Validation of a Revised Version of the Physical Self-Inventory – Very Short form for Youth with Intellectual Disabilities (PSI-VS-ID-*R*): A Bayesian Structural Equation Modeling Approach

Christophe Maïano[a,b*], Alexandre J. S. Morin[b*], Cynthia Gagnon[c], Danielle Tracey[d], Céleste Dubé[b], Rhonda G. Craven[e]

[a]Cyberpsychology Laboratory and Department of Psychoeducation and Psychology, Université du Québec en Outaouais (UQO), Saint-Jérôme, Canada.
[b]Substantive-Methodological Synergy Research Laboratory, Department of Psychology, Concordia University, Montreal, Canada.
[c]Department of Psychoeducation and Psychology, Université du Québec en Outaouais (UQO|Campus de Saint-Jérôme), Canada.
[d]School of Education, Translational Health Research Institute, Western Sydney University, Sydney, Australia
[e]Institute for Positive Psychology and Education, Australian Catholic University, Sydney, Australia

* The order of appearance of the first and second authors (C.M. and A.J.S.M.) was determined at random: Both should be considered first authors.

Disclosure statement: No potential conflict of interest was reported by the author(s).

Acknowledgements: This study was supported by grants from the Australian Research Council (DP140101559) and from the Social Sciences and Humanities Research Council of Canada (430-2012-0091, 435-2014-0909). The authors also want to warmly thanks Jiesi Guo for invaluable advice in the preliminary stages of our analyses.

Corresponding author: Christophe Maïano, Université du Québec en Outaouais, Campus de Saint-Jérôme, Département de Psychoéducation et de Psychologie, 5 rue Saint-Joseph, Saint-Jérôme, Québec, J7Z 0B7, email: christophe.maiano@uqo.ca

This is the prepublication version of the following manuscript:

Maïano, C., Morin, A. J. S., Gagnon, C., Tracey, D., Dubé, C., & Craven, R. G. (In Press, 2021). Validation of a revised version of the Physical Self-Inventory – Very Short form for youth with intellectual disabilities (PSI-VS-ID-R): A Bayesian structural equation modeling approach. *Psychology of Sport & Exercise*. Doi: 10.1016/j.psychsport.2021.102029

Highlights

- We validate English and French revised version of the Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities (PSI-VS-ID-*R*);
- Result support the superiority of a Bayesian structural equation modeling representation of PSI-VS-ID-*R* ratings incorporating cross-loadings;
- Ratings obtained on the English and French versions of the PSI-VS-ID-*R* are equivalent;
- The PSI-VS-ID-*R* factor structure is equivalent as a function youth's age, body-mass index, ID level, sex and frequency of sport involvement;
- The convergent validity of the PSI-VS-ID-*R* ratings is supported in relation to measures of self-concept and physical fitness.

Abstract

The objective of the current study was to investigate the psychometric properties of scores obtained on a revised version of the Physical Self-Inventory – Very Short form for youth with intellectual disabilities (PSI-VS-ID-R). A sample of 351 youth (M = 15.81 years) with mild to moderate levels of Intellectual Disabilities (ID) from Australia (N = 230) and Canada (N = 121) participated in this study. They respectively completed the English and French versions of the PSI-VS-ID-R, as well as additional measures of self-concept and physical fitness. Bayes structural equation modeling (BSEM) analyses provided support for the validity and reliability of the a priori nine-factor structure of scores obtained on the PSI-VS-ID-R. Moreover, additional BSEM analyses provided support for the comparability of responses obtained on the PSI-VS-ID-R (i.e., lack of differential item functioning or measurement equivalence) irrespective of linguistic versions and of participants' age, body-mass index, ID level, sex and frequency of sport involvement. Additionally, latent mean level differences in PSI-VS-ID-R's factors scores were also found across linguistic versions and as a function of youth's ID level and frequency of sport involvement. Finally, the convergent validity of scores on the PSI-VS-ID-R factors was supported in relation to measures of self-concept and physical fitness.

Keywords: Physical Self-Inventory; Intellectual Disabilities; Bayesian Structural Equation Modeling; Differential Item Functioning; Inclusive Health; Special Education Needs.

Introduction

In psychology, the self-concept arguably represents a psychological construct with one of the longest research tradition (Byrne, 2002; Craven & Marsh, 2008; Marsh et al., 2017). According to Byrne (2002), the importance of the self-concept is linked to the fact that it "*not only is a desirable outcome in many psychological and educational situations but also is often posited as a mediating variable that facilitates the attainment of other desired outcomes, such as academic performance and social competence*" (p. 897). Stemming from that long-standing research tradition, the physical self-concept has long been recognized by sport and exercise psychologists as a critically important determinant and outcome of involvement, performance, and enjoyment in sports and physical activities among youth with a typical development (e.g., Babic et al., 2014).

Inspired by Shavelson et al.'s (1976) seminal work on the structure of the self-concept, Fox and Corbin (1989) proposed a multidimensional and hierarchical representation of the physical self-concept that has since become the benchmark for research conducted in sport and exercise psychology seeking to understand the role played by physical self-perceptions for involvement in physical activities and sports (e.g., Marsh & Cheng, 2012). In their theoretical model Fox and Corbin (1989) represented the physical self-concept as a pyramid, with a general construct representing individual's global self-conceptions across all life domains (i.e., global self-worth) assumed to occupy the top of this pyramid. The middle of the pyramid is then assumed to be occupied by another general construct reflecting individuals' global physical self-conceptions across all types of physical characteristics and attributes (i.e., physical self-worth). Finally, the lowest level of this pyramid is assumed to reflect more specific constructs representing distinct physical characteristics and attributes. In Fox and Corbin's (1989) original model, this lower level thus encompassed individuals' perceptions of their own sport competence, physical condition, physical attractiveness, and physical strength.

In research conducted among typically developing populations, the short (18 items), and very short (12 items) forms of physical self-inventory (PSI-S and PSI-VS: Maïano et al., 2008) have been progressively gaining prominence as providing the shortest validated multidimensional measure of Fox and Corbin's (1989) multidimensional and hierarchical representation of the physical self-concept (Marsh & Cheng, 2012). From an initial validation study focused on samples of French-speaking youth, the PSIs have since been cross-validated in additional samples of French-speaking participants (Morin & Maïano, 2011a, 2011b), as well as in English, Turkish, Dutch, Italian, and Arabic (Aşçı et al., 2017; Maïano et al., 2015b; Morin et al., 2016, 2018; Scalas et al., 2013).

As a result of its simplicity, the PSI-VS also provides a strong starting point for the development of a measure suitable for the assessment of physical self-conceptions among youth with intellectual disabilities (ID). Indeed, in populations with ID, self-concept measurement faces the ubiquitous challenge of achieving reliable and valid measures of internal states among respondents presenting lower levels of cognitive and verbal skills than those typically required for such assessments (Scott & Havercamp, 2018; Turk et al., 2012). To address these challenges, Maïano et al. (2009) developed and validated a new version of the PSI-VS specifically adapted for youth with ID, the PSI-VS-ID. The PSI-VS-ID includes 12 items (2 items per subscale) measuring: (1) global self-worth, (2) physical self-worth, (3) sport competence, (4) physical condition, (5) physical attractiveness, and (6) physical strength. Participants with ID answer each item using a six-point graphical "facial" rating scale (ranging from a very unhappy to a very happy face). These authors examined the psychometric properties of the PSI-VS-ID among a sample of 342 French-speaking youth (12 to 18 years) with ID. Their results provided support for the *a priori* six-factor structure of scores obtained on the PSI-VS-ID and for the measurement invariance (i.e., lack of measurement bias or differential item functioning) of these scores as function of age categories (12-14 years vs. 15-18 years), ID level (mild vs. moderate), type of school placement (self-contained classes vs. specialized schools), and sex (boys vs. girls). Finally, their results also supported the composite reliability (ω = .70 to .74) and test-retest stability (r = .72 to .93) of PSI-VS-ID ratings.

These results were later replicated by Maïano et al. (2011) in a second study realized among a sample of 248 youth (12 to 20 years) with ID. In this second study, the authors found support for the *a priori* six-factor structure of PSI-VS-ID ratings, and for their scale score reliability (α = .67 to .82). Despite some minor exceptions, subsequent analyses also generally supported the equivalence (lack of differential item functioning) of PSI-VS-ID ratings in relation to participants' sex and weight status (underweight-normal weight vs. overweight-obese), age categories (12-17 years vs. 18-20 years), and

ID level (mild vs. moderate-severe). However, their results also revealed some latent mean differences on the scores obtained on the PSI-VS-ID subscales as function of age categories (12-17 years > 18-20 years), ID level (moderate-severe ID > mild ID), sex (boys > girls), and weight status (underweight-normal weight > overweight-obese).

Despite these promising results, four main limitations remain associated with this instrument. First, the factor correlations reported by Maïano et al. (2009, 2011) between the PSI-VS-ID subscales were elevated enough to suggest possible redundancies (i.e., >.50 and even >.90), particularly among subscales located at different levels of the self-concept hierarchy (i.e., between the global self-worth, physical self-worth, and physical attractiveness subscales). This issue has frequently been reported in physical self-concept research (e.g., Hagger et al., 2004; Marsh et al., 2002, 2006), more specifically when short scales attempt to cover a broad range of subscales with few items (Marsh et al., 2002), and has been found to be pervasive in research relying on the various PSIs (e.g., Maïano et al., 2008, 2009, 2011; Maïano et al., 2015b; Morin & Maïano, 2011a, 2011b; Morin et al., 2016, 2018). However, research has demonstrated that confirmatory factor analyses in which cross-loadings are constrained to be exactly zero tend to be far too restrictive for many multidimensional measures, leading to inflated estimates of factor correlations (e.g., Marsh et al., 2009, 2010a, 2014). Importantly, this assertion has been supported by statistical research showing that excluding cross-loadings present in the population model from the estimated model, even when they are as low as .100, tends to result in inflated estimates of factor correlations, whereas including unnecessary cross-loadings still results in accurate parameter estimates (Asparouhov et al., 2015; Mai et al., 2018). On this basis, Morin and Maïano (2011a) and Morin et al. (2016, 2018) have adopted the exploratory structural equation modeling (ESEM) framework (Asparouhov & Muthén, 2009) to incorporate cross-loadings to PSI-S measurement model defined in an *a priori* manner. These studies have demonstrated the superiority of this approach, which made it possible to identify factors that were more independent from one another and yet defined as well as in confirmatory factor analyses. Unfortunately, the ESEM approach remains unidentified (i.e., impossible to estimate) when using only two indicators per construct, which is the case for the PSI-VS and PSI-VS-ID. Fortunately, the recently developed Bayesian structural equation modeling (BSEM; Muthén & Asparouhov, 2012) framework provides a solution to this problem, making it possible to estimate measurement models including only two indicators per construct, and using small variance priors for the estimation of cross-loadings estimated to remain as close to zero as possible (Asparouhov et al., 2015).

Second, when compared to other physical self-concept questionnaires commonly used among typically developing youth (Dreiskämper et al., 2015; Lohbeck et al., 2017; Marsh et al., 1994; for a review, see Marsh & Cheng, 2012) the PSI-VS-ID does not assess perceptions of balance, flexibility, and running speed. This is unfortunate given that several studies have shown that youth with ID display lower levels of performance in these physical fitness areas (Blomqvist et al., 2013; Hartman et al., 2015; Kioumurtzoglou et al., 1995; Klavina et al., 2017; Mac Donncha et al., 1999; Protić-Gava & Uskoković, 2016) relative to their typically developing peers. Given the importance of these dimensions for various types of physical activities (e.g., Catuzzo et al., 2016; Hulteen et al., 2018; Utesh et al., 2019), the ability to reliably and validly assess perceptions of balance, flexibility, and running speed among youth with ID appears to be important for research and intervention purposes.

