European Psychomotricity Journal, 2019; 11, 1, 50-69
SciPsyMot-Hellas

EPJ

The long and short form of the physical self-description questionnaire: psychometric properties in an adolescent sample

J. Simons[1] & T. Van Damme[1,2]

[1]Faculty of Kinesiology and Rehabilitation Sciences, Department of Rehabilitation Sciences, KU Leuven, O & N IV, Bus 1510, B 3000 Leuven, Belgium
[2]University Psychiatric Centre, Campus Leuven, Gasthuisberg, Child & Adolescent Psychiatry, Herestraat 49, B 3000 Leuven, Belgium,
Johan.simons@kuleuven.be,

Abstract

The objective of this study is to examine empirical evidence for the psychometric properties of the long and short form of the Dutch PSDQ in a Flemish adolescent sample. The sample used in the present study consisted of 1417 participants (687 boys, 730 girls). The study provided satisfactory Cronbach α estimates and test-retest stabilities for both the long and the short form. Confirmatory factor analysis did not provide conclusive evidence for the construct validity of the long and short form, but the exploratory factor analysis confirmed the original structure of the long version. Further research concerning the validity is needed before the short form of the Dutch PSDQ can be widely used in practice with Flemish adolescents.

Keywords: physical self-concept, physical self-description questionnaire, psychometric properties, gender, Flemish adolescents

Introduction

According to Shavelson, Hubner and Stanton (1976) self-concept or self-esteem are the terms used to describe how a person perceives and evaluates him or herself within the context of their experiences and the environment in which they live. They developed a hierarchical and multidimensional model, in which self-concept, at the apex of the hierarchy, is divided into academic and non-academic self-concept. Physical self-concept (PSC) is a part of the non-academic self-concept. The bottom of the hierarchy describes the actual behavior in specific situations. Physical self-concept is a widely used and investigated construct in physical education, sports and psychology (Marsh & Cheng, 2012). This subjective feeling of competence is the basis for internal motivation and the onset of learning and development. Following the research of Shavelson, Hubner and Stanton (1976) various instruments measuring self-concept were developed such as the Physical Self Concept Inventory (Chung, 1996), Children's Physical Self-Concept Scale (Stein, Bracken, haddock, & Shadish, 1998), Children and Youth Physical Self Perception Profile (Whitehead, 1995) and Physical Self-Description Questionnaire (Marsh, Hau, Sung, & Yu, 2007) (see for further information: Sypsa & Simons, 2008). Out of these instruments the self-description questionnaires (SDQ) became widely recognized as among the strongest multidimensional self-concept instruments in terms of psychometric properties and construct validation research (Boyle, 1994; Hattie,

1992; Wylie, 1989). These instruments are designed to assess different aspects of self-concept, namely academic, social, physical and emotional aspects (Marsh, Ellis, Parada, Richards, & Heubeck, 2005; Marsh & O'Neill, 1984).

Further research of Marsh and Shavelson (1985) emphasized the multidimensionality of self-concept rather than its hierarchy, which led to a revision of the original model of self-concept (Marsh, Byrne, & Shavelson, 1988; Marsh & Shavelson, 1985). In a review of earlier self-concept instruments by Byrne (1966) it was concluded that many of these measures either ignored some or summed up all the aspects of physical self-concept into one single, one-dimensional score. As a result, Marsh recommended that, instead of only measuring the overall self-concept, every aspect of self-concept (i.e. the physical self-concept) should also be measured separately. According to these recommendations Marsh, Richards, Johnson, Roche and Tremayne (1994) developed the physical self-description questionnaire (PSDQ) for adolescents 12 years of age and older. Other instruments, specifically designed for measuring physical self-concept in adolescents, are available, such as the Physical Self-Perception Profile (PSPP; Fox & Corbin, 1989) and the Richards Physical Self-Concept Scale (RPSCS; Maïano et al., 2008; Richards, 1988; Sypsa & Simons, 2008). However, in comparison to these instruments, the PSDQ includes more dimensions of the physical self-concept and its psychometric properties have been more extensively researched (Marsh & Cheng, 2012).

The PSDQ has been translated into various languages such as French, Greek, Italian, Spanish, Turkish, Flemish and Hebrew (Guérin, Marsh & Famose, 2004; Marsh, Asci & Marco, 2002; Marsh, Bar-Eli, Zach & Richards, 2006; Marsh, Marco & Hiilya, 2002; Simons, Capio, Adriaenssens, Delbroek & Vandenbussche, 2012; Tsorbatzoudis, 2005). However, a disadvantage of the PSDQ is the large number of items included in the questionnaire and consequently considerable administration time is needed. This is impractical for usage on either occasions or in a battery of tests, which stresses the importance of the development of a more concise version of the PSDQ. A short version of the PSDQ (PSDQ-S) was introduced by Marsh, Martin and Jackson (2010). Although this short form consists of 40 items, instead of 70, it is designed to question the same aspects of physical self-concept as the long form of the questionnaire without sacrificing reliability or validity. The applicability of the PSDQ-S for an Australian sample has been confirmed in previous research (Marsh, Martin & Jackson, 2010), but there is a lack of research conducted in a Flemish adolescent sample using the translated Dutch PSDQ-S.

