

Running head: Music Self-Concept

The Music Self-Perception Inventory: Development of a Short Form

Alexandre J.S. Morin¹

L. Francesca Scalas²

Walter Vispoel³

Herbert W. Marsh¹

Zhonglin Wen⁴

1Australian Catholic University, Australia

2Università degli Studi di Cagliari, Italy

3University of Iowa, USA

4South China Normal University, China

This is the prepublication version of the following manuscript:

Morin AJS, Scalas LF, Vispoel W, Marsh HW, Wen Z. The Music Self-Perception Inventory: Development of a short form. *Psychology of Music*. 2016;44(5):915-934.

Copyright © 2016 (Sage publishing).

This paper is not the copy of record and may not exactly replicate the authoritative document published in *Psychology of Music* at:

<http://journals.sagepub.com/doi/full/10.1177/0305735615592690>

Funding

Preparation of this article was supported in part by grants from the Australian Research Council (DP130102713; DP140101559) and the National Natural Science Foundation of China (31271116).

Note

The first two authors (A. J. S. M. & L. F. S.) contributed equally to the preparation of this article, so their order of appearance was determined at random: all should be considered first authors.

The Music Self-Perception Inventory: Development of a Short Form

Abstract

Music self-concept integrates perceptions, beliefs and self-schemas about a person's musical abilities and potential. Like other self-concept dimensions, it is multifaceted, hierarchically organized and has implications for motivation toward musical practice. The Music Self-Perception Inventory (MUSPI) is a theoretically based instrument assessing six specific music self-concept dimensions, as well as the global music self-concept. Nonetheless, its applicability is limited by its length (84 items). In this study, we developed and validated a 28-item short form of the MUSPI, and showed that the short form yielded equivalent psychometric properties as the original. We cross-validated the original MUSPI on a first sample and used these results to develop a shorter version (MUSPI-S), which we then cross-validated using a new independent sample. We also tested whether the MUSPI-S psychometric properties generalized (were invariant) across gender and grade-differentiated subgroups. Finally, we examined the convergent validity of the MUSPI and MUSPI-S. Results highlighted the psychometric soundness of the MUSPI-S on all criteria, and showed that it presented pattern of associations with other constructs equivalent to that observed with the original MUSPI.

Key words. Musical Self-Concept, Measure, MUSPI, Short-Form, Validation.

Shavelson, Hubner, and Stanton (1976) defined self-concept as a person's self-perceptions in multiple domains, encompassing feelings of self-confidence, self-worth, and ability. Self-concept represents a powerful predictor of multiple positive outcomes across the lifespan (e.g., Marsh, 2007a; Marsh & Craven, 2006), due to its major implications for motivational processes (e.g., Eccles & Wigfield, 2002; Wigfield & Eccles, 2000). In line with Shavelson et al. (1976), contemporary research emphasizes the multidimensional and hierarchical nature of self-concept (e.g., Marsh, 2007a; Marsh & Craven, 2006; Marsh & Scalas, 2010). Self-concept encompasses multiple sub-selves or domains organized hierarchically, with global self-concept at the higher level, general self-domains at the next level (e.g., academic, social, or physical selves), and more specific self-components at the next lower level (e.g., math self-concept, familial self-concept) (e.g., Marsh & Shavelson, 1985; Shavelson et al., 1976). In this study, we focused on the music self-concept.

Music Self-Concept

Shavelson et al.'s model (1976) has guided the development of models of domain-specific self-concepts, such as physical (Fox & Corbin, 1989; Marsh, Richards, Johnson, Roche, & Tremayne, 1994), academic (Marsh, 1990, 1993), and social (Byrne & Shavelson, 1996) selves. This framework has been also applied to the music self-concept (Vispoel, 1994, 2003), which integrates perceptions, beliefs and self-schemas about a person's musical abilities and potential (Schnare, MacIntyre, & Doucette, 2011). Like other self-domains, music self-concept appears to be *"multifaceted in the sense that individuals code their experiences with music into categories or facets that facilitate their understanding of themselves and their environment. Facets of music self-concept, however, are not necessarily universal or context-free; they may be specific to an individual and/or shared by a group. Music self-concept is hierarchically structured in that individuals differentiate their perceptions of music skill according to levels of abstraction that move from specific to general and vice versa"* (Vispoel, 1994, p. 54).

Correlates and Specificity of Music Self-Concept

Music self-concept has been found to be related to motivation toward musical practice (West, 2013), particularly intrinsic motivation (Sandene, 1997; Schmidt, 2005). In education settings, students with more positive music self-concepts tend to use more internal attributions for musical outcomes and learning strategies, to invest more efforts and perform better in musical activities according to self-reported and teacher ratings (Austin & Vispoel, 1998; Schmidt, 2005). Vispoel (1994) found that music self-concept was more strongly related to theoretically-connected (such as artistic and verbal academic self-concepts) and hierarchically-related (e.g., global self-concept) self-components, than to theoretically-distinct self-components (e.g., physical self-concept). Music self-concept is thus part of a self-concept hierarchy involving a hierarchically-superordinate artistic domain, and hierarchically-subordinate components (music composition, instrument playing, reading music, etc.; Vispoel, 1995). When researchers look more carefully at multiple subdomains (or facets) of music self-concept, these facets appear well-differentiated and distinct from theoretically-related constructs, such as music achievement, interest, or attributions for success and failure, with which they share well-differentiated relations (Vispoel, 1994, 2003). These facets also show only modest direct relations with global self-concept but stronger indirect relations whereby music self-concept facets predict music involvement, aspirations, and interest, which together better predict global self-concept (Forte & Vispoel, 1995; Vispoel, 2003).

Music Self-Concept, Gender and Grade

Generally, females report higher motivational beliefs and participation in music activities than males (Evans, Schweingruber, & Stevenson, 2002; Jacobs, Lanza., Osgood, Eccles, & Wigfield 2002; Simpkins, Vest, & Becnel, 2010; Wigfield et al., 1997). Research also highlights gender differences, favoring females, in most dimensions of music self-concept (Austin, 1991; Austin & Vispoel, 2000; Eccles, Wigfield, Harold, & Blumenfeld, 1993; Forte & Vispoel, 1993, 1995; Vispoel, 1993a; Vispoel & Forte, 1994, 2000; Vispoel & Rizzo, 2003). Vispoel (1994) theorized that music self-concept becomes

increasingly multifaceted as one grows older and gains experience with music. Nonetheless, empirical research has not yet established the specific age at which this differentiation occurs. Although previous research has found some age or grade-related differences in self-perceptions, values and interests in music (e.g., Austin & Vispoel, 2000; Eccles et al., 1993), it has also shown that the structure of music self-concept and the pattern of interrelations between music self-concept components did not change between junior high school and college (Vispoel, 2003). However, no systematic test of the measurement invariance across gender or age/grade of music self-concept measures has ever been conducted. Because measurement invariance is a critical pre-requisite to any valid group-based comparisons (e.g., Meredith, 1993; Millsap, 2011) this is clearly an important limitation of previous research in this area.

Measurement of Music Self-Concept: The Music Self-Perception Inventory

Any systematic investigation of psychological constructs requires access to strong and well-validated measures. The Music Self-Perception Inventory (MUSPI; Vispoel, 1993b, 1993c, 1994, 2003) assesses a multidimensional and hierarchical conception of the music self-concept. The MUSPI includes 84 items focusing on musical skills covering both general and specific areas. The MUSPI includes one subscale assessing perceptions of global music self-concept (Global MSC), and six subscales assessing more subdomain-specific perceptions of music skills (singing, instrument playing, reading, composing, listening, and dancing). Each scale includes 12 items, 6 of which are negatively worded. Previous research (Vispoel, 1994, 2003) has supported the seven-factor structure of the MUSPI through confirmatory factor analyses, and revealed high estimates of scale score reliability ($\alpha = .92$ to $.98$) and test-retest coefficients in the .80s and .90s over one to four month intervals. These results showed that Singing and Dancing self-concepts are relatively distinct from other music self-concept facets, and contributed less to overall perceptions of musical ability (i.e., global music self-concept Global MSC; Vispoel, 1994, 2003). The remaining facets (Instrument Playing, Reading, Composing, Listening) proved to be more strongly related to one another – albeit still reasonably distinct – and to global music self-

concept Global MSC. Research also showed that MUSPI scores could reliably differentiate individuals with noteworthy achievements in each targeted artistic domain from other individuals. Similarly, correlations of subscales scores with external criteria revealed a logical pattern of relations consistent with the facets of music performance being measured (Vispoel, 1994, 2003).

