

Improved Representation of the Self-Perception Profile for Children Through Bifactor Exploratory Structural Equation Modeling

A. Katrin Arens

German Institute for International Educational Research

Alexandre J. S. Morin

Australian Catholic University

This study illustrates an integrative psychometric framework to investigate two sources of construct-relevant multidimensionality in answers to the Self-Perception Profile for Children (SPPC). Using a sample of 2,353 German students attending Grades 3 to 6, we contrasted: (a) first-order versus hierarchical and bifactor models to investigate construct-relevant multidimensionality related to the hierarchical nature of multidimensional self-conceptions and (b) confirmatory factor analyses (CFA) and exploratory structural equation models (ESEM) to investigate construct-relevant multidimensionality related to the assessment of conceptually related constructs. The bifactor-ESEM solution provided the best fit, suggesting the presence of both sources of construct-relevant psychometric multidimensionality. The results supported measurement invariance of the SPPC across gender and grade level and showed latent mean differences mostly supporting results from previous research.

A. KATRIN ARENS is a postdoctoral researcher at the Department of Education and Human Development and Centre for Research on Individual Development and Adaptive Education (IDeA) of the German Institute for International Educational Research, Schloßstr. 29, D-60486 Frankfurt am Main, Germany; e-mail: arens@dipf.de. Her research focuses on the structure, development, and formation of students' multidimensional self-concept. On a more general level, she is interested in motivation in education, including various cognitive, affective, and behavioral motivational facets.

ALEXANDRE J. S. MORIN is a research professor at the Institute for Positive Psychology and Education at the Australian Catholic University, where he heads the Positive Substantive Methodological Synergy research program. He is a developmental and organizational psychologist with broad research interests anchored in the exploration of the social determinants of psychological wellbeing and psychopathologies at various life stages and in various settings, such as schools and organizations. He has authored or co-authored over 120 articles and book chapters.

KEYWORDS: bifactor models, exploratory structural equation modeling (ESEM), psychometric multidimensionality, self-concept

The Self-Perception Profile for Children (SPPC; Harter, 1985) is one of the most popular measures of preadolescents' multidimensional self-concepts. The SPPC belongs to a series of Self-Perception Profiles (SPP) designed to assess multidimensional self-concepts in different age groups, which also includes the Self-Perception Profile for Adolescents (SPPA; Harter, 1988) and the Adult Self-Perception Profile (ASPP, Messer & Harter, 1986). The original English SPPC integrates five scales for assessing domain-specific self-concepts related to scholastic competence, peer acceptance, physical appearance, athletic competence, and behavioral conduct, in addition to a scale assessing students' global self-concept. The SPPC has been extensively cross-culturally validated in multiple languages: Finnish (Miller, 2000), French (Boivin, Vitaro, & Gagnon, 1992), German (Asendorpf & van Aken, 1993), Chinese (Chan, 2002), and Dutch (Muris, Meesters, & Fijen, 2003). However, despite several studies on the psychometric properties of the SPPC based on exploratory (EFA) and confirmatory factor analyses (CFA), the SPPC's structure remains unclear, and the optimal way to model the global self-concept scale remains an open question. These issues will be further focused in the present study by applying and comparing alternative modeling approaches.

The Internal Structure of the SPPC

EFA studies have provided some support to the a priori five-factor structure for the domain-specific self-concept facets of the SPPC (e.g., Miller, 2000). However, these studies have tended to demonstrate substantial cross-loadings between items and non-target factors. Research has shown this to be the case for the items designed to measure the scholastic competence and behavioral conduct factors (Stewart, Roberts, & Kim, 2010), the physical appearance and peer acceptance factors (Chan, 2002; Stewart et al., 2010), and the behavioral conduct and athletic competence factors (Miller, 2000). CFA investigations of the SPPC and of other SPP-derived instruments (e.g., the Physical Self-Perception Profile, PSPP; Fox & Corbin, 1989) generally report correlations among the self-concept factors that are so high as to call into question their discriminant validity (Boivin et al., 1992; Cole et al., 2001; Maïano et al., 2008; Stewart et al., 2010).

In order to explain these apparently surprising results (e.g., EFA cross-loadings, high CFA correlations, and even low goodness of fit for CFA solutions), the idiosyncratic response scale used in the SPPC (and other SPP-based instruments) has often been invoked. The SPPC and other SPP-based instruments include a structured alternative response format according to which each item encompasses two opposite statements (e.g., "Some kids

often forget what they learn. But other kids can remember things easily”). One statement reflects a high level of self-concept (scholastic competence self-concept in this example) while the other one reflects a low level of self-concept. In a first step, the students are asked to choose the statement that best describes themselves. In a second step, they have to indicate whether this statement is “really true” or “sort of true” for them. This response format was originally established to prevent social desirability when answering the self-concept items. However, it has been criticized as being too complex and difficult to understand, to be associated with substantial method effects, and has generally been used to explain the previously described unexpected results in relation to the SPPC factor structure (Eiser, Eiser, & Havermanns, 1995; Marsh & McDonald Holmes, 1990; Marsh, Richards, Johnson, Roche, & Tremayne, 1994; Wichstrøm, 1995; Wylie, 1989). These criticisms have even led to the development of alternative versions of the SPPC relying on Likert-type response scales (Graziano, Jensen-Campbell, & Finch, 1997; Graziano & Ward, 1992).

There is, however, evidence that the problems associated with the SPP-based instruments do not exclusively originate from the idiosyncratic response scale. Wichstrøm (1995) compared the original SPPA with a revised version using a Likert-type response scale. Although EFA cross-loadings were reduced in the revised version, they still remained high. Similar conclusions came from studies using SPP-based physical self-concept instruments (e.g., Fox & Corbin, 1989; Maïano et al., 2008; Morin & Maïano, 2011), showing that EFA cross-loadings and CFA factor correlations tended to remain high irrespective of the type of response scale. Still, when EFA cross-loadings are examined, many appear to be theoretically meaningful given that self-concept is assumed to be a multidimensional construct consisting of separate facets sharing some level of conceptual overlap.

Methodologically, the existence of cross-loadings challenges the independent cluster model (ICM) assumption of CFA (McDonald, 1985), which assumes that each item loads only on a single factor, without cross-loadings. If true cross-loadings are forced to be zero, the only way for them to be expressed is through inflated factor correlations. Indeed, simulation studies revealed that EFA tends to produce more exact estimates of the true population correlations and to remain unbiased even when the true population model meets ICM-CFA assumptions (i.e., includes no cross-loading; for a review, see Asparouhov, Muthén, & Morin, 2015). Research also shows that even when the ICM-CFA model provides a satisfactory fit to the data, factor correlations will be at least somewhat inflated unless all cross-loadings are close to zero in the true population model (Marsh, Liem, Martin, Morin, & Nagengast, 2011; Marsh, Nagengast, et al., 2011). Studies based on simulated data show that constraining cross-loadings with a population value as low as .100 was sufficient to induce biased estimates of factor correlations (Asparouhov et al., 2015; Marsh, Lüdtke, Nagengast, Morin, & Von

Davies, 2013). These observations should not be taken as a justification for test developers to stop trying to create items that provide a “pure” reflection of a single construct or to mean that unexpectedly large cross-loadings should be accepted without questioning the relative value of the item. Rather, following Morin, Arens, and Marsh (2016), we argue that the aspiration to develop strong questionnaire items needs to coexist with a realistic approach to model estimation that considers the fallible nature of typical psychometric items.

These observations suggest that ICM assumptions may be too restrictive for the SPPC and may be responsible for some observations previously attributed to the SPPC response scale. Importantly, these observations argue in favor of EFA models incorporating cross-loadings as providing a more realistic measurement model for the SPPC. Unfortunately, EFA has been outshined by the methodological advances associated with CFA and its integration into the structural equation modeling (SEM) framework (e.g., modeling of latent constructs adjusted for measurement error, goodness of fit evaluation, and invariance testing). Furthermore, EFA has suffered from the erroneous assumption that it was unsuited to confirmatory studies of a priori hypotheses whereas ICM-CFA was more naturally suited to confirmatory tests of the appropriateness of theoretically driven models. As noted by Morin, Marsh, and Nagengast (2013), the only “critical difference between EFA and CFA is that all cross-loadings are freely estimated in EFA” but that “statistically, nothing precludes the use of EFA for confirmatory purposes” (p. 396).