Various studies show that significant gender differences exist in the self-reported physical self-concept. In comparison to boys, girls tend to score lower on most of the subscales of physical self-concept, on both the PSDQ as well as on the PSDQ-S (Marsh, Martin & Jackson, 2010; Marsh, Richards, Johnson & Tremayne, 1994). The main objective of the present study is to discuss the reliability (internal consistency and test-retest-stability) and construct validity of the Dutch version of the PSDQ and the PSDQ-S in a Flemish adolescent sample. Furthermore, the correlation between the results on the Dutch PSDQ and PSDQ-S will be examined. Based on finding and conclusions from prior research, it is hypnotized that girls will report lower overall physical self-concepts scores in comparison to boys.

Method

Participants

A total sample of 1417 Dutch speaking adolescents, attending secondary schools distributed across Flanders, (Belgium) participated in this study. The data have been collected in 12 different secondary schools, by different researchers, and were finally outlined in one joint dataset. The age of the participants ranged from 12 years and 8 months old up to 19 years old (*M* = 15 years and 6 months old; *SD*= 1 year and 4 months). The sample of adolescents consisted of 687 boys and 730 girls, with *M*= 15 years and 6 months (*SD* = 1 year and 4 months) and *M =15* years and 6 months (*SD* = 1 year and 3 months) respectively. There was no significant mean age difference between the boys and girls ($p > .05$). The participants were pupils from the six degrees of secondary school and followed either general secondary education, technical secondary education or vocational secondary education. The participants were all informed of the study and had the right to refuse participation and remain anonymous.

Measure

According to the outline from the PSDQ package presented by the university of Sydney (Marsh, 1999), the PSDQ (Marsh, Richards, Johnson, Roche, & Tremayne, 1994) is a 70-item scale that consists of eleven subscales or factors. The PSDQ, originally developed for adolescents, measures nine specific components of the physical self-concept: Physical Activity (6 items), Appearance (6 items), Body Fat (6 items), Coordination (6 items), Endurance (6 items), Flexibility (6 items), Health (8 items), Sport Competence (6 items) and Strength (6 items). Furthermore, two global components are included. On one hand the Global

Physical Self-Concept (6 items), which can be characterized as the positive feelings about one's physical self. On the other hand, the Global Self-Esteem (8 items), which can be characterized as the overall positive feelings one has about oneself. Each factor is represented by 6-8 items and every item is a simple declarative statement to which the subjects can answer by using a 6-point true or false Likert-scale ('false', 'mostly false', 'more false than true', 'more true than false', 'mostly true', and 'true').

Previous research of the PSDQ demonstrate a good reliability across the 11 subscales of the PSDQ (Fletcher & Hattie, 2004; Marsh, 1996b; Marsh, Richards, Johnson, Roche, & Tremayne, 1994). Schipke and Freund's (2012) meta-analytic reliability generalization study determined that the mean reliability estimates for all subscales are above α of .80, with the Health subscale obtaining the lowest score (α =.82). The PSDQ has a good test-retest stability over short term (Median r = .83) and longer term (Median r = .69; Marsh, 1996b), a well-defined, replicable factor structure determined with a confirmatory factor analysis (Marsh, 1996b; Marsh, Richards, Johnson, Roche & Tremayne,1994), a gender invariant factor structure determined by a multiple-group confirmatory factor analysis (Marsh et al., 1994), convergent and discriminant validity of responses to three PSC instruments as shown by multitrait-multimethod studies and convergent and discriminant validity as shown by PSDQ in relations with external criteria (Marsh, 1993; Marsh et al., 1994; Marsh, 1996a; Marsh, 1997). Tietjens, Freund, Büsch and Strauss (2012) conclude in their research that, while an acceptable goodness of fit for the 11-factor solution is shown in the results of the confirmatory factor analysis, the problematic use for a yet undefined, larger minority of subjects is discovered by using a mixed Rasch model. Because some scales suffer from differential item functioning, it cannot be assumed that every subject uses these scales in the same manner. The PSDQ is originally designed for adolescents, yet statistical support in a wide variety of Australian samples confirmed its appropriateness for application within an adult population (Marsh, 2002; Marsh, Asci & Marco, 2002) and for adolescent elite and non-elite athletes (Marsh, Hey, Roche & Perry, 1997). The PSDQ has been demonstrated to be applicable in different cross-cultural settings, in adults and adolescents, using translated versions (i.a. Guérin, Marsh, & Famose, 2004; Marsh, Asci, & Marco, 2002; Marsh, Bar-Eli, Zach, & Richards, 2006; Marsh, Marco, & Hiilya, 2002; Tsorbatzoudis, 2005;). The wide range of previous research described above proved the PSDQ to be a psychometrically robust instrument.

