

Running head: Examining Self-Concept with ESEM

Examination of the Structure and Grade-Related Differentiation of Multidimensional Self-Concept Instruments for Children using ESEM

A. Katrin Arens*, German Institute for Educational Research
&

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

* The first two authors (A.K.A & A.J.S.M) contributed equally to this article and their order was determined at random: both should thus be considered first authors.

Author Note

This article was prepared when the first author was a visiting scholar at the Centre for Positive Psychology and Education, University of Western Sydney, Australia. The research period was funded by a scholarship of the German Academic Exchange Service (DAAD) to A. Katrin Arens. Correspondence regarding this investigation should be sent to A. Katrin Arens, German Institute for International Educational Research, Centre for Research on Education and Human Development and Centre for Research on Individual Development and Adaptive Education of Children at Risk (IDeA), Schloßstr. 29, D-60486 Frankfurt am Main, Germany, Email: arens@dipf.de

This is the prepublication version of the following manuscript:

Arens, A.K., & Morin, A.J.S. (2016). Examination of the structure and grade-related differentiation of multidimensional self-concept instruments for children using ESEM. *The Journal of Experimental Education*, 84, 330–355. <http://doi.org/10.1080/00220973.2014.999187>

This paper is not the copy of record and may not exactly replicate the authoritative document published in *The Journal of Experimental Education*.

Abstract

This study is a substantive-methodological synergy in which Exploratory Structural Equation Modeling (ESEM) is applied to investigate the factor structure of multidimensional self-concept instruments. Based on a sample of German students ($N = 1958$) who completed the Self-Description Questionnaire I (SDQ-I) and the Self-Perception Profile for Children (SPPC), the results supported the superiority of ESEM compared to confirmatory factor analyses (CFA) for both instruments. ESEM resulted in lower factor correlations and substantively meaningful cross-loadings. We finally proposed, and contrasted, three mechanisms for testing grade-related differences in the differentiation of self-concept facets and found no evidence of increased differentiation between grades 3 to 6.

Key Words: self-concept measurement, ESEM, age differences, self-concept differentiation.

Shavelson, Hubner, and Stanton (1976) conceptualized self-concept as a multidimensional and hierarchical construct. They depicted self-concept as a pyramid with general self-concept located at the apex, followed by global academic self-concept and global non-academic self-concept at the next lower level. Academic self-concept is then further divided into domain-specific self-concepts related to specific school subjects (e.g., self-concepts in math or English) whereas non-academic self-concept is divided into physical, social, and emotional self-concept domains. This self-concept hierarchy then becomes even more specific moving downward as the physical, social, and emotional self-concepts are themselves assumed to include subdomain-specific self-concepts (e.g., physical ability and physical appearance). By providing an empirically testable model of self-concept, Shavelson et al. (1976) seminal work marked the beginning of sophisticated self-concept research based on instruments measuring this proposed hierarchy. In this study, we focus on two of these instruments, the Self-Description Questionnaire I (Marsh, 1990b) and the Self Perception Profile for Children (Harter, 1985) (for an overview see Byrne, 1996).

Self-Description Questionnaire (SDQ)

The Self-Description Questionnaires (SDQs) are arguably the most complete and well-validated instruments available to date to assess multidimensional self-concepts in different age groups [SDQ-I for preadolescents (Marsh, 1990b), SDQ-II for adolescents (Marsh, 1990c), and SDQ-III for older adolescents and young adults (Marsh & O'Neill, 1984); for an overview see Byrne, 1996]. Given the age group considered here, we focus on the SDQ-I, but most of the conclusions also apply to the other SDQs. The SDQ-I includes scales for measuring academic (math, reading, general school) and non-academic (physical appearance, physical ability, parent-relations, peer-relations) self-concepts as well as a global self-concept scale. In general, exploratory factor analyses (EFA) and confirmatory factor analyses (CFA) conducted on the SDQ-I provided consistent evidence of the multidimensional nature of self-concept (e.g., Byrne, 1996; Marsh, 2007). However, research conducted on the SDQ-I also demonstrated the need for revisions, primarily related to the hierarchical nature of self-concept which has been found to be less pronounced than originally assumed. With regard to non-academic self-concept, physical appearance self-concept was found to be only weakly related to physical ability self-concept, yet strongly related to peer and global self-concepts (Arens, Yeung, Craven, & Hasselhorn, 2013; Marsh & Ayotte, 2003). Similarly, peer self-concept was found to be more highly related to physical self-concept facets (physical ability, physical appearance) than to parent self-concept (Arens et al., 2013; Marsh & Ayotte, 2003). In the academic area, numerous studies demonstrated only weak, or even negative, relations between math and verbal self-concepts (Marsh, 1986b, 1990a). Theoretically, the between-domain differentiation of academic self-concept (i.e., into math and verbal subdomains) has been explained by the internal/external (I/E) frame of reference model according to which academic self-concept is formed through a combination of social (external) and dimensional (internal) comparison processes (Marsh, 1986b, 1990a; Möller & Marsh, 2013; Möller, Pohlmann, Köller, & Marsh, 2009). In the social comparison process, students compare their own achievement in one school subject to the achievement of their classmates in the same subject. Given that achievements in multiple domains are positively related, the social comparison process is assumed to lead to positive relations between academic self-concept subdomains. In the dimensional comparison process, students compare their own accomplishments across subjects (e.g., math versus English), leading to negative relations between academic self-concept subdomains. The social and dimensional comparison processes balance each other resulting in negligible relations between math and verbal self-concepts. Recent research extended this domain-specificity of academic self-concept into an additional differentiation between competence and affect. The SDQ-I scales for math and verbal self-concepts comprise competence (e.g., "I learn things quickly in mathematics") and affect (e.g., "I look forward to mathematics") items that have been found to define separate, yet strongly interrelated, constructs (Arens, Yeung, Craven, & Hasselhorn, 2011; Marsh & Ayotte, 2003; Marsh, Abduljabbar, et al., 2013; Pinxten, Marsh, De Fraine, Van Den Noortgate, & Van Damme, 2014). Taken together, the findings presented thus far apparently challenge the assumption of a hierarchically structured self-concept, while providing strong support for the multidimensionality of self-concept.

Self-Perception Profiles

Harter's Self-Perception Profiles (SPPs) are the second most common set of instruments available to assess multidimensional self-concept among children (SPPC; Harter, 1985), adolescents (SPPA; Harter, 1988), and adults (ASPP; Messer & Harter, 1986). We focus here on the SPPC. The

original SPPC includes five domain-specific self-concept scales (scholastic competence, peer acceptance, physical attractiveness, athletic competence, and behavioral conduct) in addition to a global self-concept scale. All items are rated on a structured alternative response format based on two extreme statements reflecting high and low self-concepts (e.g., “Some kids often forget what they learn. But other kids can remember things easily”). Children first decide which alternative best reflect how they see themselves, and then indicate whether the descriptive is “really true” or “sort of true”.

Like the SDQ-I, the SPPC has also been extensively validated. However, research on the factor structure of the SPPC has provided mixed results. A series of CFA studies have generally tended to support the factor validity of the SPPC, as well as its invariance across gender and grade levels, within samples of Dutch (e.g., Muris, Meesters, & Fijen, 2000; Van den Bergh, & Van Rats, 1998), French-Canadian (e.g., Boivin, Vitaro, & Gagnon, 1992; Gavin, & Herry, 1996), and Irish (Shelvin, Adamson, & Collins, 2003) children. Using EFA, Miller (2000) managed to replicate the original factor structure of the Finnish SPPC, while Stewart, Roberts, and Kim (2010) failed to do so with a sample of African American females. Similarly, Schumann et al. (1999) showed that whereas the a priori factor structure of the SPPC was well supported among white American girls, a more complex factor structure including cross-loadings was required for African American girls. Finally, using a combination of EFA and CFA, Chan (2002) found evidence of a four-, instead of five-factor solution underlying the five domain-specific SPPC scales with a sample of Hong Kong boys.

Even when the a priori factor structure is supported, some studies question the distinctiveness of the SPPC subscales due to high intercorrelations ($r > .50$ and sometimes as high as .85; Boivin et al., 1992; Chan, 2002; Cole et al. 2001; Gavin & Herry, 1996; Stewart et al., 2010). These high factor correlations might be related to the non-standard, structured alternative response format of the SPPC described above, which has been criticized as being too complex for preadolescents and associated with substantial method effects (e.g., Eiser, Eiser, & Havermanns, 1995; Marsh, Richards, Johnson, Roche, & Tremayne, 1994). Similarly, research on the SPP-derived Physical Self-Perception Profile (PSPP; Fox & Corbin, 1989), also demonstrated high factor correlations (Marsh, Asçl, & Marco, 2002; Marsh, Bar-Eli, Zach, & Richards, 2006). However, research based on the French Physical Self-Inventory (PSI; Maïano et al., 2008), an adaptation of the PSPP where the structured alternative format was replaced by a 6-point Likert scale, reveals similarly high factor correlations. Thus, the problem might not lie within the answer scale after all. Other methodological issues apparently need to be considered in the investigation of the factor structure of the SPPC.