Asparouhov and Muthén (2009) recently proposed exploratory structural equation modeling (ESEM; e.g., Marsh et al., 2009; Morin et al., 2013) as a new overarching framework incorporating EFA, CFA, and SEM. Furthermore, with target rotation, it becomes possible to specify a priori hypotheses regarding the expected factor structure, reinforcing the possibility to use ESEM for confirmatory purposes (Browne, 2001; Morin et al., 2013). With ESEM, cross-loadings between EFA factors can thus be specified based on a priori assumptions. These factors can then be included in tests of invariance or prediction. These characteristics make ESEM a more flexible approach than ICM-CFA and one that is more naturally suited to multidimensional constructs (e.g., Marsh et al., 2009; Marsh, Nagengast, et al., 2011; Marsh, Nagengast, & Morin, 2012). Applying ESEM to the SPP-based Physical Self-Inventory (PSI), which relies on a Likert-type response scale, Morin and Maïano (2011) showed that ESEM provided an improved representation of the data relative to ICM-CFA and evidenced multiple substantively meaningful cross-loadings and substantially reduced factor correlations. Arens and Morin (2016) reached similar conclusions in an ESEM analysis of responses to the SPPC obtained when using the original structured alternative response scale. These studies thus suggest that ESEM provides a viable alternative for SPP-based instruments, irrespective of the response scale that is used. Thus, the elevated CFA correlations that often

characterize the SPPC may be related to the need to rely on models providing a way to account for the cross-loadings present between the conceptually related subscales forming this instrument. In addition to providing a further test of the appropriateness of an ESEM representation of SPPC responses, the present study extends previous research by an investigation of the nature and appropriate depiction of the global self-concept scale.

The Nature of Global Self-Concept

The SPPC includes five domain-specific and one global self-concept (GSC) scales. The majority of studies conducted to investigate the factor structure of the SPPC have excluded the GSC scale from the analyses (Chan, 2002; Eiser et al., 1995; Miller, 2000; Muris et al., 2003; Shevlin, Adamson, & Collins, 2003). This decision was justified by Harter's (1985) argument that the GSC scale is qualitatively different from the multiple domain-specific self-concept scales and unlikely to "systematically emerge as a distinct factor" (p. 18). Many studies rather used the GSC scale as a dependent variable predicted by the domain-specific scales (Chan, 2002; Harter, 1999; Muldoon, 2000). However, excluding the GSC scale only provides a partial picture of the factor structure of the full instrument as it is used and administered in practice.

The GSC scale assesses children's overarching self-conceptions across multiple domains. Originally, GSC was conceptualized to encompass "affective" self-evaluations, while domain-specific self-concepts were assumed to reflect "cognitive" self-descriptions (Baumeister, Campbell, Krueger, & Vohs, 2003; Rosenberg, 1979). However, this assumption has been rejected in later empirical studies and theoretical conceptualizations (Marsh, 2007; Marsh & Hattie, 1996; Shavelson, Hubner, & Stanton, 1976), leading to the current view that both GSC and domain-specific self-concepts encompass affective as well as cognitive components that simply differ in their level of generality, suggesting alternative representations of the GSC scale.

In the first and simplest representation, GSC is assumed to form a separate domain located at the same level of specificity as the domain-specific factors. Thus, independently from their self-concepts in specific domains, individuals are assumed to make separate judgments of their GSC (e.g., Rosenberg, 1965). This representation calls for simple ICM-CFA or ESEM models where all domain-specific self-concept factors as well as the GSC factor are included as correlated first-order factors.

Alternatively, GSC has often been conceptualized as a hierarchically superordinate construct relative to domain-specific self-concepts. In Shavelson et al.'s (1976) model, GSC was presumed to be located at the apex of the self-concept hierarchy, with domain-specific self-concept facets located at the lower levels. Methodologically, this hierarchically ordered structure suggests a higher-order model in which GSC is defined by

domain-specific self-concepts facets. Studies based on the German (Asendorpf & van Aken, 1993), Dutch (Muris et al., 2003), and English (Shevlin et al., 2003) SPPCs supported this representation. However, these studies only estimated the higher-order factor using the domain-specific SPPC subscales as first-order factors without including the items designed to directly assess GSC, thus leaving out a significant part of the instrument.

Bifactor models (Chen, West, & Sousa, 2006; Reise, Moore, & Haviland, 2010) provide another potential representation of the SPPC. In bifactor models, all items are assumed to simultaneously load on a global factor (G-factor) pertaining to all items forming the instrument and on specific factors (S-factors) representing each of the a priori subscales of the instrument. In CFA-based bifactor models, the items' loadings on the G-factor and on one of the S-factors are freely estimated while their cross-loadings on the other S-factors are constrained to be zero. In bifactor models, factors are usually set to be orthogonal (i.e., the correlations between the S-factors and between the S-factors and the G-factor are all constrained to be zero) to facilitate interpretations. An orthogonal bifactor model thus partitions the total covariance among the items into a G-factor underlying all items and multiple S-factors explaining the residual covariance not explained by the G-factor. Given their greater flexibility, bifactor models usually present a greater degree of fit to the data than hierarchical models.¹ In these models, if GSC does in fact constitute an overarching construct encompassing domain-specific self-concepts, a well-defined G-factor should emerge from answers provided to all SPPC items.

Bifactor models are well established in research on the structure of intelligence (Gignac & Watkins, 2013) and personality (Chen, Hayes, Carver, Laurenceau, & Zhang, 2012) and have recently been successfully applied to other constructs such as quality of life (Reise, Morizot, & Hays, 2007), attention disorders (Morin, Tran, & Caci, 2016), or internalizing disorders (Gignac, Palmer, & Stough, 2007) but have yet to be systematically applied to SPP-based multidimensional self-concept measures. Brunner, Lüdtke, and Trautwein (2008) used a bifactor representation to investigate the factor structure of a set of academic self-concept items available in the German extension of the Programme for International Student Assessment (PISA; Organisation for Economic Co-operation and Development [OECD], 2001). Their results, which supported a bifactor representation, have been replicated among 26 countries from the PISA 2000 study (Brunner, Keller, Hornung, Reichert, & Martin, 2009) as well as using a more extensive set of academic self-concept facets (Brunner et al., 2010). More recently, Morin, Arens, et al. (2016) provided evidence supporting a bifactor representation of the more extensive Self-Description Questionnaire I (SDQ I). These studies suggest that bifactor models might provide valuable insights into the structure of multidimensional self-conceptions.

A key issue that needs to be considered when implementing a bifactor representation of multidimensional self-concept ratings is how to model the GSC items. In Brunner et al.'s (2008, 2009, 2010) studies, global academic self-concept items were only allowed to define the G-factor and not associated with a S-factor of their own. However, as noted by Morin, Arens, et al. (2016), the alternative approach through which GSC items are used to define a G-factor as well as a S-factor provides a direct test of whether the GSC items retain residual specificity over and above the G-factor. GSC items aim to reflect overarching self-conceptions across various domains. However, the SPPC only directly assesses five specific domains (scholastic competence, peer acceptance, physical appearance, athletic competence, and behavioral conduct), leaving out many potentially important ones (e.g., familial, spiritual, artistic; Marsh, 2007; Shavelson et al., 1976; Vispoel, 1995). Including a S-factor based on the GSC items provides a test of whether these items retain any meaningful specificity over and above the content already covered in the other SPPC items. Should this GSC S-factor prove to be meaningful, its inclusion in the model provides a direct way to represent omitted self-concept domains (Morin, Arens, et al., 2016). Otherwise, failure of this S-factor to retain any specificity would confirm that the self-concept facets included in the SPPC provide a parsimonious coverage of the domains most relevant to children's global self-conceptions.

Construct-Relevant Psychometric Multidimensionality

So far, we have outlined two psychometric issues in need of more systematic investigation in research on the SPPC, namely, the need to clarify its internal factor structure and the nature of the GSC scale. These issues reflect construct-relevant psychometric multidimensionality, which refers to the possibility that items may be related to more than one content domain (i.e., with more than one source of true score variance; see Morin, Arens, et al., 2016).² It is important to differentiate psychometric multidimensionality from substantive multidimensionality, with the later referring to instruments (e.g., the SPPC) in which separate sets of items are used to assess multiple dimensions. In the SPPC, two sources of construct-relevant psychometric multidimensionality are likely to be present due to (a) the hierarchical nature of the instrument where all items are expected to present a significant level of association with their own subscales (e.g., peer acceptance) as well as with the hierarchically superior GSC construct and (b) the fallible nature of indicators, as the items assess conceptually related self-concept domains (e.g., athletic competence and physical appearance). Whereas the first source of construct-relevant multidimensionality calls for a hierarchical or bifactor model, the second source calls for an ESEM representation. It has only recently become possible to combine these models. The ESEM-within-CFA method (Marsh, Nagengast, et al., 2011; Morin et al., 2013) allows

for a specific ESEM solution to be reexpressed using CFA, enabling tests of hierarchical models where the first-order structure corresponds to the initial ESEM solution. Bifactor-ESEM estimation has been recently made possible with the development of bifactor rotations (Jennrich & Bentler, 2011; Reise, Moore, & Maydeu-Olivares, 2011).

