/ΔTLI < .010; ΔRMSEA < .015). These results confirm the longitudinal equivalence of PSI-S ratings, and the possibility to use this instrument in the context of longitudinal comparisons.

Focusing on the most invariant of these measurement models (latent mean invariance), the ESEM and CFA solutions of scores on both PSI-S versions achieved a satisfactory fit to the data. However, the ESEM solutions systematically presented a substantially higher level of fit to the data for scores on both PSI-S versions, according to the ΔCFI (original: +.064; revised: +.055), ΔTLI (original: +.067;

revised: +.057), and ΔRMSEA (original: -.022; revised: -.022). Because the original and revised PSI-S versions are not based on the same items, their goodness-of-fit indices cannot be directly compared. For this reason, the relative adequacy of both versions needs to be determined based on a detailed examination of parameter estimates. Still, it is noteworthy that the CFA results generally tended to suggest the superiority of the revised PSI-S version, whereas the ESEM results converge on similar level of fit to the data for both PSI-S versions. The detailed parameter estimates of the fully invariant CFA and ESEM solutions are presented in Table 2 (original version) and Table 3 (revised version).

The CFA results revealed generally well-defined factors for scores on both the original (λ = .281 to .855; M_λ = .679) and revised versions (λ = .633 to .855; M_λ = .761) of PSI-S. However, CFA results obtained with the original version confirmed the previously noted suboptimal performance of the negatively-worded items (GSW2: .290; PA1: .317; PA3: .281). In contrast, the reformulated version of these items included in the revised PSI-S presented no such problems (λ =.652 to .855). As a result, CFA-based composite reliability coefficients associated with ratings on the GSE and PA scales tended to be higher for the revised (GSW: ω = .770; PA: ω = .815) relative to the original PSI-S (GSW: ω = .574; PA: ω = .458). In contrast, the CFA-based composite reliability coefficients of ratings on the remaining subscales were satisfactory and equivalent for both versions (ω = .789 to .851, M_ω = .814).

The ESEM results also showed relatively weak target loadings associated with the negatively-worded items of the original PSI-S, and further revealed problematic GSW and PA factors. Indeed, rather than the a priori PA factor, ESEM suggested the presence of a negatively-worded item factor, characterized by higher target loadings for the first negatively-worded PA item (PA1: .550), a high cross-loading for the negatively-worded GSW item (GSW2: .567), and a moderately weak loading for the remaining negatively worded PA item (PA3: .332). In contrast, the GSW factor appears to be mainly defined by a single item (GSW1: .828). Although ESEM results associated with ratings on the revised PSI-S revealed additional concerns (to be discussed shortly), they revealed more adequately-defined GSW and PA factors. In accordance with the CFA results, ESEM-based composite reliability coefficients associated with ratings on the GSW and PA scales tended to be higher for the revised PSI-S (GSW: ω =.591; PA: ω = .804) than for the original PSI-S (GSW: ω = .486; PA: ω = .403). For ratings on the other subscales, ESEM-based composite reliability coefficients were satisfactory for the revised (ω = .736 to .807, M_{ω} = .770) and original PSI-S (ω = .755 to .789, M_{ω} = .769).

These results thus confirm Morin and Maïano's (2011a) observations regarding the suboptimal performance of the negatively-worded items included in the original PSI-S and the challenges posed by negatively-worded items in the context of cross-cultural adaptations (e.g., Aşçı et al., 2009; Watkins & Cheung, 1995). Still, following Morin and Maïano (2011a; also see Marsh, Scalas et al., 2010), we verified whether the suboptimal performance of these negatively-worded items could simply be related to the presence of an unmodeled method factor associated with these items. The results from CFA and ESEM models of the original PSI-S including such a method factor are reported in Tables S3 and S4 of the online supplements. These results showed that the addition of a method factor, although associated with a slight increase in model fit, was not sufficient to explain the poor performance of these items. Taken together, these results support the superiority of the revised PSI-S, when compared to the original PSI-S. The revised PSI-S was thus retained for all further analyses.

The revised PSI-S ESEM solution resulted in a substantial increase in model fit relative to the CFA solution, supporting Morin and Maïano's (2011a) conclusion regarding the superiority of the ESEM solution. However, a detailed examination of parameter estimates is critical to the decision to select an ESEM versus a CFA solution (Morin et al., 2013, 2016). So far, we have presented evidence showing that both the CFA and ESEM solutions resulted in the estimation of strongly defined factors and satisfactory composite reliability estimates. Statistical simulation studies and studies of simulated data (for a review, see Asparouhov et al., 2015) have shown ESEM to result in more accurate estimates of factor correlations whenever cross-loadings are present in the population model, whereas they remain unbiased otherwise. The observation of reduced factor correlations associated with ESEM, relative to CFA, would thus represent strong evidence in favor of the ESEM solution. In the present study, the revised PSI-S factor correlations proved to be much lower in the ESEM (r = .310 to .710; M_r = .505) relative to the CFA (r = .570 to .920; M_r = .734) solution. This observation supports the superiority of the ESEM solution, which was retained for further analyses. It is noteworthy that the 7-8 months test-retest latent correlations associated with scores on this version are also fully satisfactory for all PSI-S factors (r = .528 to .751; M_r = .665).

As noted above, the revised PSI-S ESEM solution still revealed some concerning observations.

First, as already noted by Morin and Maïano (2011a), the first item from the GSW subscale (GSW1: *I have a good opinion of myself*) appeared to contribute mainly to the definition of the PSW factor (.554), rather than of its own factor (.088). This strong cross-loading apparently created some confusion in the definition of the PSW factor as illustrated by the fact that the third PSW item (PSW3: *I'm confident about my physical self-worth*) appeared to contribute as much to the definition of PSW (.268) as of GSW (.281) and even PS (.273). These observations are consistent with the wording of these items and with the hierarchically-ordered nature of these three subscales suggesting that more specific items may still contribute to the definition of more global constructs. Still, they suggest that the first GSW item may need to be targeted for re-assessment, and possibly reformulated to tap even more clearly into global non-physical self-perceptions. Interestingly, this problem only appeared once the difficulties posed by negatively-worded items had been solved.

Measurement Invariance across Linguistic Versions

Starting with the retained ESEM representation of ratings on the revised PSI-S estimated based on Time 1 responses, we conducted systematic tests of measurement invariance across the English and French linguistic samples to assess the equivalence of the newly developed English linguistic version of the PSI-S with the more established French linguistic version. These results are reported in the top section of Table 4, and support the adequacy of the measurement model across samples (CFI/TLI $\geq$ .95; RMSEA $\leq$ .06), as well as its complete invariance (loadings, intercepts, uniquenesses, latent variances and covariances, latent means) across samples (ΔCFI/ΔTLI < .010; ΔRMSEA < .015). These results confirm the cross-linguistic equivalence of ratings on the revised PSI-S, and the possibility of using this instrument for cross-cultural or cross-linguistic comparisons.

DIF and Latent Mean Differences: Gender, Age, BMI, and Physical Activity

Prior to testing the relations between the continuous predictors and the PSI-S factors, as well as their equivalence across gender, we estimated a multi-group model to assess the measurement invariance of the PSI-S ratings across genders. These results are reported in the middle of Table 4 and support the adequacy of the measurement model (CFI/TLI $\geq$ .95; RMSEA $\leq$ .06), as well as the invariance of the loadings, intercepts, uniquenesses, and latent variances and covariances across genders (ΔCFI/TLI < .010; ΔRMSEA < .015). Although the changes in fit indices associated with the invariance of the latent means remained under the recommended guidelines, these changes remained higher for both the CFI (-.008) and TLI (-.009) than those observed so far in this study. Thus, based on our theoretical expectation that gender differences in mean levels should be present on the PSI-S factors, as well as on statistical evidence suggesting that typical guidelines may not be stringent enough for tests of latent mean differences (Fan & Sivo, 2009), we decided to interpret latent mean differences obtained from the model of latent variance and covariance invariance. In multi-group models, latent means are constrained to be zero in a referent group for identification purposes, so that latent means can be freely estimated in the other groups (e.g., Morin et al., 2013). These freely estimated latent means provide a direct estimation of the magnitude of the difference between the target group and the referent group, expressed in SD units, and are accompanied by tests of the statistical significance of these differences. In the current study, when girls' latent means were constrained to be zero for identification purposes, boys' latent means proved to be significantly higher ($p \leq 0.01$) on all PSI-S factors (GSW: .220; PSW: .291; PA: .479; PS: .510; PC: .552; SC: .418).