Methodological Issues

Evidence of construct validity for the SDQ-I and SPPC primarily originates from CFA studies (Byrne, 1996; Marsh, 2007). However, CFA relies on the highly restrictive independent cluster model (ICM) in which items are only allowed to load on a single factor, with no cross-loadings. It has recently been argued that the ICM assumption underlying CFA might be too restrictive for many multidimensional constructs, especially those based on complex instruments (Marsh, Lüdtke, et al., 2010; Marsh et al., 2009; Morin, Marsh, & Nagengast, 2013), such as the SDQ-I and SPPC. In fact, when examined outside of the ICM-CFA framework, factor structures based on multidimensional measures are found to include cross-loadings that can either be justified by theory or anticipated in relation to the nature or wordings of the items. More generally, all items are known to incorporate a part of random measurement error, which is traditionally assessed as part of reliability analyses and modelled as the items' uniquenesses in EFA or CFA. Cross-loadings address a more systematic type of measurement error traditionally assessed in validity analyses, and account for the fact that items might be fallible indicators of the construct they are supposed to measure and include at least some degree of association with other constructs. These issues are particularly relevant to multidimensional and hierarchical self-concept measures where all items are purported to measure partially overlapping self-concept domains and subdomains. Most of the previous EFA investigations of the SDQ-I or SPPC revealed multiple cross-loadings (e.g., Chan, 2002; Miller, 2000; Stewart et al., 2010; Watkins & Akande, 1992; Watkins & Dong, 1994), many of which make sense in light of the logical overlap between conceptually adjacent (e.g., physical appearance and physical ability) or hierarchically-ordered (e.g. physical appearance and global physical) constructs.

When real cross-loadings are forced to be zero, the only way for them to be expressed is through the inflation of the factor correlations. Indeed, even when the ICM-CFA model fits well in the first place (see Marsh, Liem, Martin, Morin, & Nagengast, 2011; Marsh, Nagengast, et al., 2011), factor

correlations will be at least somewhat inflated unless all cross-loadings are close to zero. Although the issue of high factor correlations has been primarily discussed in relation to the SPPC (Boivin et al., 1992; Gavin & Harry, 1996), the elevated factor correlations presented by Abu-Hilal and Aal-Hussain (1997) with the Arab version of the SDQ-I ($r = .41$ to $.91$ for boys and $r = .26$ to $.82$ for girls) raise similar concerns about the SDQ-I. Furthermore, even if the SDQ-I correlations generally remain reasonable, it does not mean that they are not still somehow inflated by the reliance on rigid ICM-CFA assumptions. These observations argue in favor of EFA as providing a more realistic measurement model for multidimensional self-concept instruments.

We note here that it is also possible to incorporate some cross-loadings in CFA measurement models. However this inclusion needs to rely on strong a priori assumptions about the specific items for which cross-loadings should be included or, alternatively, to rely on post hoc suboptimal two-stage procedures (i.e. based on the inspection of modification indices). Globally, the critical difference between CFA and EFA measurement models is that EFA measurement models allow for the direct estimation of all possible cross-loadings in a one-step process (something that is impossible to do in CFA), while apparently resulting in more accurate parameter estimates. Indeed, statistical simulation studies show that EFA usually results in more exact estimates of the true population values for latent correlations and that, even when the true population model corresponds to ICM-CFA, EFA still results in adequate, unbiased, parameter estimates (Asparouhov & Muthén, 2009; Marsh, Lüdtke, Nagengast, Morin, & Von Davier, 2013; Schmitt & Sass, 2011). Unfortunately, EFA has been outshined by the methodological advances associated with CFA and by the erroneous assumption that EFA was unsuitable to confirmatory studies based on a priori hypotheses. However:

This assumption still serves to camouflage the fact that the critical difference between EFA and CFA is that all cross loadings are freely estimated in EFA. Due to this free estimation of all cross loadings, EFA is clearly more naturally suited to exploration than CFA. However, statistically, nothing precludes the use of EFA for confirmatory purposes (...). (Morin et al., 2013, p. 396)

Recently, Asparouhov and Muthén (2009) integrated EFA with CFA and SEM into a new overarching Exploratory Structural Equation Modeling (ESEM) framework (e.g., Marsh et al., 2009; Morin et al., 2013). When ESEM is based on target rotation, a priori hypotheses can be specified regarding the expected factor structure, reinforcing the confirmatory nature of ESEM (Asparouhov & Muthén, 2009; Browne, 2001). With respect to self-concept, Morin and Maïano (2011) applied ESEM to the SPP-based PSI and found clear evidence that it provided a much better fit to the data than ICM-CFA and resulted in reduced factor correlations due to the presence of cross-loadings, most of which made sense substantively. Morin and Maïano further showed that this conclusion was unrelated to the specific rotation procedure used. The present study aims to test the suitability of ESEM for the SDQ-I and SPPC. The application of ESEM to the SDQ-I and SPPC might also provide a new perspective into possible developmental changes in the differentiation of self-concept.

Differentiation of Self-concept

Shavelson et al. (1976) proposed that self-concept facets will become increasingly differentiated with age or grade level (i.e., more independent from one another and thus less correlated). Studies on the SDQ-I provided some support to this conjecture. For instance, Marsh (1989) observed that the mean correlation between self-concept factors decreased with age but that this differentiation process leveled out in preadolescence. Research based on the SPPC also generally supports this hypothesis showing decreased factor correlations as a function of age (Boivin et al., 1992; Gavin & Herry, 1996; Van den Bergh & Van Rast, 1998).

Previous research of age/grade-related differentiation strictly focused on changes in the correlations between self-concept facets. However, at least three distinct processes can explain the observation of a developmental change in the size of the factor correlations, all with drastically different meanings. The only way to systematically contrast these three complementary processes is through tests of measurement invariance (e.g., Millsap, 2011) conducted within the ESEM framework. First, developmental differences in the correlation between self-concept facets might result from changes in the *reliability of self-concept measures*. Low reliability (i.e., high levels of random measurement error at the item level) is known to result in downwardly biased correlations between constructs when these correlations are assessed on the basis of observed scale scores (i.e., mean or sum of the items forming a scale). Although random measurement error can be controlled within a latent variable model as part of the item uniquenesses, studies comparing correlations based

on scale scores may report changes in correlations between self-concept facets simply related to the presence of greater, or lower, levels of measurement errors in the various age groups. However, this mechanism is unlikely to result in increased differentiation (lower correlation) of self-concept, given that older students should answer the items with greater precision (due to higher self-knowledge, richer vocabulary, etc.) and thus a lower level of random measurement error (e.g., Marsh, Debus, & Bornholt, 2005; Marsh, Craven, & Debus, 1991). Consequently developmental differences in reliability should lead to higher instead of lower correlations for older students and thus may not account for increasing differentiation of self-concept facets. Nonetheless, the invariance of items' uniquenesses should be systematically investigated as the increased differentiation of self-concept may in fact be more pronounced than initially believed.

Second, older children may answer the items with a lower level of *systematic measurement error* (i.e., higher validity), based on a clearer understanding of their domain-specificity. Although the cognitive processes at play here may be similar to those underlying the previous mechanism, the consequences of this greater validity in self-concept ratings would be translated in terms of reduced cross-loadings between items and non-target factors. In other words, older children may more closely associate the items to one specific self-domain relative to another. Thus, older children might display "purer" domain-specific self-concept factors which are less blurred by cross-loadings. Knowing that forcing true cross-loadings to be zero tends to result in artificially inflated correlations among self-concept factors when tested with ICM-CFA (or scale scores where the cross-loadings are simply ignored), this process could thus explain the previously observed reduced correlations between self-concept facets in older children. Indeed, if younger children rate self-concept items with more systematic error (i.e., higher cross-loadings), the factor correlations will appear to be more differentiated for older children in ICM-CFA. However, in this scenario, the mechanisms underlying these lower correlations have nothing to do with self-concept differentiation. The only way for this possibility to be systematically investigated is through an ESEM model allowing for the incorporation of cross-loadings and tests of their invariance across age groups or grade levels.

The third and most commonly invoked possibility for developmental changes in self-concept differentiation relates to the factor correlations themselves which should decrease if there are indeed higher levels of differentiation between self-concept facets. Studies based on ICM-CFA or scale scores cannot properly differentiate between the second and third mechanisms. To clearly disentangle these two possibilities [i.e., increasing differentiation due to (1) lower cross-loadings and (2) to lower correlations], ESEM is required to systematically investigate the invariance of the factor loadings and cross-loadings before moving on to tests of the invariance of the factor variance-covariances.

The Present Investigation

This study is a substantive-methodology synergy (Marsh & Hau, 2007) that connects the new and evolving ESEM approach to substantively important research questions related to grade-related differences in self-concept. First, this study aims to investigate the factor structure of the German versions of the SDQ-I and SPPC by contrasting the relative adequacy of ICM-CFA and ESEM. Then, we turn to investigations of grade-related differences in the factor structures of these instruments to probe potential grade-related differences in the differentiation of self-concept facets.