Third, apart from a recent linguistic adaption of the PSI-VS-ID in research focusing on English and Chinese Special Olympics athletes which tentatively supported the scale score reliability of PSI-VS-ID ratings (Pan et al., 2018), no attempt has been made to systematically develop, and validate, alternative linguistic versions of the PSI-VS-ID. Therefore, the generalizability of the internal factor structure of scores obtained on the PSI-VS-ID within other linguistic or cultural groups remains unknown. Likewise, although previous studies (Maïano et al., 2009, 2011) supported the generalizability (i.e., measurement equivalence) of the PSI-VS-ID internal factor structure as a function of youth's age, ID level, school placement, sex, and weight status, similar verifications have yet to be conducted as a function of youth's levels of physical activity/sport practice. This lack of information precludes the use of this instrument in research seeking to assess associations between physical activity/sport practice and physical self-conceptions among youth with ID.

Fourth, as already mentioned by Maïano et al. (2011), the convergent validity of PSI-VS-ID ratings has yet to be examined in relation to other self-concept instruments and physical fitness measures. Therefore, it is currently unknown whether responses provided to the PSI-VS-ID by youth with ID are significantly related to their responses to other questionnaires measuring similar physical self-concept subscales and to more objective measures of physical fitness. In contrast, results obtained among samples of typically developing youth generally show that scores on physical self-concept subscales tend to be more strongly related to scores on scales measuring similar physical self-perceptions than to scores on scales measuring other types of physical self-perceptions (e.g., Marsh et al., 2002; 2010b; Maïano et al., 2015a), with the occasional exception of: (a) global self-worth and physical appearance measures which sometimes share fairly high correlations with measures of physical self-worth, and (b) physical condition/endurance measures which sometimes also share fairly high correlations with measures of perceived strength, sport competence, and coordination. Likewise, these studies also generally show that physical fitness measures of endurance/physical condition, strength, and flexibility tend to be more strongly related with measures of matching physical self-perceptions than with measures of other types of physical self-perceptions (e.g., Guérin et al., 2004; Marsh & Redmayne, 1994), with the exception of balance self-perceptions ratings which are not typically correlated with physical fitness measures. Evidence is far more limited when samples of individuals with ID are considered. Indeed, the very few studies available to date have focused on the relations between actual and perceived physical (Yun & Ulrich, 1997) or basketball (Shapiro & Dummer, 1998) skill competence. In the first of those studies, Yun and Ulrich (1997) showed that actual and perceived motor skill competence were only significantly correlated with measures of dribbling, jumping rope, running, and skipping, but not to measures of batting, catching, kicking, long jump, shooting, and throwing. In the second study, Shapiro and Dummer (1998) similarly showed that actual and perceived basketball skill competence were related to measures of passing, jumping, and shooting, but not to measures of dribbling.

Objectives of the Study

The main objective of the present study was to develop and validate a revised version of the PSI-VS-ID, the PSI-VS-ID-*R*. In addition, to maximise the utility and applicability of this revised measure, we simultaneously develop and validate English and French versions of the PSI-VS-ID-*R*. More specifically, in this revised version: (a) three additional physical self-concept subscales measuring balance, flexibility, and running speed were included; (b) the original purely graphical response scale was replaced by a revised response scale via the addition of verbal descriptors; and (c) the words used in all items were associated with pictograms. To investigate the psychometric properties of scores obtained on the PSI-VS-ID-*R*, we first assess the internal factor structure and reliability of these scores among a sample of youth with ID using Bayesian measurement models including, or not, cross-loadings. Second, we seek to replicate and extend previous validation studies (Maïano et al., 2009, 2011) by examining the presence of measurement biases (i.e., differential item functioning) and latent mean differences of scores obtained on the PSI-VS-ID-*R* across linguistic versions and as a function of the frequency of sport involvement and other characteristics of youth with ID (i.e., age, body mass-index, ID level, and sex). Third, we verify whether and how scores on the PSI-VS-ID-*R* subscales will be related (convergent validity) to scores on other self-concept (i.e., global self-esteem, perceived physical appearance, and perceived physical abilities) and physical fitness measures (i.e., flexibility, running speed, strength, and functional walking capacity).

Method

Participants

A sample of 351 youth (60.7% boys; aged 11.92 to 21.52 years; M_{age} = 15.81 years; body mass-index: 14.50 to 50.11 kg/m², $M_{body\ mass\text{-}index}$ = 23.69, $SD_{body\ mass\text{-}index}$ = 6.45) with ID participated in this study. These participants were recruited in secondary schools or community organizations located in Australia (English speaking; *N* = 230; 67% boys; M_{age} = 15.21 years; $M_{\text{body mass-index}}$ = 23.49 kg/m²) and Canada (French-speaking; *N* = 121; 48.8% boys; M_{age} = 16.70 years; $M_{\text{body mass-index}}$ = 24.06 kg/m²). Of them 51.1% (Australian: 60.3%; Canadian: 31.7 %) had a mild level of ID and 48.9% (Australian: 39.7 %; Canadian: 68.3 %) had a moderate level of ID. Participants were involved in a sport outside of school for an average of 1.75 weekly sessions (Australian: *M* = 1.60; Canadian: *M* = 2.05).

Procedures

Permission to conduct the study was obtained from the research ethics committees of the first, fourth, and last authors' Universities. Participants were recruited in schools or community organizations that agreed to support this proposal. No compensation was offered for participation in Australia, whereas Canadian participants were eligible to win one out of 40 gift certificates ($30 CAD). Parents (or legal representatives) of all participating youth actively provided signed informed consent for their children's participation. For parents of youth recruited in schools, this consent form was directly sent to the parents by the school, with an information letter, and the signed consent form was returned to the school where members of the research team recuperated it. Parents recruited outside of the participating schools received this material directly from the research team and returned the signed consent form to the researchers using a reply-paid envelope.

The consent procedure granted the researchers access to school records, including youth's most recent level of intellectual functioning (only youth with an official school-based ID classification were recruited). The Weschler (2003) Intelligence Scale for Children – Fourth Edition (WISC-IV) was the IQ test most frequently used by the schools in both countries. When the last IQ assessment in the school records was older than four years, a new IQ assessment was conducted by a registered psychologist using the WISC-IV, the Weschler Adult Intelligence Scale-IV, or the Leiter international performance scale-revised (Roid & Miller, 1997), depending on age and verbal ability.

Participating youth were met at their school (or at a time and location most convenient for the parents of participants recruited outside of schools) by members of the research team or trained research assistants who explained the goals and procedures of the study, as well as youth's right not to participate or to withdraw from the study without any consequences. Thus, youth were asked to actively and voluntarily consent to the study. The PSI-VS-ID-*R* and physical fitness tests were administered, at school or at a location most convenient for them (for those recruited outside of schools), by members of the research team. Testing was realized in small groups including up to 8 youth with mild ID or including 1 or 2 youth with moderate ID. A read-aloud assisted procedure was utilized for the questionnaires to maximize understanding, and youth were encouraged to ask questions. The trained research assistants, using sample questions for each questionnaire, explained how to use the response scale using a template comprising a graphical displays and pictograms. Sometimes, despite the support, youth remained unable to understand an item's sentence. In these instances, they were instructed to select the "do not understand the statement" option. Those responses (.28% to 3.13%; M=1.42%, see Table S1 of the online supplements for additional details) were treated as missing values.

Measures

Youth's Characteristics. Youth's age, sex, and ID level were obtained from school records. Information about involvement in the week frequency of sport practice outside of the school context were obtained directly from the youth (i.e., *Do you practice sport when you are not at school (for example, in the evenings or on weekends?)*; If yes, *Last week, which days did you practice sport?*). Finally, youth's height and weight were measured using a stadiometer (Tanita HR200) and a scale (Tanita BF-350), respectively. These measures were then used to calculate their body mass-index [Weight/(Height^2)] in kg/m^2.

Physical Self-Concept. The procedures used to develop the PSI-VS-ID-*R* are reported in section S1 in the online supplements. This questionnaire includes 18 items measuring nine subscales (2 items per subscale): balance, flexibility, global self-worth, physical attractiveness, physical condition, physical strength, physical self-worth, sport competence, and running speed. Participants indicated whether they disagreed or agreed with each item using a six-point response scale, ranging from "No, I totally disagree" associated with a very unhappy face to "Yes, I totally agree" associated with a very happy face. The items and response scales of the PSI-VS-ID-*R* are presented in Appendix A (the complete questionnaire is available upon request from the corresponding author).

Self-Concept. English and French versions of the global self-esteem (or global self-concept), perceived physical appearance, and perceived physical abilities scales of the Self-Description Questionnaire I – Individual Administration for people with intellectual disabilities (SDQ-IA-ID; Marsh et al., 2006) were used. Each scale includes eight items measuring how participants perceive themselves in general (e.g., *I like being the way I am*), their physical appearance (e.g., *I have a nice looking face*), or their physical abilities (e.g., *I enjoy sports and games*). For purposes of this study, the original answer scale (i.e., *No, always* to *Yes, always*) was replaced by a six-point graphical answer scale matching that used for the PSI-VS-ID-*R* (i.e., *No, I totally disagree* associated with a very unhappy face to *Yes, I totally agree* associated with a very happy face).

Physical Fitness. Four tests were used to measures youth's flexibility, strength, speed and functional walking capacity. The *flexibility* of youth's posterior muscle chain was measured using the sit and reach test (Council of Europe Committee for the Development of Sport [CECDS], 1988).

Participants were asked to sit with their legs extended, their toes pointing up and the sole of their feet placed against the base of a step over which is placed a panel. They were then asked to push a ruler on the panel as far as they can go (past their toes) without bending the knees. This test was performed twice, and a mean flexibility score was calculated across the two trials reflecting the average distance (in centimetres) covered by the ruler. Higher scores indicate greater flexibility.

The *strength* of youth's upper limbs was measured using the handgrip test (CECDS, 1988). Participants were asked to squeeze the handle of a dynamometer using their preferred hand (established before testing) as hard as possible while holding their arm straight beside their leg without touching it. This test was performed twice, and a mean strength score was calculated across the two trials reflecting youth's static force (in kilograms). Higher scores indicate greater strength.

Youth's *running* s*peed* was measured using the 25-meter running speed test (Skowroński et al., 2009). Youth were asked to run as fast as possible along a 25-meter distance. Running speed scores reflect the time (in seconds) they took to run the 25 meters, with higher scores reflecting less speed.

Youth's *functional walking capacity* was measured using the six-minute walk-test (American Thoracic Society, 2002). Participants were instructed to walk (without running) back and forth as often as a possible along a 20-meter corridor for a total of 6 minutes. Functional walking capacity scores reflect the total distance walked (in meters) during this period, with a higher score indicating greater functional walking capacity.