The PSDQ-S, developed by Marsh, Martin, and Jackson (2010), consists of 40 items selected from the 70 items of the original PSDQ. Marsh et al. (2010) provided support for the psychometric properties of the PSDQ-S. The PSDQ-S has 2 items (Strength), 3 items (Body Fat, Sport Competence, Appearance, Flexibility, Endurance), 4 items (Physical Activity and Global Physical Self-Concept), or 5 items (Health, Coordination and Global Self-Esteem) for each subscale resembling the same subscales as in the PSDQ. Marsh, Martin and Jackson (2010) verified the factor validity and measurement invariance of the PSDQ-S through a confirmatory factor analysis of a normative archival sample of Australian high school students and cross validated this in five different samples (Australian students, Australian elite adolescent athletes, Spanish adolescents, Israeli university students and Australian older adults). Different CFAs also proved the factor validity and measurement invariance of the PSDQ-S across the different versions, hence the short and long form, across different age-groups, and across different sex groups. Marsh, Martin & Jackson (2010) also provided support for the convergent and discriminant validity of the PSDQ-S across a sample of Australian and Israeli students. For the eleven subscales, a sufficient internal consistency for the scale scores was found across the six samples (α = .77 - .94), as well as a sufficient test-retest stability tested with a sample of Australian adolescents and older adults (median r = .80; with a lower score of .59 for Health).

Procedure

All the study procedures were reviewed and approved by the institutional review board of the university, and were consistent with ethical principles of the Declaration of Helsinki. Parents or caregivers returned informed consent, and each participant gave verbal assent prior to testing. The sample of Flemish adolescents was asked to report their physical self-concept by independently and silently filling in the translated Dutch PSDQ (Simons, Capio, Adriaenssens, Delbroek, & Vandenbussche, 2012) during school time. Incompletely filled out questionnaires were excluded from the sample.

Of that sample, a subset of 354 boys and girls did a retest of the Dutch PSDQ within 10 days. Since the items in the PSDQ-S also exist in the PSDQ, the data needed for research of the PSDQ-S were extracted from the corresponding items in the administered long form according to the PSDQ-S model provided by Marsh, Martin & Jackson (2010). Similarly, the data needed for the retest of the PSDQ-S were extracted from the retests of the PSDQ taken by the 354 boys and girls mentioned above. Filling out the questionnaire took about 15 till 20 minutes.

Statistics

To check whether parametric statistical procedures are permitted or not, the normal distribution of the test scores across the different subscales was prospected. This was done separately for the test scores on the PSDQ and the PSDQ-S, and the test and retest. In this study a Kolmogorov-Smirnov test, a Lilliefors corrected Kolmogorov-Smirnov test, and a Shapiro-Wilk test was used to test the hypothesis of a normal distribution. When the p value is smaller than .05 a normal distribution cannot be assumed (Ghasemi, & Zahediasl, 2012). All statistical analyses were performed using Statistica 12In order to find support for the psychometrical strength of the Dutch PSDQ and PSDQ-S this study performs a number of statistical analysis, using Statistica 12 (Marsh, 1996b; Marsh, Martin, & Jackson, 2010; Statsoft Inc., 2016).

To prospect the internal consistency Cronbach α was calculated. Cronbach coefficient α range from 0 to 1, whereas values above .7 and .9 are considered an adequate internal reliability and excellent internal reliability respectively (Sijtsma, 2009; Morera & Stokes, 2016). To investigate the test-retest reliability of both the PSDQ and PSDQ-S a Pearson correlation coefficient was used. To interpret the values of this correlation coefficient the following guidelines can be applied (Portney & Watkins, 2009): (a) r lower than .25 means little to no correlation, (b) r between .25 and .5 means a fair correlation, (c) r between .5 and .75 means a moderate to good correlation and (d) r larger than .75 a good to excellent correlation.

A confirmatory factor analysis (CFA) was used to test the construct validity. Construct validity refers to the operationalization, which is the quality of the translation or transformation process of a previously defined construct into an applicable concept. Chi square must be a non-significant score ($p > .05$), and has a value of zero until infinity with zero indicating a perfect fit (Portney & Watkins, 2009). The standardized root mean square residual (SRMR) and the Bentler comparative fit index (CFI) both range from 0 to 1. According to Hu and Bentler (1999) a good fit is demonstrated with the following cut-off values: SRMR of .08 or less (acceptable model), and CFI larger than .90 (acceptable fit) or larger than .95 (good fit). Marsh, Morin, Parker and Kaur (2014) suggest that researchers should compare exploratory factor analysis (EFA) solutions with CFA solutions. If CFA does not provide a similarly acceptable fit as EFA, then EFA is a proper alternative for obtaining

acceptable model fit. Since CFA is a more specific version of a more general EFA, a direct comparison of the factor solutions is possible.