Method

Sample and Procedure

This study was specifically designed to assess the psychometric properties of the German versions of the SDQ-I and SPPC. The sample includes 1958 German students attending grades 3 to 6 in 19 randomly selected co-educational public schools located in middle Germany (50.5% boys; $N = 989$; age range = 7 to 15 years, $M = 10.66$, $SD = 1.30$). In Germany, 3rd grade [$N = 299$; mean age: 8.71 ($SD = 0.611$)] and 4th grade students [$N = 318$; mean age: 9.67 ($SD = 0.601$)] attend primary school where students of all ability levels are grouped together. The transition to secondary school takes place after grade 4 and goes along with an ability tracking for the students. Hence, according to their achievement during the four years of primary school, students are allocated to different ability tracks (high ability/academic middle ability/intermediate, low ability/vocational, and mixed ability/comprehensive tracks). Students from all ability tracks were included in this sample in grade 5 [high ability: $N = 240$; middle: $N = 35$; low: $N = 48$; mixed; $N = 307$; mean age: 10.74 ($SD = .637$)] and 6 [high ability: $N = 196$; middle: $N = 134$; low: $N = 82$; mixed: $N = 299$; mean age: 11.85 ($SD = 0.713$)]. Parental consent was obtained for all participants. Once parental consent was obtained, the

students themselves were advised of the anonymous treatment of their answers and of the voluntary nature of their participation. They thus had the opportunity to consent or not to this study and to withdraw participation at any time.

Instruments

SDQ-I. A validated German version of the SDQ-I (Arens et al., 2013) was used in the present study. The German SDQ-I incorporates eleven subscales, which all showed good scale score reliability (see Arens et al., 2013): physical appearance ($\alpha = .88$), physical ability ($\alpha = .89$), peer-relations ($\alpha = .86$), parent-relations ($\alpha = .86$), math competence ($\alpha = .93$), math affect ($\alpha = .94$), German competence ($\alpha = .91$), German affect ($\alpha = .92$), general school competence ($\alpha = .83$), general school affect ($\alpha = .86$), and global self-concept ($\alpha = .85$). The 76 items are rated on a 5-point Likert-type scale (true, mostly true, sometimes false and sometimes true, mostly false, or false). A total of 12 items were negatively worded and were reversed scored prior to the analyses.

SPPC. The German version of the SPPC (Asendorpf & van Aken 1993) includes scales for measuring students' global self-concept ($\alpha = .83$; also labelled as self-worth) as well as domain-specific self-concepts in scholastic competence ($\alpha = .80$), peer-acceptance ($\alpha = .82$), athletic competence ($\alpha = .78$), and physical appearance ($\alpha = .83$). The behavioral conduct subscale from the original English instrument was discarded from the German instrument because the authors assumed the items of this scale to be judgemental in nature, and because low reliability estimates have been reported for this subscale in the adolescent version (SPPA; Harter, 1988; Trent, Russell, & Cooney, 1994). Each subscale includes 6 items rated on the structured alternative format described above.

In line with the numerous studies supporting the validity of the original English SDQ-I and SPPC (for reviews, see Byrne, 1996; Marsh, 2007), previous investigations of the psychometric properties of the German version of these instruments (e.g., Arens et al., 2011, 2013; Asendorpf & van Aken 1993) also showed the expected patterns of associations between the various subscales and student achievement. In general, the SDQ-I and SPPC academic self-concept scales were found to be more highly related to achievement than non-academic self-concept scales. In addition, these relations were found to be even more specific for the SDQ-I as math achievement was found to be more highly related to math self-concept, followed by general school self-concept and German self-concept, with the competence components displaying higher relations with achievement than the affect components (Arens et al., 2011). Furthermore, SDQ-I and SPPC scales assessing similar self-concept domains (e.g., SDQ-I physical ability and SPPC athletic competence) were found to be more highly correlated than scales measuring different self-concept domains (for similar findings with the original instruments see for example Marsh & MacDonald-Holmes, 1990).

Statistical Analyses

In a first step, we compared CFA and ESEM representations of the SDQ-I to test whether ESEM might provide a more adequate representation of the data than the more restrictive ICM-CFA model (Model 1 in Table 1). Given that this instrument includes 12 negatively worded items, which are known to be associated with a method artefact (Marsh, 1986a), correlated uniquenesses were included between these 12 items in all models (e.g., Marsh, Scalas, & Nagengast 2010). Correlated uniquenesses were also included between items with parallel wording across academic domains (e.g., "I am good at all school subjects/German/math") (e.g., Marsh, Abduljabbar, et al., 2013). In the ICM-CFA model, each item was only allowed to load on the factor it was assumed to measure and no cross-loadings on other self-concept factors were allowed. This model included 11 distinct factors representing each of the previously described SDQ-I subscales. In the next step, an 11-factor ESEM model of the SDQ-I (Model 2 in Table 1) was estimated with oblique target rotation (Asparouhov & Muthén, 2009; Browne, 2001; Marsh, Lüdtke, et al., 2013). Target rotation seemed particularly appropriate in this study as it allows for the a priori specification of target and non-target factor loadings in a confirmatory manner (see Marsh, Morin, Parker, & Kaur, 2014). In line with common specifications of target rotation, cross-loadings were "targeted" to be close to zero, while the main loadings were freely estimated. Hence, the CFA and ESEM representations of the SDQ-I assumed the same 11 self-concept factors with the main difference that all cross-loadings were freely estimated in the ESEM solution. The same sequence of analyses was replicated with the SPPC. Models 11 and 12 in Table 1 compare the fit of CFA and ESEM representations of the SPPC. The ICM-CFA model (Model 11) estimated the a priori 5-factor model, with no cross-loadings, whereas ESEM (Model 12) estimated the same 5-factor structure, allowing cross-loadings between items and non-target factors.

All analyses were conducted with Mplus 7.0 (Muthén & Muthén, 1998-2012), based on the robust maximum likelihood (MLR) estimator providing standard errors and fit indices that are robust to the Likert nature of the items and to violations of normality assumptions. Full Information robust Maximum Likelihood (FIML) estimation was used to handle missing data. This procedure has been found to result in generally unbiased parameter estimates under even very high levels of missing data (e.g., 50%) and is generally recognized as having comparable efficacy as more intensive multiple imputation procedures (Enders, 2010; Graham, 2009). With only 0.646% items of SDQ-I and 1.709% items of the SPPC displaying missing values, the amount of missing values was very small.

Measurement Invariance. For testing invariance across grade level, we stated a series of nested models in which the measurement model was first freely estimated among students from the different grade levels (3 to 6) treated as the grouping factor, followed by models in which parameters were progressively constrained to be equal across grade levels (Meredith 1993; Millsap, 2011). In the least restrictive model (configural invariance, Model 3 for the SDQ-I, Model 13 for the SPPC in Table 1), the same pattern of associations between items and factors, and the same number of factors, were estimated in all grade levels, with no added equality constraints. The next model of weak measurement invariance (Model 4 for the SDQ-I, Model 14 for the SPPC in Table 1) assumes the invariance of factor loadings (and cross-loadings) which is the prerequisite of all further invariance models. The assumption of invariant factor loadings and cross-loadings provides a direct test of one mechanism (i.e., differences in systematic measurement error) through which grade-related differences in the differentiation of self-concept can be expressed. Models of strong measurement invariance (Model 5 for the SDQ-I, Model 15 for the SPPC in Table 1) assume the invariance of factor loadings, cross-loadings, and item intercepts. Next, Model 6 for the SDQ-I and Model 16 for the SPPC are models of strict measurement invariance that add the invariance of item uniquenesses. Thus, these models allowed for a direct test of grade-related differences in the precision (i.e., reliability) of self-concept ratings. To further scrutinize the issue of grade-related differences in the reliability of self-concept ratings of the SDQ-I, we explored the invariance of the a priori correlated uniquenesses: Model 7 assumes the grade-level invariance of the correlated uniquenesses between negative items; Model 8 assumes the grade-level invariance of the correlated uniquenesses between parallel-worded items; Model 9 assumes the grade-level invariance of all correlated uniquenesses. In the next step, factor loadings, item intercepts, item uniquenesses, and factor variances/covariances were constrained to be invariant across grade levels (Model 10 for the SDQ-I, Model 17 for the SPPC in Table 1) serving to directly compare the relative size of the relations between self-concept factors across grade levels. This step provides the final, and most direct, comparison of possible grade-related differences in the differentiation of self-concept factors. For detailed specifications of invariance tests, the reader is referred to Morin et al. (2013) and Millsap (2011).

Model evaluation and comparison. Given the known oversensitivity of the chi-square test of exact fit and of chi-square differences tests to sample size and minor model misspecifications (e.g., Marsh, Hau, & Grayson, 2005), we relied on common goodness-of-fit indices and information criteria to describe the fit of the alternative models (e.g., Browne & Cudeck, 1993; Hu & Bentler, 1999; Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004): The comparative fit index (CFI), the Tucker-Lewis index (TLI), and the root mean square error of approximation (RMSEA) with its confidence interval. Values greater than .90 and .95 for both the CFI and TLI are considered to be respectively indicative of adequate and excellent fit to the data. Values smaller than .08 or .06 for the RMSEA support respectively acceptable and excellent model fit. To compare nested models, similar to the chi-square test of exact fit, chi-square difference tests are known to be sensitive to sample size and minor model misspecification (Marsh et al., 2005). Cheung and Rensvold (2002, also see Chen, 2007) thus recommend using changes in goodness-of-fit indices for testing invariance. According to their proposed guidelines, invariance can be seen as established as long as the CFI does not decrease more than .01 and as long as the RMSEA does not increase by more than .015 between less and more restrictive models. With regard to ESEM, fit indices that correct for parsimony (TLI and RMSEA, as well as changes in these indices) might be particularly important given that the total number of estimated parameters is typically much larger than in CFA (Marsh, Lüdtke, et al., 2010; Marsh et al., 2009; Morin et al., 2013). So far, these proposed cut-of values have been established for CFA and research is still lacking regarding their adequacy for ESEM—albeit previous ESEM applications have generally relied on similar criteria (for a related discussion, see Morin et al., 2013). Thus, these

interpretation criteria should not be treated as “golden rules” but only as rough guidelines for facilitating model evaluation which should include the simultaneous consideration of a wide range of fit indices, parameters estimates, statistical conformity and theoretical adequacy (Marsh et al., 2004).