Starting with the final model of latent variance-covariance invariance across genders, predictors (age, BMI, BMI^2, physical activity) were directly integrated to the model through a MIMIC approach. The results from these models are reported in the bottom section of Table 4. First, the null effects model (forcing all relations between the predictors and the PSI-S responses and latent factors to be zero) resulted in an acceptable level of fit (CFI and TLI $\geq$ .90; RMSEA $\leq$ .06). When compared to the null effects model, the saturated model (allowing for the free estimation of the effects of the predictors on the PSI-S item responses freely in both gender groups) resulted in a substantial improvement in model fit (ΔCFI/TLI $\geq$ .010; ΔRMSEA $\geq$ .015), supporting the idea that the predictors have an effect on PSI-S responses. However, this saturated model resulted in an almost identical level fit to the data (ΔCFI/TLI $\leq$ .01; ΔRMSEA $\leq$ .015) than the far more parsimonious factors-only model (allowing the free estimation of the relations between the predictors and the PSI-S latent factors, but not item responses, freely in both gender groups). This finding supported the idea that the relations between the predictors and the PSI-S responses can be fully explained by their effects on the latent factors. Finally, starting with the factors-only model, relations between the predictors and the PSI-S factors were constrained to be equal across genders. This final model and the model in which these relations were freely estimated in all samples produced almost identical fit to the data (ΔCFI/TLI ≤ .01; ΔRMSEA ≤ .015). This finding supports the generalizability of the relations between age, BMI, BMI^2, and physical activity and PSI-S ratings across genders.

The results from this final model are reported in Table 5. These results first reveal a lack of effects of age on any of the PSI-S factors. The results also demonstrate significant positive associations between involvement in physical activity and adolescents' physical self-concept on all of the physical dimensions of the PSI-S (PSW, PA, PS, PC, SC) but not on the GSW scale, which is not specific to the physical area. This observation supports the discriminant validity of scores on the GSW scale, despite the weaker performance of item GSW1. Finally, the results revealed significant curvilinear effects of BMI on most physical self-concept dimensions (GSW, PSW, PA, PC, SC), with the sole exception of PS or which the relation with BMI was linear and positive in accordance with prior research results (Griffiths et al., 2010; Hau et al., 2005; Marsh et al., 2007; Morin & Maïano, 2011a; Sung et al., 2005). To facilitate interpretations, these relations are graphically presented in Figures 1 (girls) and 2 (boys). Although these relations are fully invariant across gender, the higher levels of physical self-perceptions observed in boys lead to intercept differences for these regressions, leading to the need to graphically present these effects separately for each gender. Overall, these relations show that levels of GSW, PC, and PSW tend to be equally positive for low (i.e. underweight) to average (i.e., normal weight) BMI levels, and to decrease almost linearly as BMI levels increase from average to high (i.e., overweight and obesity). In contrast, levels of PA and SC appear to be highest at average BMI levels (i.e., normal weight), slightly lower at low BMI levels (i.e. underweight), and to substantially decrease as BMI levels increase from average to high (i.e., overweight and obesity). Finally, PS levels increase linearly as a function of BMI levels.

Discussion

Researchers have noted that the PSI-S shows great promise as a short, yet comprehensive, measure of multidimensional physical self-conceptions (Maïano et al., 2008, 2009, 2011; Morin & Maïano, 2011a). Critical examination of the PSI-S, however, led to the identification of a series of challenges to the more widespread use of this instrument, including the high correlations reported between the PSI-S factors, the questionable performance of the negatively-worded items, and the need to develop and validate an English version of this instrument. This study addressed these challenges.

The first challenge is not specific to the PSI-S but characterizes most PSPP-based instruments, and stems from the observation of factor correlations that are high enough to call into question the distinct nature of the subscales forming these instruments (Marsh & Cheng, 2012). This challenge was first addressed by Morin and Maïano (2011a) through the application of ESEM to PSI-S responses, following reports showing that ESEM tends to provide more accurate estimates of factor correlations in instruments tapping into conceptually-related constructs (e.g., Asparouhov et al., 2015; Marsh et al., 2014; Morin et al., 2013, 2016). Morin and Maïano's (2011a) ESEM results provided strong construct validity evidence based on the factor structure of the PSI-S ratings, which supported the distinct nature of the PSI-S subscales. However, their results also called into question the adequacy of the three negatively-worded items included in the PSI-S. Coupled with the fact that negatively-worded items pose particular challenges in cross-cultural research (Aşçı et al., 2009; Watkins & Cheung, 1995), Morin and Maïano (2011a) proposed a positively-worded formulation of these items.

The results from the present study replicated Morin and Maïano's (2011a) results with the new English version of the PSI-S. More precisely, our results showed that the ESEM solution, when compared to the CFA solution, provided a substantially better fit to the data and yielded drastically reduced estimates of the latent factor correlations (r = .310 to .710; M_r = .505). This finding supports the distinct nature of the PSI-S subscales, and confirmed the questionable performance of the negatively-worded items. Similarly, the results supported the superiority of a revised PSI-S in which the negatively-worded items were replaced by positively-worded alternatives. This ESEM representation of responses to the revised PSI-S also proved completely invariant across a 7-8 month period, and resulted in satisfactory estimates of composite (M_ω = .770) and test-retest (M_r = .665) reliability – particularly considering that these factors are defined by sets of three items. These results are particularly important regarding the longitudinal stability and test-retest reliability of the PSI-S responses. Up until now, evidence of test-retest reliability of the PSI-S responses had only been demonstrated among French speaking adolescents across a two-week interval (Maïano et al., 2008).

In addition to the identification of difficulties involving the negatively-worded items, Morin and Maïano's (2011a) results also alluded to possible difficulties involving the first item from the GSW subscale (GSW1: *I have a good opinion of myself*), which might have been caused by the interference posed by the negatively-worded items. However, our results also revealed problems associated with this item (showing conceptual overlap between this item and the PSW factor), which were made worse once the problems associated with the negatively-worded items were solved in the revised PSI-S. Although this result may possibly reflect simple sampling variation, we recommend that caution should be exerted when using the English GSW scale of the PSI-S. We believe that it might be too early to recommend definitive changes to this subscale, and suggest that future research should still maintain the integrity of the PSI-S version. Given that prior research has shown that the assessment of GSW can still be performed reliably through the use of a two-item subscale solely formed of items GSW1 and GSW3 (Maïano et al., 2008; Morin & Maïano, 2011b), we believe that the risks associated with maintaining the scale integrity are minimal. Given the challenges posed by the assessment of two-item scales (see Morin & Maïano, 2011b), our recommendation is thus to retain the full set of PSI-S items, unless there is a need for a very-short version of the instrument. Still, we propose possible reformulations of the problematic item GSW1 (GSW1R: *Overall, I have a good opinion of myself* or *Overall, I like myself*) which could be added to future PSI-S studies and assessed as possible replacements for the current problematic PSI-S item. A key advantage of the first of these reformulations is an alternative adaptation for the French item, which has found to perform relatively well (Maïano et al., 2008; Morin & Maïano, 2011a, 2011b), meaning that no change is required to the original version. For this reason, we propose that the first of those should be tested initially, and followed by the more generic second formulation should the first fail to solve the problems identified in this study. We reinforce that item reformulations should not be proposed lightly, but rather based on a systematic assessment of item content. The reformulations proposed here were formulated by a panel of bilingual experts of self-concept measurement with the objective of preserving the integrity of the scale content, and ideally the linguistic equivalence with the original French version. Alternatively, researchers could also replace the complete GSW subscale with the comparable, yet longer, 5-item subscale from the PDQ-S, which remains relatively short and presents a viable, and cross-culturally valid, alternative (Maïano et al., 2015; Marsh et al., 2010).

The third challenge facing the PSI-S is related to the need to move beyond a French version to facilitate the use of this instrument in non-French speaking populations (Marsh & Cheng, 2012). In the present study, we developed an English version of the PSI-S, and tested whether scores on this version retained the psychometric properties of the French version. The present results supported the complete measurement invariance (i.e., loadings, thresholds, uniquenesses, latent variances and covariances, latent means) of scores on the revised PSI-S across samples of French- and English- speaking participants, supporting the possibility to use this instrument in the context of cross cultural or cross linguistic comparisons involving samples of French- and English- speaking adolescents.