If this model was not confirmed, according to Marsh, Morin, Parker and Kaur (2014) an exploratory factor analysis was carried out. Hereby a principal factor analysis with varimax rotation and 11 fixed number of factors to extract, was performed. To investigate gender differences between the results of boys and girls, an independent two-sample t-test was carried out. Effect size (*d*) was calculated, using the standard deviations and means to establish the magnitude of the difference (Durlak, 2009). Effect sizes were interpreted in compliance with Cohen's (1988) recommendations for effect size i.e. small $d < 0.20$, medium $d = 0.20–0.50$, large $d = 0.50–0.80$, very large $d > 0.80$. To investigate the relation between the different subscales of the Dutch PSDQ and PSDQ-S in the Flemish sample a series of Pearson correlation coefficients were calculated.

Results

Results of the PSDQ

Estimates of reliability, more specifically internal consistency and test-retest stability, of the PSDQ are displayed in Table 1. Sport Competence, Global Physical Self-Concept and Endurance have excellent reliability estimates ($\alpha > .90$). Seven of the subscales (Coordination, Global Self-Esteem, Body Fat, Appearance, Strength, Flexibility & Physical Activity) have reliability estimates between .80 and .90, which is considered as a good internal consistency. The subscale Health ($\alpha = .79$) yielded the lowest reliability estimate, yet this result is still considered as an acceptable level and is in concordance with the result of the original Australian version (Morera & Stokes, 2016; Sijtsma, 2009).

The Pearson correlation coefficients showed significantly related results ($p < .05$) between the test and retest for all the subscales. The stability coefficients are higher than .75 for all the subscales, which resembles a good to excellent relationship between the results on the test and retest (Portney & Watkins, 2009). The lowest stability coefficient is obtained on the subscale Coordination ($r = .78$) and the highest on the subscales Endurance and Sport Competence ($r = .92$).

To find statistical support for the construct validity of the Dutch PSDQ, a confirmatory factor analysis is conducted. This study analyses whether the covariance matrix of the Dutch PSDQ deviates from the covariance matrix of the original 11-factor theoretical model of the PSDQ by Marsh (1996b). The goodness of fit indices for the PSDQ yielded a

result indicating a lack of fit with the original factor structure (chi square = 15,824.5, p < .001) (Portney & Watkins, 2009). The other goodness of fit statistics, also gave no indication for an acceptable model fit (SRMR = .24 and CFI = .77) (Hu & Bentler, 1999). On the contrary, according to Marsh, Morin, Parker and Kaur (2014) exploratory factor analysis, with 11 fixed number of factors to extract, reveals the original solution, explaining 62,433 % of the variance.

The results of the independent samples t-test for the different subscales of the PSDQ are displayed in Table 2, the independent samples being a sample of boys (N = 687) and a sample of girls (N = 730). This test showed significant differences between the mean results of boys and girls for 10 of the subscales of the PSDQ (p < .001), with boys having a consistently higher average mean score for all the subscales. All the effects varied from medium till very large. However, the subscale Flexibility does not display significantly different mean results between boys and girls (p = .101), with small effect.

Results of the PSDQ-S

The values of internal consistency and test-retest stability of the PSDQ-S are presented in Table 3. The subscales Health, Coordination, Strength and Global Self-Esteem have adequate estimates of internal consistency varying from .70 until .79. Moreover, even the lowest internal consistency score, on the subscale Strength, is still of an acceptable level (α =.70). The estimates of internal consistency of the other seven subscales range from .81 till .89, which is good (Morera & Stokes, 2016; Sijtsma, 2009).

The Pearson correlation coefficients showed significantly related results (p < .05) between the test and retest for all the subscales of the PSDQ-S. The correlation coefficients vary from .76 till .89 across the different subscales, which all resemble a good to excellent relationship between the factor scores on the test and retest (Portney & Watkins, 2009). A confirmatory factor analysis conducted on the covariance matrix of the Dutch PSDQ-S indicated no sufficient evidence for the confirmation of the original factor structure (chi square = 10,971.7, p < .001) (Portney & Watkins, 2009). The SRMR of .25 and CFI of.64 also gave no indication for an acceptable fit (Hu & Bentler, 1999). The exploratory factor analysis with 11 fixed factors to extract explains 62,06 % of the variance with no clear structure. The analysis identified only eight components with an eigenvalue greater than one.

Table 4 displays the results of the independent samples t-test for the different subscales of the PSDQ-S. All the subscales, with exception of Flexibility, had a p value lower than .001 which indicates significant differences in the mean results between boys and girls in fever of the

boys. The effects varied from medium till large. The subscale Flexibility has a p value of .705, with small effect, thus indicating no significant difference in the mean results between boys and girls. Boys have a consistently higher score on every subscale of the PSDQ-S.