Development of the Short Form of the MUSPI

Although the MUSPI presents strong psychometric properties and solid theoretical bases, its length (i.e., 84 items) limits its applicability when it needs to be administered with other instruments. Thus, the purpose of this study is to develop and validate a short form (28-item, including four items for each of the original 7 subscales; see Online Supplements) of the MUSPI. Although short instruments have clear practical advantages, they also present limitations in terms of construct coverage and often fall short of reasonable psychometric standards when evaluated rigorously (Marsh, Ellis, Parada, Richards, & Heubeck, 2005; Smith, McCarthy, & Anderson, 2000). For this reason, guidelines have been proposed to develop psychometrically strong short measures (Mañano et al., 2008; Marsh et al., 2005; Myers, McCarthy, MacPherson, & Brown, 2003; Smith et al. 2000); and previous studies have shown that is possible to reliably assess complex multidimensional constructs with as few as one or two items per dimension (e.g., DeSalvo et al., 2006; Morin & Mañano, 2011; Moullec et al., 2011). These guidelines state that test developers should start with a strong long form of the instrument and show within independent cross-validation samples that (Mañano et al., 2008; Marsh et al., 2005; Myers et al., 2003; Smith et al. 2000): (a) the short form retains the content coverage of each factor; (b) the short form retains the factor structure of the original instrument; (c) the factor structure of the short form meets acceptable goodness of fit standards; (d) the short form provides adequate scale score reliability; (e) the short form relates to external criteria in the same manner as the long-form (i.e., equivalent convergent validity).

Furthermore, these guidelines also address the manner in which items should be selected from the long form to create the short form (Marsh et al., 2005; Myers et al., 2003; Smith et al. 2000). Starting

with a factor model estimated using the long form, short form items should: (a) present high factor loadings and low uniquenesses; (b) present low correlated uniquenesses and cross-loadings (as shown by modification indices); (c) seldom be missing; (d) receive a positive subjective evaluation of their content.

The MUSPI has a very complex structure, including negatively worded items, as well as items with parallel wording across each factors, both of which have been shown to result in methodological artifacts that need to be controlled for in the estimation of measurement models (Marsh 2007b; Marsh, Abduljabbar et al., 2013; Marsh, Scalas, & Nagengast, 2010; Morin, Arens, & Marsh, 2014). This control reflects the fact that the unique variance of these items is likely to be shared among items with negative, or parallel, wordings. However, previous evidence shows that the content of latent constructs was unlikely to be biased by the inclusion, or exclusion, of negatively-worded items (DiStephano & Motl, 2006; Marsh et al., 2010; Tomás & Oliver, 1999). Thus, to maximally simplify the MUSPI structure, we built the short form solely from positively-worded items. However, this decision makes it critical to ascertain that the convergent validity of the short MUSPI remains unaffected in comparison with the long form (e.g., Quilty, Oakman, & Risko, 2006; Tomás, Oliver, Galiana, Sancho, & Lila, 2013).

The Present Study

The MUSPI is a psychometrically sound instrument built on solid theoretical ground. Nonetheless, its length limits its applicability. To overcome this limitation, the current study aims to develop, and validate a valid short form of the MUSPI. Because starting with a strong long form has been identified as a critical first step in the development of a short form, we start by cross-validating the original MUSPI on a first sample of participants. Then, using the results obtained with this sample, we create a short and simplified 28-item version of the MUSPI (MUSPI-S). Then, in order to systematically test the generalizability of the obtained factor structure beyond this first sample, we tested its measurement invariance with a new independent sample, as well as across gender and grade-differentiated subgroups. These tests provide a strong test of the extent to which the MUSPI-S factor structure generalizes across males and females

participants of different grade levels. Finally, using the combined sample to maximize statistical power, we examined the convergent validity of the MUSPI and MUSPI-S with multiple external criteria. Since previous studies have highlighted the ability of MUSPI to differentiate individuals with high achievement and accomplishments in music activities from other individuals (Vispoel, 2003), our criterion included self-reported grades and ability in the areas measured by the MUSPI. We expected that grades in Music would be more strongly correlated with MUSPI dimensions than with grades in other areas. Moreover, we expected that self-reported abilities in specific dimensions would be more strongly associated with the corresponding facet of the MUSPI than with other facets (e.g., the self-reported rating in dance would be more strongly associated with Dancing self-concept than with Listening self-concept). Previous studies have also differentiated music self-concept from other dimensions of the artistic self-concept. Thus, here we also included past experiences in various artistic activities (performing music, dancing, acting, displaying art works), with the expectation that only experiences in music activities should be related to MUSPI dimensions.

Method

Participants

Two independent samples of students were used. The first sample included 304 students (12-16 years old; $M_{\text{age}} = 13.14$; $SD_{\text{age}} = .71$), all recruited within a single school, including 195 (64.1%) 7th graders and 109 (35.9%) 8th graders, with a slightly higher percentage of females ($n = 169$; 55.6%) than males ($n = 135$; 44.4%). The second sample of 208 participants (11-16 years old; $M_{\text{age}} = 13.21$; $SD_{\text{age}} = .70$) was recruited from a different school, and included 133 (63.9%) 7th graders and 75 (36.1%) 8th graders, also with a higher percentage of females ($n = 142$; 68.3%) than males ($n = 66$; 31.7%). Participation was voluntary, and participants were informed that their answers would be completely anonymous and invited to ask any questions that they wanted about the study. Consent procedures complied with the Iowa Fair Information Practices Act. Questionnaires were administered in quiet classroom conditions.

Measures

The MUSPI includes 84 items that measure music self-concept at both general and specific levels. One subscale assesses perceptions of general music ability (Global MSC, e.g., *I am good at doing most music-related activities*), and six additional ones assess perceptions of skill in the music subdomains of Singing (e.g., *I am better than most people my age at singing*), Instrument Playing (e.g., *I am good at playing a musical instrument*), reading (e.g., *I am skilled at reading music*), Composing (e.g., *I am good at making up music*), Listening (e.g., *I am good at identifying characteristics of music by ear*), and Dancing (e.g., *Creating dance movements to music is easy for me*). Each scale includes 12 items, six of which are negatively worded. Items are rated on a 6-point scale, ranging from False (1) to True (6).

Students were asked to report their last grade in: English, Math, Physical education, Dance, Music, Art, Drama on a 13-point scale (with 1 = F, 2 = D-, ..., 13 = A+). The self-perceived ability on the seven MUSPI dimensions was rated on a 6-point scale (1-Poor to 6-Outstanding). Past experiences in artistic activities were assessed through four Yes/No items about past experiences in performing music, dancing, acting, and art works (e.g., *Have you ever performed music in a band, orchestra, or choir at school, church or in your community?*). One additional question assessed the number of years playing a music instrument on a 7-point scale (1 = 0 to less than one year; 2 = 1 to 2 years; 3 = 3 to 4 years; 4 = 5 to 6 years; 5 = 7 to 8 years; 6 = 9 to 10 years; 7 = 11 or more years).