Analyses

All analyses were conducted using the Bayesian estimator (Muthén & Muthén, 2019) implemented in Mplus 8.5 (Muthén & Muthén, 2019), allowing us to rely on full information algorithms asymptotically equivalent to full-information maximum likelihood to handle the few missing responses (.85%-4.27%, M = 2.28%, see Table S1 of the online supplements for a description). These analyses relied on a Gibbs (PX3) Markov Chain Monte Carlo (MCMC) sampler (Liu & Daniels, 2006), four independent MCMC chains, a seed value of 100 (Bseed = 100) in the generation of the first MCMC chain, 150,000 iterations (Biterations; 75,000 for the burn-in phase and 75,000 for the posterior distribution), a thinning every 10th iteration in order to reduce autocorrelations, and a convergence criterion of .01 (Bconvergence; requiring thus that the Gelman-Rubin convergence diagnostic produce potential scale reduction [PSR] factor values < 1.02). We first estimated our *a-priori* nine-factor model using a classical confirmatory factor analytic (hereafter referred as the CFA model) representation without cross-loadings. No priors were used in the estimation of this model. A second Bayesian structural equation modeling (hereafter referred as the BSEM model) representation incorporating cross-loadings (Asparouhov et al., 2015) was estimated using small variance informative priors with a mean of 0 and a variance of .01 for the cross-loadings, resulting in a 95% prior distribution of 0 ± .20 for the cross-loadings estimates (Muthén & Asparouhov, 2012). As recommended by Asparouhov et al. (2015), we conducted a sensitivity analysis using very small variance informative priors with a mean of 0 and a variance of .001 (smaller by a factor of 10 compared to the prior retained in our main analyses) and .005 (five times higher than the previous .001 prior), resulting in a 95% prior distribution of respectively 0 ± .06 and 0 ± .014 for the cross-loadings (Muthén & Asparouhov, 2012). As suggested by Depaoli and van de Schoot (2017), for the best model, Bayesian posterior parameter trace plots and autocorrelation plots were inspected to monitor chain convergence. To ensure that chain convergence was obtained and avoid local identification, the best model was then re-estimated while doubling the number of iterations (Asparaouhov et al., 2015; Depaoli & van de Schoot, 2017). Small relative biases (i.e., < |5|%) between the initial solution and the solution involving more iterations supports the adequacy of the solution (Depaoli & van de Schoot, 2017).

The composite reliability of scores on PSI-VS-ID-*R* factors was estimated with McDonald's (1970) omega (ω). Model fit was assessed by considering the posterior predictive p-value (values >.05 and .50 respectively indicate acceptable and excellent fit to the data), the 95% confidence intervals for the difference between the observed and replicated χ^2 values (the posterior predictive p-value 95% CI must include zero to indicate adequate model fit), and the following fit indices: the deviance information criterion (a lower value indicate a better fitting model relative to alternative models), the comparative fit index (CFI ≥ .90 or >. 95 respectively indicate acceptable and excellent fit to the data), the Tucker-Lewis index (same thresholds as for CFI) and the root mean square error of approximation (RMSEA ≤. 08 or <. 06, respectively indicate acceptable and excellent fit to the data) (e.g., Asparouhov & Muthén, 2021; Hu & Bentler, 1999; Marsh et al., 2005; Muthén & Asparouhov, 2012). Additionally, the prior posterior predictive *p* value (Asparouhov & Muthén, 2017; Hoijtink & van de Schoot, 2018), was used to examine whether minor parameters assigned a small variance prior of 0 can be considered to be approximatively zero (higher prior posterior predictive *p* values indicate that the prior are appropriate).

Once the optimal representation of responses to the PSI-VS-ID-*R* was identified, this representation was used to estimate a multiple indicators multiple causes multiple-group model (Marsh et al., 2013; Morin et al., 2013) to examine: (a) the associations between scores on the PSI-VS-ID-*R* latent factors and the predictors [i.e., age, body mass-index, linguistic versions (French-Canadian coded 0 and English-Australian coded 1), ID level (mild coded 0 and moderate coded 1), sex (girls coded 0 and boys coded 1), and frequency of sport involvement]; and (b) the possible presence of differential item functioning (i.e., direct associations between predictors and PSI-VS-ID-*R* item response over and above the association between the predictors and the PSI-VS-ID-*R* latent factors). More precisely, these models were estimated in the following sequence (Marsh et al., 2013; Morin et al., 2013): (a) null effects model (the paths from the predictors to the PSI-VS-ID-*R* latent factors and item responses were constrained to be zero); (b) saturated model (the paths from the predictors to the PSI-VS-ID-*R* item responses were freely estimated, while the paths from the predictors to the PSI-VS-ID-*R* latent factors were constrained to be zero); and (c) factors-only model (the paths from the predictors to the PSI-VS-ID-*R* latent factors were freely estimated, while the paths from predictors to the PSI-VS-ID-*R* item responses were constrained to be zero). As recommended (Marsh et al., 2013; Morin et al., 2013), a substantial improvement in model fit (ΔCFIs-TLIs $\geq$ .01, ΔRMSEAs $\geq$ .015, and lower deviance information criterion values) associated with the factors-only and saturated models relative to the null effects model support the presence of an association between the PSI-VS-ID-*R* responses and the predictors. However, an improvement in model fit associated with the saturated model relative to the factors-only model supports the presence of differential item functioning (Marsh et al., 2013; Morin et al., 2013).

The convergent validity of responses to the PSI-VS-ID-*R* was examined in relation to self-concept (i.e., global self-esteem, perceived physical appearance and perceived physical abilities) and physical fitness (i.e., flexibility, strength, running speed, and functional walking capacity) measures. All physical fitness measures were standardized prior to the analyses. To facilitate the estimation of this model of convergent validity, plausible values (i.e., similar to factor scores obtained using a multiple imputation process) reflecting participants' scores on the latent PSI-VS-ID-*R* factors were first extracted from the factor distribution of the retained solution using a series of 20 draws and a thinning of 10, leaving one final set of plausible values (Muthén & Muthén, 2019). Additionally, plausible values reflecting participants' scores on the latent self-concept factors (from the SDQ-IA-ID) were extracted from the factor distribution of retained solution described in Table S2 in the online supplements. These plausible values where then used to assess the correlations between participants' scores on the PSI-VS-ID-*R* factors and convergent self-concept measures, and the observed scores of the physical fitness measures.

A final consideration, when interpreting and reporting Bayesian statistics is that, when compared to frequentist approaches, Bayesian analyses do not rely on point estimates of the parameter of interests and do not rely on statistical significance testing (e.g., Depaoli & van de Schoot, 2017). Indeed, in Bayesian analyses, uncertainty in the estimation of the parameter value is captured by the estimation of a posterior distribution of possible values. Whereas the median of this distribution is reported in Mplus to provide an idea of the parameter value, this value is also reported with a 95% credibility interval (CI) reporting the range of likely values taken by this parameter. When this credibility interval excludes 0, then the parameter can be considered to represent a meaningful effect.

Results

Internal Factor Structure and Reliability

The goodness-of-fit of the alternative (CFA and BSEM) solutions are reported in Table 1 (models 1-1 to 1-4). The results from the CFA solution resulted in excellent fit indices (CFI-TLI $\geq$.95 and RMSEA $\leq$.06), except for the posterior predictive *p*-value that was lower than .05 and was associated with a 95% CI excluding zero. In contrast, the BSEM solution (Model 1-2; with a prior distribution of with a mean of 0 and a variance of .01 for the cross-loadings) achieved an excellent level of fit according to all model fit statistics (CFI-TLI >.95 and RMSEA <.06; posterior predictive *p*-value value of .100; posterior predictive *p*-value 95% CI including zero), resulted in a meaningful improvement in model fit relative to the previous model (ΔCFI = +.028, ΔTLI = +.041, ΔRMSEA = -.034; and lower DIC value), and converged on a prior posterior predictive *p* value of .796.

As suggested by Asparouhov et al. (2015), we conducted sensitivity analyses in which this model (Model 1-2) was contrasted to two alternative models estimated with smaller variance priors for the cross-loadings using a prior distribution with a mean of 0 and a variance of .001 (Model 1-3) or .005 (Model 1-4). The results revealed that the Model 1-3 (0, .001) resulted in an excellent level of fit (CFI-TLI >.95 and RMSEA <.06), but in a posterior predictive *p*-value <. 05, a 95% CI excluding zero, and a very low prior posterior predictive *p* value. In contrast, Model 1-4 (0, .005) achieved an excellent level of fit according to all model fit statistics (CFI-TLI >.95 and RMSEA <.06; posterior predictive *p*-value of .059; posterior predictive *p*-value 95% CI including zero), and a prior posterior predictive *p* value (.600). Importantly, both of these models resulted in a decrease in model fit relative to Model 1-2 (Model 1-3: ΔCFI = -.011, ΔTLI = -.018, ΔRMSEA = +.018 and higher DIC value; Model 1-4: ΔCFI = -.002, ΔTLI = -.005, ΔRMSEA = +.006, and higher DIC value), although the difference was not as marked for Model 1-4. Examination of the PSR values (all < 1.02) from the Gelman-Rubin convergence diagnostic confirmed the MCMC chain convergence for all BSEM models. The convergence criteria was also rapidly reached in all models with cross-loadings (Models 1-2 to 1-4). Finally, the results also showed that the posterior predictive *p*-value 95% CI went further under 0 in the BSEM Model 1-2 (0, .01) relative to Model 1-4 (0, .005), and that the prior posterior predictive *p* value is substantially higher in the BSEM Model 1-2 (0, .01) relative to the BSEM Model 1-4 (0, .005). Consequently, our initial BSEM solution (Model 1-2 with 0, .01) satisfied all criteria for model fit evaluation and appropriateness of the prior used. Thus, the Bayesian posterior parameter trace plots and the autocorrelation plots associated with this model were inspected, and no bias was found in parameters between models estimated with 150,000 and 300,000 iterations. This model was retained for further analyses. The plots associated with all items as part this model are presented in Table S3 of the online supplements, and are all consistent with a proper chain convergence.

The detailed parameter estimates from the CFA (Model 1-1) and retained BSEM (Model 1-2 with 0, .01) solutions are reported in Tables 2 and 3, respectively. In the CFA solution (Model 1-1), the main factors loadings (λ = .642-.901, M_λ = .828) were all reasonably high, and resulted in acceptable to excellent coefficients of composite reliability for all factors when considering that each factors is defined by two items (Streiner, 2003; ω = .646-.895, M_ω = .813). However, the latent factor correlations obtained in this model were high enough to call into question the discriminant validity of some of these factors (r = .430-.912, M_r = .636). In particular, and as expected (e.g., Maïano et al., 2009, 2011), the highest factor correlations were found between the global self-worth, physical self-worth, and physical attractiveness factors.

The retained BSEM solution (Model 1-2 with 0, .01) also resulted in reasonably high main factor loadings (λ = .634-.999, M_λ = .843) and acceptable to excellent composite reliability (ω = .756-.909, M_ω = .839) for the global self-worth, physical self-worth, physical condition, sport competence, physical attractiveness, flexibility, balance, and running speed factors. However, the physical strength factors appeared to be slightly more weakly defined (ω = .601), mainly as a result of the lower factor loading associated with the second (PS2: λ = .282; *I can carry heavy things*), but not the first (PS1: λ = .903; *I am stronger than others*) item from the physical strength subscale. As a result, the item PS2 might be targeted for re-assessment in future studies. This solution also revealed slightly reduced factor correlations than the CFA solution (r = .410-.847, M_r = .595), supporting more differentiated, yet inter-related, factors. Additionally, although the BSEM solution revealed reasonably small cross-loadings, these cross loadings were small enough not to interfere with factors definition ($|\lambda|$= .001-.163, $M_{|\lambda|}$ = .031). Therefore, as recommended in the literature (e.g., Asparouhov et al., 2015; Morin et al., 2020) the BSEM solution was retained.

Differential Item Functioning and Latent Mean Differences as Function of Predictors

The results from the multiple indicators multiple causes models are presented in Table 1. In relation to the two linguistic versions of the questionnaire, the results showed that both the saturated (model 2-2) and factors-only (model 2-3) models resulted in a substantial improvement in model fit relative to the null effects model (model 2-1). These results thus support the idea that the linguistic versions are associated with PSI-VS-ID-*R* responses. In addition, the factors-only model resulted in similar level of fit than the saturated model (ΔCFI = -.006, ΔTLI = -.007, ΔRMSEA = +.008), suggesting a lack of differential item functioning. Results from this model showed that English-Australian youth with ID tended to score lower on global self-worth (-.337, 95%CI = -.454 to -.220), physical self-worth (-.265, 95%CI = -.380 to -.144), physical attractiveness (-.366, 95%CI = -.489 to -.244), physical strength (-.187, 95%CI = -.327 to -.037), flexibility (-.188, 95% CI = -.307 to -.065), and running speed (-.189, 95%CI = -.309 to -.063) factors relative to French-Canadian youth with ID.