Table 1: Reliability coefficients for the 11 PSDQ subscales: internal consistency (Cronbach alpha) and test-retest reliability (Pearson product moment correlation)

Subscales	PSDQ Internal consistency (N=1417)	Test-retest reliability (N=354)
Health	.79	.84
Coordination	.82	.78
Physical activity	.89	.89
Body fat	.87	.89
Sport competence	.92	.92
Appearance	.87	.84
Strength	.87	.89
Flexibility	.88	.87
Endurance	.91	.92
Global physical self-concept	.92	.86
Global self-esteem	.83	.81

Table 2: Independent samples t-test comparing the mean results of girls and boys on the 11 PSDQ subscales

	Boys (N=687)		Girls (N=730)				
Subscales	Mean	Standard deviation	Mean	Standard deviation	t-value	p-value	ES
Health	4.85	.80	4.41	.98	9.19	<.001	.49
Coordination	4.50	.87	4.10	.99	8.02	<.001	.43
Physical activity	4.25	1.42	3.43	1.41	10.89	<.001	.58
Body fat	5.01	1.12	4.23	1.38	11.61	<.001	.62
Sport competence	4.39	1.13	3.52	1.26	13.63	<.001	.73
Appearance	4.24	1.00	3.73	1.11	9.11	<.001	.48
Strength	4.36	1.1	3.65	1.08	12.25	<.001	.65
Flexibility	4.01	1.15	3.91	1.27	1.64	.101	.08
Endurance	4.31	1.34	3.16	1.34	16.09	<.001	.86
Global physical self-concept	4.81	1.06	4.08	1.25	11.69	<.001	.63
Global self-esteem	4.78	.80	4.45	.92	7.22	<.001	.38

Table 3: Reliability coefficients for the 11 PSDQ-S subscales: internal consistency (Cronbach alpha) and test-retest reliability (Pearson product moment correlation)

Subscales	PSDQ Internal consistency (N=1417)	Test-retest reliability (N=354)
Health	.75	.81
Coordination	.79	.77
Physical activity	.85	.87
Body fat	.85	.88
Sport competence	.89	.88
Appearance	.81	.79
Strength	.70	.77
Flexibility	.86	.85
Endurance	.84	.89
Global physical self-concept	.89	.84
Global self-esteem	.75	.76

Table 4: Independent samples t-test comparing the mean results of girls and boys on the 11 PSDQ-S subscales

	Boys (N=687)		Girls (N=730)				
Subscales	Mean	Standard deviation	Mean	Standard deviation	t - value	p	ES
Health	5.12	.84	4.73	1.03	7.86	<.001	.41
Coordination	4.62	.87	4.20	.99	8.36	<.001	.45
Physical activity	4.28	1.48	3.54	1.46	9.53	<.001	.50
Body fat	5.12	1.26	4.11	1.59	13.20	<.001	.70
Sport competence	4.59	1.18	3.70	1.35	13.25	<.001	.70
Appearance	3.96	1.10	3.35	1.19	9.86	<.001	.53
Strength	4.40	1.19	3.68	1.23	11.21	<.001	.59
Flexibility	3.60	1.36	3.63	1.45	-.38	.705	-.02
Endurance	4.39	1.32	3.35	1.39	14.37	<.001	.77
Global physical self-concept	4.85	1.08	4.16	1.29	10.97	<.001	.58
Global self-esteem	4.63	.84	4.30	.90	7.07	<.001	.38

Table 5: Pearson product moment correlation between the 11 subscales of the PSDQ and PSDQ-S

	Correlation coefficients (N=1417)
Health	.94
Coordination	.98
Physical activity	.97
Body fat	.96
Sport competence	.96
Appearance	.94
Strength	.87
Flexibility	.95
Endurance	.96
Global physical self-concept	.98
Global self-esteem	.94

Comparison of the PSDQ and PSDQ-S

The internal consistency estimates of the PSDQ-S (Table 3) are systematically lower for all the subscales compared to those of the PSDQ (Table 1). Table 5 represents the Pearson correlation coefficients between the PSDQ and PSDQ-S for the 11 subscales. The correlation between all these different subscales of the PSDQ and PSDQ-S is significant ($p < .05$). Ten subscales have correlation coefficients varying from .94 until .98, which resembles a very good relationship between the results of the PSDQ and the PSDQ-S. The subscale Strength is the only subscale with a score lower than .90 ($r = .87$), yet this still resembles a good correlation (Portney & Watkins, 2009).

Discussion

The first objective was to describe the reliability and validity of the translated PSDQ in a Flemish adolescent sample. Overall, the reliability of the Dutch PSDQ was high and showed high internal consistencies. None of the reliability estimates increase when specific items were extracted from a subscale. The lowest Cronbach α was reported on the subscale Health which is in accordance with the findings of Marsh (1996b). The Dutch version of the PSDQ demonstrated lower reliability for all the subscales compared to the PSDQ tested a priori on the Australian sample (Marsh, 1996b). The original version of the PSDQ was translated into Dutch using a back to back translation procedure (Simons et al., 2012). Schipke and Freund (2012) state that the use of translated versions consistently has a negative

impact on the reliability, because of the possible changes of meaning during the translation process. These findings may suggest that more research and thorough effort should be invested in the translation process, whereby misinterpretations can be diminished. On the other hand, the PSDQ has been translated into different languages (e.g. Guérin, Marsh, & Famose, 2004; Marsh, Asci, & Marco, 2002). In these previous studies, as well as in the present study, acceptable reliability estimates were obtained for all the translated versions.