To further test the extent to which the linguistic adaptation of the revised PSI-S would preserve the properties of the original French PSI-S, we followed Morin and Maïano (2011a) and investigated the effects of gender, age, and BMI on PSI-S responses. We also extended Morin and Maïano's (2011a) analyses to see whether the current results would replicate the well-established positive relation between physical activity involvement and physical self-perceptions (Babic et al., 2014; Bowker, 2006; Findlay & Bowker, 2007; Schmalz & Davison, 2006). However, Morin and Maïano (2011a) relied on a rough categorization of continuous age and BMI information, which is likely to have reduced the power of the analyses to detect meaningful latent mean differences (MacCallum et al., 2002). The multiple-group MIMIC approach used here allowed us to test for possible measurement biases (DIF) in item responses as a function of these covariates as well as for latent mean differences, while providing a test of the extent to which the results generalized across gender.

Our results revealed that PSI-S responses presented no bias (DIF) in relation to gender, age, BMI or physical activity involvement. Furthermore, our results generally supported prior research in showing that physical and global self-perceptions tended to be higher among males relative to females (Aşçı, 2002; Hagger et al., 2005; Marsh et al., 2006, 2007; Morin & Maïano, 2011a), and that physical (but not global) self-perceptions also tended to be higher among participants more frequently involved in physical activity (Babic et al., 2014; Bowker, 2006; Findlay & Bowker, 2007; Schmalz & Davison, 2006). Supporting Morin and Maïano's (2011a) results, but contrasting with prior research conducted in this area (e.g., Hagger et al., 2005; Marsh et al., 2007), the present findings failed to reveal significant relations between age and physical self-perceptions – although this result is most likely explained by the very limited age range considered in the present study (12-14 years) given that age-related differences should become more apparent as youth move beyond puberty.

Finally, our results also support prior research (Griffiths et al., 2010; Hau et al., 2005; Marsh et al., 2007; Morin & Maïano, 2011a; Sung et al., 2005) of significant associations between BMI levels and physical self-conceptions. Most of these relations proved to be curvilinear, showing that higher levels of BMI (overweight and obesity) tend to be associated with much lower physical self-perceptions than moderate levels of BMI (normal weight), while low levels of BMI (underweight) only tended to be associated with lower levels of PA and SC. The only exception to these curvilinear patterns also supports prior research and shows that PS presents positive linear associations with BMI levels. This result is consistent with the observation that high levels of BMI might not only be a function of body fat, but also a function of stronger levels of muscular or bone structure (Morin & Maïano, 2011a), suggesting that future research would do well to investigate the relation between physical self-conceptions (especially PS) and objective measures of body fat and physical fitness. All of these relations between age, BMI and physical activity, and the PSI-S factors proved to be fully equivalent across genders, attesting to the generalizability of the observed relations. However, this equivalence of these relations across genders is also slightly counter-intuitive, given that physical standards are known to differ as a function of gender-differentiated sociocultural norms and socialization processes (e.g., Klomsten et al., 2004; Smith et al., 1999; McCabe, & Ricciardelli, 2003). Still, these differences are also assumed to emerge during puberty, which results in undesired body fat accumulation in girls versus a more desirable muscular increase in boys (Morin, et al., 2011; Siegel et al., 1999; Stice & Bearman, 2001). In the current study, the ability to identify gender-differentiated patterns of relations between these covariates and the PSI-S factors might also be potentially explained by the reduced age-range considered. Future research covering a greater age range should devote more attention to the possible emergence of gender-differentiated differences across adolescence.

Some limitations must be taken into account. For instance, the current study relies on two convenient samples of normally achieving French- and English- speaking adolescents, which cannot be considered to be representative of the targeted populations or equivalent across linguistic groups. Although it remains possible that the smaller sample size of French participants might have reduced our ability to detect significant differences in item responding across the two linguistic versions, we believe this is unlikely given that sample size remained fully satisfactory in both groups. Still, future research is needed to establish the conditions in which scores on these linguistic versions will preserve their psychometric properties. Furthermore, the use of this instrument should for the moment be limited to normally achieving adolescents from the targeted linguistic groups from cultural backgrounds similar to that of the current participants. The next step in evaluating the robustness and generalizability of the PSI-S responses should be to test its linguistic adaptation to additional linguistic groups (e.g., Spanish, Chinese, German) in order to maximize its generalizability. Another limitation to the generalizability of the current results is related to the fact that participants were tested in the context of typical classroom settings, a context which may have reduced the salience of their physical self-concept. In particular, an interesting direction for future research would be to test the ability of the PSI-S to provide satisfactory ratings of physical self-conceptions among elite athletes. To our knowledge, the only measure of physical self-conception to have been validated to date in this population is the PSDQ (Marsh, 1998). In addition, although we provided some evidence of the convergent validity of PSI-S scores in relation to age, gender, BMI, and physical activity, additional tests are needed in relation to other physical self-concept instruments, and a variety of external criterions (physical fitness, body fat, body image disturbances, etc.).

Despite these limitations, the current study provides strong evidence in support of the psychometric properties of ratings on the English version of PSI-S. This instrument shows much promise a measure of physical self-concept that is short enough to be included in multi-section questionnaire designed to test a range of constructs.

References

Arens, A. K., & Morin, A. J. S. (2016). Examination of the structure and grade-related differentiation of multidimensional self-concept instruments for children using ESEM. *Journal of Experimental Education*, *84,* 333-355. doi: 10.1080/00220973.2014.999187

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. doi: 10.1016/j.psychsport.2003.10.003

Aşçı, F. H., Fletcher, R., & Çağlar, E. (2009). Differential item functioning analysis of the PSDQ with Turkish and New Zealand/Australian adolescents. *Psychology of Sport & Exercise*, *10,* 12-18. doi: 10.1016/j.psychsport.2008.05.001

Asparouhov, T. (2005). Sampling weights in latent variable modeling. *Structural Equation Modeling*, *12,* 411-434. doi: 10.1207/ s15328007sem1203_4

Asparouhov, T., & Muthén, B.O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397-438. doi: 10.1080/10705510903008204

Asparouhov, T., Muthén, B. O., & Morin, A. J. S. (2015). Bayesian Structural equation modeling with cross-loadings and residual covariances. *Journal of Management*, *41*, 1561-1577. doi: 10.1177/0149206315591075

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.

Babic, M. J., Morgan, P. J., Plotnikoff, R. C., Lonsdale, C., White, R. L., & Lubans, D. R. (2014). Physical activity and physical self-concept in youth: Systematic review and meta-analysis. *Sports Medicine, 44*, 1589-1601. doi: 10.1007/s40279-014-0229-z

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. doi: 10.1080/02673843.1993.9747728

Booth, M. L., Okely, A. D., Chey, T., & Bauman, A. (2001). The reliability and validity of the physical activity questions in the WHO health behaviour in schoolchildren (HBSC) survey: A population study. *British Journal of Sports Medicine, 35*, 263-267. doi: 10.1136/bjsm.35.4.263

Bowker, A. (2006). The relationship between sports participation and self-esteem during early adolescence. *Canadian Journal of Behavioural Science*, 38, 214-229. doi: 10.1037/cjbs2006009

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111–150. doi: 10.1207/ S15327906MBR3601_05

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*, 464–504. doi:10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*, 233–255. doi: 10.1207/S15328007SEM0902_5

Cole, T. J., & Lobstein, T. (2012). Extended international (IOTF) body mass index cut-offs for thinness, overweight and obesity. *Pediatric obesity*, *7*, 284-294. doi:10.1111/j.2047-6310.2012.00064.x

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

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects with negatively worded items on self-report surveys. *Structural Equation Modeling*, *13*, 440-464. doi: 10.1207/s15328007sem1303_6

Eiser, C., Eiser, J. R., & Havermans, T. (1995). The measurement of self-esteem: Practical and theoretical considerations. *Personality & Individual Differences*, *18*, 429-432. doi: 10.1016/0191-8869(94)00179-V

Eklund, R. C., Whitehead, J., & Welk, G. (1997). Validity of the children physical self-perception profile: A confirmatory factor analysis. *Research Quarterly for Exercise & Sport*, *68*, 249-256. doi: 10.1080/02701367.1997.10608004

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

Fan, X., & Sivo, S. A. (2009). Using goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling*, *16*, 54–69. doi: 10.1080/10705510802561311

Findlay, L. C., & Bowker, A. (2007). The link between competitive sport participation and self-concept in early adolescence: A consideration of gender and sport orientation. *Journal of Youth and Adolescence*, *38*, 29-40. doi: 10.1007/s10964-007-9244-9

Flegal, K. M., & Cole, T. J. (2013). Construction of LMS parameters for the centers for disease control and prevention 2000 growth charts. *National Health Statistics Report*, *63*, 1-4.