For age, body mass-index and sex, the results showed that both the saturated (models 3-2, 4-2, and 6-2) and factors-only (models 3-3, 4-3, and 6-3) models did not result in a substantial improvement in model fit relative to the null effects model (models 3-1, 4-1, and 6-1). These results thus indicate a lack of differential item functioning as well as a lack of association between these predictors and scores on the PSI-VS-ID-*R* latent factors.

For ID level and frequency of sport involvement, the results showed that both the saturated (models 5-2 and 7-2) and factors-only (models 5-3 and 7-3) models resulted in a substantial improvement in model fit relative to the null effects model (models 5-1 and 7-1). These results thus support the idea that ID level and the frequency of sport involvement were associated with responses from the PSI-VS-ID-*R*. Additionally, the saturated (models 5-2 and 7-2) and factors-only (models 5-3 and 7-3) models were found to present a comparable level of fit to the data (CFI-TLI $\leq$ .01 and ΔRMSEA $\leq$. 015), suggesting a lack of differential item functioning (i.e., measurement equivalence). Results associated with the effects of the predictors on the latent factors from the factors-only model for which the 95% CI excluded zero showed that: (a) youth with moderate levels of ID tended to score higher on the global self-worth (.192, 95%CI = .059-.319), physical self-worth (.214, 95% CI = .090-.333), physical attractiveness (.330, 95% CI = .208-.447), physical strength (.282, 95% CI = .150-.413), and flexibility (.159, 95% CI = .038-.281) factors relative to youth with mild levels ID; and (b) youth with ID more frequently involved in sport practice tended to score higher on global self-worth (.252, 95%CI = .093-.390), physical self-worth (.298, 95% CI = .169-.418), physical condition (.313, 95% CI = .193-.428), sport competence (.299, 95% CI = .186-.407), physical attractiveness (.253, 95% CI = .100-.388), physical strength (.272, 95% CI = .112-.418), flexibility (.261, 95% CI = .136-.377), balance (.246, 95% CI = .116-.368), and running speed (.278, 95% CI = .162-.391) factors relative to youth with ID less frequently involved in sport practice.

Convergent validity

As reported in Table 1 (Model 1-5), the BSEM model used to assess the convergent validity of scores on the PSI-VS-ID-*R* factor scores resulted in an adequate level of fit according to all model fit statistics (CFI-TLI >.95 and RMSEA <.06; posterior predictive *p*-value of .294; and posterior predictive *p*-value 95% CI including zero). Examination of the PSR values (all < 1.02) from the Gelman-Rubin convergence diagnostic confirmed the MCMC chain convergence. The results from this model are reported in Table 4 and correlations coefficients associated with a 95% CI excluding zero reveal that scores on all of the PSI-VS-ID-*R* factors were positively related to scores on all self-concept measures. In addition, results showed that: (a) flexibility capacity was positively related to scores of perceived flexibility; (b) running speed capacity (a higher score indicates lower running speed capacity) was negatively related to scores of perceived physical condition and perceived running speed; (c) strength capacity was negatively related to scores of perceived physical attractiveness; and (d) no correlations were found between functional walking capacity and scores on the PSI-VS-ID-*R* subscales.

Discussion

The first objective of the present study was to examine the internal factor structure and reliability of scores obtained on the English and French versions of the PSI-VS-ID-*R* among youth with ID. Results from Bayesian analyses supported the internal factor structure and reliability of the *a priori* nine-factor structure of the PSI-VS-ID-*R*, although they also suggested that one item from the physical strength subscale (PS2: *I can carry heavy things*) did not perform as well in the current study and should be targeted for re-assessment in future research. We come back to this item when discussing convergent validity and propose a possible reformulation. Importantly, our reliance on Bayesian estimation procedures to incorporate cross-loadings made it possible to estimate reasonably differentiated latent factors without resulting in the estimation of unreasonably high cross-loadings which would have interfered with the interpretation of the factors. Observing that all cross-loadings were relatively small provided further support to the ability of most PSI-VS-ID-*R* items to properly define their *a priori* constructs (Morin & Maïano, 2011a).

The second objective was to examine the possible presence of differential item functioning (i.e., measurement biases) and latent mean differences in PSI-VS-ID-*R* scores as a function of the linguistic version of the PSI-VS-ID-*R* and of several predictors. In this regard, our analyses first revealed that PSI-VS-ID-*R* responses were equivalent as a function of the linguistic version and of the predictors (i.e., age, body mass-index, ID level, sex, and frequency of sport involvement). Overall, the present results thus suggest that the PSI-VS-ID-*R* can be confidently used when measuring the physical self-concept among youth with ID irrespective of their linguistic (English or French) background, ID level, age, body mass-index, sex, and frequency of sport involvement.

In terms of discriminant validity, the results revealed latent mean differences in scores on the PSI-VS-ID-*R* subscales as function of the linguistic version, ID level, and frequency of sport involvement. First, the results revealed that English-Australian youth with ID tended to score lower on several physical self-concept scales (i.e., global self-worth, physical self-worth, physical attractiveness, physical strength, flexibility, and running speed) relative to their French-Canadian counterparts. These unexpected results thus suggest that some cultural, or educational, difference might be present between Canada and Australia in relation to the relative importance attributed to different aspects of the physical self-concept. For instance, positive physical self-perceptions might be harder to achieve and maintain for Australian youth due to the warmer Australian (relative to Canadian) climate, leading to higher levels of skin exposure (i.e., making ones' body shape and attributes more visible to others; Maïano et al., 2006). However, future studies are needed to clearly identify the mechanisms explaining this difference. Second, matching results from previous research conducted on the PSI-VS-ID (Maïano et al., 2011), the present results showed that youth with moderate levels of ID tended to score higher on various PSI-VS-ID-*R* factors (i.e., global self-worth, physical self-worth, physical attractiveness, physical strength, and flexibility) relative to youth with mild levels of ID. Third, and in accordance with previous results obtained in research conducted among typically developing youth (e.g., Marsh et al., 1997; Maïano et al., 2015a; Morin et al., 2018; Schmalz & Davison, 2006), our results revealed that youth more frequently involved in sport practice tended to score higher on all PSI-VS-ID-*R* factors relative to youth less frequently involved in sport practice. However, and contrasting with previous results obtained in relation to the PSI-VS-ID (Maïano et al., 2011), no associations were found between age, body mass-index, sex, and any of the PSI-VS-ID-*R* factors. Although methodological differences might explain part of this difference in results [i.e., we relied on continuous, rather than categorical, measures of age and body mass-index, and relied on a larger and more diversified sample of French-Canadian and Australian participants whereas Maïano et al. (2011) relied on a sample of French participants], additional studies would be needed to better document the conditions to which these differences generalize, or fail to do so.

The third objective was to examine the convergent validity of responses to the PSI-VS-ID-*R* in relation to other self-concept measures and to objective measures of physical fitness. In this regard, and in accordance with the results obtained among studies of typically developing youth (e.g., Maïano et al., 2015a; Marsh et al., 2002; 2010b), the present results revealed that scores on the global self-worth, physical self-worth, and physical attractiveness subscales from the PSI-ID-VS-*R* were strongly correlated with matching subscales from the SDQ-IA-ID. In addition, the results also showed that scores on the global self-worth subscale from the PSI-ID-VS-*R* displayed a similar pattern of correlations with scores on the global self-esteem and perceived physical appearance subscales of the SDQ-IA-ID, whereas scores on the physical self-worth subscale were more strongly related to scores on the global self-esteem and perceived physical appearance subscales than to scores on the perceived physical abilities subscale of the SDQ-IA-ID. These associations are all consistent with the elevated correlations often observed among ratings of physical self-concept dimensions reflecting the different levels of Fox and Corbin's (1989) representation, more specifically global self-worth, physical self-worth, and physical attractiveness (e.g., Marsh et al., 2002; 2010b; Maïano et al., 2008, 2009, 2011; 2015a, 2015b; Morin & Maïano, 2011a, 2011b; Morin et al., 2016, 2018).

In relation to the objective measures of physical fitness, the results were also partially consistent with those obtained in previous studies conducted among samples of typically developing youth (e.g., Guérin et al., 2004; Marsh & Redmayne, 1994) and of youth with ID (Shapiro & Dummer, 1998; Yun & Ulrich, 1997). More precisely, the results revealed that scores on the perceived measures of running speed and flexibility were related to their physical fitness counterparts. Additionally, the results also revealed a: (a) negative correlation between ratings of perceived physical condition and actual running speed (i.e., youth perceiving themselves to be in better physical condition tended to display higher levels of actual running speed); and (b) negative correlations between ratings of perceived physical attractiveness and actual strength capacities (i.e., youth perceiving themselves to be more physically attractive tended to display lower levels of actual strength). However, no correlations were found between scores of perceived and actual measures of physical strength, and between scores of perceived physical condition and actual functional walking capacity.

Many of these unsupported correlations can be explained by the degree of alignment (or lack thereof) between the PSI-VS-ID-*R* factors and the physical fitness measures. For instance, and contrasting with flexibility and running speed, the perceived and actual measures of physical condition and physical strength were not completely aligned. Indeed, the new subscales of perceived flexibility and running speed both include one item (SPE2: *I would be good in a test where you have to run fast*; FL2: *I would be good in a test where you have to be flexible*) that is directly aligned with the content of the tests used to measure flexibility (i.e., be flexible) and running speed (i.e., running fast). In contrast, the actual measure of physical condition involved functional walking capacity, rather than directly addressing the endurance component of both items used in this subscale (PC1: *I can run a long time without getting tired*; PC2: *I can run 10 times around the basketball court without stopping*). This observation suggests that stronger convergent associations should be observed with objective tests of endurance not limited to running. Alternatively, it also suggests that the physical condition subscale might be improved by replacing one of these two items (perhaps PC2 who has a lower factor loading on the physical condition factor) with a new item focusing on walking endurance (i.e., *I can walk for a long time without stopping / Je peux marcher longtemps sans m'arrêter*). Similarly, the actual measure of physical strength involved hand strength whereas the items used in the physical strength subscales were related either to general strength (PS1: *I am stronger than others*) or to carrying heavy things (PS2: *I can carry heavy things*). In this context, the second item, which also did not perform as well as the first one (as evidenced by its weak factor loading on the physical strength factor) might be replaced by an item focusing more directly on youth performances on tests designed to measure strength (i.e., *I would be good in a test measuring strength* / *Je serai bon(ne) dans un test mesurant la force*). Indeed, this option seems to be more appropriate than refocusing this item specifically on hand strength, which lacks real world generalizability. Alternatively, as for physical condition, relying on objective tests of strength not limited to the hand might result in increased estimates of convergent validity for this factor.

Despite the strengths of our study, some limitations still must be taken into account when considering our results. The PSI-VS-ID-*R* was validated using a single sample of English-Australian and French-Canadian youth with ID. It thus remains unknown whether and how the present results would generalize to samples of youth with ID from other English (e.g., Americans, Canadians, English, New Zealanders, South-African) and French speaking countries (e.g., Algerians, Belgians, Moroccans, Swiss, Tunisians), or to other linguistic versions. Additionally, the test-retest reliability of the PSI-VS-ID-*R* was not examined and should thus be investigated in future research. This limitation is concerning given that the test retest reliability of the original PSI-VS-ID has not been re-examined since Maïano et al.'s (2009) original study, making it hard to clearly anticipate the extent to which scores on this instrument could be expected to be stable, or to fluctuate, over different periods.

