Received: 27 February 2023 | Accepted: 2 August 2023

DOI: 10.1111/bjep.12635

ARTICLE

the british psychological society
promoting excellence in psychology

Revisiting the academic self-concept transcultural measurement model: The case of Spain and China

Igor Esnaola[1] | **Albert Sesé**[2] | **Lorea Azpiazu**[1] | **Yina Wang**[3]

[1]University of the Basque Country (UPV/EHU), Bilbao, Spain

[2]University of Balearic Islands, Palma, Spain

[3]Myda Educational Consulting, Kingston, Canada

Correspondence
Igor Esnaola, Faculty of Education, Philosophy and Anthropology, University of the Basque Country, Avenida de Tolosa 70, 20018 San Sebastian, Spain.
Email: igor.esnaola@ehu.eus

Abstract

Background: Modelling academic self-concept through second-order factors or bifactor structures is an important issue with substantive and practical implications; besides, the bifactor model has not been analysed with a Chinese sample and cross-cultural studies in the academic self-concept are scarce. Likewise, latent structure validity evidence using network psychometrics has not been carried out.

Aims: The aim of this study is twofold: to analyse (1) the internal structure of ASC through the Self-Description Questionnaire II-Short (SDQII-S) in Chinese and Spanish samples using two approaches, structural equation modelling and network psychometrics conducting an exploratory graph analysis; and (2) the measurement invariance of the best model across countries and investigate the cross-cultural differences in ASC.

Sample: The sample was composed by 651 adolescents. Seven models of ASC were tested.

Results: Results supported the multi-dimensional nature of the data as well as the reliability. The best-fitted model for the two subsamples was the three-factor ESEM model, but only the configural invariance of this model was supported across countries. The graph function shows that the *school* dimension appears more related to the *verbal* factor in the Spanish subsample and to the *math* dimension in the Chinese subsample. Likewise, the relationship between *verbal* and *math* factors in Spanish students is non-existent, but this connection is more relevant for Chinese students.

Conclusion: These two differences may be behind the difficulty in finding invariance using SEM models. It is a question of the construct's nature, less related to analytical phenomena, and deserves deeper discussion.

KEYWORDS
academic self-concept, cross-cultural, measurement invariance, network psychometrics, structural equation modelling

INTRODUCTION

Shavelson et al. (1976) developed a complete and unified self-concept theory, in which a general facet at the apex of the self-concept hierarchy is divided into *academic* and *non-academic* components of the self-concept. *Non-academic* self-concept is divided into social, emotional, and physical self-concepts; and *academic* self-concept into self-concepts in particular subject areas (Mathematics, English, etc.).

Academic self-concept (ASC) is an important domain of general self-concept and refers to learners' knowledge and perceptions about themselves in an overall academic domain (Wigfield & Karpathian, 1991). Students' ASC has received considerable attention in educational research (Arens et al., 2011) due to its relationship to desirable educational outcomes, such as interest, persistence, coursework selection, and academic achievement (Parker et al., 2014). Despite the relevance of this psychological construct, it has proven a particular challenge to find a structural model of ASC capable of accounting for domain specificity, hierarchy, and/or relations between mathematics and verbal self-concepts.

Academic self-concept models

The academic section of the Shavelson et al.'s (1976) model (Figure 1a) was modified into the "Marsh/Shavelson model" (Marsh, 1990; Figure 1b), which distinguished between general mathematics and general verbal self-concepts, both represented as higher-order factors that influence ASCs in related domains and subjects. The reason for this proposal was that a structural model with one higher-order factor representing general ASC failed to explain the pattern of intercorrelations among the domain-specific ASCs. Moreover, general ASC, represented as a first-order factor, was assumed to be simultaneously influenced by mathematics and verbal self-concepts, although they were either uncorrelated or showed only a modest positive correlation (Marsh, 1990). In order to test the multi-dimensionality of ASC, Marsh's group found empirical support for a first-order factor model (Figure 1c) and showed that it was reasonably invariant across 25 countries (Marsh et al., 2006) based on PISA database, but they did not compare several models. Therefore, although this model fitted the data adequately, it should be evaluated whether it is the most plausible model of those available in the literature. Likewise, Spain and China were not included. Although the first-order factor model has been extensively studied, the current research aims to go further and analyse a first-order factor model that allows cross-loadings estimation (ESEM 3F, Figure 1d) for the first time.