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. (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. doi: 10.1123/jsep.11.4.408

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. doi: 10.3109/17477160903473697

Guay, F., Morin, A. J .S , Litalien, D., Valois, P., & Vallerand, R. J. (2015). Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of Experimental Education*, *83*, 51-82. doi: 10.1080/00220973.2013.876231

Gudmundsson, E. (2009). Guidelines for translating and adapting psychological instruments. *Nordic Psychology, 61*, 29-45. doi: 10.1027/1901-2276.61.2.29

Hagger, M. S., Aşçı, F. H., & Lindwall, M. (2004). A cross-cultural evaluation of a multidimensional and hierarchical model of physical self-perceptions in three national samples. *Journal of Applied Social Psychology*, *34*, 1075-1107. doi: 10.1111/j.1559-1816.2004.tb02584.x

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. doi: 10.1080/1740898970020208

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. doi: 10.1177/0013164404272484

Hambleton, R. K. (2005). Issues, designs, and technical guidelines for adapting tests to languages and cultures. In R. K. Hambleton, P. Merenda, & C. Spielberger (Eds.), *Adapting educational and psychological tests for cross-cultural assessment* (pp. 3-38). Mahwah, NJ: Erlbaum.

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 Marsh, H. W., Craven, R. G., & McInerney, D. M., (Eds.), *The new frontiers of self research* (pp. 259-272). Charlotte, NC: Information Age.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria vs new alternatives. *Structural Equation Modeling, 6,* 1-55. doi: 10.1080/10705519909540118

Klomsten, A. T., Shaalvik, E. M., & Espnes, G. A. (2004). Physical self-concept and sports: Do gender differences still exist? *Sex Roles, 50,* 119–127. doi: 10.1023/B:SERS.0000011077.10040.9a

Lindwall, M., Aşçı, F. H., Palmeira, A., Fox, K. R., & Hagger, M. S. (2011). The importance of importance in the physical self: Support for the theoretically appealing but empirically elusive model of James. *Journal of Personality, 79,* 303-333. doi: 10.1111/j.1467-6494.2010.00678.x

Lindwall, M., Aşçı, F. H., & Hagger, M., (2011). Factorial validity and measurement invariance of the Revised Physical Self-Perception Profile (PSPP-R) in three countries. *Psychology, Health & Medicine, 16*, 115-128. doi: 10.1080/13548506.2010.521567

Lonsdale, C., Lester, A., Owen, K. B., White, R. L., Moyes, I., Peralta, L., . . . Lubans, D. R. (2016). An internet-supported physical activity intervention delivered in secondary schools located in low socio-economic status communities: Study protocol for the activity and motivation in physical education (AMPED) cluster randomized controlled trial. *BMC Public Health, 16*, 1-15. doi: 10.1186/s12889-015-2583-7

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. doi: 10.1037//1082-989X.7.1.19

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. doi: 10.1007/s10803-008-0686-z

Maïano C., Morin A. J. S., Bégarie J., & Ninot G. (2011). The intellectual disability version of the very short form of the physical self-inventory (PSI-VS-ID): Cross-validation and measurement invariance across gender, weight, age and intellectual disability level. *Research in Developmental Disabilities*, 32, 1652-1662. doi: 10.1016/j.ridd.2011.02.019

Maïano, C., Morin, A. J. S., & Mascret, N. (2015). Psychometric properties of the short form of the physical self-description questionnaire in a French adolescent sample. *Body Image*, *12*, 89-97. doi: 10.1016/j.bodyim.2014.10.005

Maïano, C., Morin, A. J. S., Ninot, G., Monthuy-Blanc, J., Stephan, Y., Florent, J.-.F., & Vallée, P. (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. doi: 10.1016/j.psychsport.2007.10.003

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., Abduljabbar, A. S., Abu-Hilal, M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker, P. (2013). Factorial, Convergent, and Discriminant Validity of TIMSS Math and Science Motivation Measures: A Comparison of Arab and Anglo-Saxon Countries. *Journal of Educational Psychology*, *105*, 108-128. doi: 10.1037/a0029907

Marsh, H. W., Aşçı, F.H., & Marco, I. T. (2002). Multitrait-multimethod analyses of two physical self-concept: A cross-cultural perspective. *Journal of Sport & Exercise Psychology*, *24*, 99-119. doi: 10.1123/jsep.24.2.99

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. doi: 10.1123/jsep.28.3.310

Marsh, H. W., & Cheng, J. H. S. (2012). Physical self-concept. In G. Tenenbaum, R. Eklund, & A. Kamata (Eds), *Handbook of measurement in sport and exercise psychology* (pp. 215–226). Champaign, IL: Human Kinetics.

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation in structural equation modeling. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics. A*

Festschrift for Roderick P. McDonald. Mahwah, 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. doi: 10.1037/0012-1649.43.3.647

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999). *Structural Equation Modeling, 11*, 320-341. doi: 10.1207/s15328007sem1103_2

Marsh, H. W., Martin, A. J., & Jackson, S. (2010). Introducing a short version of the physical self-description questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport & Exercise Psychology, 32*, 438-482. doi: 10.1123/jsep.32.4.438

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of Big-Five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and La Dolce Vita effects. *Developmental Psychology, 49*, 1194-1218. doi: 10.1037/a0026913

Marsh, H. W., Morin, A. J. S., Parker, P. D., & Kaur, G. (2014). Exploratory structural equation modelling: An integration of the best features of exploratory and confirmatory factor analyses. *Annual Review of Clinical Psychology, 10*, 85-110. doi: 10.1146/annurev-clinpsy-032813-153700

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. doi: 10.1111/j.1745-3984.1984.tb00227.x

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. doi: 10.1037/a0019225

Marsh, H. W., Richards, G., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical self-description questionnaire: Psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport & Exercise Psychology, 16*, 270-305. doi: 10.1123/jsep.16.3.270

McCabe, M. P., & Ricciardelli, L. A. (2003). Sociocultural influences on body image and body changes among adolescent boys and girls. *Journal of Social Psychology*, *143*, 5-26. doi: 10.1080/00224540309598428

McDonald, R. P. (1970) Theoretical foundations of principal factor analysis and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*, 1-21. doi: 10.1111/j.2044-8317.1970.tb00432.x

Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York, NY: Taylor & Francis.

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., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, *23*, 116-139. doi: 10.1080/10705511.2014.961800

Morin, A. J. S., & Maïano, C. (2011a). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport & Exercise*, *12*, 540-554. doi: 10.1016/j.psychsport.2011.04.003

Morin A. J. S., & Maïano C. (2011b). Cross-validation of the Very Short Form of the Physical Self-Inventory (PSI-VS): Invariance across genders, age groups, ethnicities, and weight statuses. *Body Image*, *8*, 404-410. doi: 10.1016/j.bodyim.2011.06.005

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. doi: 10.1080/00273171.2010.546731

Morin, A. J. S., Marsh, H.W., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. R. Hancock, & R. O. Mueller (Eds.). *Structural equation modeling: A second course* (2nd ed., pp. 395-436). Charlotte, NC: Information Age Publishing, Inc.

Muthén, L. K., & Muthén, B. O. (2016). *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.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. New York, NY: McGraw-Hill.

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. doi: 10.1016/0191-8869(93)90151-r

Prochaska, J. J., Sallis, J. F., & Long, B. (2001). A physical activity screening measure for use with adolescents in primary care. *Archives of Pediatrics and Adolescent Medicine, 155*, 554-559. doi: 10.1001/archpedi.155.5.554

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

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*, 407-411. doi: 10.2307/1170010

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. doi: 10.1016/s1054-139x(98)00160-8

Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha [Introduction to a special issue]. *Psychometrika*, *74*, 107-120. doi: 10.1007/s11336-008-9101-0

Schmalz, D., & Davison, K. (2006). Differences in physical self-concept among pre-adolescents who participate in gender-typed and cross-gendered sports. *Journal of Sport Behavior*, *29*, 335-352.