Statistical research has shown that these alternative models are quite efficient at absorbing unmodeled sources of construct-relevant psychometric multidimensionality. For this reason, whenever a measure is expected to include both sources of construct-relevant psychometric multidimensionality, it becomes critical to embark on a rigorous comparison of these different solutions. Unmodeled cross-loadings tend to result in inflated estimates of factor correlations in CFA or of the global factor in bifactor-CFA, whereas an unmodeled global factor tends to result in inflated cross-loadings in EFA/ESEM or inflated factor correlations in CFA (e.g., Asparouhov et al., 2015; Morin, Arens, et al., 2016; Murray & Johnson, 2013). Morin, Arens, et al. (2016) provide a vivid illustration of this phenomenon, showing that goodness-of-fit assessment is not sufficient to recover the underlying population-generating model but needs to be accompanied by a complete examination of the parameter estimates generated by these alternative models. They recommend starting this process by a comparison of ESEM and CFA. An ESEM representation should be retained when it results in well-defined factors and reduced factor correlations. The second criterion is particularly important to consider given the observation that unmodeled cross-loadings tend to result in inflated CFA factor correlations (see Asparouhov et al., 2015). The ESEM representation then needs to be contrasted with a bifactor-ESEM representation whenever there are reasons to expect the presence of an overarching global factor. In this solution, observing a well-defined G-factor coupled with slightly reduced cross-loadings would argue in favor of a bifactor-ESEM representation. The second criterion is aligned with the observation that an unmodeled G-factor is likely to result in inflated cross-loadings (Morin, Arens, et al., 2016).

Gender and Grade-Level Differences

Studies exploring gender differences in multidimensional self-concepts based on the SPPC (e.g., Eiser et al., 1995; Muldoon, 2000) or other instruments (Kling, Hyde, Showers, & Buswell, 1999; Marsh, 1989; Wigfield, Eccles, MacIver, Reuman, & Midgley, 1991) generally revealed stereotypical gender differences showing that boys tend to present higher levels of athletic competence, physical appearance, math self-concepts, and GSC relative to girls, whereas girls tend to display higher levels of self-concept in the verbal and behavioral conduct areas. Only a few studies have examined age- or grade-related differences in the mean levels of self-concept facets using the SPPC. Whereas Muldoon (2000) reported decreasing levels for some of

the SPPC self-concept facets between the age of 8 and 11 years, Eiser et al. (1995) did not find any mean level differences on any of the SPPC self-concept facets between grade levels seven and eight. Other studies based on alternative self-concept measures have shown that secondary school students tend to present lower levels of academic self-concept, peer-related self-concept, and GSC relative to elementary school students (Cantin & Boivin, 2004; Eccles et al., 1989; Marsh, 1989; Wigfield et al., 1991).

The investigation of mean-level differences in self-concept facets as a function of age or grade level should go along with the examination of gender differences as boys and girls tend to react differently to the multiple biopsychosocial transformations that mark their early adolescence and secondary school years. With respect to academic and physical competence self-concepts, previous research has documented that adolescents seem to present lower levels relative to children and that this difference appears to be similar in magnitude for boys and girls (Cole et al., 2001; Jacobs, Lanza, Osgood, Eccles, & Wigfield, 2002). Similarly, research has demonstrated lower levels of GSC among adolescents relative to children, with some studies showing this difference to be similar for boys and girls (Huang, 2010; Rhodes, Roffman, Reddy, & Fredriksen, 2004; Seidman, Allen, Aber, Mitchell, & Feinman, 1994; Steiger, Allemand, Robins, & Fend, 2014), whereas other suggest that it might be more pronounced for girls (Diseth, Meland, & Breidablik, 2014; Robins & Trzesniewski, 2005; Zimmerman, Copeland, Shope, & Dielman, 1997). Finally, additional studies show that adolescents tend to present slightly lower levels of physical appearance self-concept relative to children, a difference that appears particularly marked among girls relative to boys (e.g., Eisenberg, Neumark-Sztainer, & Paxton, 2006; Frisén, Lunde, & Berg, 2015; Lunde, Frisén, & Hwang, 2007; Morin, Maïano, Marsh, Janosz, & Nagengast, 2011).

The Present Study

We noted that two sources of construct-relevant psychometric multidimensionality are expected to be present in answers provided to the SPPC given that SPPC items are likely to display meaningful associations with multiple (hierarchically superior or conceptually related) self-concept facets. In this study, we contrast alternative representations (ICM-CFA, hierarchical-CFA, bifactor-CFA, ESEM, hierarchical-ESEM, bifactor-ESEM) of responses to the SPPC to disentangle these two sources of construct-relevant psychometric multidimensionality. In addition to these methodological considerations, this study aims to test mean level differences in the SPPC self-concept factors across subgroups of students formed on the basis of gender, grade level, and combinations of gender and grade level. Examination of meaningful gender and grade-level differences in self-concept levels should be conditional on the a priori demonstration of the psychometric

equivalence (or invariance) of the measurement model (configural invariance, weak invariance of the factor loadings, strong invariance of the factor loadings and item intercepts, and strict invariance of the factor loadings, item intercepts, and item uniquenesses; Millsap, 2011).

Method

Sample

The sample includes 2,353 German students ($M_{\text{age}} = 10.73$ years; $SD_{\text{age}} = 1.268$) from the German federal state of Lower Saxony, attending Grade 3 to 6 (Grade 3: $N = 324$; Grade 4: $N = 356$; Grade 5: $N = 793$; Grade 6: $N = 880$). The total sample ($N = 1,158$ boys: 49.21%) as well as the subsamples specific to various grade levels (Grade 3: 159 boys, 165 girls; Grade 4: 170 boys, 186 girls; Grade 5: 389 boys, 404 girls; Grade 6: 440 boys, 440 girls) were equally distributed across gender. In Germany, students of Grades 3 and 4 attend elementary school ($N = 680$) where students of all ability levels are taught together. Commonly starting in Grade 5, students are allocated to different secondary school achievement tracks depending on their accomplishments in elementary school. The subsamples attending Grades 5 and 6 ($N = 1,673$) include students of all achievement tracks: Grade 5: academic track: $N = 365$; intermediate track: $N = 75$; low achievement track: $N = 48$; comprehensive (i.e., mixed-ability) track: $N = 305$; Grade 6: academic track: $N = 296$; intermediate track: $N = 195$; low achievement track: $N = 91$; comprehensive track: $N = 298$. Only students with parental consent were included. Before the students completed the self-concept measures, they were advised of the voluntary and anonymous nature of participation. The instruments were administered during regular school lessons by the first author or a trained research assistant following standardized administration guidelines. To circumvent confounding effects of reading ability, the items were read aloud.

SPPC

The German version of the SPPC (Asendorpf & van Aken, 1993) was used in this study. This instrument includes scales for measuring GSC (example item: "Some kids are very happy being the way they are, but other kids wish they were different"; $\alpha = .828$; see Supplementary Table S1 in the online version of the journal for all items) as well as four domain-specific self-concepts: scholastic competence (example item: "Some kids feel that they are very good at their schoolwork, but other kids worry about whether they can do the schoolwork assigned to them"; $\alpha = .798$), peer acceptance (example item: "Some kids have a lot of friends, but other kids don't have very many friends"; $\alpha = .818$), athletic competence (example item: "Some kids wish they could be a lot better at sports, but other kids feel they are

good enough at sports”; $\alpha = .787$), and physical appearance (example item: “Some kids are happy with the way they look, but other kids are not happy with the way they look”; $\alpha = .831$). The subscale related to children’s self-concept in the behavioral conduct domain was discarded from the German instrument as the authors (i.e., Asendorpf & van Aken, 1993) assumed the items of this scale to be judgmental in nature. Further, low reliability estimates have been reported for this SPPA subscale (Harter, 1988; Trent, Russell, & Cooney, 1994). Each scale consists of six items rated on the structured alternative response format described previously.

Analyses

Analyses were conducted with the robust maximum likelihood (MLR) estimator available in Mplus 7.0 (Muthén & Muthén, 2012), which provides standard errors and fit indices that are robust to the Likert nature of the items and to non-normality. The small amount of missing data ($M = 1.99\%$) was handled with full information maximum likelihood (FIML) estimation (Enders, 2010), as implemented in Mplus in conjunction with the MLR estimator (Muthén & Muthén, 2012).

We contrasted ICM-CFA, hierarchical-CFA (H-CFA), bifactor-CFA (B-CFA), ESEM, hierarchical ESEM (H-ESEM), and bifactor-ESEM (B-ESEM) representations of answers provided to the SPPC. The ICM-CFA model included five first-order factors representing the SPPC subscales (scholastic competence, peer acceptance, athletic competence, physical appearance, and GSC). Each item was associated with the factor it was assumed to measure, and no cross-loadings were allowed. In the H-CFA model, these five factors were specified as indicators of a single higher-order factor. In the B-CFA model, all items were simultaneously used to define one G-factor as well as five S-factors corresponding to the a priori SPPC scales. All factors were specified as orthogonal (Chen et al., 2006; Reise, 2012). The same series of models (i.e., first-order, hierarchical, bifactor) was then estimated using ESEM rather than CFA. These ESEM models were estimated in a confirmatory manner using target rotation (Browne, 2001) by “targeting” all cross-loadings to be as close to zero as possible while freely estimating the main loadings. The first-order ESEM model was estimated using oblique target rotation in which the main loadings were specified as in ICM-CFA. A H-ESEM model was derived from this model using ESEM-within-CFA (Morin et al., 2013). Finally, a B-ESEM model was estimated using orthogonal bifactor target rotation (Reise, 2012; Reise et al., 2011). In this model, all items contributed to define the G-factor as well as the five S-factors. These S-factors were specified using the same pattern of target and non-target factor loadings as in the first-order ESEM model. Figure 1 presents a graphical representation of this model.