In conclusion, results from the present study support the psychometric properties of the PSI-VS-ID-*R*. This questionnaire can be used in the context of group-based comparisons between English and French-speaking youth with ID and as functions of youth with ID's characteristics (i.e., age, body mass-index, ID level, sex, and frequency of sport involvement), at least among cultural and linguistic samples similar to those used in the present study. However, in line with the results from the convergent validity with physical fitness some reformulations were proposed for two items (PC2 and PS2). The performance of these reformulations should be investigated in research before their use can be recommended.

References

Aşçı, H.F., Maïano, C., Morin, A.J.S., Çağlar, E., & Bilgili, N. (2017). Validity and reliability of the very short form of the Physical Self-Inventory among Turkish adolescents. *Journal of Sports Sciences, 35*, 2060-2066.

American Thoracic Society Committee on Proficiency Standards for Clinical Pulmonary Function Laboratories (2002). American Thoracic Society Statement: Guidelines for the six minute walk test. *American Journal of Respiratory and Critical Care Medicine*, *166*, 111–117.

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

Asparouhov, T., & Muthén, B. (2017). Prior-Posterior Predictive P-Values. *Mplus Web Notes*, 22, version 2.

Asparouhov, T., & Muthén, B. (2021). Advances in Bayesian model fit evaluation for structural equation models. *Structural Equation Modeling, 28*, 1-14.

Asparouhov, T., Muthén, B., & Morin, A.J.S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances. *Journal of Management*, *41*, 1561-1577.

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.

Blomqvist, S., Olsson, J., Wallin, L., Wester, A., & Rehn, B. (2013). Adolescents with intellectual disability have reduced postural balance and muscle performance in trunk and lower limbs compared to peers without intellectual disability. *Research in Developmental Disabilities*, *34*), 198-206.

Byrne, B.M. (2002). Validating the measurement and structure of self-concept: Snapshots of past, present, and future research. *American Psychologist, 57*, 897–909.

Cattuzzo, M.T., dos Santos Henrique, R., Ré, A.H.N., de Oliveira, I.S., Melo, B.M., de Sousa Moura, M., ... & Stodden, D. (2016). Motor competence and health related physical fitness in youth: A systematic review. *Journal of Science and Medicine in Sport*, *19*, 123-129.

Council of Europe Committee for the Development of Sport (1988). *Eurofit: Handbook for the EUROFIT Tests of Physical Fitness*. Rome, Italy: Edigraf editorial grafica.

Craven, R. G., & Marsh, H. W. (2008). The centrality of the self-concept construct for psychological wellbeing and unlocking human potential: Implications for child and educational psychologists. *Educational and Child Psychology*, *25*, 104-118.

Dreiskämper, D., Tietjens, M., Honemann, S., Naul, R., & Freund, P.A. (2015). PSK-Kinder–Ein Fragebogen zur Erfassung des physischen Selbstkonzepts von Kindern im Grundschulalter. Grundschulalter [PSK – A questionnaire for assessing the physical self-concept of primary school children] *Zeitschrift für Sportpsychologie*, *22*, 97-111.

Depaoli, S., & Van de Schoot, R. (2017). Improving transparency and replication in Bayesian statistics: The WAMBS-Checklist. *Psychological Methods*, *22*, 240-261.

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

Guérin, F., Marsh, H.W., & Famose, J.-P. (2004). Generalizability of the PSDQ and its relationship to physical fitness: The European French connection. *Journal of Sport & Exercise Psychology*, *26*, 19-38.

Hartman, E., Smith, J., Westendorp, M., & Visscher, C. (2015). Development of physical fitness in children with intellectual disabilities. *Journal of Intellectual Disability Research*, *59*, 439-449.

Hagger, M.S., Asç1, 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.

Hoijtink, H., & van de Schoot, R. (2018). Testing small variance priors using prior-posterior predictive p values. *Psychological Methods, 23*, 561-569.

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

Hulteen, R.M., Morgan, P.J., Barnett, L.M., Stodden, D.F., & Lubans, D.R. (2018). Development of foundational movement skills: A conceptual model for physical activity across the lifespan. *Sports Medicine*, *48*, 1533-1540.

Kioumourtzoglou, E., Batsiou, S., & Theodorakis, Y. (1995). Age difference and physical fitness levels of mentally retarded and nonretarded individuals. *International Journal of Physical Education, 32*, 24-28.

Klavina, A., Zusa-Rodke, A., & Galeja, Z. (2017). The assessment of static balance in children with hearing, visual and intellectual disabilities. *Acta Gymnica*, *47*), 105-111.

Liu, X., & Daniels, M.J. (2006) A new algorithm for simulating a correlation matrix based on parameter expansion and reparameterization. *Journal of Computational and Graphical Statistics*, *15*, 897-914.

Lohbeck, A., Tietjens, M., & Bund, A. (2017). A short German Physical-Self-Concept Questionnaire for elementary school children (PSCQ-C): Factorial validity and measurement invariance across gender. *Journal of sports sciences, 35*, 1691-1696.

Mac Donncha, C., Watson, A.W., McSweeney, T., & O'Donovan, D.J. (1999). Reliability of Eurofit physical fitness items for adolescent males with and without mental retardation. *Adapted Physical Activity Quarterly*, *16*), 86-95.

Mai, Y., Zhang, Z., & Wen, Z. (2018). Comparing exploratory structural equation modeling and existing approaches for multiple regression with latent variables. *Structural Equation Modeling*, *25*, 737-479.

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

Maïano, C., Morin, A.J.S., 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.

Maïano, C., Morin, A.J.S., & Mascret, N. (2015a). Psychometric properties of the short form of the Physical Self-Description Questionnaire in a French adolescent sample. *Body Image*, *12*, 89-97.

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.

Maïano, C., Morin, A.J.S., & Probst, M. (2015b). Cross-linguistic validity of the French and Dutch versions of the very short form of the Physical Self-Inventory among adolescents. *Body Image*, *15*, 35-39.

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

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

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

Marsh, H.W., & Cheng, J.H.S. (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* (pp. 275-340). Hillsdale, NJ: Erlbaum.

Marsh, H.W., Hey, J., Roche, L.A., & Perry, C. (1997). The structure of physical self-concept: Elite athletes and physical education students. *Journal of Educational Psychology*, *89*, 369-380.

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

Marsh, H. W., Martin, A. J., & Jackson, S. (2010b). Introducing a short version of the Physical Self-Description Questionnaire: New strategies, short-form evaluative criteria, and applications of factor analyses. *Journal of Sport and Exercise Psychology*, *32*, 438-482.

Marsh, H.W., Martin, A.J., Yeung, A., & Craven, R. (2017). Competence self-perceptions. In A. J. Elliot., C. Dweck., & D. Yeager (Eds.), *Handbook of Competence and Motivation* (2nd edition; pp. 85-115). New York, NY: Guilford Press.

Marsh, H.W., Morin, A.J.S., Parker, P.D., & Kaur, G. (2014). Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology*, *10*, 85-110.

Marsh, H.W., Muthén, B.O., Asparouhov, T., Lüdtke, O., Robitzsch, A., Morin, A.J.S., & Trautwein, U. (2009). Exploratory Structural Equation Modeling, Integrating CFA and EFA: Application to Students' Evaluations of University Teaching. *Structural Equation Modeling*, *16*, 439-476.

Marsh, H.W., Nagengast, B., & Morin, A.J.S. (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.

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

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

McDonald, R.P. (1970). Theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical & Statistical Psychology, 23*, 1-21.

Morin A.J.S. & Maïano C. (2011a). Cross-validation of the short form of the Physical Self-Inventory (PSI-18) using exploratory structural equation modeling (ESEM). *Psychology of Sport and Exercise, 12*, 540-554.

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.

Morin, A.J.S, Maïano, C., Scalas, L. F., Aşçı, F. H., Boughattas, W., Abid, S., ... & Probst, M. (2018). Cross-cultural validation of the short form of the Physical Self Inventory (PSI-S). *Sport, Exercise, and Performance Psychology, 7*, 60-79.

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

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. 395-438). Charlotte, NC: Information Age.

Morin, A.J.S., Myers, N.D., & Lee, S. (2020). Modern factor analytic techniques: Bifactor models, exploratory structural equation modeling (ESEM) and bifactor-ESEM. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of Sport Psychology, 4th Edition* (pp. 1044-1073). London, UK. Wiley.

Muthén, B., & Asparouhov, T. (2012). Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods, 17*, 313-335.

Muthén, L.K., & Muthén, B. (2019). *Mplus user's guide.* Los Angeles, CA: Muthén & Muthén.

Pan, C.-C., Maïano, C., & Morin, A. J. S. (2018). Physical self-concept and body dissatisfaction among Special Olympics athletes: A comparison between sex, weight status, and culture. *Research in Developmental Disabilities, 76*, 1-11.

Protić-Gava, B.V., & Usković, F.R. (2016). Differences between the motor abilities of students attending a regular secondary school and those attending a secondary school for the education of children with disabilities. *Facta Universitatis, Series: Physical Education and Sport*, 43-50.

Roid, G.H., & Miller, L.J. (1997). *Leiter International Performance Scale-Revised*. Stoelting.

Scalas, L.F., Morin, A.J.S., Maïano, C., & Fadda, D. (2013). Contributo alla validazione italiana delle versioni breve e molto breve del Physical Self Inventory (PSI) per adolescenti [A contribution to the Italian validation of the short and very short versions of the Physical Self Inventory (PSI) for adolescents]. *Ricerche di Psicologia, 3*, 385-408.

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

Scott, H.M. & Havercamp, S.M. (2018). Comparisons of self and proxy report on health-related factors in people with intellectual disability. *Journal of Applied Research in Intellectual Disabilities, 31*, 927-936.

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

Shapiro, D.R., & Dummer, G.M. (1998). Perceived and actual basketball competence of adolescent males with mild mental retardation. *Adapted Physical Activity Quarterly, 15*, 179-190.

Skowroński, W., Horvat, M., Nocera, J., Roswal, G., & Croce, R. (2009). Eurofit special: European fitness battery score variation among individuals with intellectual disabilities. *Adapted Physical Activity Quarterly, 26*, 54-67.

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

Turk, V., Khattran, S., Kerry, S., Corney, R., & Painter, K. (2012). Reporting of health problems and pain by adults with an intellectual disability and by their carers. *Journal of Applied Research in*

Intellectual Disabilities, *25*, 155-165.

Utesch, T., Bardid, F., Büsch, D., & Strauss, B. (2019). The relationship between motor competence and physical fitness from early childhood to early adulthood: A meta-analysis. *Sports Medicine*, *49*, 541-551.

Weschler, D. (2003). *Weschler Intelligence Scale for Children-Fourth Edition*. Pearson.

Yun, J., & Ulrich, D. A. (1997). Perceived and actual physical competence in children with mild mental retardation. *Adapted Physical Activity Quarterly*, *14*, 285–297.