Results

SDQ-I

Table 1 presents the goodness-of-fit indices for the various CFA and ESEM solutions. The CFA solution (Model 1) provided an acceptable fit to the data (CFI and TLI > .900; RMSEA < .060), but the fit of the ESEM solution (Model 2) is excellent and substantially improved (CFI and TLI > .900; RMSEA < .060; Δ CFI = +.043; Δ TLI = +.035; Δ RMSEA = -.008). The CFA and ESEM solutions also differed in their factor correlations (see Table 2) with much lower factor correlations (i.e. clearer differentiation) for ESEM ($r = -.061$ to $.637$, $M = .263$) than CFA ($r = .105$ to $.814$, $M = .399$). The ESEM solution was thus retained for the SDQ-I.

It is interesting to note that the observed decrease in factor correlations does not necessarily change the interpretation of which factors are most, or least, correlated with one another. However, even when the interpretation of the relative size of all correlation does not change, there is still a critical advantage in using ESEM to deflate the relative size of the factor correlations, especially at the highest end of the range. Knowing that ESEM has been showed to result in more exact estimate of true correlation values (Marsh, Lüdtke, et al., 2013), this issue becomes intimately related to the discriminant validity of the factors, given that inflated factor correlations create higher risks of multicollinearity (or redundancy) when these factors are in turn used together in predictions. It is also interesting to note that the observed decrease in factor correlations is most pronounced for the initially highest factor correlations. As we will see below, the relative amplitude of the observed decrease in the size of the factor correlations appears to be intimately related to the number and size of the cross-loadings observed between the items forming the various subscales.

A more detailed examination of the parameter estimates (see Table 3) reveals that the ESEM factors were well-defined through to substantial factor loadings. Multiple item cross-loadings were also present, providing additional support to the ESEM solution. For example, items 37 (“Overall, I am no good” - reversed; loading = .253, cross-loading = .193), 45 (“In general, I like being the way I am”; loading = .331, cross-loading = .290), and 72 (“A lot of things about me are good”; loading = .617, cross-loading = .239) from the global self-concept scale displayed high cross-loadings on the physical appearance scale, which in some cases are even of similar magnitude as their main loadings on the global self-concept factor. Similarly, item 62 (“I have nice features like nose, and eyes, and hair”; loading = .369, cross-loading = .326) displayed a cross-loading on the global self-concept scale which was of similar magnitude as its main loading on the physical appearance subscale. These results, in conjunction with the fact that both factors remained well-defined by their main loadings (global self-concept: .244 to .648, $M = .461$; physical appearance: .369 to .907, $M = .605$), replicated Morin and Maïano’s (2011) observation of substantial relations between physical appearance and global self-concepts. It should be noted however, that at least two items from the global self-concept scale (item 29: “I do lots of important things”; main loading = .244; item 37: “Overall, I am no good” - reversed; main loading = .253) present a relatively low level of association with the global self-concept factor and should be targeted for re-examination or reformulation in future studies.

Supporting the results from previous studies suggesting some conceptual overlap between these factors (Arens et al., 2013; Marsh & Ayotte, 2003): (a) one item from the physical appearance scale showed a substantial cross-loading on the physical ability scale (Item 46: “I have a good looking body”; cross-loading = .244); (b) one item from the peer self-concept scale displayed a substantial cross-loading on the physical appearance scale (Item 36: “I am easy to like”; cross-loading = .217). In light of the often cited dichotomy between physical and academic abilities, it is also interesting to note that two items from the physical ability scale displayed substantial negative cross-loadings on the general school competence scale (Item 24: “I enjoy sports and games”; cross-loading = -.318; Item 17: “I hate sports and games” - reversed; cross-loading = -.384).

Another noteworthy pattern of results pertains to the strong level of interrelations observed between the various academic self-concepts subscales. Although the competence and affect factors were relatively well defined within the academic self-concept domain (i.e., general school self-concept) and subdomains (i.e., Math and German), multiple items from the competence subscales presented small to moderate cross-loadings on their affect counterparts, and vice-versa [e.g., items of

math competence revealed cross-loadings on math affect between .040 and .211 ($M = .144$), and items of math affect demonstrated cross-loadings on math competence between .087 and .156 ($M = .120$]. This observation suggests that the distinction between competence and affect components does indeed exist within general school, math, and German academic self-concepts, but that the items still present a high level of content specificity over and above their competence or affect nature, explaining the previously reported elevated factor correlations between these subscales (Arens et al., 2011; Marsh & Ayotte, 2003). Interestingly, no such pattern of cross-loadings between the competence and affect factors could be observed across math and German self-concepts supporting the strong differentiation of academic self-concept into math and verbal facets (Marsh, 1986b; Möller et al., 2009). Second, the general school competence (main loadings = .047 to .346, $M = .263$) factor seemed somewhat less clearly defined than the other academic self-concept factors. Furthermore, the items forming the general school competence factor presented substantial cross-loadings on the general school affect factor (-.014 to .377, $M = .244$) as well as on the German (.126 to .378, $M = .269$) and math (.155 to .261, $M = .204$) competence factors. This observation suggests that the content of the general school competence factor may need to be revised in the context of future studies. The fact that the German and math self-concept items present only small cross-loadings across subdomains and with general school self-concept is fully consistent with the I/E model (Marsh, 1990a) addressing the expected low level of interrelations between subdomain-specific academic self-concepts.

Retaining the ESEM model for the SDQ-I, we turned to invariance tests with students' grade level treated as a grouping factor. Assuming an invariant factor pattern across grade levels, the model of configural invariance (Model 3) had an acceptable level of fit to the data (CFI and TLI > .900; RMSEA < .060). When loadings and cross-loadings were constrained to invariance (weak measurement invariance; Model 4), an increase in fit due to the greater model parsimony could be observed ($\Delta\text{CFI} = +.005$; $\Delta\text{TLI} = +.025$; $\Delta\text{RMSEA} = -.005$)¹. This result clearly supports the weak invariance of the model across grade level. Thus, variations in the pattern of item loadings and item cross-loadings do not seem to play a role in possible grade level differences in the differentiation between self-concept factors. The next models (Models 5 to 9) similarly support the invariance (ΔCFI and $\Delta\text{TLI} < .01$; $\Delta\text{RMSEA} < .015$) of the intercepts (strong invariance), uniquenesses (strict invariance), and correlated uniquenesses across grade levels. Combined with the finding of invariant factor loadings and cross-loadings, these findings suggest that children attending grades 3 to 6 rate the SDQ-I self-concept items with a similar level of random (i.e. reliability) and systematic (i.e. validity) error variance. Thus, differences in the psychometric properties of self-concept ratings do not seem to explain developmental differences in self-concept differentiation. Finally, Model 10 also provides evidence that the variance-covariances of the SDQ-I self-concept factors is fully invariant across grade levels. This finding stands in stark contrast with previous studies and theoretical assumptions suggesting an increasing differentiation between self-concept facets as a function of age (Marsh, 1989; Marsh & Ayotte, 2003; Shavelson et al., 1976).

SPPC

The CFA solution for the SPPC (Model 11 in Table 1) provided an acceptable level of fit to the data, while the ESEM solution (Model 12) provided a much improved level of fit to the data ($\Delta\text{CFI} = +.031$; $\Delta\text{TLI} = +.016$; $\Delta\text{RMSEA} = -.004$). Further, ESEM led to more differentiated (less correlated) self-concept factors (Table 4: $r = .309$ to $.557$, $M = .391$) in comparison with the ICM-CFA solution ($r = .366$ to $.827$; $M = .499$). Once again, these results supported the superiority of ESEM. Similar to the observations made for the SDQ-I, the observed decrease in factor correlations does not necessarily change the interpretation of which factors are most, or least, correlated with one another, and seems to be most pronounced for the highest factor correlations.

A detailed examination of the parameter estimates from this model (Table 5) reveals that all factors are well defined by their a priori items, although multiple cross-loadings on non-target factors are also present. Given the greater simplicity of the SPPC measurement model incorporating fewer items and factors, the ESEM solution appeared to be much cleaner for the SPPC than it was for the SDQ-I. In fact, only two observations are noteworthy. First, in line with the SDQ-I results (also see Morin & Maïano, 2011), two physical appearance items demonstrated substantial cross-loadings on the global self-concept scale while demonstrating relatively low main loadings (Item 4: "Some kids are happy with the way they look, but other kids are not happy with the way they look; main loading = .366, cross-loading = .340; Item 29: "Some kids think that they are good looking, but other kids

think that they are not very good looking”, main loading = .288, cross-loading = .375). Second, one item from the athletic competence scale (Item 23: “In games and sports some kids usually watch instead of play, but other kids usually play rather than watch”), presented relatively low and equivalent loadings on all domain-specific factors (i.e., cognitive competence: .117; peer acceptance: .152; athletic competence: .174, physical appearance: .188), showing no clear pattern of association with any factor. This finding suggests that this item should either be deleted, or reformulated in the context of future studies. Models 13 to 17 (Table 1) present the results from the tests of invariance. These results fully replicated those obtained with the SDQ-I and supported the invariance of the factor loadings, cross-loadings, intercepts, and uniquenesses, as well as the invariance of the factor variances and covariances across grade levels. Once again, this result argues against any form of increasing differentiation of self-concept factors as a function of grade level.