Analyses

All models were estimated using Mplus 7.2 (Muthén & Muthén, 2012) robust weighted least square estimator (WLSMV), which outperforms Maximum Likelihood estimation with ordered-categorical Likert-type items such as those used in the present study (e.g., Bandalos, 2014; Finney & DiStefano, 2006, 2013). Models were estimated based on the full available information (Asparouhov & Muthén, 2010), to take into account the very few missing responses present at the item level (Sample 1: 0 to 7.42%, M = 3.04%, SD = 1.85%; Sample 2: 0 to 3.48%, M = 1.41%, SD = 1.11%).

With the first sample, we analyzed responses to the full MUSPI using Confirmatory Factor Analyses (CFA). In this 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 7 correlated factors representing the previously described MUSPI subscales. To take into account the methodological artifact due to the negatively-worded items (reversed-coded prior to the analyses to facilitate interpretation), this model included an orthogonal method factor underlying all negatively-worded items (e.g., Marsh et al., 2010). Furthermore, to take into account the additional methodological artifact due to the parallel-worded items, a priori correlated uniquenesses among parallel-worded items were also included to the model. Negatively-worded and parallel-worded items are identified in Table 2 below.

From this model, the best four positively-worded indicators of all factors were selected to create the MUSPI-S, following the previously enumerated guidelines. The adequacy of the measurement model underlying the MUSPI-S was then verified using CFA on the data from the first sample, and cross-validated using the data from the second sample. These models were tested with, and without, the inclusion of the correlated uniquenesses used to represent items with parallel wording. From these models, composite reliability (Raykov & Grayson, 2003) was calculated with McDonald's (1970) omega (ω) coefficient: $\omega = (\sum |\lambda_i|)^2 / ((\sum |\lambda_i|)^2 + \sum \delta_{ii})$ where λ_i are the factor loadings and δ_{ii} , the error variances. Compared to Cronbach's α , ω has the advantage of taking into account the strength of association between items and constructs as well as item-specific measurement errors (see Sijtsma, 2009).

Tests of measurement invariance of the MUSPI-S across samples were then conducted in the following sequence (Meredith, 1993; Millsap, 2011), adjusted for ordered-categorical items (Millsap & Tein, 2004; Morin et al., 2011): (1) configural invariance; (2) weak invariance (invariance of the factor loadings); (3) strong invariance (loadings and thresholds); (4) strict invariance (loadings, thresholds and uniquenesses); (5) invariance of the variance/covariance matrix (loadings, thresholds, uniquenesses, and latent variances-covariances); (6) latent mean invariance (loadings, thresholds, uniquenesses, latent

variances-covariances, and latent means). Similar tests of measurement invariance were then conducted across gender, and grade levels. These additional tests of measurement invariance were conducted on the combined sample to maximize sample size and statistical power to detect latent mean differences.

The fit of all models was evaluated based on the following indices as operationalized in Mplus 7.2 in conjunction with WLSMV estimation (Hu & Bentler, 1999; Yu, 2002): The Chi-square (χ^2), the Comparative Fit Index (CFI), the Tucker-Lewis Index (TLI), the Root Mean Square Error of Approximation (RMSEA) and its 90% confidence interval. Values greater than .90 and .95 for both the CFI and TLI are considered to indicate adequate and excellent fit to the data, respectively, while values smaller than .08 or .06 for the RMSEA reflect acceptable and excellent model fit (Hu & Bentler, 1999; Yu, 2002). WLSMV chi-square values are not exact, but "*estimated*" as the closest integer necessary to obtain a correct *p*-value. Thus, in practice, only the *p*-value should be interpreted. This explains why sometimes the chi-square values and resulting CFI values can be non-monotonic with model complexity. For the CFI, any increase when constraints are added should thus simply be interpreted as random, rather than as an improvement in fit. This also means that the chi square difference tests cannot be computed by hand but need to be conducted via Mplus' DIFFTEST function ($MD\Delta\chi^2$; Asparouhov & Muthén, 2006; Muthén, 2004). However, as with the χ^2 , $MD\Delta\chi^2$ are oversensitive to sample size and minor misspecifications so that nested model comparisons (e.g., test of invariance) generally rely on examinations of changes in fit indices (Chen, 2007; Cheung & Rensvold, 2002). A CFI decline of .01 or less, and a RMSEA increase of .015 or less, between nested models indicates that the more parsimonious model (e.g., invariant) should be retained. With complex models, it has been suggested that the inspection of fluctuations in fit indices that correct for parsimony (TLI and RMSEA) may be important given the large number of estimated parameters and the fact that these indices can improve when constraints are added to a model (Marsh, Hau, & Grayson, 2005; Marsh, Hau, & Wen, 2004). We emphasize that these proposed cut-off scores should only be considered as rough guidelines (Marsh et al., 2004, 2005).

Results

Sample 1: Analyses of the Full MUSPI and Development of the MUSPI-S

The goodness of fit indices for the a priori CFA model estimated on the original 84 items of the MUSPI are reported in **Table 1** and all indicate an excellent fit to the data (CFI and $TLI \geq .95$; $RMSEA \leq .06$). These results also support the need to incorporate a priori controls for wording effects, as shown by a substantial improvement in fit indices. The parameter estimates from this model are reported in **Table 2** (factor loadings, uniquenesses, and composite reliability) and **Table 3** (factor correlations) and show that all factors present a fully satisfactory level of composite reliability ($\omega = 0.96$ to 0.97 ; $M_\omega = 0.97$) and are well defined by high factor loadings ($\lambda = 0.66$ to 0.94 ; $M_\lambda = 0.83$). Supporting previous research (Vispoel, 1994, 2003; Vispoel & Forte, 1994), these results show that Singing ($M = 0.44$) and Dancing ($M = 0.37$) are not as strongly related to other dimensions as are Instrument Playing, Reading, Composing, and Listening ($M = 0.75$). Similarly, in accordance with Shavelson et al.'s (1976) model, music self-concept dimensions present generally stronger relations with Global MSC ($M = 0.73$) than with the other music self-concept dimensions ($M = 0.54$).

[Insert Table 1-2-3 here]

Using the aforementioned guidelines, we then selected the four best positively worded items for each factor to create the MUSPI-S. The resulting factor model also provided an excellent fit to the data (CFI and $TLI \geq .95$; $RMSEA \leq .06$, see Table 1), although it was less clear from this model whether the methodological control for parallel wording needed to be retained. For consistency with the previous model, and in line with previous studies showing the importance to incorporate these controls to obtain proper parameter estimates (Marsh 2007b; Marsh, Abduljabbar et al., 2013; Marsh et al., 2010), these controls were kept in the model (see Table 2 for details). Results from analyses excluding these controls remain substantively identical to the results reported here, and can be consulted in **Tables S1 to S4 of the online supplements**. The parameter estimates from this model are reported in Table 2 and 3 and show

that all factors from the MUSPI-S present a fully satisfactory level of composite reliability ($\omega = 0.91$ to 0.95 ; $M_\omega = 0.93$) and are well defined by high factor loadings ($\lambda = 0.82$ to 0.94 ; $M_\lambda = 0.88$).

Importantly, factors loadings, reliability estimates, and factor correlations seem unaffected by the change from a long to a short form. To estimate more directly the similarity of the parameter estimates obtained for the MUSPI and MUSPI-S, we calculated a profile similarity index (PSI). The PSI is simply an estimate of the correlations between parameter estimates obtained for different models. Here, we calculated the correlation between the loadings of the 28 items retained in the MUSPI-S across the measurement models estimated for the MUSPI and MUSPI-S. Similarly, we calculated PSI for the composite reliability estimates and factor correlations from the MUSPI and MUSPI-S measurement models. For all parameters, the PSI confirm the very high level of similarity between the long and short forms of the MUSPI ($PSI_\lambda = .75$; $PSI_\omega = .91$; $PSI_r = .99$).