Overall, the test-retest stability for every subscale displays a high or very high correlation, from which can be concluded that the responses given on every factor is stable. Compared with the 3-month interval test-retest stabilities of the research of Marsh (1996b), this study finds either equal or higher test-retest stabilities on the PSDQ subscales, with exception of Coordination and Global Self-Esteem. The test-retest stabilities are consistently higher compared to the 14-month interval correlations of Marsh (1996b), which is probably because PSC is subject to possible changes in a larger time interval.

The absolute fit indices (chi square and SRMR) and relative fit indices (CFI), following the CFA, provided no standing evidence that the translated PSDQ has a similar factor structure as the a priori model, and therefore this study does not offer enough empirical evidence for the construct validation of the Dutch version of the PSDQ. Some important annotations following these findings can be discussed. Chi square is very sensitive for sample size and number of variables, and will increase proportionally with those two features (Marsh, Hau, & Grayson, 2005). Also, large factor structures (with 50 items or more, and 5 factors or more) are regularly unable to meet the minimal terms for acceptable fit in relation to an a priori factor structure, because CFA is often too restrictive for psychometric research (Marsh 2007; Marsh, Hau & Grayson, 2005). The use of an exploratory factor analysis, reveals the same eleven factor structure as the original version.

Furthermore, this research studied the mean gender differences of PSC in a Flemish sample. Regarding the translated PSDQ, boys and girls differed significantly in mean subscale scores for 10 of the subscales, with Flexibility being the only subscale without significant mean differences. All the effects varied from medium till very large, thus, this study states that boys in a Flemish adolescent sample report higher levels of PSC on the Dutch PSDQ in comparison to girls. These findings agree with previous research on gender-based differences of the original PSDQ (Marsh et al., 1994). Although the mean differences in gender have been discussed in the present study, implications for further research comprise the possible effects of age, BMI and socio-cultural aspects influencing PSC in a Flemish sample. For instance, in

ongoing research the measurement invariance between gender subgroups should be conducted to investigate whether the questionnaire measures the same constructs, and does not acts differently, across these subgroups.

The second objective was to examine the reliability and validity of the translated PSDQ-S in a Flemish adolescent sample. The data for the PSDQ-S were extracted from the corresponding items of the assessed PSDQ. Reliability estimates of the Dutch PSDQ-S are consistently high or adequate for all the subscales. although the subscale Strength has reached the cut-off point for adequate reliability. The subscale Strength demonstrated the lowest indices (α = .70). This finding can be explained by the fact that in the Dutch PSDQ-S, the subscale strength is comprised of only two items and reliability estimates are partly influenced by the number of items. The Cronbach alpha does not increase when one of the questionnaire items is deleted, this result applies to every subscale. Most of the PSDQ-S subscales in the present study have lower reliabilities coefficients than those tested in an Australian adolescent group (Marsh, Martin, & Jackson, 2010), except for Body Fat, Sport Competence and Endurance. According to Schipke and Freund (2012) this can partially be explained by the imperfect translation process and the misinterpretations following this process. Yet, the decrease in reliability is rather small and does not compromise the internal consistency of the questionnaire.

The magnitude of the test-retest stability coefficients of the translated PSDQ-S subscales was high and they were all in accordance with those found in Marsh, Martin and Jacksons (2010) research, except for the Health subscale which has a considerably lower correlation coefficient (r = .59) than in the present study (r = .81).

The goodness of fit statistics provided insufficient support for concluding an equal factor structure as the original PSDQ-S, nor it can be presumed that the corresponding items on the original and translated PSDQ-S measure the same concepts. However, regarding the goodness of fit statistics, the above-mentioned annotations concerning the full form of the translated PSDQ also apply. Annotations concerning the sensitivity of chi square for sample size and number of variables (Marsh, Hau, & Grayson, 2005), and the restrictiveness of CFA in psychometric studies (Marsh 2007; Marsh, Hau, & Grayson, 2005). The exploratory factor analysis looking for 11 factors, reveals no clear structure. The scree-test indicates that the PSDQ-S had only 8 components with an eigenvalue higher than one. This indicates that more thorough future research is advisable, focusing on finding evidence for the construct validity of the translated PSDQ-S version.

The independent two-sample t-test for gender differences in the translated PSDQ-S demonstrates the following results, boys and girls differed significantly in mean subscale scores for 10 of the subscales, with effects varied from medium till large. Thus, this research states that boys in a Flemish adolescent sample report higher levels of PSC with the Dutch PSDQ-S in comparison to girls. This concords with previous research on gender-based differences of the original PSDQ (Marsh, Martin, & Jackson, 2010). Implications for further research for the Dutch PSDQ-S contain the study of measurement invariance across gender subgroups, to prove that the questionnaire behaves similarly across these subgroups. Furthermore, the possible influences and mean differences of age groups, BMI groups and socio-cultural aspects should be a topic of research.