Discussion

The Underlying Structure of Multidimensional Self-Concept Instruments

In this study, we relied on a confirmatory approach to ESEM based on target rotation to investigate the factor structure of the SDQ-I and SPPC, two prominent measures of preadolescents’ multidimensional self-concept, in attempts to more precisely investigate the often observed elevated factor correlations and expected cross-loadings between the subscales of these instruments. Our study fits well with the current state of research using ESEM for the study of multidimensional constructs (e.g., personality: Marsh, Lüdtke, et al., 2010; Marsh, Nagengast, & Morin, 2012; students’ evaluation of university teaching: Marsh et al., 2009; bullying and victimization: Marsh, Nagengast, et al., 2011; motivation: Guay, Morin, Litalien, Valois, & Vallerand, 2014; Marsh, Liem, et al., 2011). So far, in self-concept research, ESEM had only been applied a measure of physical self-concept (Morin & Maïano, 2011), but not yet to more extensive multidimensional self-concept measures. Furthermore, grade-related differences in self-concept differentiation (e.g., Marsh, 1989; Marsh & Ayotte, 2003) were reexamined with ESEM as a potentially more appropriate methodological framework for self-concept measures.

In the present study, ESEM provided a superior representation of the underlying measurement structure of both the SDQ-I and SPPC, both in terms of a better fit to the data as well as lower estimates of the factor correlations consistent with a greater differentiation between self-concept facets. For both instruments, ESEM supported the a priori factor structures (i.e., 11-factor model for the SDQ-I, 5-factor model for the SPPC) with well-defined factors for the multiple self-concept facets. However, in addition to these well-defined factors, various cross-loadings were observed, further supporting the usefulness of ESEM for the study of multidimensional self-concept measures. In fact, many of the most noteworthy cross-loadings are compatible with previous research suggesting possible conceptual overlaps between self-concept facets, such as between global and physical appearance self-concepts (Harter, 1999; Marsh & Ayotte, 2003; Morin & Maïano, 2011).

The investigation of cross-loadings of the SDQ-I provided additional insights into the recently proposed differentiation of general school, math, and verbal self-concepts into competence and affect components (Arens et al., 2011; Marsh & Ayotte, 2003; Pinxten et al., 2014). Within the academic domain (i.e., general school self-concept) and subdomains (i.e., math and German self-concepts), competence- and affect-related items were found to form separate factors, but still demonstrated substantial cross-loadings with one another. This finding might account for the previously found high CFA correlations between the competence and affect components of academic self-concept facets (Arens et al., 2011; Marsh & Ayotte, 2003; Pinxten et al., 2014). However, in this study, no marked cross-loading was apparent between the math and German competence factors, between the math and German affect factors, or between the competence and affect factors associated with these subdomains (e.g., math competence and German affect). This finding implies the existence of competence and affect components that remain dominated by a strong math/verbal-specificity.

This ESEM investigation of item cross-loadings provided further insight into required refinements for the SDQ-I and SPPC. Regarding the SDQ-I, the items designed to measure general school competence demonstrated relatively low main loadings on their a priori factor but relatively high cross-loadings on other academic factors such as math and German competence, or general school affect. Further studies should thus focus on possible modifications regarding the wordings of the general school competence items in order to improve the validity of its assessment. Regarding the SPPC, item 23 did not present a clear pattern of association with any specific factor implying an

insufficient level of domain specificity for this item. Thus, the application of ESEM to self-concept measures also has implications for reliable and valid self-concept measurement.

The Usefulness of ESEM

This study clearly illustrates the usefulness of ESEM for self-concept measurement. Thus, at this stage two distinct studies (ours and Morin & Maïano, 2011), using three distinct instruments (i.e., PSI, SDQ-I, and SPPC), focusing on two different age groups (i.e., preadolescents and adolescents), and realized in two countries (Germany and France), have converged on the conclusion that ESEM provides a better and more appropriate representation of multidimensional self-concept ratings than ICM-CFA. Although we do not argue that ESEM will systematically provide a better representation of the measurement model underlying all self-concept instruments, the current body of knowledge strongly suggests that researchers should start the investigation of the measurement structure of self-concept instruments by a comparison of ESEM and CFA solutions. However, we reinforce that no decision should be purely based on statistical information so that statistical improvement (better fit, lower factor correlations) should always be balanced by theoretical meaningfulness and practical requirements. Similarly, whenever research or practice rely on scale scores, rather than latent variables, our results strongly suggest that interpretations and conclusions should be based on factor scores derived from ESEM representations, rather than scale score which ignore cross-loadings.

Although CFA models also allow for the inclusion of cross-loadings, these models typically require an a priori specification of the cross-loadings to be included, and only allow for a limited number of them. However, the a priori specification of all expected cross-loadings seldom appears to be realistic, particularly for long and complex instruments such as those considered here. Alternatively, without specific a priori expectations, one would need to embark on a suboptimal two-stage process involving the progressive inclusion of cross-loadings after the inspection of modification indices. Such a procedure is clearly susceptible to capitalization on chance. ESEM, on the other hand, provides a simpler alternative where all possible cross-loadings can be simultaneously considered. As noted before, ESEM has been shown to provide more accurate parameter estimates when cross-loadings are present and unbiased parameters estimates when cross-loadings are absent, whereas the omission of even a few small cross-loadings in CFA solutions has been shown to potentially result in biased estimates of factor correlations (Asparouhov & Muthén, 2009; Marsh, Lüdtke, et al., 2013; Morin et al., 2013; Schmitt & Sass, 2011). It can be argued, on statistical grounds, that less restrictive models such as ESEM will almost always fit better than more restrictive models such as CFA. However, we believe that the advantages of ESEM, in terms of decreased factor correlations and meaningful item cross-loadings, may often outweigh the disadvantages posed by its greater complexity.

These advantages all seem to be related to the greater realism of the ESEM approach which does not rely on the unrealistic assumption that items will be perfect indicators of the constructs they aim to assess. Rather, as we noted in the introduction, items tend to be fallible indicators including at least some degree of relevant association with constructs other than the main constructs that they are designed to measure (this relates to their imperfect validity). For this reason, items tend to present some degree of systematic association with other constructs, a phenomenon that generally tends to be reinforced for complex multidimensional instruments assessing conceptually-related and partially overlapping constructs. Unfortunately, a common idea is that “good” items provide a “perfect” reflection of a single construct, and that cross-loadings “taint” the meaning of the constructs (see Marsh et al., 2014). However, according to the reflective logic of factor analyses, it is the factors that are specified as influencing the indicators, not the reverse. Thus, small cross-loadings should be seen as reflecting the influence of the factor on the indicators, rather than as any kind of influence of the indicators on the factor. Naturally, this interpretation concerns only relatively small cross-loadings, whereas large and unexplainable cross-loadings should be re-examined. Hence, although seeking parsimony and simpler representation of the data is not inherently problematic and understandable in applied research, it remains important to keep in mind that “pure” indicators seldom exist in reality. Of course, in some cases the distortions attributable to tiny cross-loadings might be deemed ignorable. However, these decisions should be based on empirical results. In this study, empirical results as well as theoretical expectations both clearly showed that self-concept researchers and practitioners should be well aware that cross-loadings are to be expected in self-concept instruments and should not be

ignored. Ignoring these cross-loadings carries the risk of resulting in inflated correlations among self-concept dimensions, and lower discriminant validity.

Grade-Related Differences in Self-concept Differentiation

This study tested grade-related differences in self-concept differentiation based on ESEM, allowing us to disentangle three possible sources of variation in the previously reported correlations between self-concept facets. First, we tested whether children of grade levels 3 to 6 differed in the reliability of self-concept ratings. Our results suggest, for both the SDQ-I and the SPPC, that children of grades 3 to 6 did not differ in relation to the reliability of their self-concept ratings, as demonstrated by the invariance of the items' uniquenesses. This finding counters previous research suggesting that self-concept measurement may become more reliable with age (Marsh, Debus et al., 2005). However, the age-related improvement in the precision of self-concept ratings may be already completed at grade 3 (i.e., the youngest subgroup in this study). In fact, previous reports of increased reliability of self-concept measures have been predominantly based on preschool children (e.g., Marsh et al., 1991). Thus, a greater age range might be needed to identify developmental changes in the reliability of self-concept measures. Nonetheless, our results show that both the SDQ-I and SPPC can be used with equivalent reliability for children from grades 3 to 6. It should be kept in mind that we relied on the German versions of the SDQ-I and SPPC, which might differ from the English instruments in terms of reliability and sensitivity to age.

Second, age-related differences in self-concept differentiation might be due to differences in the validity of self-concept items as indicators of a single self-concept factor. This possibility is easily tested within ESEM by constraining the factor loadings and cross-loadings to be invariant across grade levels. Once again, our results showed that German students did not differ in this specific form of systematic measurement error across grades 3 to 6, as demonstrated by the invariance of factor loadings and item cross-loadings across grade levels for both instruments.