It is important to note that the interpretation of the G- and S-factors is slightly different from that of first-order factors. First-order (CFA or ESEM)

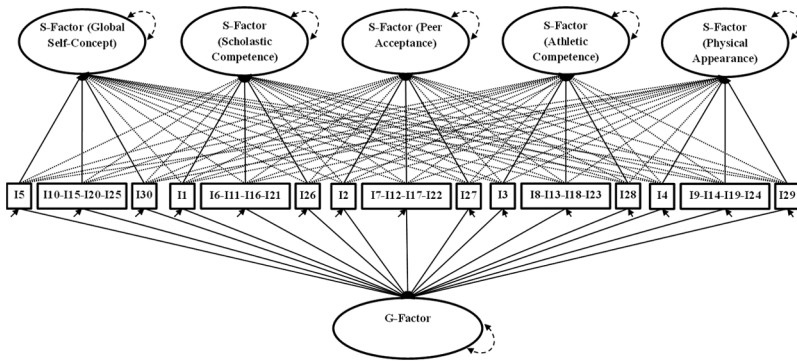


Figure 1. Bifactor exploratory structural equation modeling representation of answers to the Self-Perception Profile for Children (SPPC).

Note. I1 to I30 = SPPC items; oval = latent specific (S-factors) and global (G-factor) factors; rectangles = observed variables (four out of six SPPC items are grouped in a single rectangle for each facet simply for simplicity of illustration; these were treated as separate observed variables in the estimated models); full unidirectional arrows linking ovals and squares = main target factor loadings; dotted unidirectional arrows linking ovals and squares = non-target cross-loadings; full unidirectional arrows linked to the items = item uniquenesses; bidirectional dashed arrows connecting a single oval = factor variances.

factors simply reflect the covariance shared among a subset of items. In bifactor (CFA or ESEM) models, the G-factor reflects the covariance that is shared among all items included in the instrument, whereas the S-factors reflect the covariance shared among a subset of items but partialled out from the covariance that this set of items shares with the other sets of items (which is absorbed by the G-factor). For example, a first-order athletic competence factor will reflect the extent to which respondents feel good about themselves in the athletic area, but these ratings will also be tainted by the extent to which they feel good about themselves in general. A bifactor athletic competence S-factor will provide a purer measure of self-conceptions in the athletic domain net of global self-conceptions that participants have about themselves across domains.

To systematically test the added value of the GSC S-factor, we estimated alternative B-CFA and B-ESEM models excluding this S-factor, in which the GSC items therefore were only allowed to present main loadings on the G-factor. Within the ESEM framework where all items are allowed to be correlated with all factors, this model has identical covariance implications than, and is thus equivalent to, the first-order ESEM model since both models estimate a total of five factors (five first-order factors in the first-order ESEM model vs. one G-factor and four S-factors in this alternative B-ESEM) on which each

item is allowed to load (Hershberger, & Marcoulides, 2013; MacCallum, Wegener, Uchino, & Fabrigar, 1993). This model will always provide a fit to the data that is identical to the first-order ESEM model. Thus, relative to the complete B-ESEM model, a substantial decrease in fit associated with this alternative model would support the need to retain a GSC S-factor.³

For assessing model fit, we consider several commonly accepted goodness-of-fit indices with their respective interpretation guidelines (e.g., Hu & Bentler, 1998; 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. With regard to the RMSEA, values smaller than .08 or .06 support, respectively, acceptable and excellent model fit. For comparing the fit between the alternative representations of the SPPC, we consider the differences between the models. Since all models are nested within one another, they can be compared with the robust chi-square difference test (Satorra & Bentler, 1999). However, 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., 2004). Thus, we rely on the recommendations proposed by Cheung and Rensvold (2002) and Chen (2007) according to which nested models can be seen as equivalent as long as the CFI and TLI do not decrease more than .01 and as long as the RMSEA does not increase by more than .015. However, the cut-off values for model evaluation and model comparison only remain rough guidelines that should be complemented by an examination of parameters estimates, statistical conformity, and theoretical adequacy (Marsh et al., 2004).

Measurement invariance across subgroups formed on the basis of gender and grade level (elementary school boys, secondary school boys, elementary school girls, and secondary school girls) of the model retained as providing the best representation of answers to the SPPC was then investigated (Millsap, 2011). We started with a model of configural invariance, where the same number of factors with the same pattern of associations between items and factors was specified in all groups. In this model, all latent variances are constrained to one, and all factor means are constrained to be zero for identification purposes. Then, a model of weak invariance was estimated by constraining the factor loadings to be invariant across groups, which allowed the factor variances to be freely estimated in all groups save the first one (i.e., elementary school boys). The invariance of factor loadings is seen as the prerequisite for all further invariance models (Millsap, 2011). In a model of strong measurement invariance, both the factor loadings and the item intercepts were constrained to equality across groups. Strong measurement invariance is essential to meaningful comparisons of latent means (Millsap, 2011). Constraining the item intercepts to invariance allowed for the free estimation of the latent means in all groups save the first one.

Then, a model of strict measurement invariance was estimated based on the specification of equal factor loadings, item intercepts, and item uniquenesses across groups. This step is essential to group-based comparisons based on factor scores or scale scores but not as important in studies based on latent variables (Millsap, 2011). However, this step provides important insights into group differences in measurement error (reliability).

Subsequent to these tests of measurement invariance, we turned to substantive group-based comparisons. Building on the preceding models, we tested the invariance of the latent variances and covariances. Substantively, this step tests whether the variability (i.e., variances) of the constructs is similar across subgroups and whether the structural relations among the factors (i.e., covariances) are the same across subgroups. In line with typical bifactor assumptions of orthogonality (e.g., Chen et al., 2006; Reise, 2012), the invariance of the factor covariances cannot be verified in a B-CFA model. However, the orthogonality of a B-ESEM solution is a function of the specific rotation procedure that is retained. Given that invariance constraints are imposed on the unrotated solution (before orthogonality is imposed), invariance of the latent covariances can still be tested with B-ESEM. Although the added value of this test is debatable given the orthogonality of the final rotated solution that is retained for interpretation in all groups. However, it is part of the specificities of ESEM/B-ESEM that invariance constraints need to be simultaneously imposed on the factor variances and covariances, making it essential to constrain the unrotated covariances to equality to be able to test for the invariance of the latent variances. We finally conducted models of complete invariance with factor means additionally restricted to be equal across groups. The models forming this invariance sequence are nested and can be compared with one another. Similar to the approach of model comparison described previously, changes in goodness-of-fit indices between a model and the previous one can be seen to support measurement invariance when $\Delta\text{CFI}/\Delta\text{TLI} \leq .01$ and $\Delta\text{RMSEA} \leq .015$ (Chen, 2007; Cheung & Rensvold, 2002).

Results

Alternative Measurement Models

Table 1 displays the goodness-of-fit indices of the models estimated in this study. The first-order CFA (Model 1, also see Supplementary Table S1 in the online version of the journal) provided a higher degree of fit to the data (CFI = .926, TLI = .918, RMSEA = .039) than the corresponding H-CFA (Model 2: CFI = .917, TLI = .910, RMSEA = .041). In contrast, the fit of the B-CFA including (Model 3: CFI = .931, TLI = .920, RMSEA = .038) or excluding (Model 4: CFI = .927, TLI = .917, RMSEA = .039) a separate S-factor associated with the GSC items remained comparable to one another and mainly in line with the fit of the first-order CFA. However, all CFA-based models

Table 1
Goodness-of-Fit Indices

Model	χ^2	df	CFI	TLI	RMSEA	90% CI	Model Description			
CFA										
Model 1	1,800.437	395	.926	.918	.039	[.037, .041]	First-order CFA			
Model 2	1,967.574	400	.917	.910	.041	[.039, .043]	Higher-order CFA			
Model 3	1,669.737	375	.931	.920	.038	[.036, .040]	Bifactor-CFA (including a GSC S-factor)			
Model 4	1,758.148	381	.927	.917	.039	[.037, .041]	Bifactor-CFA (without a GSC S-factor)			
ESEM										
Model 5	1,169.909	295	.954	.932	.036	[.033, .038]	First-order ESEM			
Model 6	1,178.464	300	.953	.932	.035	[.033, .037]	Higher-order ESEM			
Model 7	717.301	270	.976	.962	.027	[.024, .029]	Bifactor-ESEM (including a GSC S-factor)			
Model 8	1,169.909	295	.954	.932	.036	[.033, .038]	Bifactor-ESEM (without a GSC S-factor)			
Model	χ^2	df	CFI	TLI	RMSEA	90% CI	Δ CFI	Δ TLI	Δ RMSEA	Model Description
Invariance across gender and grade levels										
Model 9	1,793.138	1,080	.963	.941	.034	[.031, .036]	.000	+.017	-.006	Configural invariance
Model 10	2,226.297	1,512	.963	.958	.028	[.026, .031]	.000	-.006	+.002	Weak invariance (loadings)
Model 11	2,433.833	1,584	.956	.952	.030	[.028, .033]	-.007	-.006	+.002	Strong invariance (loadings, intercepts)
Model 12	2,684.018	1,674	.948	.946	.032	[.030, .034]	-.008	-.006	+.002	Strict invariance (loadings, intercepts, uniq.)
Model 13	2,861.356	1,737	.942	.942	.033	[.031, .035]	-.006	-.004	+.001	Full invariance (loadings, intercepts, uniq., v/c)
Model 14	3,132.721	1,755	.929	.930	.037	[.034, .039]	-.013	-.012	+.005	Latent mean invariance (loadings, intercepts, uniq., v/c, means)