Smith, C. J., Noll, J. A., & Bryant, J. B. (1999). The effect of social context on gender self-concept. *Sex Roles*, *40*, 499–512. doi: 10.1023/a:1018879811991

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. doi: 10.1037/0012-1649.37.5.597

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. doi: 10.1038/sj.ejcn.1602044

Tiggemann, M. (2003). Media exposure, body dissatisfaction and disordered eating: Television and magazines are not the same. *European Eating Disorders Review*, *11*, 418–430. doi:

10.1002/erv.502

Van de Vliet, P., Knapen, J., Onghena, P., Fox, K., Van Coppenolle, H., David, A., Pieters, G., & Peuskens, J. (2002). Assessment of physical self-perceptions in normal Flemish adults versus depressed psychiatric patients. *Personality & Individual Differences*, *32*, 855-863. doi: 10.1016/s0191-8869(01)00091-5

Van de Vrijver, F. J. R., & Hambleton, R. K. (1996). Translating tests: Some practical guidelines. *European Psychologist*, *1*, 89-99. doi: 10.1027/1016-9040.1.2.89

Watkins, D., & Cheung, S. (1995). Culture, gender, and response bias: An analysis of responses to the self-description questionnaire. *Journal of Cross-Cultural Psychology*, *26*, 490-504. doi: 10.1177/0022022195265003

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. doi: 10.1207/s15327752jpa6501_8

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. doi: 10.1123/pes.7.2.132

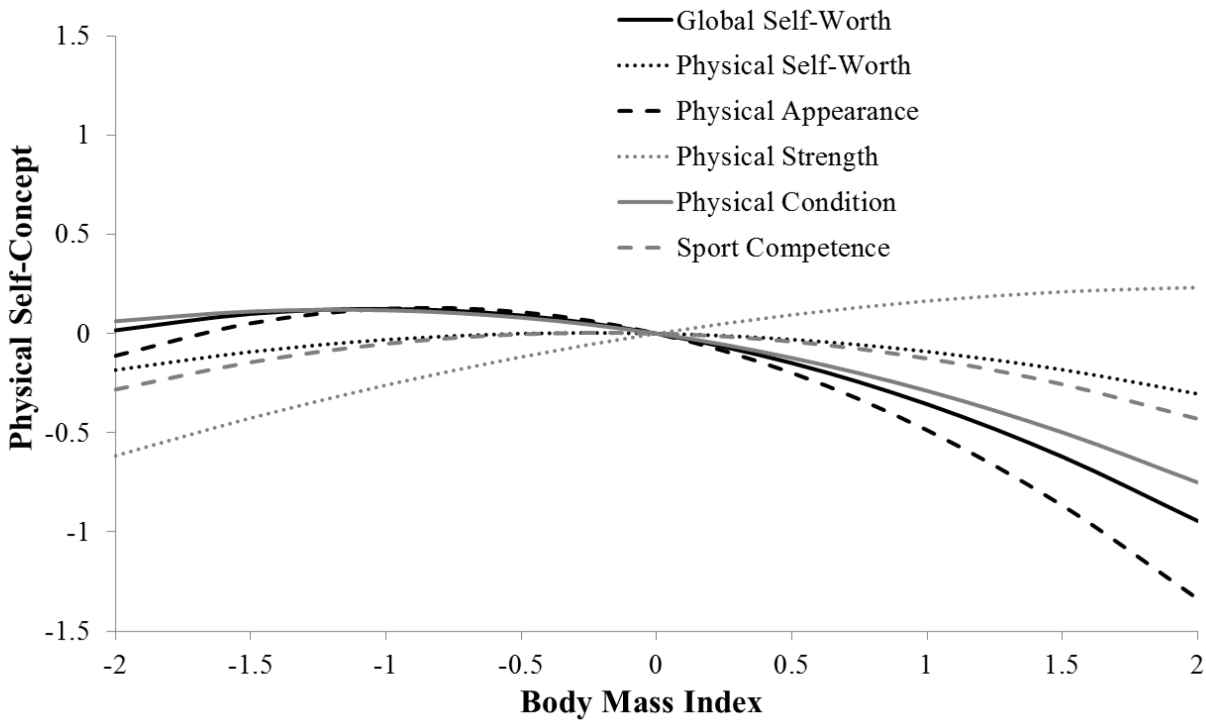

Figure 1. Relations between Normalized Body Mass Index (Z-Scores) and Girls' Physical Self-Concept

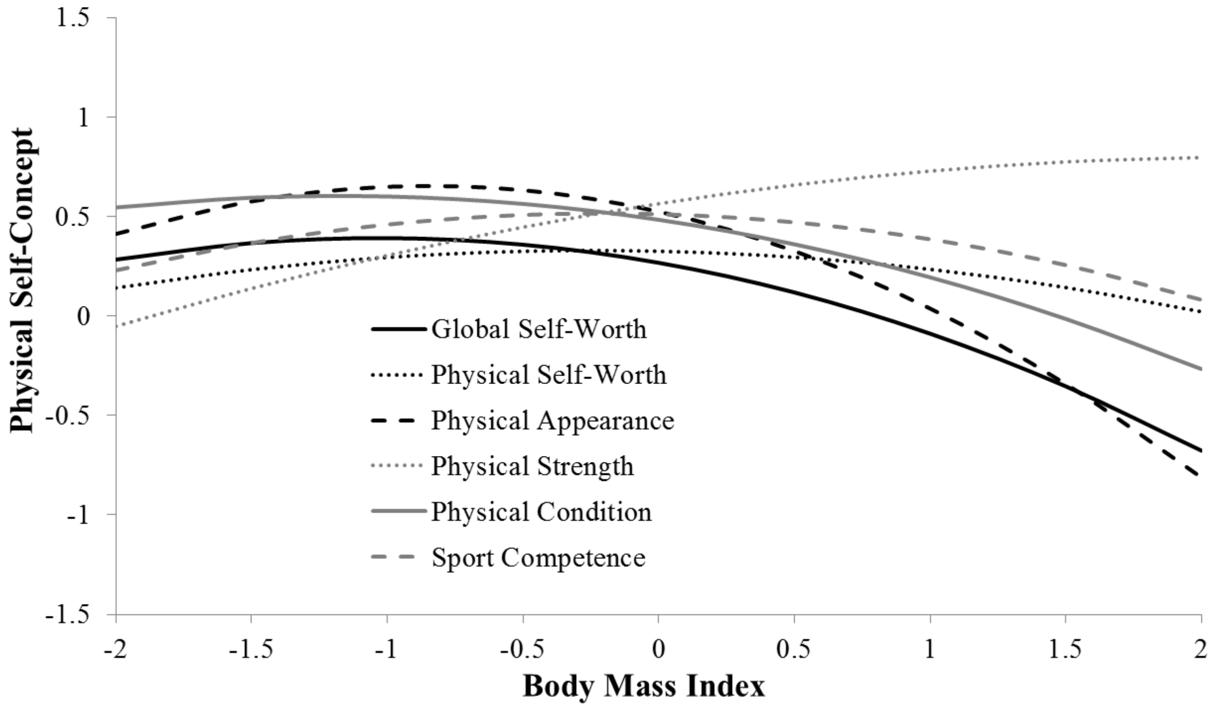

Figure 2. Relations between Normalized Body Mass Index (Z-Scores) and Boys' Physical Self-Concept