Third, the most direct way to examine self-concept differentiation involves testing whether the covariances between the various self-concept factors are invariant across grade levels once the other sources of differentiation have been shown, and constrained, to be invariant. Our results clearly showed that the latent covariances between the self-concept factors measured with either the SDQ-I or the SPPC were invariant across grades 3 to 6. In other words, these findings showed no sign of differences in self-concept differentiation across grade levels. This finding stands in stark contrast to previous studies and theoretical perspectives suggesting an increasing differentiation between self-concept facets as a function of age (Marsh, 1989; Marsh & Ayotte, 2003; Shavelson et al., 1976). Once again, it might be that the differentiation process is already completed or not yet set in motion for our sample so that future studies should consider a wider age range. Furthermore, given that this study relied on a sample of German students, cultural diversities or divergences in the school system might account to some extent for these differences with previous studies. However, given that our results could be replicated with two different self-concept measures, it might also be that previously found age differences in self-concept are less pronounced than originally assumed. Nonetheless, further longitudinal data conducted with students from other countries, educational systems, and age groups would provide an even more sophisticated approach to study developmental changes in the self-concept construct.

Endnote

¹ Although the CFI associated with regular ML estimation is monotonic with model complexity (i.e. cannot increase when constraints are added), we rely on the MLR estimator for which corrections factors are taken into account in the adjustments of the chi-square tests and resulting CFIs. When these correction factors changes across nested models, the CFI can be non-monotonic, which should simply be ignored and interpreted as supporting equivalent levels of fit.

References

- Abu-Hilal, M. M., & Aal-Hussain, A.-Q. A. (1997). Dimensionality and hierarchy of the SDQ-I in a non-western milieu: A test of self-concept invariance across gender. *Journal of Cross-Cultural Psychology*, 28, 535-553.
- Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept: Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, 103, 970-981.
- Arens, A. K., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2013). A short German version of the

- Self-Description Questionnaire I: Theoretical and empirical comparability. *International Journal of Research & Method in Education*, 36, 415-438.
- Asendorpf, J. B., & van Aken, M. A. G. (1993). Deutsche Version der Selbstkonzeptskalen von Harter [German version Harter's self-concept scales]. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 25, 64-86.
- Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397-438.
- Boivin, M., Vitaro, F., & Gagnon, C. (1992). A reassessment of the Self-Perception Profile for Children: Factor structure, reliability, and convergent validity of a French version among second through six grade children. *International Journal of Behavioral Development*, 15, 275-290.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111-150
- Browne, M. W., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. A. Bollen, & J. S. Long (Eds.), *Testing structural equation models* (pp. 136-162). Newbury Park, CA: Sage.
- Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, DC: American Psychological Association.
- Chan, D. W. (2002). Perceived domain-specific competence and global self-worth of primary students in Hong Kong. *School Psychology International*, 23, 355-368.
- Chen, F. F. (2007). Sensitivity of goodness of fit indices to lack of measurement invariance. *Structural Equation Modeling*, 14, 464-504.
- Cheung, G. W. & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, 9, 233-255.
- Cole, D. A., Maxwell, S. E., Martin, J. M., Peeke, L. G., Seroczynski, A. D., Tram, J. M., et al., (2001). The development of multiple domains of child and adolescent self-concept: A cohort sequential longitudinal design. *Child Development*, 72, 1723-1746.
- Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford.
- Eiser, C., Eiser, J.R., & Havermanns, T. (1995). The measurement of self-esteem: Practical and theoretical considerations. *Personality and Individual Differences*, 18, 429-432.
- Fox, K. R., & Corbin, C. B. (1989). The physical self-perception profile: development and preliminary validation. *Journal of Sport & Exercise Psychology*, 11, 408-430.
- Gavin, D.A.W., & Herry, Y. (1996). The French Self-Perception Profile for Children: Score validity and reliability. *Educational and Psychological Measurement*, 56, 678-700.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60, 549-576.
- Guay, F., Morin, A. J. S., Litalien, D., Valois, P., & Vallerand, R. J. (2014). Application of exploratory structural equation modeling to evaluate the academic motivation scale. *The Journal of Experimental Education*. Early View: DOI: 10.1080/00220973.2013.876231
- Harter, S. (1985). *Self-Perception Profile for Children: Revision of the Perceived Competence Scale for Children*. Denver, CO: University of Denver.
- Harter, S. (1988). *Self-Perception Profile for Adolescents*. Denver, CO: University of Denver.
- Harter, S. (1999). *The construction of the self. A developmental perspective*. New York: Guilford.
- Hu, L.-T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Marsh, H. W., & MacDonald-Holmes, I. W. (1990). Multidimensional self-concepts: Construct validation of responses of children. *American Educational Research Journal*, 27, 89-117.
- Maïano, C., Morin, A. J. S., Ninot, G., Monthuy-Blanc, J., Stephan, Y., Florent, J.-F., et al., (2008). A short and very short form of the physical self-inventory for adolescents: Development and factor validity. *Psychology of Sport & Exercise*, 9, 830-847.
- Marsh, H.W. (1986a). Negative item bias in rating scales for preadolescent children: A cognitive-developmental phenomenon. *Developmental Psychology*, 22, 37-49.
- Marsh, H. W. (1986b). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, 23, 129-149.
- Marsh, H. W. (1989). Age and sex effects in multiple dimensions of self-concept: Preadolescence to early adulthood. *Journal of Educational Psychology*, 82, 417-430.
- Marsh, H. W. (1990a). Influences of internal and external frames of reference on the formation of

- math and English self-concepts. *Journal of Educational Psychology*, 82, 107-116.
- Marsh, H. W. (1990b). *Self-Description Questionnaire – I (SDQ-I). Manual*. Macarthur, NSW. Australia: University of Western Sydney.
- Marsh, H. W. (1990c). *Self-Description Questionnaire – II (SDQ-II). Manual*. Macarthur, NSW. Australia: University of Western Sydney.
- Marsh, H. W. (2007). *Self-concept theory, measurement and research into practice: The role of self-concept in educational psychology*. Leicester, UK: British Psychological Society.
- Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker, P. (2013). Factor structure, discriminant and convergent validity of TIMSS math and science motivation measures: A comparison of USA and Saudi Arabia. *Journal of Educational Psychology*, 105, 108-128.
- Marsh, H. W., Asçl, F. H., & Marco, I. T. (2002). Multitrait-multimethod analyses of two physical self-concept instruments. *Journal of Sport & Exercise Psychology*, 24, 99-119.
- Marsh, H. W., & Ayotte, V. (2003). Do multiple dimensions of self-concept become more differentiated with age? The differential distinctiveness hypothesis. *Journal of Educational Psychology*, 95, 687-706.
- 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,
- Marsh, H. W., Craven, R. G., & Debus, R. (1991). Self-concepts of young children aged 5 to 8: Their measurement and multidimensional structure. *Journal of Educational Psychology*, 83, 377-392.
- Marsh, H. W., Debus, R., & Bornholt, L. J. (2005). Validating young children's self-concept responses: Methodological ways and means to understand their responses (pp.138-160). In D. M. Teti (Ed.), *Handbook of Research Methods in Developmental Science*. Oxford, UK: Blackwell.
- Marsh, H.W., & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, 32, 151-171.
- Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit evaluation. In A. Maydeu-Olivares & J. McArdle (Eds.), *Contemporary psychometrics*. Mahwah NJ: Erlbaum.
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999). *Structural Equation Modeling*, 11, 320-341.
- Marsh, H.W., Liem, G.A.D., Martin, A.J., Morin, A. J.S., & Nagengast, B. (2011). Methodological measurement fruitfulness of exploratory structural equation model (ESEM): New approaches to key substantive issues in motivation and engagement. *Journal of Psychoeducational Assessment*, 29, 322-346.
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22, 471-491.
- Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & Von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, 18, 257-284.
- Marsh, H. W., Morin, A. J. S., Parker, P., & 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., Asparouhov, A., 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. (2012). 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., Nagengast, B., Morin, A. J. S., Parada, R. H., Craven, R. G., & Hamilton, L. R. (2011). Construct validity of the multidimensional structure of bullying and victimization: An application of exploratory structural equation modeling. *Journal of Educational Psychology*, 103, 701-732.
- Marsh, H. W., & O'Neill, R. (1984). Self-Description Questionnaire III: 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., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg self-esteem scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22*, 366-381.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.
- Messer, B., & Harter, S. (1986). *Manual for the Adult Self-Perception Profile*. Denver, CO: University of Denver.
- Miller, H.-M., (2000). Cross-cultural validity of a model of self-worth application to Finnish children. *Social Behavior and Personality, 28*, 105-118.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Möller, J., & Marsh, H. (2013). Dimensional comparison theory. *Psychological Review, 120*, 544-560.
- Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). Meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research, 79*, 1129-1167.
- Morin, A. J. S., & Maïano, C. (2011). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM). *Psychology of Sport & Exercise, 12*, 540-554.
- Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling. In Hancock, G. R., & Mueller, R. O. (Eds.). *Structural equation modeling: A second course* (2nd ed., pp. 395-436). Charlotte, NC: Information Age Publishing, Inc.
- Muris, P., Meesters, C., & Fijen, P. (2003). The Self-Perception Profile for Children: Further evidence for factor structure, reliability, validity. *Personality & Individual Differences, 35*, 1791-1802.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.
- Pinxten, M., Marsh, H. W., De Fraine, B., Van Den Noortgate, W., & Van Damme, J. (2014). Enjoying mathematics or feeling competent in mathematics? Reciprocal effects on mathematics achievement and perceived math effort expenditure. *British Journal of Educational Psychology, 84*, 152-174.
- Schmitt, T. A., & Sass, D. A. (2011). Rotation criteria and hypothesis testing for exploratory factor analysis: implications for factor pattern loadings and interfactor correlations. *Educational & Psychological Measurement, 71*, 95-113.
- Schumann, B. C., Striegel-Moore, R. H., McMahon, R. P., Waclawiw, M. A., Morrison, J. A., & Schreiber, G. B. (1999). Psychometric properties of the Self-Perception Profile for Children in a biracial cohort of adolescent girls: The NHLBI growth and health study. *Journal of Personality Assessment, 73*, 260-275.
- Shavelson, R. J., Hubner, J. J. & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Journal of Educational Research, 46*, 407-441.
- Shelvin, M., Adamson, G., & Collins, K. (2003). The Self-Perception Profile for Children (SPPC): A multiple-indicator multiple-wave analysis using LISREL. *Personality and Individual Differences, 35*, 1993-2005.
- Stewart, P. K., Roberts, M. C., & Kim, K. L. (2010). The psychometric properties of the Harter Self-Perception Profile for children with at-risk African American Females. *Journal of Child and Family Studies, 19*, 326-333.
- Trent, L. M. Y., Russell, G., & Cooney, G. (1994). Assessment of self-concept in early adolescence. *Australian Journal of Psychology, 46*, 21-28.
- Van den Bergh, B. R. H., & Van Rast, N. (1998). Self-concept in children: Equivalence of measurement and structure across gender and grade of Harter's Self-Perception Profile for Children. *Journal of Personality Assessment, 70*, 564-582.
- Watkins, D., & Akande, A. (1992). Internal structure of the Self-Description Questionnaire: A Nigerian investigation. *British Journal of Educational Psychology, 62*, 120-125.
- Watkins, D., & Dong, Q. (1994). Assessing the self-esteem of Chinese school children. *Educational Psychology, 14*, 129-137.