Note. ESEM models were conducted with target oblique rotation. All models were estimated with the MLR estimator. For all χ^2 values, $p < .01$. CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = root mean square error of approximation; CI = RMSEA 90% confidence interval; Δ = change in fit indices relative to the previous model in the invariance testing sequence; ESEM = exploratory structural equation models; CFA = confirmatory factor analysis; uniq. = uniquenesses; v/c = latent variances and covariances; GSC = global self-concept.

presented a lower level of fit to the data than the corresponding ESEM models. The first-order ESEM (Model 5) provided a better fit to the data (CFI = .954, TLI = .932, RMSEA = .036) than the first-order CFA. The results from this model (Supplementary Table S2 in the online version of the journal) demonstrated the presence of construct-relevant psychometric multidimensionality related to the items' associations with multiple substantively related self-concept constructs assessed within the SPPC. For example, items pertaining to the GSC subscale showed substantial cross-loadings on the physical appearance factor ($\lambda = .102$ to $.279$; $M = .174$), while items designed to measure physical appearance displayed high cross-loadings on the GSC factor ($\lambda = .025$ to $.368$; $M = .167$). Although less pronounced, cross-loadings were also present between the other SPPC subscales ($|\lambda| = .002$ to $.203$; $M = .062$) and remained of a magnitude sufficient to induce biases if constrained to be zero (e.g., Asparouhov et al., 2015; Marsh et al., 2013). The adequacy of considering cross-loadings is further corroborated by the observation of lower factor correlations (Supplementary Table S3 in the online version of the journal) resulting from the first-order ESEM (for correlations involving the GSC factor: $r = .328$ to $.564$; $M = .413$; for other correlations: $r = .313$ to $.451$; $M = .378$) compared to the first-order CFA (for correlations involving the GSC factor: $r = .438$ to $.837$; $M = .592$; for other correlations: $r = .356$ to $.507$; $M = .436$). These results support the superiority of an ESEM, relative to CFA, representation of SPPC responses and confirm the presence of construct-relevant psychometric multidimensionality due to the measure of conceptually related constructs.

The most pronounced cross-loadings involved the GSC factor, which also presented the lowest factor loadings ($\lambda = .309$ to $.594$; $M = .433$) and the highest correlations with the other factors. This finding suggests the presence of construct-relevant psychometric multidimensionality due to the hierarchical nature of the self-concept construct assessed by the SPPC. We thus contrasted the first-order ESEM (Model 5) with the H-ESEM (Model 6) and B-ESEM (Models 7 and 8). In contrast to the H-ESEM (CFI = .953, TLI = .932, RMSEA = .035) and the B-ESEM without a GSC S-factor (CFI = .954, TLI = .932, RMSEA = .036), which fit the data as well as the first-order ESEM, the B-ESEM including a GSC S-factor (CFI = .976, TLI = .962, RMSEA = .027) demonstrated the best fit of all models and showed a substantial improvement in fit compared to the alternative ESEM models ($\Delta\text{CFI} > .020$; $\Delta\text{TLI} > .030$). This model (Model 7) was thus retained as the final model and serves as the basis for the tests of invariance across gender and grade level reported below.

The results of this model (Model 7) are reported in Table 2. In this model, the G-factor is well defined by substantial loadings ($\lambda = .200$ to $.716$; $M = .464$). The integration of a G-factor had few effects on the definition of the S-factors representing the domain-specific SPPC scales (scholastic competence, peer acceptance, athletic competence, and physical appearance), which all remained well defined through high main factor loadings

Table 2
Standardized Factor Loadings of the Bifactor-ESEM Model Including the Global Self-Concept S-Factor (Model 7)

Item	G-Factor	GSC	Scholastic Competence	Peer Acceptance	Physical Appearance	Athletic Competence	Residual
5	.607**	.253**	.052	.070**	.062*	-.028	.555**
10	.477**	.392**	.079**	.082**	.036	-.017	.604**
15	.716**	.209**	-.018	-.079**	.004	-.064*	.433**
20	.605**	.227**	.001	-.014	.065*	-.057*	.575**
25	.712**	.219**	-.063**	-.067**	.105**	-.057**	.423**
30	.510**	.375**	.131**	.003	.003	-.007	.582**
1	.371**	-.004	.414**	-.067**	-.042	.038	.683**
6	.450**	-.032	.471**	-.056**	-.102**	-.023	.561**
11	.282**	.012	.545**	.116**	-.010	.057**	.607**
16	.325**	.086**	.505**	.060**	.042	.047*	.625**
21	.401**	-.063*	.568**	-.032	-.092**	-.050*	.501**
26	.341**	.131**	.602**	.088**	.060*	.041*	.491**
2	.364**	-.009	.018	.527**	.016	.073**	.584**
7	.451**	-.095**	-.018	.586**	-.150**	.063**	.417**
12	.364**	.099**	.064**	.608**	.100**	.061**	.471**
17	.363**	.021	-.036	.375**	-.161**	.095**	.691**
22	.401**	.089**	.055**	.525**	.065*	.022	.548**
27	.507**	-.103**	.042*	.467**	-.093**	.045**	.502**
3	.407**	-.133**	-.038*	.072**	-.060**	.662**	.369**
8	.342**	.078**	.007	.131**	.144**	.546**	.541**
13	.443**	-.070**	.014	.038	-.094**	.588**	.442**
18	.382**	-.139**	-.011	.012	-.066**	.565**	.510**
23	.200**	.186**	.101**	.182**	.142**	.266**	.791**
28	.335**	.139**	.111**	.128**	.038	.437**	.647**
4	.688**	-.079**	-.115**	-.038	.182**	-.050*	.470**
9	.477**	-.087	.000	-.071**	.286**	.067**	.674**
14	.551**	.040	-.004	-.063**	.479**	.034	.459**
19	.634**	.056*	-.056**	-.012	.568**	-.022	.269**
24	.513**	.189**	.005	-.009	.353**	-.031	.576**
29	.703**	-.066*	-.091**	-.017	.138**	-.012	.473**

Note. ESEM = exploratory structural equation model; S-factor = specific factor; G-factor = global factor; GSC = global self-concept. Bold items represent target loadings.
 * $p < .05$. ** $p < .01$.

($\lambda = .138$ to $.662$; $M = .469$) and reasonably low cross-loadings ($\lambda = -.161$ to $.189$; $M = .010$). This observation confirms that these domain-specific S-factors all assess relevant specificity over and above the G-factor. Furthermore, the inclusion of the G-factor resulted in a substantial decrease in the magnitude of the most elevated cross-loadings identified in the

first-order ESEM (e.g., in the first-order ESEM, 6 of the cross-loadings were higher than $\lambda = .200$, and 28 were between $\lambda = .100$ and $.200$ while comparable figures are 0 and 20 in B-ESEM), confirming that at least a part of them include construct-relevant psychometric multidimensionality due to the hierarchical nature of the SPPC. Furthermore, in the retained B-ESEM solution (Model 7), the GSC S-factor became defined with less precision (main loadings: $\lambda = .209$ to $.392$; $M = .279$) than in the first-order ESEM (main loadings: $\lambda = .309$ to $.594$; $M = .433$), indicating that a substantial part of the GSC items' variability is related to the G-factor. However, the loadings of the GSC items on the GSC S-factor remain significant and of a non-negligible magnitude, suggesting that at least some relevant variance remains specific to the GSC items and not explained by the overarching G-factor, explaining the superior fit of this model relative to the fit of alternative ESEM models. For comparison purposes, the results from the alternative B-ESEM excluding the GSC S-factor (Model 8) are reported in the Supplementary Table S4 in the online version of the journal. These additional results confirm that the inclusion of the GSC S-factor only had a minimal impact on the definition of the G-factor and of the remaining S-factors.

As a further test of the appropriateness of the retained B-ESEM solution (Model 7), we calculated omega (ω) estimates of composite reliability (McDonald, 1970; Reise, 2012) for the G- and S-factors based on standardized model parameters. The reliability of the GSC S-factor remained marginal ($\omega = .47$), consistent with the idea that a substantial part of the information included in the GSC items contributed to the definition of the G-factor but also with the idea that meaningful information remained in these items—possibly reflecting the self-concept domains not covered in the SPPC. Still, we note that although the reliability of this S-factor remains low, in the context of latent variable models corrected for measurement errors, scores on the latent factors can be considered to be perfectly reliable. Importantly, the composite reliability of the remaining S-factors proved to be adequate (scholastic competence: $\omega = .74$; peer acceptance: $\omega = .75$; physical appearance: $\omega = .58$; athletic competence: $\omega = .74$) and that of the G-factor to be fully satisfactory ($\omega = .92$).