Table 1
Longitudinal Measurement Models

Model	Description	χ^2 (df)	CFI	TLI	RMSEA	90% CI	CM	$\Delta^S\chi^2$(df)	ΔCFI	ΔTLI	ΔRMSEA
CFA	1-1-Configural invariance	2217.124 (510)*	.919	.900	.049	.047-.052	–	–	–	–	–
Original version	1-2-λ invariant	2246.811 (522)*	.918	.901	.049	.047-.051	1-1	28.347 (12)*	-.001	+.001	.000
	1-3-λ, τ invariant	2268.188 (534)*	.918	.903	.049	.047-.051	1-2	21.422 (12)	.000	+.002	.000
	1-4-λ, τ, δ invariant	2339.224 (552)*	.915	.903	.049	.047-.051	1-3	71.569 (18)*	-.003	.000	.000
	1-5-λ, τ, δ, ξ/φ invariant	2375.806 (573)*	.915	.906	.048	.046-.050	1-4	30.648 (21)	.000	+.003	-.001
	1-6-λ, τ, δ, ξ/φ, η invariant	2398.194 (579)*	.914	.906	.048	.046-.050	1-5	22.342 (6)*	-.001	.000	.000
CFA	2-1-Configural invariance	2240.094 (510)*	.929	.912	.050	.048-.052	–	–	–	–	–
Revised version	2-2-λ invariant	2267.752 (522)*	.928	.913	.049	.047-.052	2-1	24.870 (12)	-.001	+.001	-.001
	2-3-λ, τ invariant	2289.666 (534)*	.928	.915	.049	.047-.051	2-2	18.244 (12)	.000	+.002	.000
	2-4-λ, τ, δ invariant	2340.381 (552)*	.926	.916	.049	.047-.051	2-3	57.115 (18)*	-.002	+.001	.000
	2-5-λ, τ, δ, ξ/φ invariant	2382.296 (573)*	.925	.918	.048	.046-.050	2-4	37.955 (21)	-.001	+.002	-.001
	2-6-λ, τ, δ, ξ/φ, η invariant	2403.597 (579)*	.925	.918	.048	.046-.050	2-5	21.562 (6)*	.000	.000	.000
ESEM	3-1-Configural invariance	724.812 (390)*	.984	.974	.025	.022-.028	–	–	–	–	–
Original version	3-2-λ invariant	832.150 (462)*	.982	.976	.024	.022-.027	3-1	108.772 (72)*	-.002	+.002	-.001
	3-3-λ, τ invariant	848.720 (474)*	.982	.976	.024	.021-.027	3-2	16.440 (12)	.000	.000	.000
	3-4-λ, τ, δ invariant	896.163 (492)*	.981	.975	.025	.022-.027	3-3	43.873 (18)*	-.001	-.001	+.001
	3-5-λ, τ, δ, ξ/φ invariant	963.839 (513)*	.979	.974	.025	.023-.028	3-4	76.267 (21)*	-.002	-.001	.000
	3-6-λ, τ, δ, ξ/φ, η invariant	988.563 (519)*	.978	.973	.026	.023-.028	3-5	21.126 (6)*	-.001	-.001	+.001
ESEM	4-1-Configural invariance	781.509 (390)*	.984	.974	.027	.024-.030	–	–	–	–	–
Revised version	4-2-λ invariant	883.231 (462)*	.983	.976	.026	.023-.028	4-1	102.695 (72)	-.001	+.002	-.001
	4-3-λ, τ invariant	898.278 (474)*	.983	.977	.026	.023-.028	4-2	13.806 (12)	.000	+.001	.000
	4-4-λ, τ, δ invariant	962.110 (492)*	.981	.975	.026	.024-.029	4-3	55.775 (18)*	-.002	-.002	.000
	4-5-λ, τ, δ, ξ/φ invariant	992.866 (513)*	.980	.976	.026	.024-.029	4-4	30.820 (21)	-.001	+.001	.000
	4-6-λ, τ, δ, ξ/φ, η invariant	1015.827 (519)*	.980	.975	.026	.024-.029	4-5	21.490 (6)*	.000	-.001	.000

Note. CFA = confirmatory factor analytic model; ESEM = exploratory structural equation modeling; χ^2 = scaled chi-square test of exact fit; 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; λ = factor loadings; τ = intercepts; δ = uniquenesses; ξ = factor variances; φ = factor covariances; η = factor means; CM = comparison model; $\Delta^S\chi^2$ = scaled chi square difference tests (calculated from models loglikelihoods 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 (Original)

	Confirmatory factor analysis							Exploratory structural equation modeling						
Standardized factor loadings (λ) and uniquenesses (δ)														
Items	GSW (λ)	PSW (λ)	PA (λ)	PS (λ)	PC (λ)	SC (λ)	δ	GSW (λ)	PSW (λ)	PA (λ)	PS (λ)	PC (λ)	SC (λ)	δ
GSW1	.727						.472	.828	*.112*	-.134	*.023*	*-.024*	.049	*.207*
GSW2	.290						.916	.141	*.004*	.567	-.096	*.018*	*.008*	.599
GSW3	.624						.611	.174	.229	.303	*-.004*	*.073*	.120	.574
PSW1		.830					.312	.320	.540	-.137	-.038	.074	.120	.268
PSW2		.800					.360	*-.077*	.816	*.008*	*.024*	.043	*.075*	.252
PSW3		.801					.359	.169	.390	.137	.252	.099	*.011*	.363
PA1			.317				.899	*.043*	*.025*	.550	-.071	*.040*	*.045*	.652
PA2			.773				.402	.217	*-.025*	.271	.259	.262	*.019*	.509
PA3			.281				.921	.135	.116	.332	*-.004*	-.154	*.060*	.804
PS1				.716			.487	*.026*	.070	*-.011*	.699	*-.018*	*-.026*	.481
PS2				.855			.270	*-.028*	*.036*	.077	.827	-.067	.091	.254
PS3				.657			.568	*.002*	*-.040*	-.055	.485	.130	.140	.565
PC1					.814		.338	*-.060*	.342	*-.054*	.091	.456	.092	.369
PC2					.779		.393	*.025*	*.038*	*.010*	*-.024*	.790	*.024*	.323
PC3					.655		.571	*.007*	-.128	*.019*	*-.025*	.795	*.050*	.419
SC1						.839	.297	*-.035*	*.042*	*-.016*	.114	.132	.619	.341
SC2						.630	.603	*.055*	.117	*-.013*	.107	.197	.262	.617
SC3						.841	.293	*.009*	*-.061*	*-.003*	*-.046*	-.101	.961	*.143*
ω	.574	.851	.458	.789	.795	.817		.486	.776	.403	.757	.789	.755	
T-R *r*	.705	.651	.813	.730	.754	.743		.575	.563	.628	.706	.757	.744	

Factor correlations (95% Confidence intervals)

Factors	PSW		PA		PS		PC		SC		PSW		PA		PS		PC		SC	
GSW	.96	(.92-.99)	.87	(.81-.94)	.59	(.54-.63)	.64	(.58-.70)	.70	(.65-.76)	.62	(.52-.72)	.35	(.21-.48)	.40	(.31-.49)	.38	(.29-.47)	.46	(.38-.55)
PSW			.73	(.66-.80)	.72	(.68-.76)	.77	(.73-.81)	.83	(.80-.87)			.24	(.13-.34)	.54	(.48-.60)	.50	(.42-.58)	.72	(.66-.78)
PA					.64	(.60-.67)	.71	(.66-.76)	.68	(.61-.74)					*.01*	(-.07-.09)	.18	(.09-.26)	.18	(.10-.26)
PS							.74	(.69-.78	.80	(.77-.82)							.63	(.59-.67)	.69	(.63-.76)
PC									.83	(.79-.86)									.68	(.60-.75)

Note. GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA= physical attractiveness; PS= physical strength; T-R *r* = test-retest stability correlation coefficient over an 8 months period; Greyscale = main loadings. Non-significant parameters ($p \geq .05$) are marked in italics.

Table 3

Standardized Parameters Estimates from the Confirmatory Factor Analytic and Exploratory Structural Equation Models of the PSI-S (Revised)

	Confirmatory factor analysis							Exploratory structural equation modeling						
Standardized factor loadings (λ) and uniquenesses (δ)														
Items	GSW (λ)	PSW (λ)	PA (λ)	PS (λ)	PC (λ)	SC (λ)	δ	GSW (λ)	PSW (λ)	PA (λ)	PS (λ)	PC (λ)	SC (λ)	δ
GSW1	.705						.503	.088	.554	.245	.078	*-.010*	*-.066*	.435
GSW2	.766						.413	.641	.138	*.033*	.083	.105	*-.009*	.325
GSW3	.706						.502	.564	*.016*	.258	*-.073*	*-.023*	.145	.397
PSW1		.812					.341	*-.046*	.831	.056	*-.030*	.101	.109	.125
PSW2		.806					.350	.241	.420	-.073	.087	.057	.265	.352
PSW3		.811					.343	.281	.268	.097	.273	.125	*.012*	.351
PA1			.855				.269	.354	*-.008*	.605	*-.032*	.046	*.020*	.250
PA2			.798				.363	-.037	*-.022*	.760	.081	.091	*.010*	.298
PA3			.652				.576	-.112	*.004*	.705	.071	*-.019*	*.052*	.499
PS1				.717			.485	*-.055*	.048	*.056*	.718	*-.007*	*-.038*	.474
PS2				.851			.275	*.003*	*-.014*	*.015*	.838	*-.016*	*.045*	.257
PS3				.660			.564	*-.041*	-.055	*.055*	.492	.087	.140	.575
PC1					.811		.342	*.026*	.171	*-.040*	.118	.454	.191	.391
PC2					.781		.391	.048	*.014*	-.063	*-.010*	.949	-.090	.239
PC3					.659		.566	-.128	-.078	.152	-.069	.725	*.050*	.456
SC1						.836	.300	-.081	*.034*	.068	.047	.071	.746	.269
SC2						.633	.599	.154	.056	*-.008*	.126	.184	.474	.599
SC3						.841	.292	.059	.083	*.003*	*.039*	*-.015*	.762	.270
ω	.770	.851	.815	.789	.796	.818		.591	.736	.804	.763	.807	.775	
T-R *r*	.669	.650	.686	.730	.755	.744		.562	.528	.693	.720	.751	.735	

Factor correlations (95% Confidence intervals)

Factors	PSW	PA	PS	PC	SC	PSW	PA	PS	PC	SC
GSW	.92 (.89-.94)	.86 (.82-.89)	.57 (.52-62)	.64 (.59-.70)	.69 (.63-.76)	.47 (.41-.53)	.54 (.47-.60)	.31 (.23-.39)	.37 (.29-.45)	.39 (.31-.48)
PSW		.72 (.68-.76)	.73 (.69-.77)	.77 (.74-.81)	.83 (.80-.87)		.43 (.40-.46)	.50 (.45-.55)	.43 (.38-.48)	.53 (.44-.61)
PA			.60 (.56-.63)	.67 (.63-.71)	.64 (.60-.68)			.51 (.43-.59)	.56 (.51-.61)	.49 (.44-.54)
PS				.74 (.70-.78)	.80 (.77-.83)				.64 (.60-.69)	.71 (.68-.74)
PC					.83 (.79-.86)					.70 (.66-.73)

Note. GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA= physical attractiveness; PS = physical strength; T-R r = test-retest stability correlation coefficient over an 8 months period; Greyscale = main loadings. Non-significant parameters ($p \geq .05$) are marked in italics.