The third objective of the present study encompasses the comparison of the psychometric properties of the Dutch PSDQ and PSDQ-S in a Flemish adolescent sample. As described by Smith, McCarthy and Anderson (2000) developing a short version of an existing instrument invokes a series of issues that endanger the psychometric characteristics of the instrument. Although, fewer items for every subscale of the questionnaire diminishes the administration time, it also evokes a decrease in the reliability estimates. The Cronbach alpha for the PSDQ are consistently higher than those for the PSDQ-S, however the decrease in reliability is reasonable and the PSDQ-S still has acceptable reliability estimates for all subscales.

All the correlations between the corresponding subscales in the translated PSDQ and PSDQ-S suggest very high relations, which implicates covariance between the subscales in the different forms of the translated questionnaire. Some different subscales also correlated significantly, for example Sport Competence and Endurance (r = .70), however those two factors of PSC hold some corresponding aspects, so some degree of relationship is reasonable, and the correlations were all consistently lower than those of the corresponding subscales.

Limitations of the study

A first possible bias of the study, relates to the methodology used. As indicated, the data for the PSDQ-S were obtained by extraction from the PSDQ-data. For this method was chosen for practical reasons. It did not seem appropriate to have the same subjects - and some of them to do it twice - filled in the long and short version. A possible bias is that the young people could remember their answers and because of the ‘déjà vu effect’ they would no longer respond realistically. By collecting the data on a separate pilot group with only the PSDQ-S,

several socioeconomic parameters could be strongly influenced, which would make it difficult to compare the results. We have chosen for a test-retest reliability with an interval of 14 days, because a longer interval is measuring more the stability.

A second limitation of this study concerns the choice of age of subjects (12:8 till 19:0 years). This group was chosen because this is the age of young people in secondary education in Flanders. However, this implies that developmental differences within this age group may be ignored. In the construction of the Self-Description Questionnaires (SDQ), Marsh (1989) takes an analogous age group of 13 to 17 years as the norm group. Ideally, sub-groups could be created in this regard, such as gender, the level of education, the Body Mass Index, the socio-economic background and so on. To make this possible, the pilot group should considerably be extended, since 1417 young people from all over the various Flemish regions have already participated.

A third limitation concerns the purpose of this study. We wanted to check the original factor structure of the PSDQ and the PSDQ-S. Therefore, a confirmatory factor analysis was used, followed by an exploratory factor analysis, like Marsh, Morin, Parker and Kaur (2014) suggested. Seen the results on the PSDQ-S it seems necessary to search for a new structure by means of other techniques like for example a Rasch-models.

Conclusion

The translated PSDQ, as well as the translated PSDQ-S, have demonstrated an acceptable to excellent level of internal consistency and test-retest stability. The differences in gender were in correspondence with those found in the original PSDQ and PSDQ-S, which supports the validity of the Dutch PSDQ and PSDQ-S in a Flemish adolescent sample (Marsh et al., 1994; Marsh, Martin, & Jackson, 2010). Nevertheless, CFA of the translated PSDQ and PSDQ-S in a Flemish adolescent sample could not provide enough empirical support to confirm that the translated model adequately fits the theoretical model. For the translated version of the PSDQ an exploratory factor analysis with extraction of 11 factors, confirmed the original structure of Marsh (1997). But this was not the case for the PSDQ-S. Further research regarding the construct validity is indicated for confirming whether the translated questionnaire indeed measures the same concepts as the corresponding items in the original questionnaire.

The decrease in reliability from the translated PSDQ to the translated PSDQ-S was small enough to maintain acceptable reliability. Furthermore, the corresponding factors in the

translated PSDQ and PSDQ-S correlate highly. So far, this suggests that the Dutch PSDQ-S is a good alternative for the Dutch PSDQ in a Flemish sample. However, the present study extracted data for investigating the Dutch PSDQ-S from the 40 corresponding items of the full Dutch PSDQ filled out by the participants. Moreover, the present study could not provide empirical evidence for the structure of the translated PSDQ-S version. Guidelines on short form development state the importance of finding evidence for the factor structure of the short form by comparing samples that complete the PSDQ, from which the short form items are subtracted, with samples that only complete the PSDQ-S (Marsh, Martin & Jackson, 2010; Smith, McCarthy & Anderson, 2000). This is an important implication for ongoing research concerning the psychometrical properties of the Dutch PSDQ-S in the Flemish adolescent sample.

Although the present study provides empirical evidence for some of the important psychometric properties, our findings suggest that the PSDQ-S does not yet meet sufficient psychometric properties for confident use in practice on Flemish adolescent population and therefore the use of the full form of the PSDQ is preferable at this moment.

References

Boyle, G. J. (1994). Self-Description Questionnaire II: A review. *Test Critiques, 10*, 632-643.