It should be noted that SPPC Item 23 (“In games and sport some kids usually watch instead of play but other kids usually play rather than watch”), which in theory should be associated with the athletic competence subscale, systematically showed low factor loadings on its target factor (CFA: $\lambda = .352$; ESEM: $\lambda = .198$; B-ESEM Model 7: $\lambda = .266$; B-ESEM Model 8: $\lambda = .206$), equivalent cross-loadings on non-target factors in ESEM ($\lambda = -.156$ to $.172$) and B-ESEM (Model 7: $\lambda = .101$ to $.186$; Model 8: $\lambda = .109$ to $.172$), and a similarly low factor loading on the G-factor in B-ESEM (Model 7: $\lambda = .200$; Model 8: $\lambda = .234$). This item should be targeted for reexamination in future studies.

Measurement Invariance and Latent Mean Differences Across Grade Level and Gender

The measurement invariance of the final B-ESEM Model 7 was tested across groups formed on the basis of students' gender and grade level (elementary school boys, elementary school girls, secondary school boys, and secondary school girls) to investigate potential interactive effects of grade level and gender (Models 9 to 14). Tests of measurement invariance across a single grouping variable are reported in Supplementary Table S9 in the online version of the journal.

We started with a model of configural invariance (Model 9) that assumes an invariant factor pattern (i.e., the same number of factors and the same pattern of main loadings and cross-loadings specified for the target rotation procedure) across subgroups. This model presented a satisfactory level of fit to the data (CFI = .963, TLI = .941, RMSEA = .034). We then successively constrained the factor loadings (Model 10), item intercepts (Model 11), and item uniquenesses (Model 12) to be equal across subgroups. These models never showed a decrement in fit exceeding the recommended guidelines, thus supporting the strict measurement invariance of the SPPC across gender and grade level. The results also supported the full invariance of the latent variances and covariances (Model 13). Finally, the results from Model 14 clearly demonstrated latent mean differences ($\Delta\text{CFI} = -.013$; $\Delta\text{TLI} = -.012$). To explore these differences, we relied on Model 13, where all factor variances were fixed to be 1, allowing for mean differences to be expressed in *SD* units.

When latent mean differences are examined using elementary school boys as the referent group: (a) Secondary school boys had significantly lower mean levels on the G-factor ($-.226, p < .01$) and on the S-factor for scholastic competence ($-.211, p < .05$) but higher levels on the S-factor for peer acceptance ($.265, p < .01$), (b) elementary school girls had lower mean levels on the G-factor ($-.284, p < .01$) and the S-factor for athletic competence ($-.359, p < .01$) but higher levels on the S-factors relating to physical appearance ($.224, p < .05$) and GSC ($.299, p < .05$), and (c) secondary school girls displayed lower mean levels on the G-factor ($-.643, p < .01$) and on the S-factor for athletic competence ($-.421, p < .01$) while showing higher mean levels for the S-factors addressing peer acceptance ($.662, p < .01$) and GSC ($.380, p < .01$). When latent mean differences are examined using elementary school girls as the referent group: (a) Secondary school girls had lower mean levels on the G-factor ($-.359, p < .01$) as well as lower mean levels on the S-factors for scholastic competence ($-.215, p < .01$) and physical appearance ($-.363, p < .01$) but higher mean levels on the S-factor for peer acceptance ($.507, p < .01$), and (b) secondary school boys displayed higher mean levels on the S-factor for athletic competence ($.325, p < .01$) but lower mean levels on the S-factors for scholastic competence ($-.281, p < .01$), physical appearance ($-.244, p < .01$), and GSC ($-.305, p < .01$). Finally, when latent means differences

are examined using secondary school boys as the referent group, secondary school girls had lower mean levels on the G-factor ($-.418, p < .01$) and the athletic competence S-factor ($-.387, p < .01$) while displaying higher mean levels on the peer acceptance S-factor ($.397, p < .01$) and GSC ($.385, p < .01$).

Discussion

We investigated two sources of construct-relevant psychometric multidimensionality in answers provided to the German SPPC, related to the assessment of conceptually related and hierarchically ordered constructs. To do so, we relied on a psychometric framework involving the comparison of first-order, hierarchical, and bifactor CFA and ESEM models. The results supported the superiority of a bifactor-ESEM representation of answers provided to the SPPC. This finding suggests that both sources of construct-relevant psychometric multidimensionality are present in answers to the SPPC.

First, the superiority of ESEM over ICM-CFA suggests that cross-loadings are needed to reflect the fact that the SPPC assesses conceptually related constructs. The observation of multiple cross-loadings remaining within an acceptably low level (i.e., most of them below $.200$) corresponds to the findings from previous EFA investigations of the SPPC (Chan, 2002; Miller, 2000; Stewart et al., 2010). The strongest cross-loadings emerged between GSC and physical appearance, in line with previous research (Frost & McKelvie, 2004; Harter, 1999; Morin & Maïano, 2011). The observation of reduced factor correlations when using ESEM relative to CFA also supports the superiority of ESEM. These findings attest to the presence of construct-relevant multidimensionality related to the fact that SPPC items tend to present associations with more than one conceptually related construct.

Second, the superiority of a bifactor-ESEM over a first-order ESEM representation supports the presence of construct-relevant multidimensionality due to the hierarchical nature of multidimensional self-conceptions. This conclusion was supported by a variety of results. First, the finding of multiple, relatively small, but significant cross-loadings between items and multiple factors in the first-order ESEM solution supported the idea that a bifactor representation could be useful by suggesting that all items might reflect a global overarching construct. Second, the observation that the most pronounced cross-loadings involved the GSC items also supports this interpretation. Third, cross-loadings were generally reduced in B-ESEM, suggesting that the inclusion of a global factor was efficient at reducing ESEM cross-loadings. Finally, the bifactor-ESEM solution demonstrated that items from all specific self-concept scales (including the specific GSC scale) contributed to the definition of both a G-factor as well as domain-specific S-factors, suggesting that the inclusion of a global factor did not detract from the proper measurement of the specific constructs.

Prior CFA research on the factor structure of the SPPC and other SPP-based instruments has led to some surprising results (e.g., high factor

correlations; Boivin et al., 1992; Cole et al., 2001; Maïano et al., 2008; Stewart et al., 2010) that have often been attributed to its structured alternative response format (Eiser et al., 1995; Marsh et al., 1994; Wichstrøm, 1995; Wylie, 1989). The present study shows that this might not be the case. Rather, our results add to accumulating research evidence showing the need to rely on measurement models providing a proper representation of the construct-relevant psychometric multidimensionality present in self-concept measures and related to the assessment of conceptually related (Arens & Morin, 2016; Morin & Maïano, 2011) and hierarchically ordered (Brunner et al., 2008, 2009, 2010; Morin, Arens, et al., 2016) constructs. Still, it must be kept in mind that irrespective of the fact that the current results support the psychometric properties of the original SPPC response scale, it does not offset criticisms in regards to the fact that this structured alternative response scale remains more time consuming and complex than alternatives, which may affect compliance and missing responses (e.g., Eiser et al., 1995; Marsh & McDonald Holmes, 1990).

As noted in the introduction, including a S-factor based on the GSC items directly tests whether these items retain any meaningful specificity over and above the content already covered in the other SPPC items. The findings of the present study support the idea that GSC items present some form of specificity over and above the shared content of all SPPC items reflected in the G-factor. First, our main conclusions remained unchanged when the GSC S-factor was taken out of the model. Yet, the superior fit of the model including the GSC S-factor and the substantial level of specificity associated with the GSC items over and above their association with the G-factor support the need to maintain the GSC S-factor in the model. Second, gender-related latent mean level differences show an opposite pattern of differences on the G-factor (with boys presenting higher levels than girls) versus the GSC S-factor (with girls presenting higher levels than boys). Although this finding seems to lend credence to the idea that GSC may reflect a distinct domain of self-conceptions, alternative interpretations appear more plausible. For instance, results from previous studies typically show that boys tend to present higher levels of global self-concepts than girls (e.g., Kling et al., 1999; Marsh, 1989; Wigfield et al., 1991), suggesting that the G-factor (on which boys present higher levels), rather than the GSC S-factor (on which girls present a higher level), provides a more exact representation of global self-conceptions. The GSC S-factor may be more properly considered as a reflection of the multiple self-concept areas (e.g., math and verbal, parental, spiritual, or artistic) that are excluded from the SPPC (Byrne & Shavelson, 1996; Marsh, 2007; Shavelson et al., 1976; Vispoel, 1995; for a similar interpretation in relation to the SDQ I, see Morin, Arens, et al., 2016). Indeed, girls are known to present higher levels on some of these excluded self-concept facets relative to boys (e.g., Eccles et al., 1989; Marsh, 1989; Wigfield et al., 1991). Thus, the SPPC does not seem to

cover the major facets of students' self-concept exhaustively but rather leaves out important facets. With respect to the academic component of self-concept, the SPPC only addresses a global academic self-concept (scholastic competence), neglecting the consistently demonstrated separation of math and verbal self-concepts (Möller, Pohlmann, Köller, & Marsh, 2009). With respect to the nonacademic self-concept, it does not consider social self-concepts in relation to parents and fails to take emotional self-concept facets into account (Byrne & Shavelson, 1996; Shavelson et al., 1976). Hence, an interesting perspective for future research would be to replicate the current finding using the SPPA (Harter, 1988) and the ASPP (Messer & Harter, 1986) as these instruments cover a broader range of self-concept facets than the SPPC.