Sample 2: Cross-Validation of MUSPI-S

The psychometric properties of the MUSPI-S were then cross validated using data from the second sample. The results from this model fully replicated those from Sample 1 and showing a fully satisfactory level of fit to the data (CFI and $TLI \geq .95$; $RMSEA \leq .06$, see Table 1), factors that were well-defined through high factor loadings ($\lambda = 0.82$ to 0.95 ; $M_\lambda = 0.89$; see Table 2), and fully satisfactory estimates of composite reliability ($\omega = 0.92$ to 0.95 ; $M_\omega = 0.94$) in line with estimates from Study 1 ($PSI_\lambda = .57$; $PSI_\omega = .88$). The pattern of factor correlation, reported in **Table 4**, remained fully in line with the correlations observed in Study 1 for both the long ($PSI_r = .95$) and short ($PSI_r = .95$) forms of the MUSPI.

[Insert Table 4 here]

Tests of measurement invariance across samples were conducted to systematically investigate the degree to which the measurement model was replicated across samples. These results (see Table 1) supported the complete invariance (i.e., configural, loadings, thresholds, uniquenesses, latent variance-covariance, latent means) of the MUSPI-S across samples. Indeed, the ΔCFI , ΔTLI , and $\Delta RMSEA$ all

remained well under the recommended cut-off scores, and changes in fit indices including an adjustment for model parsimony (i.e., TLI and RMSEA) showed an improvement in fit at most steps of the invariance sequence. These results suggest that the psychometric properties of the MUSPI-S, as well as the estimated latent factor correlations and latent means, were fully replicated across samples. Both samples could thus be combined to maximize the power of further analyses.

Combined Sample: Invariance Across Gender and Grade Levels.

Tests of measurement invariance were conducted on the combined sample to systematically investigate the degree to which the measurement model was replicated across genders and grade levels, and to investigate possible latent mean differences across these subgroups of participants. The results from these tests are reported in Table 1. These results support the strict measurement invariance (i.e. configural, loadings, thresholds, uniquenesses), as well as the invariance of the latent variances and covariances, of the MUSPI-S across genders and grade levels. The results also support the latent mean invariance of the MUSPI-S across grade levels, suggesting that previously reported (e.g., Austin & Vispoel, 2000; Eccles et al., 1993) or suggested (Vispoel, 1994) age/grade-related differences might have been due to the lack of control for possible non-invariance of the measures, or to the possibility that these differences may occur at younger ages. In contrast, when latent means are constrained to invariance across gender, the ΔCFI and ΔTLI were very close to the recommended cut-off score of .01 and the $\Delta RMSEA$ proved greater than the recommended cut-off score of .015. These results, coupled with Fan and Sivo's (2009) observation that changes in goodness-of-fit indices tend to be less trustworthy indicators of latent mean invariance, suggests that latent means may differ across gender. Because of their theoretical relevance, we thus examined possible latent mean differences across gender. The results showed that when the latent means of males were fixed to zero for identification purposes, the latent means (expressed in standard deviation units) of females were significantly ($p \leq .05$) higher on the Dancing (.77), Instrument Playing (.34), Reading (.21), Singing (.48), and Global MSC (.41) factors, but not

significantly different on the Composing (.03) and Listening (.03) factors.

Combined Sample: Convergent Validity.

Finally, still using the combined sample to maximize statistical power, we investigated the convergent validity of the MUSPI and MUSPI-S factors with an array of theoretically-relevant external criteria. These results are reported in **Table 5**. Before moving on to a detailed examination of these results, it is first noteworthy that the pattern of correlations observed for the MUSPI is almost identical to the pattern of correlations observed for the MUSPI-S. For both versions, the subscales showed low correlations with school grades (with the exception of music and English) and past experience in art (outside of music), low to moderate correlations with past music experiences and self-perceived ability in non-matching musical domains, and moderate to high correlations with self-perceived ability in the same musical domains. Indeed, the difference in the sizes of these correlations remain tiny and vary between 0 and .06 ($M = .01$). Furthermore, the PSI calculated between both sets of correlations is near perfect ($PSI_r = .99$). These results clearly show that the convergent validity of the instrument seems unaffected by the process that we followed to retain only a subset of positively-worked items from the original MUSPI.

[Insert Table 5 here]

When we look more specifically at the correlations between the MUSPI-S and external criteria, the results show interesting observations. First, correlations with school grades were almost null, or at least very small ($r \leq .20$) for the school subjects not involving Musical abilities (Math, Physical Education, Art, Drama). However, correlations between musical self-concept dimensions were higher with Music course grades. Dancing self-concept is the only exception, showing stronger correlations with Dance course grades than Music course grades. Finally, in accordance with Vispoel (1994) results, musical skills are not completely disconnected from other skills, as illustrated by larger correlations between English grades and the Global MSC, Instrument Playing, Reading, and Composing subscales. This result is likely due to the fact that verbal skills may be involved in the process of learning and nurturing musical skills.

Second, apart from Dancing, all other MUSPI-S subscales were positively and significantly correlated with past experiences of involvement in musical groups and number of years of instrument playing. In contrast, the Dancing scale showed a specific positive correlation with previous experiences of involvement in dance groups. Interestingly, the Singing subscale also showed a substantial positive relation with previous experiences of involvement in dance groups, a result which should be more thoroughly investigated in future studies. As expected, for almost all MUSPI-S subscales, correlations between students' levels of performance in acting or art work remained small. In fact, only the Composing subscale correlated more substantially with art work, suggesting that the creative skills involved in music composition may be shared with other artistic areas.

Third, and again as expected, students' self-concepts and self-perceived abilities in the same areas presented a pattern of relationship that was clearly differentiated across musical domains, with within-domain correlations (e.g., Dancing self-concept with self-perceived ability in dance) systematically higher than non-matching correlations (e.g., Dancing self-concept with self-perceived ability in singing, and Singing self-concept with self-perceived ability in dance). Moreover, the Global MSC subscale showed moderate to high correlations with all areas of self-perceived musical abilities (with the only exception of dance). Similarly, self-perceived ability in music-related activities also showed moderate to high correlations with all MUSPI-S subscales.

Discussion

The main objective of this study was to develop and evaluate the factor validity and reliability of a 28-item short form version of the MUSPI for use in research and practice contexts where the regular 84-item version is too long. Starting from the 84-item long form version of the MUSPI, the results from CFAs conducted on a first sample of adolescents showed that the a priori 7-factor model provided a satisfactory degree of fit to the data, and resulted in well-defined, and highly reliable ($\omega = 0.96$ to 0.97) factors. From this model, the best 28 positively worded items (with 4 items per dimension) were selected

for the creation of the MUSPI-S following guidelines previously proposed for the development of short forms (Mañano et al., 2008; Marsh et al., 2005; Myers et al., 2003; Smith et al. 2000). We thus retained items presenting high factor loadings, low uniquenesses, low correlated uniquenesses and cross-loadings (shown by modification indices), few missing data, and good content validity. The results from a CFA model conducted on the same sample to assess the psychometric properties of the MUSPI-S fully supported the adequacy of this model in terms of goodness-of-fit, well-defined factors, and high estimates of composite reliability ($\omega = 0.91$ to 0.95). More importantly, the parameter estimates from the MUSPI-S CFA model were almost identical from those obtained on the original MUSPI ($\text{PSI}_k = .75$; $\text{PSI}_\omega = .91$; $\text{PSI}_r = .99$). Finally, the MUSPI-S was cross-validated on a new independent sample. The CFA model estimated on this new sample again supported the psychometric properties of the MUSPI-S in terms of goodness-of-fit, well-defined factors, and high estimates of composite reliability ($\omega = 0.92$ to 0.95). Attesting to the robustness of the MUSPI-S factor structure across sample, the estimated measurement model was fully invariant (i.e., configuration, loadings, thresholds, uniquenesses, latent variance-covariance, and latent means) across these two samples.

It is interesting to note that the factor correlations obtained in the current study (which was essentially the same across versions of the MUSPI and samples, as shown by PSI_r varying between .95 and .99) generally supported the results from previous studies on the MUSPI (Forte & Vispoel, 1995; Vispoel, 1994, 2003; Vispoel & Forte, 1994) and our theoretical expectations (e.g., Marsh, 2007a; Shavelson et al., 1976). More precisely, these results showed stronger associations between music self-concept facets and Global MSC than among music self-concept facets, which were still significantly related to one another. In particular, Singing and Dancing self-concepts were more distinct from the other facets of music self-concept, showing that these two abilities tap into relatively distinct domains of performing abilities.