Table 1
Goodness-of-fit indices for CFA and ESEM Measurement Models

	χ^2	df	SCF	CFI	TLI	RMSEA	RMSEA 90% CI	Model description
SDQ-I								
Model 1	8183.740	2629	1.3782	.924	.917	.033	.032 - .034	CFA: Total sample
Model 2	4417.119	1979	1.3291	.967	.952	.025	.024 - .026	ESEM: Total sample
Model 3	13059.275	7916	1.4289	.935	.907	.036	.035 - .038	ESEM: Configural invariance
Model 4	14833.678	10061	1.3779	.940	.932	.031	.030 - .032	ESEM: Weak invariance (loadings)
Model 5	15236.366	10256	1.4254	.937	.930	.031	.030 - .033	ESEM: Strong invariance (loadings, intercepts)
Model 6	16021.395	10485	1.3524	.930	.924	.033	.032 - .034	ESEM: Strict invariance (loadings, intercepts, uniq.)
Model 7	16175.453	10682	1.3587	.931	.926	.032	.031 - .033	ESEM: Invariance of loadings, intercepts, uniq., and CUs (negative items)
Model 8	16051.825	10556	1.3523	.931	.925	.033	.032 - .034	ESEM: Invariance of loadings, intercepts, uniq., and CUs (parallel-worded items)
Model 9	16247.576	10754	1.3596	.931	.927	.032	.031 - .033	ESEM: Invariance of loadings, intercepts, uniq., and CUs (all).
Model 10	16681.408	10952	1.3381	.928	.925	.033	.032 - .034	ESEM: Invariance of loadings, intercepts, uniq., CUs, and variances/covariances
SPPC								
Model 11	1578.361	395	1.3035	.923	.916	.039	.037 - .041	CFA: Total sample
Model 12	1006.025	295	1.2415	.954	.932	.035	.033 - .038	ESEM: Total sample
Model 13	2118.179	1180	1.1534	.942	.914	.041	.038 - .043	ESEM: Configural invariance
Model 14	2567.390	1555	1.2105	.937	.930	.037	.034 - .039	ESEM: Weak invariance (loadings)
Model 15	2720.609	1630	1.2010	.933	.928	.037	.035 - .040	ESEM: Strong invariance (loadings, intercepts)
Model 16	2885.633	1720	1.2204	.928	.927	.037	.035 - .040	ESEM: Strict invariance (loadings, intercepts, uniq.)
Model 17	2960.185	1765	1.2249	.926	.927	.037	.035 - .040	ESEM: Invariance of loadings, intercepts, uniq., and variances/covariances

Note. SCF = scaling correction factor associated with the χ^2 ; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; CUs = correlated uniquenesses. ESEM models were conducted with target oblique rotation. All models integrated correlated uniquenesses between negative and parallel-worded items and were conducted with the MLR estimator. All χ^2 values are all significant ($p < .01$).

Table 2

Factor Correlations for the SDQ-I: CFA (Model 1; above the diagonal) and ESEM (Model 2; under the diagonal)

	Appearance	Physical Ability	Peer	Parents	School Comp.	School Affect	German Comp.	German Affect	Math Comp.	Math Affect	General
Appearance		.342**	.620**	.356**	.348**	.310**	.281**	.240**	.203**	.170**	.725**
Physical Ability	.300**		.408**	.209**	.234**	.248**	.131**	.139**	.207**	.214**	.397**
Peer	.492**	.367**		.355**	.334**	.260**	.300**	.210**	.207**	.119**	.727**
Parents	.322**	.211**	.274**		.347**	.373**	.249**	.301**	.264**	.261**	.550**
School Comp.	.155**	.122**	.160**	.063		.732**	.720**	.505**	.637**	.449**	.548**
School Affect	.131**	.146**	.100**	.263**	.160**		.506**	.695**	.487**	.610**	.465**
German Comp.	.140**	.044	.158**	.203**	.229**	.404**		.780**	.252**	.105**	.422**
German Affect	.127**	.110**	.128**	.184**	.275**	.350**	.397**		.149**	.212**	.346**
Math Comp.	.104**	.166**	.113**	.217**	.257**	.363**	.277**	-.019		.814**	.353**
Math Affect	.151**	.209**	.071**	.230**	.213**	.350**	-.061**	.170**	.637**		.270**
General	.601**	.320**	.562**	.443**	.244**	.228**	.260**	.205**	.255**	.191**	

Note. * $p < .05$. ** $p < .01$.

Table 3
Standardized ESEM Factor Loadings, Cross-Loadings and Residuals for the SDQ-I items (Model 2)

Item	Global	Peer	Physical Appearance	Physical Ability	Parent	School Competence	School Affect	German Competence	German Affect	Math Competence	Math Affect	Residual
29	.244**	.084*	.055	.068**	.024	.075*	.108**	-.042	.155**	-.042	.038	.752**
37	.253**	.048	.193**	-.033	.152**	-.060	.024	.029	-.100**	.045	-.033	.747**
45	.331**	-.102**	.290**	.069**	.191**	-.079	.014	-.084*	.016	.025	.007	.613**
53	.458**	.082**	.016	.070**	.186**	-.013	.033	-.003	.003	-.011	.017	.564**
61	.374**	.003	.017	-.018	.218**	.017	.000	.205**	-.144**	.010	-.001	.663**
67	.590**	.106**	-.040	.046*	.016	.043	.018	.027	.001	.110**	-.017	.488**
70	.457**	.268**	.010	-.053*	.015	-.033	.057*	.016	-.017	-.138**	.101**	.588**
72	.617**	.012	.239**	.027	-.032	.001	.005	-.003	-.034	-.102**	.072*	.399**
74	.642**	.050	.010	.021	-.051*	.017	-.008	.070*	.016	.095**	-.066	.497**
76	.648**	.002	-.010	.049*	.039	.074**	.035	-.033	.036	.056	-.038	.484**
7	-.147*	.639**	-.003	.055*	.120**	-.071	.021	-.002	-.006	.031	-.044	.614**
14	-.087*	.640**	.002	.006	-.015	.030	.012	.053	.006	.029	-.035	.622**
21	-.133**	.691**	-.021	.045*	.043	-.050	.003	.088**	-.061*	.072*	-.091**	.565**
28	-.060	.582**	-.021	.065**	.139**	-.002	.038	.010	.002	-.028	.045	.607**
36	.122**	.373**	.217**	-.025	-.007	.017	.021	.012	-.047	-.032	.030	.648**
44	.102**	.620**	.078**	-.039*	.003	-.012	.000	-.011	.020	-.040	.053	.489**
52	.058	.584**	.022	.002	-.039	.015	-.029**	-.099**	.066*	.083*	-.027	.606**
60	.204**	.567**	.037	.022	-.041	.047	-.090*	.009	-.018	.041	.003	.470**
69	.220**	.649**	.038	.013	-.020	-.028	-.008	-.008	-.011	-.068*	.054	.356**
1	-.036	.015	.767**	.023	-.034	-.018	.001	.020	-.032	-.015	.031	.437**
8	.020	-.033	.702**	.067**	.044*	-.068*	.001	-.014	.003	.017	.045	.454**
15	-.094**	.031	.872**	-.031*	-.013	.006	.021	.018	.058*	.025	-.037	.304**
22	-.057	.019	.907**	-.055**	-.005	.011	.044*	-.017	.035	.039	-.050*	.236**
30	.127**	-.081**	.612**	-.001	.102**	-.089*	-.031	-.019	.005	.071	.001	.519**
38	.026	.336**	.418**	.015	-.073**	.014	.026	.009	-.002	-.051	.018	.566**
46	.118*	.004	.371**	.244**	.045	.126**	-.004	.032	-.011	-.033	.014	.593**
54	.160**	.111**	.424**	.050*	-.095**	.111*	-.053	-.095**	.005	.076	-.105*	.632**
62	.326**	.020	.369**	-.026	.006	-.027	-.032	.056	.013	-.089*	.070	.602**
3	-.089*	.033	.102**	.694**	-.004	.202**	-.131**	.065*	-.003	-.085**	.066*	.427**
10	-.059	-.027	.002	.560**	.088**	.028	.025	-.040	.053	-.027	.003	.677**
17	.015	-.001	-.065**	.784**	-.042	-.384**	.215**	.051	-.115**	.045	-.079*	.274**
24	.007	.007	-.056**	.832**	-.043*	-.318**	.171**	.008	-.024	.037	-.069*	.257**
32	.121*	.049	.058	.383**	-.051	.101*	-.109**	-.151**	.102**	.053	.010	.720**
40	-.017	-.013	.012	.892**	-.006	.027	-.054*	.027	-.032	-.028	.012	.229**
48	.033	.018	-.037	.657**	.031	.180**	-.109**	.013	.034	.008	.054	.473**
56	.013	-.017	.034	.882**	-.013	.140**	-.073**	.022	-.004	-.071*	.022	.188**
64	.005	.036	.039	.534**	.022	.120**	-.058	-.140**	.040	.079*	.025	.612**