In addition to aiming to replicate the present results regarding the internal factor structure of SPP-based instruments, it would also be important for future research to examine potentially differential relations to outcome criteria in order to get further insight into the meaning of the GSC S-factor relative to the G-factor (Byrne, 1996). In this regard, it is important to keep in mind that bifactor models are essentially designed to represent theoretically meaningful G- and S-factors estimated from the same set of items whenever there are reasons to expect construct-relevant multidimensionality due to the presence of hierarchically ordered constructs. As such, it is part of the inherent theoretical logic of bifactor models that all factors need to be incorporated in predictive models. In contrast, alternative models are available whenever there is a need to control for theoretically meaningless, or construct-irrelevant, sources of multidimensionality in a measure, such as models incorporating correlated method factors (Eid, 2000) or models incorporating a global factor aiming to control for shared responses tendencies in the estimation of meaningful correlated factors (Podsakoff, MacKenzie, & Podsakoff, 2003).

Our results supported the strict measurement invariance of the bifactor-ESEM model as well as the invariance of the latent variances and covariances across subgroups formed on the basis of gender and grade level, supporting the applicability of the SPPC to male and female elementary and secondary school German students. With respect to latent mean differences, our results replicated previous studies in demonstrating higher levels of athletic competence and lower levels of peer acceptance among boys relative to girls (Kling et al., 1999; Marsh, 1989; Wigfield et al., 1991). Our findings further revealed lower levels of GSC, scholastic competence, and physical appearance and a higher level of peer acceptance in secondary school students relative to elementary school students (Cantin & Boivin, 2004; Eccles et al., 1989; Wigfield et al., 1991).

Regarding the G-factor, our results showed both a main effect of grade level (with lower levels in secondary school students) and a main effect of gender (with girls presenting lower levels). The finding of a lower level of scholastic competence self-concept in secondary school, relative to elementary school, of a similar magnitude for boys and girls replicated previous findings (e.g., Cole et al., 2001; Jacobs et al., 2002), suggesting that

differences in learning environments between elementary and secondary schools may account for this difference (e.g., Eccles et al., 1993). Both boys and girls displayed higher levels of peer acceptance self-concept in secondary relative to elementary school, countering previous findings of lower levels of peer-related self-concepts in secondary school students (Cantin & Boivin, 2004; Wigfield et al., 1991). Additionally, our results also showed an interactive effect of gender and grade in relation to physical appearance self-concept. Although girls' level of physical appearance self-concept was significantly higher than boys' in elementary schools, their level was equivalent to boys' in secondary schools. This finding replicates results from previous research (e.g., Eisenberg et al., 2006; Frisé et al., 2015; Lunde et al., 2007; Marsh, 1989; Morin et al., 2011) and might be explained by girls' increasing sensitivity to physical appearance when they get older.

Conclusion and Directions for Future Research

Prior research on the SPPC and other SPP-based measures of students' multidimensional self-concept has raised a series of important questions regarding the psychometric properties of these instruments, revealing elevated cross-loadings and high factor correlations, as well as regarding the most appropriate way to represent the GSC subscale. The present study provided new insights into the SPPC factor structure, supporting the need to rely on a bifactor-ESEM representation. This representation allowed for a proper representation of the various sources of construct-relevant psychometric multidimensionality present in the SPPC due to the assessment of conceptually related and hierarchically ordered constructs. This study also shows mean level differences in self-concept facets between elementary and secondary school boys and girls. These insights should be extended in future longitudinal studies that would provide a more adequate approach to study changes in self-concept occurring as a function of grade level. Given that this study only considers the German version of the SPPC administered to German students attending Grades 3 to 6, it would be interesting for future research to systematically explore whether similar findings emerge with populations from other educational systems and cultures. The psychometric framework presented here has a wide applicability for the investigation of the internal structure of many multidimensional instruments and could thus advantageously be incorporated in future psychometric research.

Notes

The two authors (A.K.A. and A.J.S.M.) contributed equally to this article, and their order was determined at random: Both should thus be considered first authors. This paper was made possible in part by grants from the Australian Research Council awarded to the second author (DP130102713; DP140101559).

¹A Schmid-Leiman (Schmid & Leiman, 1957) transformation procedure (SLP) can be applied to convert a hierarchical factor model into a bifactor approximation (see Gignac, 2007; Reise, 2012). Although this suggests similarities between higher-order and bifactor models, the SLP shows that hierarchical factor models implicitly rely on more stringent assumptions (Chen, West, & Sousa, 2006; Reise, 2012). When a SLP is applied to a hierarchical model, the relation between an item and the global factor (G-factor) from the resulting bifactor approximation is represented as the indirect effect of the higher-order factor on the item, as “mediated” by the first-order factor. More precisely, each item’s first-order factor loading is multiplied by the loading of this first-order factor on the second-order factor, which yields the loadings of this item on the SLP-estimated G-factor. The second term in this multiplication is thus a constant as far as the items associated with a single first-order factor are concerned. Relations between the items and the SLP-estimated specific factors (S-factors) are similarly reflected by the product of their loadings on their first-order factor by the squared root of the disturbance of this first-order factor (corresponding to the regression path associated with the unique part, or disturbance, of the first-order factor). This second term is also a constant and reflects the unique part of the first-order factor that remains unexplained by the higher-order factor. In other words, each item’s association with the SLP G-factor and S-factors are obtained by multiplying their first-order loadings by constants, resulting in a ratio of G-factor loadings to S-factors loadings that will be exactly the same for all items associated with the same first-order dimension. These constraints are one of the reasons why true bifactor models tend to provide a much better fit to the data than hierarchical models (Chen et al., 2006; Reise, 2012). Jennrich and Bentler (2011) showed that while bifactor models were able to properly recover true higher-order factor structures, higher-order factor models could not always properly recover true bifactor structures. Bifactor models should thus be preferred over higher-order models unless strong theoretical reasons are present to support the need to model the relations between the indicators and the global factors as indirect and the presence of the implicit proportionality constraints (for a more extensive discussion of these issues, see Gignac, 2016).

²Construct-irrelevant psychometric multidimensionality may also be present (informants, wording, etc.) and controlled with correlated uniquenesses/method factors (Marsh, Scalas, & Nagengast, 2010).

³We also estimated models excluding the global self-concept (GSC) items. Results from these models are reported in Supplementary Tables S5 to S8 in the online version of the journal. These results remain in line with the results reported here.

References

- Arens, A. K., & Morin, A. J. S. (2016). Examination of the structure and grade-related differentiation of multidimensional self-concept instruments for children using ESEM. *The Journal of Experimental Education*, *84*, 330–355.
- 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. O. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, *16*, 397–438.
- Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances. *Journal of Management*, *41*, 1561–1577.
- Baumeister, R. F., Campbell, J. D., Krueger, J. I., & Vohs, K. D. (2003). Does high self-esteem cause better performance, interpersonal success, happiness, or healthier lifestyles? *Psychological Science in the Public Interest*, *4*, 1–44.
- 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.
- Brunner, M., Keller, U., Dierendonck, C., Reichert, M., Ugen, S., Fischbach, A., & Martin, R. (2010). The structure of academic self-concepts revisited: The nested Marsh/Shavelson model. *Journal of Educational Psychology*, 102, 964–981.
- Brunner, M., Keller, U., Hornung, C., Reichert, M., & Martin, R. (2009). The cross-cultural generalizability of a new structural model of academic self-concepts. *Learning and Individual Differences*, 19, 387–403.
- Brunner, M., Lüdtke, O., & Trautwein, U. (2008). The internal/external frame of reference model revisited: Incorporating general cognitive ability and general academic self-concept. *Multivariate Behavioral Research*, 43, 137–172.
- Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research*, 36, 111–150.
- Byrne, B. M. (1996). *Measuring self-concept across the life span: Issues and instrumentation*. Washington, DC: American Psychological Association.
- Byrne, B. M., & Shavelson, R. J. (1996). On the structure of social self-concept for pre-, early, and late adolescents: A test of the Shavelson, Hubner, and Stanton (1976) model. *Journal of Personality and Social Psychology*, 70, 599–613.
- Cantin, S., & Boivin, M. (2004). Change and stability in children's social network and self-perceptions during transition from elementary to junior high school. *International Journal of Behavioral Development*, 28, 561–570.
- 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.
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of Personality*, 80, 219–251.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, 41, 189–255.
- 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., . . . Maschman, T. (2001). Age and gender effects in multiple domains of child and adolescent self-concept: A cohort sequential longitudinal design. *Child Development*, 72, 1723–1746.
- Diseth, Å., Meland, E., & Breidablik, H. J. (2014). Self-beliefs among students: Grade level and gender differences in self-esteem, self-efficacy and implicit theories of intelligence. *Learning and Individual Differences*, 35, 1–8.
- Eccles, J. S., Midgley, C., Wigfield, A., Buchanan, C. M., Reuman, D., Flanagan, C., & MacIver, D. (1993). Development during adolescence: The impact of stage-environment fit on young adolescents' experiences in schools and in families. *American Psychologist*, 48, 90–101.
- Eccles, J. S., Wigfield, A., Flanagan, C. A., Miller, C., Reuman, D. A., & Yee, D. (1989). Self-concepts, domain values, and self-esteem: Relations and changes at early adolescence. *Journal of Personality*, 57, 283–310.
- Eid, M. (2000). Multitrait-multimethod model with minimal assumptions. *Psychometrika*, 65, 241–261.
- Eisenberg, M. E., Neumark-Sztainer, D., & Paxton, S. J. (2006). Five-year change in body satisfaction among adolescents. *Journal of Psychosomatic Research*, 61, 521–527.