A critical component of the present study was to assess the convergent validity of the MUSPI-S, and

to show that it remained unaffected by the selection of a subset of positively worded items from the original MUSPI. In this regard, our results provided strong support to the ability of the MUSPI-S factors to replicate the convergent validity of the original MUSPI, resulting in almost identical estimates of correlations with a total of 19 external criterion indices ($\text{PSI}_r = .99$). These correlations were also fully in line with results from previous research and theoretical expectations (e.g., Marsh, 2007a; Shavelson et al., 1976; Vispoel, 1994, 2003).

First, the correlations showed that MUSPI-S factors were positively and significantly related with school grades in music, but showed almost no significant associations with school grades in subjects not related to musical abilities. The only exceptions to this pattern were: (a) substantial associations between: the Global MSC, Instrument Playing, Reading, and Composing subscales of the MUSPI-S with English grades, suggesting the importance of verbal skills in the process of learning and teaching music in junior high school (see also Vispoel, 1994), and (b) more substantial associations between Dancing self-concept and Dance course grades.

Second, the correlations revealed that most MUSPI-S factors showed positive associations with past experiences of involvement in music. In contrast, Dancing (and Singing to a lesser extent) self-concept was more strongly related to dancing experiences. Interestingly, the Composing self-concept scale also presented more substantial associations with previous experiences in art work, suggesting that the creative skills involved in music composition may be shared across artistic areas. Finally, students' music self-concepts facets presented a well-differentiated pattern of association with self-perceived abilities in the same areas, as illustrated by stronger within-domain than across-domain correlations.

A last objective of the present study was to examine measurement invariance and latent mean differences of MUSPI-S scores across subgroups of adolescents formed on the basis of gender (males versus females) and grade-level (grade 7 versus 8). Analyses provided strong support for the strict measurement invariance (i.e., configuration, loadings, thresholds, uniquenesses) of the MUSPI-S across

all of these subgroups, as well as for the invariance of the latent variances and covariances. This last result confirms that the pattern of associations and differentiations between the MUSPI-S subscales stayed identical across gender and grade-level. Similarly, no mean-level differences in music self-concept facets could be identified between students attending the seventh or eighth grade. These last results apparently contradict previous research suggesting the presence of age or grade-related differences in music self-perceptions, values and interests (e.g., Austin & Vispoel, 2000; Eccles et al., 1993), and theoretical expectations that music self-concept may become increasingly multifaceted and differentiated as one grows older and gains more musical experience. However, it should be noted that Vispoel (2003) previously reported a similar result regarding the lack of differences in the observed pattern of associations between music self-facets when he compared junior high school and college students. Taken together, these observations reinforce the need to replicate the current findings on new and more diversified samples (from different school levels, age groups, cultures, language group and levels of musical proficiency), relying on systematic tests of measurement invariance. However, they also suggest that the expected differences in means and correlations may rather appear at a younger age, when students first experience music training or move from elementary school to junior high school, or at a later age, when they move to a more professional level of musical practice.

Our results revealed significant latent mean differences across genders on the MUSPI-S factors, mainly showing that females tend to present higher levels of music self-concepts (save for the most auditory-cognitive dimensions) than males. This result is fully in line with previous reports that females tend to present higher levels of motivation, skills, involvement, and self-conceptions in the music areas than males (e.g., Austin & Vispoel, 2000; Evans et al., 2002; Simpkins et al., 2010; Vispoel & Forte, 1994, 2000; Vispoel & Rizzo, 2003; Wigfield et al., 1997), save for the most auditory-cognitive dimensions, thus supporting the construct validity of the MUSPI-S. These gender differences are known to have an impact on music education (Hargreaves, Comber, & Colley, 1995; Welch et al., 2004). For

example, the choice of a preferred instrument and music style has previously been showed to be affected by gender (Harrison & O'Neill, 2000; O'Neill & Boulton, 1996), so that gender-differences may potentially limit the openness of the pupils to a variety of musical styles and instruments (Hargreaves et al., 1995). Alternatively, research also suggests that these differences may partly result from stereotypes conveyed by music teachers (for a more extensive discussion, see Maidlow, & Bruce, 1999). Future research should look more carefully into these differences and at possible interventions to increase teacher awareness of gender-differentiated processes, and of their possible role in shaping these differences.

Finally, our results not only have implications for music self-concept, confirming its multidimensional nature and the specificity of its components, but also have implications for self-concept theory and research more generally. For example, the mechanisms of interconnection between specific and global components of self-concept and the possibility that the effects of these components may be moderated by importance are still debated in the literature (Scalas, Morin, Marsh, & Nagengast, 2014). In this area, investigating narrowly defined self-domains considered unimportant for most people but very important for some people, such as the music self-concept, might help clarifying these issues. Preliminary studies seem to confirm the hypothesis that possible moderation by importance may be particularly important for these narrow self-concept domains (Vispoel, 2003), but need to be corroborated with stronger methodologies (Scalas, Marsh, Morin, Nagengast, 2013). Therefore, having a sound and short instrument to evaluate music self-concept might be useful to further expand this are of research.

In conclusion, our findings show that the MUSPI-S presents acceptable psychometric properties and provides an assessment of adolescents' music self-concept that is comparable to that achieved using the longer, 84-item version, of the MUSPI. Our results thus suggest that the MUSPI-S may be confidently used in to assess the music self-concept among samples of English speaking adolescents comparable in age to those from the current samples (grade 7 and 8). However, because the MUSPI and MUSPI-S have only been validated in English, verifying their cross-linguistic validity remains a priority for future

research, as this represents a pre-requisite to cross-cultural investigations of this construct.

References

- Asparouhov, T. & Muthén, B.O. (2006). *Robust chi-square difference testing with mean and variance adjusted test statistics*. Los Angeles, CA: Muthén & Muthén.
- Asparouhov, T., & Muthén, B.O. (2010). *Weighted Least Square estimation with missing data*. Los Angeles, CA: Muthén & Muthén.
- Austin, J.R. (1991). Competitive and noncompetitive goal structures: An analysis of motivation and achievement among elementary band students. *Psychology of Music, 19*, 142–158.
- Austin, J.R., & Vispoel, W.P. (1998). How American adolescents interpret success and failure in classroom music: Relationships among attributional beliefs, self-concept and achievement. *Psychology of Music, 26*, 26-45.
- Austin, J.R., & Vispoel, W.P. (2000). Children's ability self-perceptions and interests: Grade level, gender and race differences in music, reading, and math. In R. G. Craven & H. W. Marsh (Eds.), *Self-Concept, Theory, Research and Practice: Advances for the New Millennium* (pp. 133-142). Sydney, Australia: Self-concept Enhancement and Learning Facilitation Research Centre.
- Bandalos, D.L. (2014). Relative performance of categorical diagonally weighted least squares and robust maximum likelihood estimation. *Structural Equation Modeling, 21*, 102-116.
- 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.
- Chen, F.F. (2007). Sensitivity of goodness of fit indexes 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*, 233-55.
- DeSalvo, K.B., Fisher, W.P., Tran, K., Bloser, N., Merrill, W., & Peabody, J. (2006). Assessing