Item	Global	Peer	Physical Appearance	Physical Ability	Parent	School Competence	School Affect	German Competence	German Affect	Math Competence	Math Affect	Residual
5	.042	.008	.011	.024	.628**	-.019	.019	.003	.040	.003	.007	.543**
12	.103*	.000	.006	-.045	.356**	-.035	.021	.129**	-.059	.147**	-.068	.761**
19	-.003	-.042	.033	.022	.640**	-.042	-.027	-.039	.062*	-.004	-.006	.591**
26	.014	.009	.070*	.001	.602	-.077*	-.055	.007	.035	.043	-.037	.600**
34	.012	.088**	-.012	-.001	.489**	.025	.071*	.008	-.003	-.018	-.011	.707**
42	.059	.104**	-.036	.020	.587**	.031	.046	-.050	.022	.023	-.040	.571**
50	-.023	-.005	-.011	-.018	.827**	.024	-.022	-.052*	.054*	.041	.007	.332**
58	.005	-.016	.003	-.032*	.897**	.040	-.003	-.050	-.006	-.021	.002	.229**
66	.071	.062*	-.073**	.034*	.754**	.039	-.016	-.005	-.055*	-.095**	.071*	.393**
2	.011	.003	.103**	.015	.027	.312**	.317**	.334**	-.181**	.186**	-.051	.431**
16	-.014	.039	.064**	.029	.067**	.346**	.317**	.378**	-.206**	.155**	-.024	.370**
31	.125**	.018	-.029	.041*	.032	.274**	.224**	.227**	-.020	.261**	-.059	.495**
47	.190	-.025	.012	.037	.100	.047	-.014	.280**	.016	.169**	.004	.718**
63	.146**	.011	.006	.012	.001	.339**	.377**	.126**	-.033	.251**	-.064*	.386**
9	.008	-.003	.109**	.006	.114**	.201**	.484**	-.021	.120**	-.096**	.177**	.472**
23	.067	.013	.019	.005	.093**	-.108**	.218**	.133**	.132**	-.038	.217**	.737**
39	.080**	-.010	.020	.045**	.009	.189**	.611**	-.097**	.186**	-.026	.137**	.331**
55	.069**	.030	.010	.021	.030	.165**	.583**	-.115**	.249**	-.056	.205**	.289**
71	.070**	.018	-.013	.035*	.023	.184**	.608**	-.134**	.226**	.074*	.115**	.276**
4	.025	.027	.052*	-.023	.013	.178**	-.028	.783**	.018	-.104**	.092**	.303**
18	-.030	.065**	.037*	.007	.013	.132**	-.056**	.729**	.218**	-.010	.020	.234**
33	.070*	.004	.037	.012	.033	.032	-.057*	.667**	.168**	-.009	.066*	.394**
49	.067*	.054*	-.020	-.002	-.009	.088**	-.077**	.593**	.342**	.032	-.013	.317**
73	.170**	.040	-.047*	.052**	-.039*	.145**	-.003	.467**	.332**	.090**	-.098**	.343**
11	-.057*	.002	.098**	-.017	.038*	-.039	.118**	.275**	.624**	-.033	-.004	.311**
25	-.032	.011	.069**	.002	.054**	-.054*	.151**	.120**	.689**	.049	-.023	.323**
41	.005	.028	.009	.046**	.044**	-.039	.154**	.131**	.711**	.021	-.032	.266**
57	.019	.001	.017	.009	.049**	-.025	.166**	.149**	.726**	.015	-.005	.204**
65	.068*	-.006	-.028	.002	.036	-.073*	.109**	.224**	.602**	.027	-.027	.396**
13	-.032	.042*	.025	-.008	-.005	.048**	-.044*	-.102**	.027	.759**	.145**	.300**
27	-.038	.023	.068**	-.016	.025	.092**	-.014	.016	-.012	.783**	.040	.289**
43	.042	.061**	-.019	.039**	.030	.078**	-.045	-.037	.032	.677**	.211**	.250**
59	.024	.007	.035**	.011	.007	.061**	-.040*	-.032	.045*	.781**	.171**	.164**
75	.061*	-.014	-.004	.027	.046*	.014	-.048*	-.012	-.012	.689**	.152**	.341**
6	-.022	-.033	-.007	.007	.041*	-.039	.004	.039	-.090**	.135**	.719**	.359**
20	.007	-.013	.008	-.008	.012	-.034*	.121**	.015	-.018	.097**	.777**	.214**
35	-.025	.021	.011	.024	-.002	-.032	.096**	.010	.004	.126**	.768**	.205**
51	.036	-.024	.000	.022*	-.030	-.061**	.071**	.064**	-.057**	.087**	.879**	.101**
68	-.006	.032	-.037*	.043**	.002	.003	.083**	.001	.021	.156**	.728**	.222**

Note. * $p < .05$. ** $p < .01$.

Table 4

Factor Correlations for the SPPC: CFA (Model 9; above the diagonal) and ESEM (Model 10; under the diagonal)

	Scholastic Competence	Peer Acceptance	Athletic Competence	Physical Appearance	Global
Scholastic Competence		.417**	.366**	.403**	.549**
Peer Acceptance	.417**		.515**	.457**	.538**
Athletic Competence	.309**	.448**		.470**	.452**
Physical Appearance	.354**	.395**	.353**		.827**
General	.372**	.367**	.337**	.557**	

Note. * $p < .05$. ** $p < .01$.

Table 5
 Standardized ESEM Factor Loadings, Cross-loadings and residuals for the SPPC Items (Model 9)

	Scholastic Competence	Peer Acceptance	Athletic Competence	Physical Appearance	Global	Residuals
1	.505**	-.086**	.091**	-.014	.118**	.692**
6	.584**	-.067*	.057*	-.082	.202**	.586**
11	.628**	.078**	.018	-.025	-.107**	.605**
16	.622**	.018	-.002	.015	-.021	.607**
21	.677**	-.056*	.018	-.131**	.140**	.539**
26	.703**	.037	-.024	.082**	-.094*	.496**
2	-.011	.667**	.006	.045	-.069	.565**
7	-.062**	.749**	.070**	-.171**	.096**	.454**
12	.018	.717**	-.040	.104**	-.107**	.488**
17	-.055	.494**	.100**	-.147**	.165**	.701**
22	.032	.654**	-.051*	.080**	-.024	.548**
27	.017	.611**	.074**	-.093	.137**	.539**
3	-.041*	.016	.824**	-.034	-.039	.364**
8	.000	.076*	.571**	.194**	-.140**	.574**
13	.039	.013	.727**	-.059*	.061*	.440**
18	-.002	-.035	.705**	-.027	.009	.532**
23	.117**	.152**	.174**	.188**	-.158*	.843**
28	.138**	.115**	.418**	.090	-.080	.689**
4	-.068**	.063*	.092**	.366**	.340**	.553**
9	.013	-.052	.177**	.425**	.077	.704**
14	.014	-.047	.065	.711**	.020	.461**
19	-.036	.019	.004	.830**	.044	.274**
24	.066**	.003	-.029	.547**	.134**	.577**
29	-.035	.072*	.145**	.288**	.375**	.540**
5	.126**	.175**	-.001	.179**	.355**	.587**
10	.156**	.170**	-.075**	.144	.307*	.687**
15	.083	.039	.044*	.149*	.589**	.429**
20	.060	.087*	-.006	.182**	.471**	.566**
25	.011	.022	.038	.253**	.562**	.418**
30	.223**	.079*	-.031	.100	.360**	.662**

Note. * $p < .05$. ** $p < .01$.