Table 1

Goodness-of-Fit Statistics for the PSI-VS-ID-R

Models	Nº	Description	FP	PPP	PPP 95% CI	PPPP	DIC	CFI	CFI 90%CI	TLI	TLI 90%CI	RMSEA	RMSEA 90% CI	CM	ΔCFI	ΔTLI	ΔRMSEA
Measurement model	1-1	CFA[a]	90	.000	94.075; 189.387	-	17844.233	.965	.959-.971	.949	.941-.957	.061	.056-.066	-	-	-	-
	1-2	BSEM-CL-*N*(0, .01)	234	.100	-18.694; 85.266	.796	17722.112	.993	.987-1.00	.990	.983-1.00	.027	.006-.038	-	-	-	-
	1-3	BSEM-CL-*N*(0, .001)	234	.002	23.730; 125.488	.011	17783.848	.982	.975-.989	.972	.961-.983	.045	.036-.054	-	-	-	-
	1-4	BSEM-CL-*N*(0, .005)	234	.059	-10.537; 92.228	.600	17749.165	.991	.983-.998	.985	.974-.996	.033	.016-.043	-	-	-	-
	1-5	Convergent validity[a]	152	.294	-39.466; 64.675	-	10619.413	.998	.993-1.00	.986	.935-1.00	.044	.000-.095	-	-	-	-
DIF: Linguistic versions	2-1	Null effects	236	.000	55.144; 162.883	.660	18200.029	.974	.967-.981	.967	.958-.976	.047	.040-.053	-	-	-	-
	2-2	Saturated	254	.148	-26.273; 84.191	.800	18137.403	.994	.988-1.00	.991	.982-1.00	.025	.000-.037	2-1	+.020	+.024	-.022
	2-3	Factors only	245	.032	-3.068; 105.490	.780	18150.696	.988	.981-.996	.984	.974-.994	.033	.020-.041	2-1	+.014	+.017	-.014
DIF: Age	3-1	Null effects	236	.030	-2.279; 104.606	.767	19054.117	.989	.982-.996	.986	.977-.995	.031	.019-.039	-	-	-	-
	3-2	Saturated	254	.102	-19.786; 90.227	.798	19054.607	.993	.987-1.00	.990	.981-1.00	.026	.000-.038	3-1	+.004	+.004	-.005
	3-3	Factors only	245	.103	-19.498; 89.268	.785	19045.138	.993	.987-1.00	.990	.982-1.00	.025	.000-.036	3-1	+.004	+.004	-.006
DIF: BMI	4-1	Null effects	236	.050	-8.838; 99.537	.779	19944.522	.990	.982-.997	.987	.978-.996	.029	.017-.038	-	-	-	-
	4-2	Saturated	254	.101	-19.876; 92.066	.788	19951.363	.992	.986-1.00	.989	.980-1.00	.027	.007-.039	4-1	+.002	+.002	-.002
	4-3	Factors only	245	.073	-14.622; 94.367	.792	19945.812	.991	.984-.998	.988	.978-.998	.028	.012-.038	4-1	+.001	+.001	-.001
DIF: ID-level	5-1	Null effects	236	.019	3.548; 111.592	.747	18189.145	.987	.980-.994	.983	.974-.993	.033	.022-.041	-	-	-	-
	5-2	Saturated	254	.183	-30.933; 79.933	.801	18175.007	.995	.989-1.00	.993	.983-1.00	.022	.000-.035	5-1	+.008	+.010	-.011
	5-3	Factors only	245	.163	-28.521; 81.950	.805	18168.046	.995	.988-1.00	.993	.984-1.00	.022	.000-.035	5-1	+.008	+.010	-.011
DIF: Sex	6-1	Null effects	236	.030	-2.296; 105.344	.768	18219.220	.988	.981-.995	.985	.976-.994	.032	.021-.040	-	-	-	-
	6-2	Saturated	254	.124	-22.978; 87.395	.798	18215.611	.993	.987-1.00	.990	.981-1.00	.026	.000-.038	6-1	+.005	+.005	-.006
	6-3	Factors only	245	.117	-22.139; 86.429	.819	18208.405	.993	.987-1.00	.990	.982-1.00	.025	.000-.036	6-1	+.005	+.005	-.007
DIF: FSI	7-1	Null effects	236	.013	8.606; 116.288	.735	19156.256	.985	.978-.993	.982	.973-.991	.035	.025-.043	-	-	-	-
	7-2	Saturated	254	.170	-29.000; 82.766	.797	19143.091	.994	.988-1.00	.992	.982-1.00	.024	.000-.037	7-1	+.009	+.010	-.009
	7-3	Factors only	245	.189	-30.994; 78.514	.809	19131.318	.995	.989-1.00	.993	.985-1.00	.021	.000-.034	7-1	+.010	+.011	-.014

Notes. *p ≤ .01; PSI-VS-ID-R = Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities – Revised version; CFA = confirmatory factor analytic model (with no cross-loadings); BSEM = Bayesian structural equation model (with cross-loadings); CL = cross-loadings; *N* = normal distribution; FP = free parameters; PPP = posterior predictive *p* value; PPPP = prior posterior predictive *p* value; CI = confidence interval (90% or 95%); DIC = deviance information criterion; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CM = comparison model; Δ = change from the CM; DIF = differential item functioning; BMI = body-mass index; ID = intellectual disability; FSI = frequency of sport involvement; [a] Given that no priors were specified in this model, the PPPP was not estimated.

Table 2

Standardized Parameters Estimates from the Confirmatory Factor Analytic Model of the PSI-VS-ID-R Without Cross-Loadings (Model 1-1)

Items	GSW (λ)	PSW (λ)	PC (λ)	SC (λ)	PA (λ)	PS (λ)	FL (λ)	BAL (λ)	SPE (λ)	δ
GSW1	.835									.303
GSW2	.737									.457
PSW1		.840								.294
PSW2		.876								.233
PC1			.842							.290
PC2			.776							.398
SC1				.900						.191
SC2				.901						.188
PA1					.789					.378
PA2					.784					.385
PS1						.738				.455
PS2						.642				.588
FL1							.901			.187
FL2							.876			.233
BAL1								.848		.280
BAL2								.887		.214
SPE1									.877	.231
SPE2									.862	.257
ω	.765	.848	.792	.895	.764	.646	.883	.859	.861	
Latent Factor Correlations										
	GSW	PSW	PC	SC	PA	PS	FL	BAL	SPE	
GSW	-									
PSW	.886	-								
PC	.430	.542	-							
SC	.592	.629	.661	-						
PA	.912	.894	.521	.619	-					
PS	.741	.763	.692	.801	.837	-				
FL	.438	.478	.724	.514	.531	.550	-			
BAL	.460	.478	.613	.546	.500	.500	.697	-		
SPE	.545	.586	.835	.738	.578	.661	.721	.680	-	

Notes. BAL = balance; FL = flexibility; GSW = global self-worth; PA = physical attractiveness; PC = physical condition; PS = physical strength; PSW = physical self-worth; PSI-VS-ID-*R* = Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities – Revised version; SC = sport competence; SPE = running speed; λ = factor loadings; δ = Uniquenesses; ω = McDonald's omega coefficient of composite reliability; Parameter estimates are the median from the Bayesian posterior distribution; All parameter estimates are associated with 95% credibility intervals excluding the value of 0 and can thus be considered to be meaningful.

Table 3

Standardized Parameters Estimates from the Bayesian Structural Equation Modeling with Cross-Loadings (0, .01) of the PSI-VS-ID-R (Model 1-2)

Items	GSW (λ)	PSW (λ)	PC (λ)	SC (λ)	PA (λ)	PS (λ)	FL (λ)	BAL (λ)	SPE (λ)	δ
GSW1	.933	-.003	.004	-.011	-.008	.010	.004	-.033	-.078	.267
GSW2	.660	.006	-.018	-.018	.033	-.011	-.015	.052	.097	.449
PSW1	-.002	.999	-.016	-.008	-.048	-.032	-.048	-.014	.003	.220
PSW2	.002	.718	.005	-.037	.094	.045	.063	.040	-.019	.270
PC1	-.004	-.022	.921	-.067	-.010	.013	-.007	.006	.010	.255
PC2	.001	.018	.676	.075	.009	-.011	.031	-.003	-.001	.420
SC1	-.058	-.050	.023	.993	-.004	.006	-.025	.021	.001	.120
SC2	.073	.048	-.065	.839	.008	.011	.021	-.031	.008	.218
PA1	-.025	-.025	-.017	-.037	.877	.009	-.008	-.001	.037	.334
PA2	.060	.055	.016	.059	.634	.016	.015	.005	-.055	.404
PS1	-.010	-.024	-.019	-.034	.027	.903	-.007	-.016	-.017	.321
PS2	.038	.072	.089	*<u>.163</u>*	-.036	.282	.016	.033	.052	.611
FL1	-.011	.044	.060	.010	.007	-.012	.817	-.049	.039	.237
FL2	.006	-.043	-.033	-.020	-.004	.009	.957	.044	-.020	.152
BAL1	-.007	.020	.080	-.001	.016	.012	-.050	.780	.013	.325
BAL2	.009	-.008	-.060	.003	-.011	-.016	.043	.954	-.004	.133
SPE1	.032	.030	.019	.002	.014	-.006	-.010	-.054	.877	.208
SPE2	-.030	-.032	-.002	.008	-.011	.004	.023	.056	.849	.255
ω	.780	.857	.791	.909	.756	.601	.890	.868	.865	
Latent Factor Correlations										
	GSW	PSW	PC	SC	PA	PS	FL	BAL	SPE	
GSW	-									
PSW	.847	-								
PC	.433	.532	-							
SC	.570	.615	.663	-						
PA	.843	.824	.496	.574	-					
PS	.653	.665	.603	.684	.790	-				
FL	.430	.433	.683	.499	.501	.470	-			
BAL	.436	.433	.583	.530	.460	.410	.675	-		
SPE	.533	.563	.809	.718	.548	.561	.685	.651	-	

Notes. PSI-VS-ID-*R* = Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities – Revised version; λ = factor loadings; δ = Uniquenesses; GSW = global self-worth; PSW = physical self-worth; PC = physical condition; SC = sport competence; PA = physical attractiveness; PS = physical strength; FL = flexibility; BAL = balance; SPE = running speed; ω = composite reliability; Parameter estimates are the median from the Bayesian posterior distribution; All main loadings and correlations estimates are associated with 95% credibility intervals excluding the value of 0 and can thus be considered to be meaningful; Cross-loadings associated with 95% credibility intervals excluding the value of 0 are underlined and italicized.

Table 4

Latent Factor Correlations from the Convergent Validity Analyses of the PSI-VS-ID-R

PSI-VS-ID-R	Global self-Esteem	Perceived Physical Appearance	Perceived Physical Abilities	Physical Fitness			
				Functional Walking Capacity	Running Speed	Strength	Flexibility
Global self-worth	.743*	.767*	.638*	-.034	-.027	-.077	.004
Physical self-worth	.769*	.792*	.708*	-.022	-.048	-.055	.005
Physical condition	.540*	.556*	.763*	.052	-.156*	.067	.012
Sport competence	.605*	.577*	.819*	.008	-.101	.024	-.003
Physical attractiveness	.747*	.813*	.638*	-.098	.035	-.132*	.027
Physical strength	.695*	.715*	.741*	-.092	-.012	-.051	.018
Flexibility	.450*	.497*	.585*	-.063	-.006	-.096	.119*
Balance	.411*	.420*	.510*	-.056	.001	-.036	.108
Running speed	.575*	.594*	.805*	.028	-.167*	.030	.009

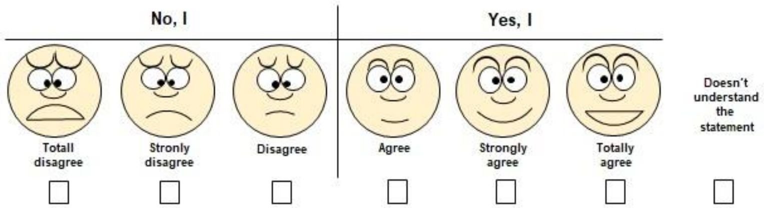
Notes. PSI-VS-ID-*R* = Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities – Revised version; Corresponding measures of physical self-perceptions and physical fitness are in greyscale; Parameter estimates are the median from the Bayesian posterior distribution; Correlations associated with 95% credibility intervals excluding the value of 0 are marked by an asterisk.