- Eiser, C., Eiser, J. R., & Havermans, T. (1995). The measurement of self-esteem: Practical and theoretical considerations. *Personality and Individual Differences, 18*, 429–432.
- Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford.
- Fox, K. R., & Corbin, C. B. (1989). The physical self-perception profile: Development and preliminary validation. *Journal of Sport & Exercise Psychology, 11*, 408–430.
- Frisén, A., Lunde, C., & Berg, A. I. (2015). Developmental patterns in body esteem from late childhood to young adulthood: A growth curve analysis. *European Journal of Developmental Psychology, 12*, 99–115.
- Frost, J., & McKelvie, S. (2004). Self-esteem and body satisfaction in male and female elementary school, high school, and university students. *Sex Roles, 51*, 45–54.
- Gignac, G. E. (2007). Multi-factor modeling in individual differences research: Some recommendations and suggestions. *Personality and Individual Differences, 42*, 37–48.
- Gignac, G. E. (2016). The higher-order model imposes a proportionality constraint: That is why the bifactor model tends to fit better. *Intelligence, 55*, 57–68.
- Gignac, G. E., Palmer, B., & Stough, C. (2007). A confirmatory factor analytic investigation of the TAS-20: Corroboration of a five-factor model and suggestions for improvement. *Journal of Personality Assessment, 89*, 247–257.
- Gignac, G. E., & Watkins, M. W. (2013). Bifactor modeling and the estimation of model-based reliability in the WAIS-IV. *Multivariate Behavioral Research, 48*, 639–662.
- Graziano, W. G., Jensen-Campbell, L. A., & Finch, J. F. (1997). The self as a mediator between personality and adjustment. *Journal of Personality and Social Psychology, 73*, 392–404.
- Graziano, W. G., & Ward, D. (1992). Probing the Big Five in adolescence: Personality and adjustment during a developmental transition. *Journal of Personality, 60*, 427–439.
- Harter, S. (1985). *Manual for the Self-Perception Profile for Children: Revision of the Perceived Competence Scale for Children*. Denver, CO: University of Denver.
- Harter, S. (1988). *Manual for the Self-Perception Profile for Adolescents*. Denver, CO: University of Denver.
- Harter, S. (1999). *The construction of the self: A developmental perspective*. New York, NY: Guilford.
- Hershberger, S. L., & Marcoulides, G. A. (2013). The problem of equivalent structural models. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 13–42). Charlotte, NC: Information Age Publishing, Inc.
- Hu, L. T., & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*, 424–453.
- Huang, C. (2010). Mean-level change in self-esteem from childhood through adulthood: Meta-analysis of longitudinal studies. *Review of General Psychology, 14*, 251–260.
- Jacobs, J. E., Lanza, S., Osgood, D. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child Development, 73*, 509–527.
- Jennrich, R. I., & Bentler, P. M. (2011). Exploratory bi-factor analysis. *Psychometrika, 76*, 537–549.
- Kling, K. C., Hyde, J. S., Showers, C. J., & Buswell, B. N. (1999). Gender differences in self-esteem: A meta-analysis. *Psychological Bulletin, 125*, 470–500.
- Lunde, C., Frisén, A., & Hwang, C. P. (2007). Ten-year-old girls and boys body composition and peer victimization experiences: Prospective relations with body satisfaction. *Body Image, 4*, 11–28.

- MacCallum, R. C., Wegener, D. T., Uchino, B. N., & Fabrigar, L. R. (1993). The problem of equivalent models in applications of covariance structure analysis. *Psychological Bulletin*, *114*, 185–199.
- 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. (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. (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., & Hattie, J. (1996). Theoretical perspectives on the structure of self-concept. In B. A. Bracken (Ed.), *Handbook of self-concept* (pp. 38–90). New York, NY: Wiley.
- 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., 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., & MacDonald Holmes, I. W. (1990). Multidimensional self-concepts: Construct validation of responses by children. *American Educational Research Journal*, *27*, 89–117.
- 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., Richards, G., Johnson, S., Roche, S., & Tremayne, P. (1994). Physical Self-Description Questionnaire: Psychometric properties and multi-trait-multimethod analysis. *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.
- McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*, 1–21.
- McDonald, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.

- 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, NY: Routledge.
- 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., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*, *23*, 116–139.
- 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., Maïano, C., Marsh, H. W., Janosz, M., & Nagengast, B. (2011). The longitudinal interplay of adolescents' self-esteem and body image: A conditional autoregressive latent trajectory analysis. *Multivariate Behavioral Research*, *46*, 157–201.
- Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2013). Exploratory structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural equation modeling: A second course* (2nd ed., pp. 395–436). Charlotte, NC: Information Age Publishing, Inc.
- Morin, A. J. S., Tran, A., & Caci, H. (2016). Factorial validity of the ADHD Adult Symptom Rating Scale in a French community sample: Results from the ChiP-ARD study. *Journal of Attention Disorders*, *20*, 530–541.
- Muldoon, O. T. (2000). Social group membership and self-perceptions in Northern Irish children: A longitudinal study. *British Journal of Developmental Psychology*, *18*, 65–80.
- Muris, P., Meesters, C., & Fijen, P. (2003). The Self-Perception Profile for Children: Further evidence for its factor structure, reliability, validity. *Personality & Individual Differences*, *35*, 1791–1802.
- Murray, A., & Johnson, W. (2013). The limitations of model fit in comparing the bifactor versus higher-order models of cognitive ability structure. *Intelligence*, *41*, 407–422.
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- Organisation for Economic Co-operation and Development (OECD) (Ed.). (2001). *Knowledge and skills for life: First results from the OECD Programme for International Student Assessment (PISA) 2000*. Paris: OECD.
- Podsakoff, P., MacKenzie, S., Lee, J., & Podsakoff, N. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, *88*, 879–903.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*, 667–696.
- Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*, 544–559.
- Reise, S. P., Moore, T. M., & Maydeu-Olivares, A. (2011). Targeted bifactor rotations and assessing the impact of model violations on the parameters of unidimensional and bifactor models. *Educational and Psychological Measurement*, *71*, 684–711.

- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31.
- Rhodes, J., Roffman, J., Reddy, R., & Fredriksen, K. (2004). Changes in self-esteem during the middle school years: A latent growth curve study of individual and contextual influences. *Journal of School Psychology, 42*, 243–261.
- Robins, R. W., & Trzesniewski, K. H. (2005). Self-esteem development across the lifespan. *Current Directions in Psychological Science, 14*, 158–162.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Middletown, CT: Wesleyan University.
- Rosenberg, M. (1979). *Conceiving the self*. New York, NY: Basic Books.
- Satorra, A., & Bentler, P. M. (1999). *A scaled difference chi-square test statistic for moment structure analysis* (Technical report). Los Angeles, CA: University of California.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika, 22*, 53–61.
- Seidman, E., Allen, L., Aber, J. L., Mitchell, C., & Feinman, J. (1994). The impact of school transitions in early adolescence on the self-system and perceived social context of poor urban youth. *Child Development, 65*, 507–522.
- Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407–441.
- Shevlin, 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.
- Steiger, A. E., Allemand, M., Robins, R. W., & Fend, H. A. (2014). Low and decreasing self-esteem during adolescence predict adult depression two decades later. *Journal of Personality and Social Psychology, 106*, 325–338.
- 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.
- Vispoel, W. P. (1995). Self-concept in artistic domains: An extension of the Shavelson, Hubner, and Stanton (1976) model. *Journal of Educational Psychology, 87*, 134–153.
- Wichström, L. (1995). Harter's Self-Perception Profile for Adolescents: Reliability, validity, and evaluation of the question format. *Journal of Personality Assessment, 65*, 100–116.
- Wigfield, A., Eccles, J., MacIver, D., Reuman, D. A., & Midgley, C. (1991). Transitions during early adolescence: Changes in children's domain-specific self-perceptions and general self-esteem across the transition to junior high school. *Developmental Psychology, 27*, 552–565.
- Wylie, R. C. (1989). *Measures of self-concept*. Lincoln, NE: University of Nebraska.
- Zimmerman, M. A., Copeland, L. A., Shope, J. T., & Dielman, T. E. (1997). A longitudinal study of self-esteem: Implications for development. *Journal of Youth and Adolescence, 26*, 117–141.

Manuscript received January 12, 2016

Final revision received June 22, 2016

Accepted July 31, 2016