Table 4

Tests of Measurement Invariance, Differential Item Functioning, and Latent Mean Differences for the Final ESEM Model

Model	Description	χ^2 (df)	CFI	TLI	RMSEA	90% CI	CM	$\Delta^S\chi^2$(df)	ΔCFI	ΔTLI	ΔRMSEA
ESEM	5-1-Configural invariance	387.815 (120)*	.981	.956	.053	.047-.059	–	–	–	–	–
Revised version	5-2-λ invariant	594.779 (192)*	.974	.957	.051	.047-.056	5-1	209.669 (72)*	-.007	+.001	-.002
Linguistic invariance	5-3-λ, τ invariant	743.601 (204)*	.966	.949	.058	.053-.062	5-2	67.672 (12)*	-.008	-.008	+.007
	5-4-λ, τ, δ invariant	752.786 (222)*	.966	.953	.055	.051-.059	5-3	74.232 (18)*	.000	+.004	-.003
	5-5-λ, τ, δ, ξ/φ invariant	831.209 (243)*	.962	.952	.055	.051-.059	5-4	79.008 (21)*	-.004	-.001	.000
	5-6-λ, τ, δ, ξ/φ, η invariant	871.626 (249)*	.960	.951	.056	.052-.060	5-5	31.918 (6)*	-.002	-.001	+.001
ESEM	6-1-Configural invariance	253.158 (120)*	.987	.968	.040	.033-.047	–	–	–	–	–
Revised version	6-2-λ invariant	319.316 (192)*	.988	.981	.031	.025-.037	6-1	77.449 (72)	+.001	.013	-.009
Gender invariance	6-3-λ, τ invariant	341.703 (204)*	.987	.980	.031	.025-.037	6-2	22.773 (12)	-.001	-.001	.000
	6-4-λ, τ, δ invariant	411.794 (222)*	.982	.975	.035	.030-.041	6-3	68.785 (18)*	-.005	-.005	+.004
	6-5-λ, τ, δ, ξ/φ invariant	466.816 (243)*	.979	.973	.037	.032-.042	6-4	52.492 (21)*	-.003	-.002	+.002
	6-6-λ, τ, δ, ξ/φ, η invariant	557.631 (249)*	.971	.964	.043	.038-.047	6-5	79.595 (6)*	-.008	-.009	+.006
ESEM	7-1- MIMIC Null	921.659 (387)*	.933	.922	.055	.051-.060	–	–	–	–	–
Revised version	7-2-MIMIC Saturated	383.634 (243)*	.980	.967	.036	.029-.043	7-1	552.948 (144)*	+.047	+.045	-.019
Multi-group	7-3-MIMIC FO	564.625 (339)*	.972	.962	.038	.033-.044	7-2	184.811 (96)*	-.008	-.005	+.002
MIMIC model	7-4-MIMIC Gender Inv.	577.454 (363)*	.973	.967	.036	.031-.042	7-3	16.187 (24)	+.001	+.005	-.002

Note. CFA = confirmatory factor analytic model; ESEM = exploratory structural equation modeling; MIMIC = multiple indicator multiple cause model; χ^2 = scaled chi-square test of exact fit; 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; λ = factor loadings; τ = intercepts; δ = uniquenesses; ξ = factor variances; φ = fctor covariances; η = factor means; FO = Effects of the covariates limited to the factors (factors only); Inv = invariant; CM = comparison model; $\Delta^S\chi^2$ = Scaled chi square difference tests (calculated from models loglikelihoods for greater precision); Δdf = change in degrees of freedom; ΔCFI = change in CFI; ΔTLI = change in TLI; ΔRMSEA = change in RMSEA; *p<.01.

Table 5

Relations between the Covariates and the Self-Concept Factors in the Final ESEM Multiple Indicator Multiple Cause (MIMIC) Model

	Body mass index (BMI)			BMI^2			Physical activity			Age		
Self-Concept	*b* (s.e.)	β_{fem}	β_{male}	*b* (s.e.)	β_{fem}	β_{male}	*b* (s.e.)	β_{fem}	β_{male}	*b* (s.e.)	β_{fem}	β_{male}
Global self-worth	-.240 (.048) **	-.222	-.240	-.116 (.037)**	-.177	-.154	.067 (.038)	.061	.065	-.087 (.052)	-.086	-.086
Physical self-worth	-.030 (.042)	-.027	-.029	-.061 (.022)**	-.091	-.080	.308 (.040)**	-.277	.295	.014 (.051)	.014	.013
Physical appearance	-.306 (.044)**	-.276	-.296	-.181 (.037)**	-.267	-.232	.196 (.054)**	.174	.185	-.051 (.029)	-.049	-.049
Physical strength	.212 (.047)**	.183	.196	-.048 (.037)	-.067	-.059	.419 (.052)**	.356	.378	.056 (.056)	.051	.051
Physical condition	-.203 (.043)**	-.181	-.193	-.086 (.030)**	-.126	-.109	.363 (.056)**	.319	.337	.026 (.052)	.025	.024
Sport competence	-.037 (.044)	-.033	-.035	-.089 (.026)**	-.127	-.110	.430 (.046)**	.372	.394	.053 (.035)	.050	.049

Note. *b* = unstandardized regression coefficient (invariant across gender); s.e. = standard error of the coefficient; β_{fem}/β_{male} = gender-specific standardized regression coefficient (although the regressions are invariant across gender, standardized coefficients are a function of within-group variability and slightly differ across gender; *p<.05; **p<.01.

Online Supplements for:

English Validation of the Short Form of the Physical Self-Inventory (PSI-S)

Table S1

English Back-Translated Items and French Original Items from the PSI-S

Items	English Items	French Items
GSW1	I have a good opinion of myself	J'ai une bonne opinion de moi-même
PSW1	Globally, I'm proud of what I can do physically	Globalement, je suis satisfait(e) de mes capacités physiques
PA1*	I don't like very much the appearance of my body	Je n'aime pas beaucoup mon apparence physique
PS1	I'm physically stronger than most people	Je suis physiquement plus fort(e) que les autres
GSW2*	There are many things in myself that I would change	Il y a des tas de choses en moi que j'aimerais changer
PSW2	I am happy with what I can do physically	Je suis content(e) de ce que je peux faire physiquement
PC1	I would be good at physical stamina exercises	Je serais bon(ne) dans une épreuve d'endurance
SC1	I find that I'm good in all sports	Je trouve que je suis bon(ne) dans tous les sports
PA2	I have a nice body to look at	J'ai un corps agréable à regarder
PS2	I would be good at exercises that require strength	Je serais bon(ne) dans une épreuve de force
PSW3	I'm confident about my physical self-worth	Je suis confiant(e) vis-à-vis de ma valeur physique
PC2	I think I could run for a long time without tiring	Je pense pouvoir courir longtemps sans être fatigué(e)
SC2	I can find a way out of difficulties in all sports	Je me débrouille bien dans tous les sports
PA3*	Nobody find me good-looking	Personne ne me trouve beau(belle)
PS3	Faced with a situation requiring physical strength, I'm the first to offer assistance	Face à des situations demandant de la force, je suis le(la) premier(ière) à proposer mes services
PC3	I could run five kilometers without stopping	Je pourrais courir 5 km sans m'arrêter
SC3	I do well in sports	Je réussis bien en sport
GSW3	I would like to stay as I am	Je voudrais rester comme je suis
GSW2R	Overall I am satisfied with being the way I am	Globalement, je m'accepte tel que je suis
PA1R	I am really pleased with the appearance of my body	J'aime beaucoup mon apparence physique
PA3R	Everybody thinks that I am good-looking	Tout le monde me trouve beau(belle)
Answer Scale	1- Not at all; 2- Very little; 3- Some;4- Enough; 5- A lot; 6- Entirely	1-Pas du tout; 2- Très peu; 3- Un peu; 4- Assez; 5- Beaucoup; 6- Tout à fait

Note. * Negatively-worded; R = reformulated version; GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA = physical attractiveness; PS = physical strength.