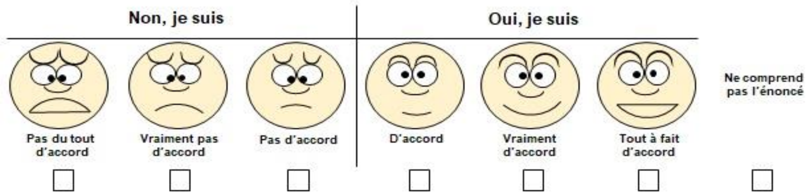
Appendix A

Items and Response Scales of the English and French Versions of the PSI-VS-ID-R

N°	Scales	English items	French items
1	GSW1	I like myself	Je m'aime bien
2	PSW1	I am happy about all the things I can do with my body	Je suis content(e) de toutes les choses que j'arrive à faire avec mon corps
3	PS1	I am stronger than others	J'ai plus de force que les autres
4	PA1	My body is nice to look at	Mon corps est beau à regarder
5	PS2	I can carry heavy things	J'arrive à porter quelque chose de lourd
6	PC1	I can run a long time without getting tired	Je peux courir beaucoup sans être fatigué
7	SC1	I am good in all sports	Je suis bon(ne) dans tous les sports
8	PA2	Everybody finds me good-looking	Les gens me trouvent beau(belle)
9	PSW2	I am happy with myself and what I can do with my body	Je suis content de moi et de ce que je peux faire avec mon corps
10	PC2	I can run 10 times around the basketball court without stopping	Je peux faire 10 tours de terrain de basket-ball sans m'arrêter
11	SC2	I do things well in sports	Je réussis bien à faire les choses en sport
12	GSW2	I want to stay as I am	Je veux rester comme je suis
13	FL1	My body is flexible	J'ai un corps souple
14	BAL1	I am good at keeping my balance	Je suis bon(ne) pour garder mon équilibre
15	SPE1	I am good at running fast	Je suis bon(ne) pour courir vite
16	FL2	I would be good in a test where you have to be flexible	Je serais bon(ne) dans un test où il faut être souple
17	BAL2	I would be good in a test where you have to keep your balance	Je serais bon(ne) dans un test où il faut garder son équilibre
18	SPE2	I would be good in a test where you have to run fast	Je serais bon(ne) dans un test où il faut courir vite

Answer scales





Notes. BAL = balance; FL = flexibility; GSW = global self-worth; PA = physical attractiveness; PC = physical condition; PS = physical strength; PSW = physical self-worth; PSI-VS-ID-*R* = Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities –Revised version; SC = sport competence; SPE = running speed.

Online Supplements for:

Validation of a Revised Version of the Physical Self-Inventory – Very Short form for Youth with Intellectual Disabilities (PSI-VS-ID-*R*): A Bayesian Structural Equation Modeling Approach

S1. Scale Development

Objectives

The first objective of this preliminary study was to develop six additional items measuring perceived balance, flexibility, and speed. The second objective of this preliminary study was to improve the response scale and format of the Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities (PSI-VS-ID). The third objective of this preliminary study was to adapt to English the resulting improved version of the PSI-VS-ID (PSI-VS-ID-*R*). The fourth objective of this preliminary study was to examine the format and clarity of the English and French versions of the PSI-VS-ID-*R* among youth with ID.

Method

Participants

A sample of 34 youth (35% girls; aged between 13 and 21) with mild to moderate-severe levels of ID participated in this preliminary study, including 20 were English-speaking Australians and 14 were French-speaking Canadians. A first subsample of 18 youth (8 in Canada and 10 in Australia) was solicited to assess the format and clarity of the PSI-VS-ID-*R*. A second subsample of 16 youth (6 in Canada and 10 in Australia) was solicited to assess the format and clarity of the final adapted version of the PSI-VS-ID-*R*. The procedures used in this pilot study were identical to those used in the main study, and were approved by the same research ethics committees. However, in order to maximise youth's understanding and to facilitate discussion, the PSI-VS-ID-*R* was administered individually (at school) by trained research assistants using a read-aloud assisted procedure. This administration was mainly focused on assessing the youth's level of understanding and the ease with which they could respond to the items.

Measures

To develop the three additional subscales, two items measuring perceived flexibility were first selected from the French version of the PSDQ (Guérin et al., 2004; Marsh et al., 1994) and adapted (i.e., the sentence was maximally simplified whilst retaining the original meaning) by two members of the research team familiar with the use of self-concept questionnaires among youth with ID. In addition, four items measuring perceived balance and running speed were developed by the same two members of the research team in a way that was consistent with the items already included in the original PSI-VS-ID. To maximize youth's understanding of all items, the wording of all items were associated with pictograms, presented above the words and illustrating their meaning.

Second, to further increase youth's understanding of the items, words were added to the original purely graphical six-point response scale (i.e., ranging from "Totally disagree" associated with a "very unhappy face" to "Totally agree" associated with a "very happy face"). The original, and revised, response scale was inspired by the Wong–Baker facial pain rating scale (Wong & Baker 1988). Additionally, a "do not understand the statement" option was added to the response scale for situations in which respondents remained unable to understand the item.

Third, the preliminary French version of the PSI-VS-ID-*R* was translated into English by two bilingual members of the research team. Then, this preliminary English version was back-translated into French by two other bilingual members of the research team. The back-translated French version was then compared with the original French version and discrepancies were resolved in committee (by adjusting the English version) until a consensus was reached. This whole process of adaptation and translation was conducted in collaboration with school personnel and professionals (i.e., teachers, psychologists and physical educators) familiar with youth with ID.

Results

Responses provided by the first subsample of participants showed that the formulation of some of the six additional items remained hard to understand (more specifically by youth with more severe levels of ID). These results also revealed that the adjusted response scale seemed easy to understand for all participants, but that some participants did not use the exact wording of the verbal anchors of this response scale when answering items, preferring to use a simpler "yes" or "no". Therefore, the problematic words were replaced by simpler words preserving the same meaning, and the verbal anchors of the response scale were revised as follow (in italic and underlined): "*No, I* totally disagree", "*No, I* strongly disagree", "*No, I* disagree", "*Yes, I* agree", "*Yes, I* strongly agree", and "*Yes, I* totally agree". Finally, a template comprising a graphical displays and pictograms was developed to explain to the youth how to use the answer scale. The updated version of the PSI-VS-ID-*R* was then administered to the second subsample of youth with ID. Results supported the adequacy and suitability of the final English and French versions of the PSI-VS-ID-*R* for use among youth with ID.

References

Guérin, F., Marsh, H.W., & Famose, J.-P. (2004). Generalizability of the PSDQ and its relationship to physical fitness: The european French connection. *Journal of Sport & Exercise Psychology, 26,* 19-38.

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

Table S1

Frequency and Percentage of Missing Responses and Doesn't Understand the Statement Option for all items from the PSI-VS-ID-R

N°	Scales	English items	Missing		DNU		Total	
			Frequency	%	Frequency	%	Frequency	%
1	GSW1	I like myself	1	.28%	2	.57%	3	0.85%
2	PSW1	I am happy about all the things I can do with my body	1	.28%	5	1.42%	6	1.71%
3	PS1	I am stronger than others	4	1.14%	3	.85%	7	1.99%
4	PA1	My body is nice to look at	4	1.14%	3	.85%	7	1.99%
5	PS2	I can carry heavy things	3	.85%	1	.28%	4	1.14%
6	PC1	I can run a long time without getting tired	3	.85%	4	1.14%	7	1.99%
7	SC1	I am good in all sports	3	.85%	2	.57%	5	1.42%
8	PA2	Everybody finds me good-looking	4	1.14%	5	1.42%	9	2.56%
9	PSW2	I am happy with myself and what I can do with my body	3	.85%	8	2.28%	11	3.13%
10	PC2	I can run 10 times around the basketball court without stopping	2	.57%	11	3.13%	13	3.70%
11	SC2	I do things well in sports	2	.57%	3	.85%	5	1.42%
12	GSW2	I want to stay as I am	2	.57%	5	1.42%	7	1.99%
13	FL1	My body is flexible	4	1.14%	8	2.28%	12	3.42%
14	BAL1	I am good at keeping my balance	4	1.14%	4	1.14%	8	2.28%
15	SPE1	I am good at running fast	4	1.14%	6	1.71%	10	2.85%
16	FL2	I would be good in a test where you have to be flexible	6	1.71%	9	2.56%	15	4.27%
17	BAL2	I would be good in a test where you have to keep your balance	2	.57%	4	1.14%	6	1.71%
18	SPE2	I would be good in a test where you have to run fast	2	.57%	7	1.99%	9	2.56%
		Mean		.85%		1.42%		2.28%
		Standard deviation		.34%		.74%		.85%

Notes. DNU = Doesn't Understand the Statement; BAL = balance; FL = flexibility; GSW = global self-worth; PA = physical attractiveness; PC = physical condition; PS = physical strength; PSW = physical self-worth; PSI-VS-ID-*R* = Physical Self-Inventory – Very Short form for youth with Intellectual Disabilities –Revised version; SC = sport competence; SPE = running speed.

Table S2

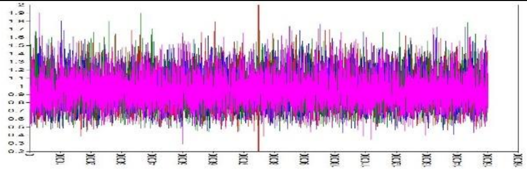
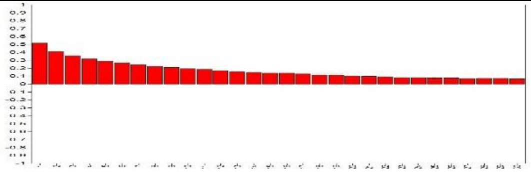
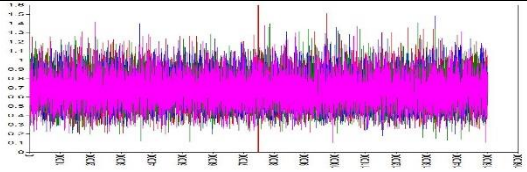
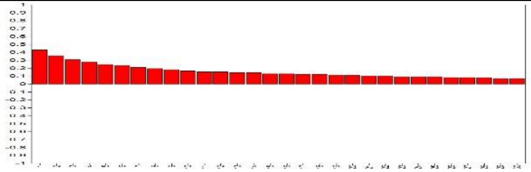
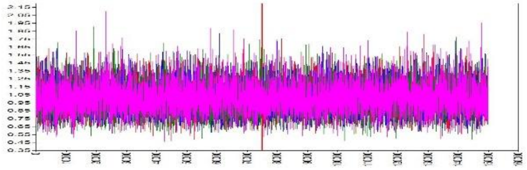
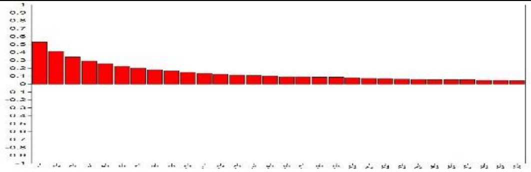
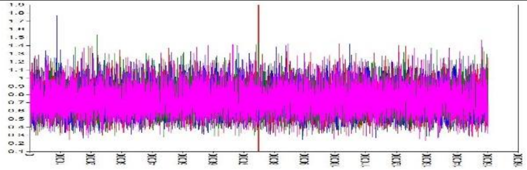
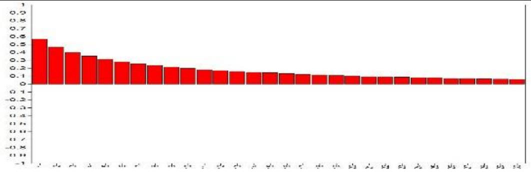
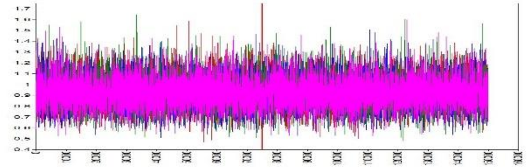
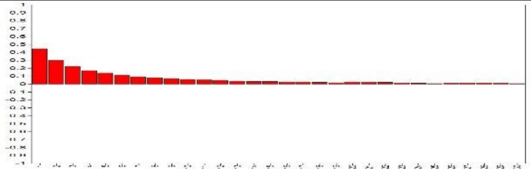
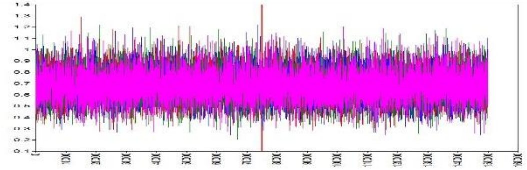
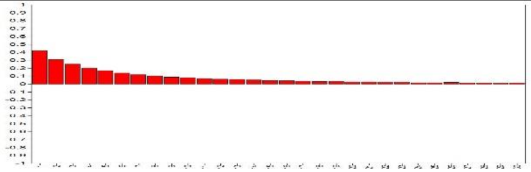
Measurement Model of the Self-Concept Scales used in the Analyses of Convergent Validity