- measurement properties of two single-items general health measures. *Quality of Life Research*, 15, 191-201.
- DiStephano, C., & Motl, R.W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling*, 13, 440–464.
- Eccles, J.S., Wigfield, A., Harold, R.D., & Blumenfeld, P. (1993). Age and gender differences in children's self- and task perceptions during elementary school. *Child Development*, 64, 830-847.
- Eccles, J.S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, 53, 109-132.
- Evans, E.M., Schweingruber, H., & Stevenson, H.W. (2002). Gender differences in interest and knowledge acquisition: The United States, Taiwan, and Japan. *Sex Roles*, 47, 153–167.
- Fan, X., & Sivo, S.A. (2009). Using goodness-of-fit indexes in assessing mean structure invariance. *Structural Equation Modeling*, 16, 54-69.
- Finney, S.J., & DiStephano, C. (2006). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R. O. Mueller (Eds), *Structural Equation Modeling: A Second Course* (pp. 269-314). Greenwich, CO: IAP.
- Finney, S.J., & DiStephano, C. (2013) Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R. O. Mueller (Eds), *Structural Equation Modeling: A Second Course, second edition* (pp. 439-492). Greenwich, CO: IAP.
- Forte, E.E., & Vispoel, W.P. (1995, April). Domain importance and involvement: Relations between domain self-concepts and general self-esteem in pre-adolescence. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.
- Fox, K.R., & Corbin, C.B. (1989). The physical self-perception profile: Development and preliminary validation. *Journal of Sport and Exercise Psychology*, 11, 408–430.
- Hardy, L., & Moriarty, T. (2006). Shaping self-concept: The elusive importance effect. *Journal of*

- Personality*, 74(2), 377–401.
- Hargreaves, D.J., Comber, C., & Colley, A. (1995). Effects of age, gender, and training on musical preferences of British secondary school students. *Journal of Research in Music Education*, 43, 242–250.
- Harrison, A.C., & O'Neill, S.A. (2000). Children's gender-typed preferences for musical instruments: An intervention study. *Psychology of Music*, 28, 81–97.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- 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.
- Maïano, C., Morin, A.J.S., Ninot, G., Monthuy-Blanc, J., Stephan, Y., Florent, J.-F., & Vallée, P. (2008). A short and very short form of the physical self-inventory for adolescents: Development and factor validity. *Psychology of Sport & Exercise*, 9, 830–847.
- Maidlow, S., & Bruce, R. (1999). The role of psychology research in understanding the sex/gender paradox in music – Plus ça change... *Psychology of Music*, 27, 147–58.
- Marsh, H.W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, 82, 623–636.
- Marsh, H.W. (1993). Academic self-concept: Theory, measurement and research. In J. Suls (Ed.), *Psychological perspectives on the self* (Vol. 4, pp. 59–98). Hillsdale, NJ: Erlbaum.
- Marsh, H.W. (2007a). *Self-concept theory, measurement and research into practice: The role of self-concept in Educational Psychology*. Leicester, UK: British Psychological Society.
- Marsh, H.W. (2007b). Application of confirmatory factor analysis and structural equation modeling in sport/exercise psychology. In G. Tenenbaum & R. C. Eklund (Eds.), *Handbook of on sport*

- psychology* (3rd ed., pp. 774 - 798). New York: Wiley.
- 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., & Craven, R.G. (2006). What comes first? A reciprocal effects model of the mutually reinforcing effects of academic self-concept and achievement. In H. W. Marsh, R. G. Craven, & D. M. McInerney, (Eds.) *International advances in self research* (pp 15–52). Greenwich, CT: Information Age.
- Marsh, H.W., Ellis, L., Parada, R., Richards, G., & Heubeck, B. (2005). A short version of the Self-Description Questionnaire II: Operationalizing criteria for short-form with new applications of confirmatory factor analyses. *Psychological Assessment, 17*, 81-102.
- Marsh, H.W., Hau, K., & Grayson, D. (2005). Goodness of Fit in Structural Equation Models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Contemporary psychometrics* (pp. 275-340). Mahwah, NJ: Erlbaum.
- Marsh, H.W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings. *Structural Equation Modeling, 11*, 320-341.
- Marsh, H.W., Richards, G.E., Johnson, S., Roche, L., & Tremayne, P. (1994). Physical self-description questionnaire: Psychometric properties and a multitrait-multimethod analysis of relations to existing instruments. *Journal of Sport and Exercise Psychology, 16*, 270–305.
- Marsh, H.W., & Scalas L.F. (2010). Self-concept in learning: Reciprocal effects model between academic self-concept and academic achievement. In P. Peterson, E. Baker, & B. McGaw, (Eds), *International encyclopedia of education*, vol. 6, pp. 660-667. Oxford: Elsevier.

- 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.
- Marsh, H.W., & Shavelson, R.J. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist, 20*, 107–125.
- McDonald, R.P. (1970). Theoretical foundations of principal factor analysis and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology, 23*, 1-21.
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58*, 525-543.
- Millsap, R.E. (2011). *Statistical approaches to measurement invariance*. New York: Routledge.
- Millsap, R.E., & Tein, J.-Y. (2004). Assessing factorial invariance in ordered-categorical measures. *Multivariate Behavioral Research, 39*, 479-515.
- Morin, A.J.S., Arens, A.K., & Marsh, H.W. (2015, In Press). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling*.
- Morin, A.J.S. & Maïano, C. (2011). Cross-validation of the very short form of the Physical Self-Inventory: Invariance across genders, age, ethnicities, and weight statuses. *Body Image, 8*, 404-410.
- Morin, A.J.S., Moullec, G., Maïano, C., Layet, L., Just, J.-L., & Ninot, G. (2011). Psychometric properties of the Center for Epidemiologic Studies Depression Scale (CES-D) in French clinical and non-clinical adults. *Epidemiology and Public Health, 59*, 327-340.
- Moullec, G., Maïano, C., Morin, A.J.S., Monthuy-Blanc, J., Rosello, L., & Ninot, G. (2011). A very short visual analog of the Center for Epidemiologic Studies Depression Scale for the idiographic measurement of depression. *Journal of Affective Disorders, 128*, 220-234.
- Muthén, B.O. (2004). *Mplus technical appendices*. Los Angeles, CA: Muthén & Muthén.

- Muthén, L.K., & Muthén, B.O. (1998–2012). *Mplus user's guide* (7th ed.). Los Angeles, CA: Author.
- Myers, M. G., McCarthy, D. M., MacPherson, L., & Brown, S. A. (2003). Constructing a short form of the smoking consequences questionnaire with adolescents and youth adults. *Psychological Assessment, 15*, 163-172.
- O'Neill, S.A., & Boulton, M.J. (1996). Boys' and girls' preferences for musical instruments: A function of gender? *Psychology of Music, 24*, 171–83.
- Quilty, L.C., Oakman, J.M., & Risko, E. (2006). Correlates of the Rosenberg Self-Esteem Scale method effects. *Structural Equation Modeling, 13*, 99–117.
- Raykov, T., & Grayson, D. (2003). A test for change of composite reliability in scale development. *Multivariate Behavioral Research, 38*, 143–159.
- Sandene, B.A. (1997). An investigation of variables related to student motivation in instrumental music. *Dissertation Abstracts International: Section A. Humanities and Social Sciences, 58*(10), 3870.
- Scalas, L.F., Marsh, H.W., Nagengast, B., & Morin, A.J.S. (2013). Latent variable approaches to the Jamesian model of importance-weighted averages. *Personality and Social Psychology Bulletin, 39*, 100-114.
- Scalas, L. F., Morin, A. J. S., Marsh, H. W., & Nagengast, B. (2014). Importance models of the physical self: Improved methodology supports a normative-cultural importance model but not the individual importance model. *European Journal of Social Psychology, 44*, 154-174.
- Schmidt, C.P. (2005). Relations among motivation, achievement, and music experience variables in secondary instrumental music students. *Journal of Research in Music Education, 53*, 134–147.
- Schnare, B., MacIntyre, P., & Doucette, J. (2012). Possible selves as a source of motivation for musicians. *Psychology of Music, 40*, 94-111.
- Shavelson, R.J., Hubner, J.J., & Stanton, G.C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407–441.