Byrne, B. M. (1966). *Measuring Self-concept across the life span, issues and Instruments.* Washington, DC: American Psychological Association.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Chung B. F. (1996) The Construct of Physical Self-Concept and its Relations to Construct of Physical Fitness among Children. Thesis, National College of Physical Education and Sports, Taiwan, China. In Chiang, L.C., Huang J. L. & Fu, L. S. (2005). Physical activity and physical self-concept: comparison between children with and without asthma. *Journal of Advanced Nursing, 54(6),* 653–662.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology, 34*(9), 917–928.

Fletcher, R. B., & Hattie, J. A. (2004). An examination of the psychometric properties of the Physical Self-Description Questionnaire using a polytomous item response model. *Psychology of Sport and Exercise, 5*, 423– 446.

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

Ghasemi, A., & Zahediasl, S. (2012). Normality tests for statistical analysis: A guide for non-statisticians. *International journal of endocrinology and metabolism, 10*(2), 486-489.

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*(1), 19-38.

Hattie, J. (1992). *Self-concept.* Hillsdale, NJ: Erlbaum.

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

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 and Exercise, 9,* 830–847.

Marsh, H. W. (1993). Physical fitness self-concept: Relations to field and technical indicators of physical fitness for boys and girls aged 9-15. *Journal of Sport & Exercise Psychology, 15,* 184–206.

Marsh, H. W. (1996a). Construct validity of Physical Self-Description Questionnaire responses: Relations to external criteria. *Journal of Sport & Exercise Psychology, 18*(2), 111–131.

Marsh, H. W. (1996b). Physical Self -Description Questionnaire: Stability and discriminant validity. *Research Quarterly for Exercise and Sport, 67,* 249–264.

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

Marsh, H. W. (1999*). Physical Self-Description Questionnaire: Package.* SELF Research Centre, Bankstown Campus, University of Western Sydney, Australia.

Marsh, H. W. (2002). A multidimensional physical self-concept: A construct validity approach to theory, measurement, and research. *Psychology: The Journal of the Hellenic Psychological Society, 9,* 459–493.

Marsh, H. W. (2007). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum, & R. C. Eklund (Eds.), *Handbook of sport psychology 3rd ed.* (pp. 774–798). New York: Wiley.

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

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

Marsh, H. W., Byrne, B. M., & Shavelson, R. J. (1988). A multifaceted academic self- concept: Its hierarchical structure and its relation to academic achievement. *Educational Psychology, 80,* 366-380.

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

Marsh, H. W., 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., Hau, K. T., Sung R. Y. T., & Yu, C. W. (2007). Childhood Obesity, Gender, Actual-Ideal Body Image Discrepancies and Physical Self-concept, in Hong Kong Children: Cultural Differences in the Value of Moderation. *Developmental Psychology, 43(3),* 647-662.

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

Marsh, H. W., Marco, I. T., & Hiilya, F.H. (2002). Cross-cultural validity of the physical self-description questionnaire: Comparison of factor structures in Australia, Spain, and Turkey. *Research Quarterly for Exercise and Sport, 73*(3), 257-270.

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

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

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.

Marsh, H. W., & Shavelson, R. J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*,107-125.

Morera, O. F., & Stokes, S. M. (2016). Coefficient α as a measure of test score reliability: Review of 3 popular misconceptions. *American journal of public health*, *106*(3), 458- 461.

Portney, L. G., & Watkins M. P. (2009). Correlation. In L.G. Portney, & M.P. Watkins (Eds.), *Foundations of clinical research: Applications to practice, 3rd ed.* (pp. 523-538). London: Pearson Education.

Richards, G. E. (1988). *Physical Self Concept Scale*. Sydney: Australian Outward-Bound Foundation.

Schipke, D., & Freund, P. A. (2012). A meta-analytic reliability generalization of the Physical Self-Description Questionnaire (PSDQ). *Psychology of Sport and Exercise, 13*, 789- 797.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407–441.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107-120.

Simons, J., Capio, C., Adriaenssens, P., Delbroek, H., & Vandenbussche, I. (2012). Self- concept and physical self-concept in psychiatric children and adolescents. *Research in Developmental Disabilities, 33*, 874- 881.

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

Statsoft Inc. (2016). Statistica 12. Retrieved from *http://www.statsoft.com*

Stein, R. J., Bracken, B. A., Haddock, C. K., Shadish, W. R. (1998). Preliminary development of the Children's Physical Self-Concept Scale. *Journal of developmental and behavioral pediatrics 19 (1)*, 1-8.

Sypsa, C, & Simons, J. (2008). Questionnaires measuring the physical self in children: A review. *European Psychomotricity Journal, 1*(2), 53-63.

Tietjens, M., Freund, P.A., Büsch, D., & Strauss, B. (2012). Using mixture distribution models to test the construct validity of the Physical Self-Description Questionnaire. *Psychology of Sport and Exercise, 13*, 598-605.

Tsorbatzoudis, H. (2005). Psychometric evaluation of the Greek Physical Self-Description Questionniare. *Perceptual and Motor Skills, 101*(1), 79-89.

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

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