Table S2

Time Specific Measurement Models

Country	Model	χ^2 (df)	CFI	TLI	RMSEA	90% CI
Australia	S1-1-CFA – Original	918.327 (120)*	.912	.887	.070	.066-.074
Time 1	S1-2-CFA – Original MF	744.026 (117)*	.931	.909	.063	.058-.067
	S1-3-CFA – Revised	907.570 (120)*	.926	.906	.069	.065-.074
	S1-4-ESEM – Original	174.925 (60)*	.987	.968	.037	.031-.044
	S1-5-ESEM – Original MF	137.218 (57)*	.991	.976	.032	.025-.039
	S1-6-ESEM – Revised	169.420 (60)*	.990	.974	.037	.030-.043
Australia	S2-1-CFA – Original	868.411 (120)*	.916	.893	.073	.069-.078
Time 2	S2-2-CFA – Original MF	691.592 (117)*	.936	.916	.065	.060-.070
	S2-3-CFA – Revised	918.913 (120)*	.922	.901	.076	.071-.080
	S2-4-ESEM – Original	200.776 (60)*	.984	.960	.045	.038-.052
	S2-5-ESEM – Original MF	134.097 (57)*	.991	.977	.034	.027-.042
	S2-6-ESEM – Revised	200.659 (60)*	.986	.965	.045	.038-.052

Note. CFA = confirmatory factor analytic model; ESEM = exploratory structural equation modeling; MF = method factor for the negatively worded items χ^2 = scaled chi-square test of exact fit; 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; *p<.01.

Table S3

Longitudinal Measurement Models Including a Method Factor for the Negatively-Worded Items

Model	Description	χ^2 (df)	CFI	TLI	RMSEA	90% CI	CM	$\Delta S\chi^2$(df)	ΔCFI	ΔTLI	ΔRMSEA
CFA	1-1-Configural invariance	1865.456 (503)*	.935	.919	.044	.042-.047	–	–	–	–	–
Original version	1-2-λ invariant	1895.278 (517)*	.935	.920	.044	.042-.046	1-1	27.995 (14)	.000	+.001	.000
Method factor	1-3-λ, τ invariant	1913.850 (528)*	.934	.922	.044	.042-.046	1-2	17.574 (11)	-.001	+.002	.000
	1-4-λ, τ, δ invariant	1985.078 (546)*	.932	.921	.044	.042-.046	1-3	70.670 (18)*	-.002	-.001	.000
	1-5-λ, τ, δ, ξ/φ invariant	2023.072 (568)*	.931	.924	.043	.041-.045	1-4	32.287 (22)	-.001	+.003	-.001
	1-6-λ, τ, δ, ξ/φ, η invariant	2048.601 (575)*	.930	.924	.043	.041-.045	1-5	25.500 (7)*	-.001	.000	.000
ESEM	3-1-Configural invariance	691.548 (383)*	.985	.976	.024	.021-.027	–	–	–	–	–
Original version	3-2-λ invariant	787.641 (457)*	.984	.978	.023	.020-.026	3-1	102.696 (74)	-.001	+.002	-.001
Method factor	3-3-λ, τs invariant	800.459 (468)*	.984	.979	.023	.020-.025	3-2	12.676 (11)	.000	+.001	.000
	3-4-λ, τs, δs invariant	877.370 (486)*	.981	.976	.024	.022-.027	3-3	72.845 (18)*	-.003	-.003	+.001
	3-5-λ, τs, δs, ξ/φ invariant	907.904 (508)*	.981	.977	.024	.021-.026	3-4	32.273 (22)	.000	+.001	.000
	3-6-λ, τs, δs, ξ/φ, ηs invariant	936.688 (515)*	.980	.976	.024	.022-.027	3-5	26.082 (7)*	-.001	-.001	.000

Note. CFA = confirmatory factor analytic model; ESEM = exploratory structural equation modeling; χ^2 = scaled chi-square test of exact fit; 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; λ = factor loadings; τ = intercepts; δ = uniquenesses; ξ = factor variances; φ = factor covariances; η = factor means; CM = comparison model; $\Delta S\chi^2$ = scaled chi square difference tests (calculated from models loglikelihoods for greater precision); Δdf = change in degrees of freedom; ΔCFI = change in CFI; ΔTLI = change in TLI; ΔRMSEA = change in RMSEA; *p<.01.

Table S4

Standardized Parameters Estimates from the PSI-S (Original) Models Including a Method Factor for the Negatively-Worded Items

	Confirmatory factor analysis							Exploratory structural equation modeling						
Standardized factor loadings (λ) and uniquenesses (δ)														
Items	GSW (λ)	PSW (λ)	PA (λ)	PS (λ)	PC (λ)	SC (λ)	δ	GSW (λ)	PSW (λ)	PA (λ)	PS (λ)	PC (λ)	SC (λ)	δ
GSW1	.734						.461	.266	.506	*.052*	.168	*-.007*	*-.044*	.465
GSW2	.237						.656	.511	*-.013*	*.028*	*-.105*	*.006*	*.029*	.610
GSW3	.619						.617	.462	.092	*.148*	*-.001*	*.054*	.181	.519
PSW1		.831					.310	*.032*	.952	-.143	*-.018*	*.060*	*.050*	.025
PSW2		.800					.361	*.038*	.404	*.214*	*.014*	.090	.263	.361
PSW3		.800					.360	*.213*	.257	.273	.253	.150	*.023*	.320
PA1			.255				.590	.397	*-.005*	***.030***	*-.109*	.064	*.050*	.597
PA2			.789				.378	.399	*.012*	***.018***	.289	.233	*.007*	.476
PA3			.228				.818	.237	*.089*	***.137***	*-.017*	-.114	*.061*	.770
PS1				.716			.487	*-.037*	.055	*-.021*	.743	*-.023*	*-.012*	.459
PS2				.854			.270	*-.026*	*.003*	*.037*	.778	*-.028*	.111	.280
PS3				.658			.567	*-.060*	*-.032*	*-.056*	.514	.111	.154	.556
PC1					.814		.338	*-.153*	.178	.194	*.039*	.545	.132	.345
PC2					.779		.394	*-.041*	*.044*	*.067*	*-.039*	.844	*-.031*	.330
PC3					.655		.570	*.051*	*-.067*	-.212	*-.004*	.814	*.030*	.347
SC1						.839	.297	*-.020*	*.022*	*-.040*	.097	.100	.697	.304
SC2						.630	.602	*.031*	.076	*.058*	.116	.192	.293	.611
SC3						.841	.293	*.031*	*.048*	*-.016*	*.017*	*-.065*	.880	.228
ω	.593	.851	.475	.789	.795	.817		.491	.787	.018	.762	.826	.754	
T-R *r*	.695	.651	.830	.730	.754	.743		.767	.518	.467	.718	.755	.735	

Factor correlations (95% Confidence Intervals)

Factors	PSW	PA	PS	PC	SC	PSW	PA	PS	PC	SC
GSW	.96 (.93-.99)	.84 (.75-.92)	.60 (.55-.65)	.64 (.58-.70)	.71 (.65-.77)	.32 (.15-.48)	*.28* (-.08-.63)	.21 (.09-.33)	.35 (.28-.43)	.28 (.16-.40)
PSW		.73 (.63-.82)	.72 (.68-.76)	.77 (.73-.81)	.83 (.80-.86)		.40 (.25-.55)	.51 (.46-.56)	.45 (.28-.63)	.59 (.53-.66)
PA			.66 (.61-.70)	.72 (.66-.79)	.68 (.61-.76)			.32 (.16-.47)	.20 (.03-.36)	.40 (.31-.49)
PS				.74 (.69-.78)	.80 (.77-.82)				.65 (.59-72)	.69 (.64-.75)
PC					.83 (.79-.86)					.70 (.63-.76)

Note. GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA= physical attractiveness; PS= physical strength; T-R *r* = Test-retest stability correlation coefficient over an 8 months period; Greyscale = main loadings. Non-significant parameters ($p \geq .05$) are marked in italics.