- Sijtsma, K. (2009). On the use, misuse, and the very limited usefulness of Cronbach's alpha [Introduction to a special issue]. *Psychometrika*, 74, 107-120.
- Simpkins, S.D., Vest, A.E., & Becnel, J.N. (2010). Participating in sport and music activities in adolescence: The role of activity participation and motivational beliefs during elementary school. *Journal of Youth and Adolescence*, 39, 1368-1386.
- Smith, G., McCarthy, D., & Anderson, K. (2000). On the sins of short-form development. *Psychological Assessment*, 12, 102-111.
- Tomás, J.M., & Oliver, A. (1999). Rosenberg's Self-Esteem Scale: Two factors or method effects. *Structural Equation Modeling*, 6, 84-98
- Tomás, J.M., & Oliver, A. Galiana, L., Sancho, P., & Lila, M. (2013). Explaining method effects associated with negatively worded items in trait and state global and domain-specific self-esteem scales. *Structural Equation Modeling*, 20, 299-313.
- Vispoel, W.P. (1993a). The development and validation of the Arts Self-Perception Inventory for adolescents. *Educational and Psychological Measurement*, 53, 1023-1033.
- Vispoel, W.P. (1993b). The Music Self-Perception Inventory (Adolescent Form). Iowa City, IA: Author.
- Vispoel, W.P. (1993c). The Music Self-Perception Inventory (Adult/College Form). Iowa City, IA: Author.
- Vispoel, W.P. (1994). Integrating self-perceptions of music skill into contemporary models of self-concept. *The Quarterly Journal of Music Teaching and Learning*, 5, 42-57.
- Vispoel, W.P. (1995). Self-concept in artistic domains: An extension of the Shavelson model. *Journal of Educational Psychology*, 85, 134-153.
- Vispoel, W.P. (2003). Measuring and understanding self-perceptions of musical ability. In H. W. Marsh, R. G. Craven & D. M. McInerney (Eds.), *International advances in self research* (Vol. 1, pp. 151-180). Greenwich, CT: Information Age.

- Vispoel, W.P., & Forte, E.E. (1994, April). Predicting global self-esteem from domain-specific self-concepts: The role of importance and involvement. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- Vispoel, W.P., & Forte, E.E. (2000). Response bias and their relation to sex differences in multiple domains of self-concept. *Applied Measurement in Education*, 13, 79-97.
- Vispoel, W.P., & Rizzo, K.M. (2003, April). *Sex differences in early adolescent music students' self-concepts*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- West, C. (2013). Motivating music students: A review of the literature. *Update*, 31, 11–19.
- Wigfield, A., Eccles, J.S., Yoon, K.S., Harold, R.D., Arbretton, A.J.A., Freedman-Doan, C., et al. (1997). Change in children's competence beliefs and subjective task values across the elementary school years: A 3-year study. *Journal of Educational Psychology*, 89, 451–469.
- Wigfield, A., & Eccles, J.S. (2000). Expectancy-value theory of achievement motivation. *Contemporary Educational Psychology*, 25, 68–81.
- Yu, C.Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes*. Los Angeles, CA: University of California.

Table 1. Goodness-of-Fit indices of the Alternative Models

	χ^2	df	CFI	TLI	RMSEA	CI 90%	$MD\Delta\chi^2$ (df)	Δ CFI	Δ TLI	Δ RMSEA
<i>Sample 1 (N = 304)</i>										
Long Version (84 items)	4800*	3381	.968	.967	.037	.035-.040				
Long Version (84 items), MF + CUs	4026*	3087	.979	.976	.032	.029-.034	1109(294)*	+.011	+.009	+.005
<i>Sample 1 (N = 304)</i>										
Short Version (28 items)	516*	329	.990	.989	.043	.036-.050				
Short Version (28 items), CUs	431*	267	.991	.988	.045	.037-.053	124 (62)*	+.001	-.001	+.002
<i>Sample 2 (N = 208)</i>										
Short Version (28 items)	566*	329	.979	.976	.059	.051-.067				
Short Version (28 items), CUs	483*	267	.981	.973	.062	.053-.071	126 (62)*	+.002	-.003	+.003
<i>Invariance Across Samples (N = 304 and 208)</i>										
Configural invariance	912*	534	.987	.982	.053	.047-.058				
Weak invariance (Loadings)	920*	555	.988	.983	.051	.045-.057	21 (21)	+.001	+.001	-.002
Strong invariance (Loadings, Thresholds)	1001*	660	.989	.987	.045	.039-.051	103 (105)	+.001	+.004	-.006
Strict invariance (Loadings, Thresholds, Uniq.)	1018*	688	.989	.988	.043	.038-.049	50 (28)*	.000	+.001	-.002
Invariance of Latent Variances-Covariances	880*	714	.994	.994	.030	.023-.037	43 (26)	+.005	+.006	-.013
Invariance of Latent Means	979*	721	.991	.991	.037	.031-.043	34 (7)*	-.003	-.003	+.007
<i>Invariance Across Gender (N = 200 and 310)</i>										
Configural invariance	901*	534	.988	.983	.052	.046-.058				
Weak invariance (Loadings)	921*	555	.988	.984	.051	.045-.057	34 (21)	.000	+.001	-.001
Strong invariance (Loadings, Thresholds)	1019*	660	.988	.987	.046	.041-.052	126 (105)	.000	+.003	-.005
Strict invariance (Loadings, Thresholds, Uniq.)	1257*	688	.981	.980	.057	.052-.062	187 (28)*	-.007	-.007	+.011
Invariance of Latent Variances-Covariances	925*	714	.993	.993	.034	.027-.040	36 (26)	+.012	+.013	-.023
Invariance of Latent Means	1201*	721	.984	.984	.051	.046-.056	77 (7)*	-.009	-.009	+.017
<i>Invariance Across Grade Level (N = 327 and 183)</i>										
Configural invariance	855*	534	.990	.985	.049	.042-.055				
Weak invariance (Loadings)	858*	555	.990	.987	.046	.040-.052	20 (21)	.000	+.002	-.003
Strong invariance (Loadings, Thresholds)	942*	660	.991	.990	.041	.035-.047	102 (105)	.001	+.003	-.005
Strict invariance (Loadings, Thresholds, Uniq.)	982*	688	.990	.990	.041	.035-.047	64 (28)*	-.001	.000	.000
Invariance of Latent Variances-Covariances	864*	714	.995	.995	.029	.021-.035	39 (26)	+.005	+.005	-.012
Invariance of Latent Means	953*	721	.993	.992	.035	.029-.041	32 (7)*	-.002	-.003	+.006

Note. N= sample size; χ^2 = WLSMV chi square test; df = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CI = confidence interval; Δ change; $MD\Delta\chi^2$: chi square difference test (Mplus DIFFTEST); MF = method factor; CUs = correlated uniquenesses; * $p \leq .01$.

Table 2. Parameter Estimates for the Measurement Models

Items	Sample 1, Long Version		Sample 1, Short Version		Sample 2, Short Version	
	λ	δ	λ	δ	λ	δ
Composing						
4 ^N	.73	.47				
11	.82	.33				
18 ^N	.85	.23				
25 ^S	.86	.27	.82	.34	.82	.32
32 ^N	.82	.21				
39 ^S	.85	.27	.84	.29	.87	.24
46 ^N	.80	.32				
53	.79	.37				
60 ^N	.76	.37				
67 ^S	.87	.25	.85	.27	.88	.22
74 ^N	.82	.27				
81 ^S	.85	.28	.86	.26	.83	.31
Reliability (ω)	.96		.91		.92	
Listening						
5	.76	.43				
12 ^N	.74	.42				
19 ^S	.84	.29	.83	.31	.84	.29
26 ^N	.83	.25				
33 ^S	.87	.24	.97	.24	.91	.18
40 ^N	.83	.20				
47 ^S	.91	.18	.91	.17	.91	.17
54 ^N	.77	.35				
61 ^S	.86	.27	.83	.32	.87	.24
68 ^N	.74	.34				
75	.76	.42				
82 ^N	.75	.33				
Reliability (ω)	.96		.92		.93	
Dancing						
6 ^N	.85	.25				
13 ^S	.88	.23	.86	.27	.86	.25
20 ^N	.78	.25				
27	.77	.41				
34 ^N	.84	.24				
41 ^S	.89	.21	.87	.24	.90	.20
48 ^N	.75	.39				
55	.79	.38				
62 ^N	.83	.26				
69 ^S	.85	.28	.85	.28	.92	.15
76 ^N	.78	.30				
83 ^S	.87	.25	.89	.20	.88	.23
Reliability (ω)	.97		.92		.94	
Instrument Playing						
2 ^N	.80	.35				
9 ^S	.90	.19	.89	.21	.91	.17
16 ^N	.87	.16				
23 ^S	.89	.20	.89	.20	.92	.15
30 ^N	.88	.19				
37 ^S	.94	.13	.94	.11	.92	.15
44 ^N	.79	.23				
51	.79	.38				
58 ^N	.84	.22				
65	.88	.22				
72 ^N	.78	.36				
79 ^S	.91	.18	.88	.22	.85	.29
Reliability (ω)	.97		.95		.95	