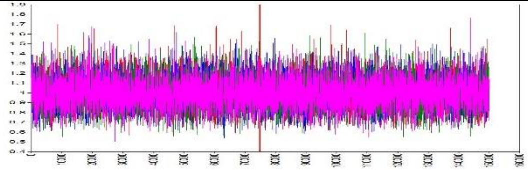
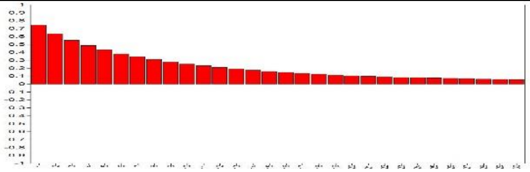
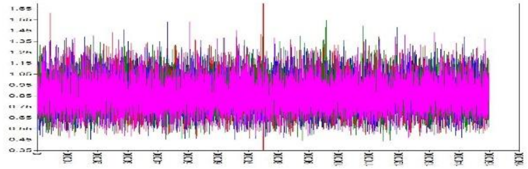
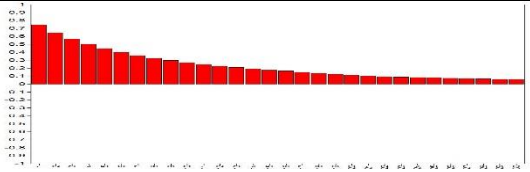
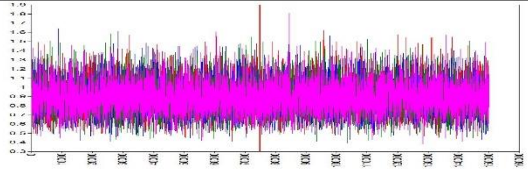
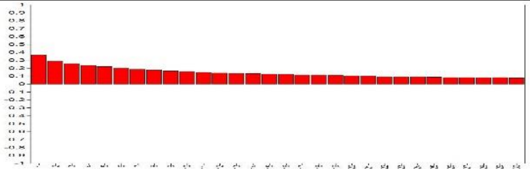
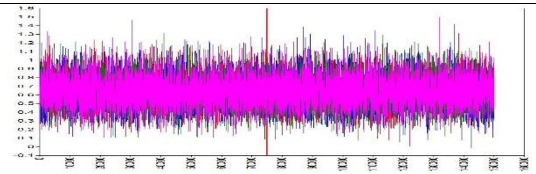
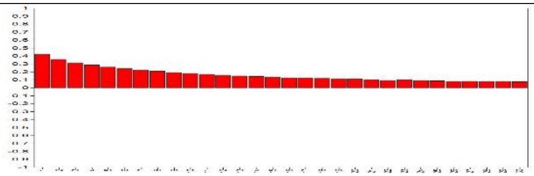
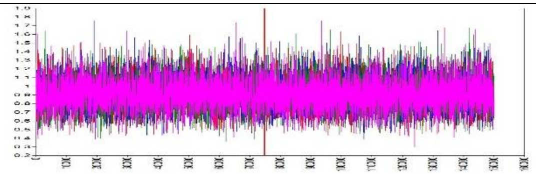
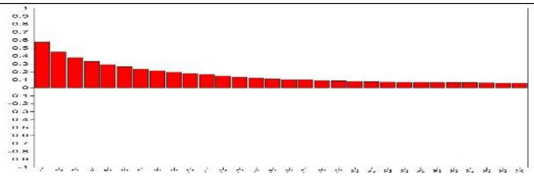
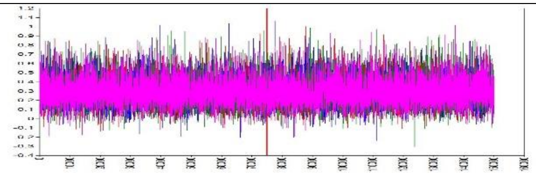
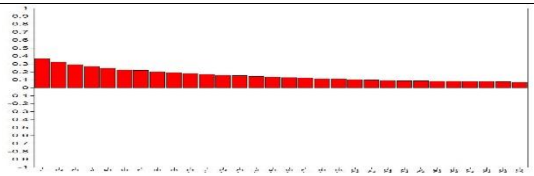
Items	GSE (λ)	PPA (λ)	PPHA (λ)	δ
GSE1	.707	-.016	-.016	.528
GSE2	.745	.020	.002	.414
GSE3	.644	.056	.046	.472
GSE4	.781	-.011	.023	.375
GSE5	.732	.007	.031	.420
GSE6	.835	-.047	-.054	.416
GSE7	.686	.071	.000	.443
GSE8	.668	.021	.052	.475
PPA1	-.067	.836	-.035	.416
PPA2	.038	.743	.035	.359
PPA3	.019	.749	.019	.390
PPA4	.058	.819	-.051	.305
PPA5	.016	.776	.045	.324
PPA6	-.029	.848	.019	.295
PPA7	-.023	.771	-.016	.446
PPA8	.041	.740	-.005	.405
PPHA1	-.004	.019	.684	.515
PPHA2	.050	.047	.648	.483
PPHA3	.035	-.008	.678	.510
PPHA4	.026	.045	.634	.530
PPHA5	-.065	-.050	.865	.371
PPHA6	-.034	-.043	.765	.484
PPHA7	-.004	.007	.758	.419
PPHA8	.073	.024	.643	.486
ω	.905	.931	.895	
Latent Factor Correlations				
GSE	-			
PPA	.790	-		
PPHA	.680	.679	-	

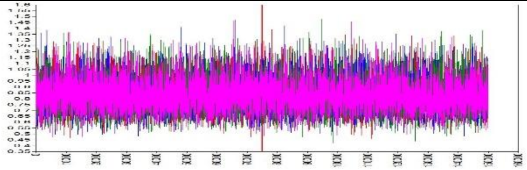
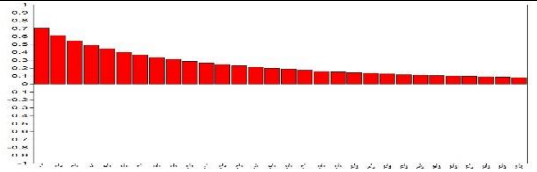
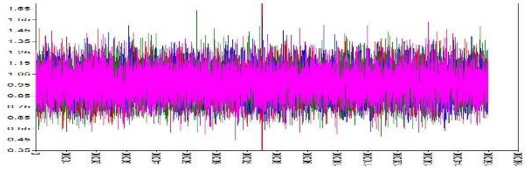
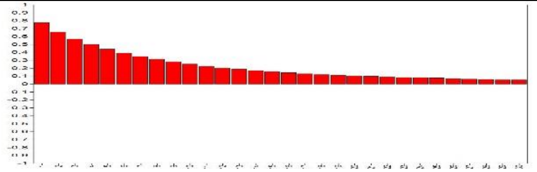
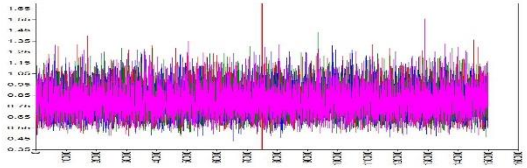
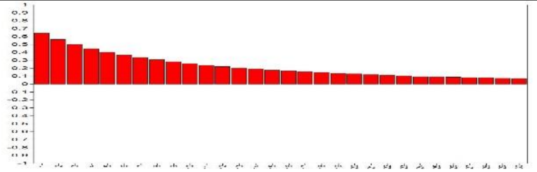
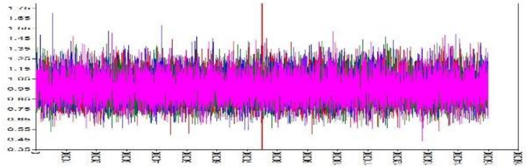
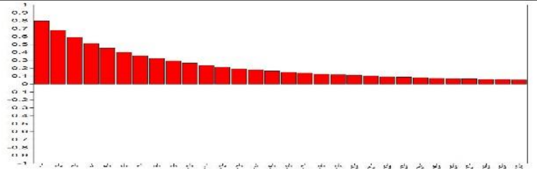
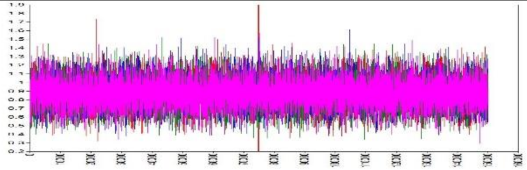
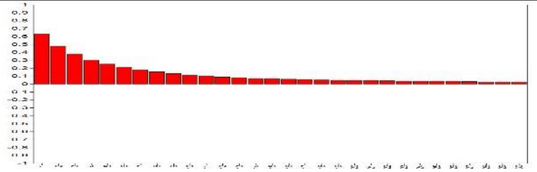
Notes. GSE = Global Self-Esteem; PPA = Perceived Physical Appearance; PPHA = Perceived Physical Abilities; λ =Factor loadings (target loadings are in greyscale); δ = Uniquenesses; ω = McDonald's omega coefficient of composite reliability. Bayesian structural equation modeling representation of responses to the SDQ-IA-ID specified following Muthén and Asparouhov's (2012) recommendations. Results: Posterior predictive *p*-value [PPP] = .577; PPP 95% confidence interval [CI] = -80.754;65.675; prior posterior predictive *p*-value= .516; free parameters = 399; deviance information criterion = 22487.481; comparative fit index [CFI] =1.00; CFI 90%CI = .994-1.00; Tucker-Lewis index [TLI] = 1.00; TLI 90%CI =.689-1.00; root mean square error of approximation [RMSEA] = .000; RMSEA 90% CI = .000-.129. Parameter estimates are the median from the Bayesian posterior distribution. All main loadings and correlations estimates are associated with 95% credibility intervals excluding the value of 0 and can thus be considered to be meaningful.

Table S3

Bayesian Posterior Parameter Trace Plot and Bayesian Autocorrelation Plots of all Items in the Bayesian Model with a Mean of 0 and Variance of .01

Items	Bayesian posterior parameter trace plots	Bayesian aurocorrelation plots
GSW1		
GSW2		
PSW1		
PSW2		
PC1		
PC2		

Items	**Bayesian posterior parameter trace plots**	**Bayesian aurocorrelation plots**
SC1		
SC2		
PA1		
PA2		
PS1		
PS2		

Items	**Bayesian posterior parameter trace plots**	**Bayesian aurocorrelation plots**
FL1		
FL2		
BAL1		
BAL2		
SPE1		
SPE2