Table 2 (continued)

Items	Sample 1, Long Version		Sample 1, Short Version		Sample 2, Short Version	
	λ	δ	λ	δ	λ	δ
Reading						
3	.79	.38				
10 ^N	.85	.25				
17	.86	.26				
24 ^N	.80	.28				
31^S	.90	.20	.89	.21	.91	.17
38 ^N	.86	.19				
45^S	.90	.18	.90	.19	.88	.23
52 ^N	.73	.31				
59^S	.89	.21	.89	.21	.90	.20
66 ^N	.86	.21				
73^S	.89	.21	.89	.21	.90	.20
80 ^N	.78	.26				
Reliability (ω)	.97		.94		.94	
Singing						
1	.87	.25				
8 ^N	.76	.35				
15^S	.87	.24	.87	.24	.88	.23
22 ^N	.81	.24				
29^S	.89	.22	.88	.23	.92	.16
36 ^N	.81	.22				
43^S	.90	.18	.89	.21	.95	.11
50 ^N	.81	.27				
57^S	.94	.12	.95	.09	.92	.15
64 ^N	.82	.29				
71	.85	.27				
78 ^N	.83	.26				
Reliability (ω)	.97		.94		.95	
Global						
7^S	.85	.29	.83	.31	.83	.31
14 ^N	.79	.31				
21^S	.86	.27	.85	.28	.89	.21
28 ^N	.80	.30				
35^S	.90	.19	.89	.21	.87	.25
42 ^N	.79	.29				
49^S	.90	.20	.88	.22	.84	.29
56 ^N	.66	.43				
63	.76	.43				
70 ^N	.78	.26				
77	.83	.32				
84 ^N	.72	.31				
Reliability (ω)	.96		.92		.92	

Note. λ = Standardized factor loading; δ = Standardized item uniqueness; ^N = Negatively-worded items; ^S: items retained in the short form (bold); ω = omega composite reliability coefficient; All factors loadings and uniquenesses significant at $p \leq .01$; Items with parallel wording are: (a) Good (1,13,25,37,49,61,73); (b) Difficult (2,14,26,38,50,62,74); (c) I am better (3,15,27,39,51,63,75); (d) Hopeless (4,16,28,40,52,64,76); (e) Do well (5,17,29,41,53,65,77); (f) Trouble (6,18,30,42,54,66,78); (g) Skilled (7,19,31,43,55,67,79); (h) Harder (8,20,32,44,56,68,80); (i) Confident (9,21,33,45,57,69,81); (j) Never good (10,22,34,46,58,70,82); (k) Easy (11,23,35,47,59,71,83); (l) Others are better (12,24,36,48,60,72,84).

Table 3. Latent Factor Correlations (Sample 1)

	Composing	Listening	Dancing	Instrument Playing	Reading	Singing	Global
Composing		.75	.48	.68	.68	.48	.76
Listening	.79		.28	.72	.81	.50	.87
Dancing	.50	.29		.34	.27	.48	.40
Instrument Playing	.70	.74	.31		.87	.30	.87
Reading	.71	.83	.25	.86		.43	.91
Singing	.53	.54	.47	.29	.43		.57
Global	.80	.90	.37	.89	.93	.56	

Note. Factor correlations for the long version above the diagonal. Factor correlations for the short version under the diagonal. All correlations significant at $p \leq .01$.

Table 4. Latent Factor Correlations (Sample 2)

	Composing	Listening	Dancing	Instrument Playing	Reading	Singing
Composing						
Listening	.66					
Dancing	.29	.23				
Instrument Playing	.61	.62	.05			
Reading	.56	.65	-.05	.82		
Singing	.46	.44	.47	.27	.28	
Global	.64	.73	.29	.83	.78	.55

Note. Most correlations are significant at $p \leq .01$; Non-significant correlations are italicized.

Table 5. Correlations with External Criteria for the Long and Short Versions (Overall Sample)

	Composing		Listening		Dancing		Instrument Playing		Reading		Singing		Global	
	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short	Long	Short
Grades														
English	.19**	.21**	.22**	.23**	.07	.07	.29**	.29**	.28**	.28**	.16**	.17**	.32**	.33**
Math	.04	.07	.05	.06	.03	.03	.21**	.19**	.18**	.18**	.01	.01	.17**	.18**
Physical education	.13*	.12*	.09*	.11*	.06	.05	.18**	.18**	.14**	.14**	.09*	.09*	.18**	.16**
Dance	.14**	.14**	.12*	.12*	.33**	.33**	.11*	.10*	.05	.05	.18**	.18**	.14**	.16**
Music	.23**	.25**	.30**	.32**	.07	.08	.28**	.28**	.30**	.29**	.18**	.17**	.34**	.36**
Art	.15**	.15**	.11*	.11**	.13**	.12**	.15**	.15**	.11**	.11**	.14**	.16**	.16**	.15**
Drama	.16**	.17**	.11*	.12*	.16**	.17**	.09	.09	.06	.08	.22**	.22**	.12*	.14**
Past Experiences														
Music group	.31**	.30**	.34**	.31**	.16**	.16**	.41**	.41**	.45**	.44**	.24**	.23**	.42**	.38**
Acting	.13**	.14**	.08	.09	.09	.08	.10*	.10*	.08	.08	.17**	.17**	.13**	.13**
Dance group	.12**	.12*	.11*	.10*	.37**	.37**	.14**	.13**	.13**	.12**	.24**	.24**	.18**	.17**
Art works	.26**	.24**	.19**	.19**	.16**	.15**	.23**	.22**	.19**	.20**	.09	.07	.17**	.15**
Years	.43**	.43**	.45**	.42**	.13**	.11*	.67**	.65**	.57**	.55**	.20**	.19**	.57**	.55**
Self-Perceived Abilities														
Singing	.35**	.37**	.36**	.36**	.37**	.38**	.20**	.20**	.27**	.24**	.80**	.81**	.40**	.41**
Instrument Playing	.53**	.52**	.53**	.52**	.19**	.16**	.80**	.79**	.68**	.67**	.24**	.22**	.67**	.65**
Reading Music	.49**	.48**	.56**	.55**	.12**	.11*	.65**	.64**	.78**	.77**	.26**	.25**	.63**	.63**
Making Up Music	.76**	.77**	.53**	.54**	.27**	.28**	.46**	.46**	.44**	.44**	.32**	.32**	.51**	.51**
Identifying Music by Ear	.56**	.54**	.75**	.73**	.17**	.15**	.50**	.48**	.54**	.53**	.35**	.34**	.59**	.57**
Creating Dance Movements	.30**	.29**	.16**	.16**	.79**	.81**	.17**	.13**	.12**	.10*	.31**	.32**	.24**	.22**
Music-Related Activities	.55**	.54**	.55**	.55**	.31**	.30**	.65**	.63**	.64**	.62**	.41**	.41**	.75**	.74**

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