More recently, Brunner et al. (2008) proposed a *nested-factor model* (Figure 1e) based on bifactor model theory. The model is an incomplete bifactor model where general ASC is assumed to directly influence domain general and domain-specific measures of ASCs. In addition to general ASC, a specific mathematics self-concept and a specific verbal self-concept are distinguished to represent the multi-faceted nature of ASC. General ASC is uncorrelated with mathematic and verbal self-concepts, and the estimated correlation parameter between domain-specific factors is negative, based on students tending to think of themselves either as a "mathematic" or "verbal" person (Marsh & Hau, 2004). The generalizability of this model was supported by Brunner et al.'s (2009) cross-cultural research across 26 countries (China and Spain were not included) and by Esnaola et al. (2018) in a Spanish sample. In the case of Brunner et al. (2009), two models of ASC were compared, the first-order factor model (without cross-loadings estimations, Figure 1c) and the incomplete bifactor model (Figure 1e). On the other hand, Esnaola et al. (2018) analysed three models: (1) a second-order model, with two first-order factors (verbal and math) and general school factor as a second-order factor (Figure 1a); (2) a second-order factor

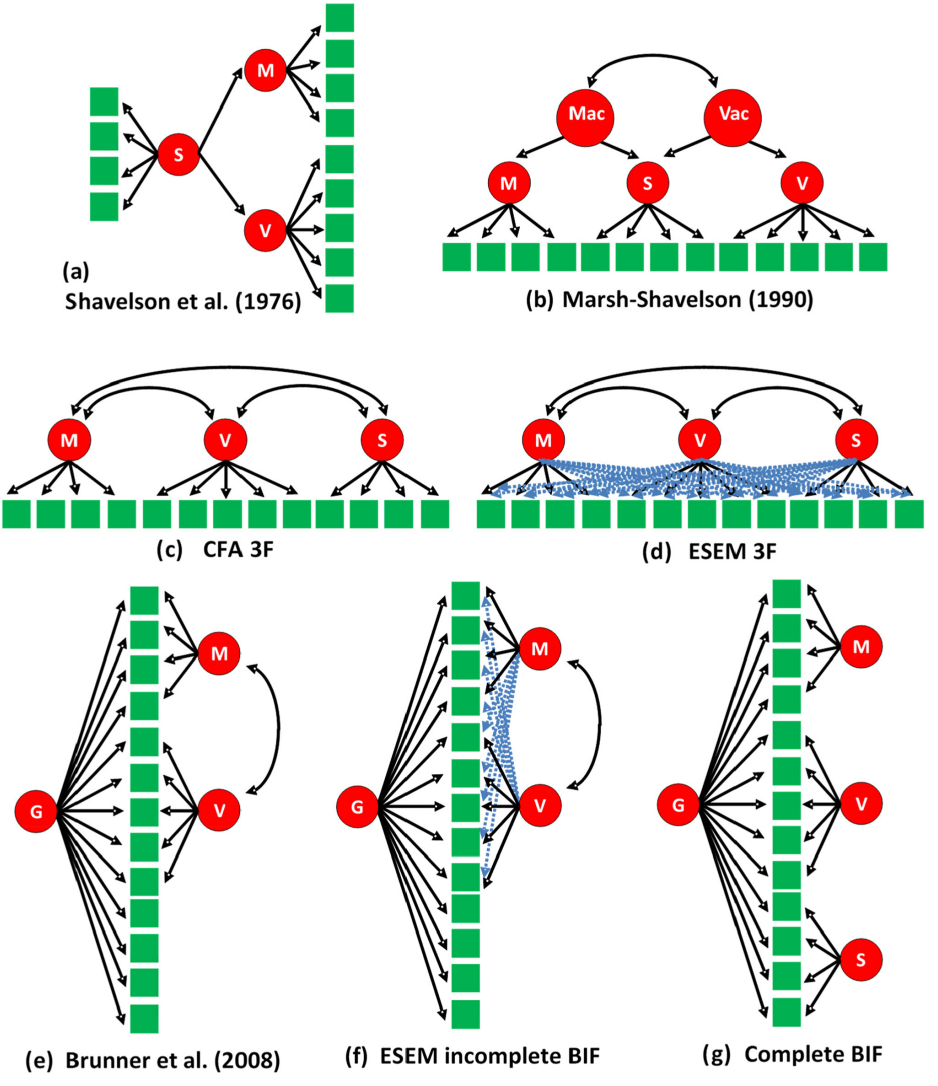

FIGURE 1 The seven models compared: (a) The Shavelson et al.'s (1976) second-order factor model (M = *math* and V = *verbal* as first-order factors and S = *general school* as a second-order factor); (b) Marsh/Shavelson model (Marsh, 1990), three-factor CFA model with two second-order factors (Mac = general mathematic self-concept; Vac = general verbal self-concept) and three first-order factors (*math*, *verbal*, and *general school*); (c) three-factor confirmatory model (CFA 3F); (d) three-factor model that allows the cross-loadings estimation (ESEM 3F); (e) Brunner et al.'s (2008) model, incomplete bifactor model with two specific and correlated factors (*math* and *verbal*) and general academic self-concept (G) as the general factor (incomplete BIF); (f) the (e) model but allowing cross-loadings between specific factors (ESEM incomplete BIF); and (g) complete bifactor model with three specific and uncorrelated factors and G as the general factor (complete BIF).

model with three first-order factors (verbal, math, and general school) and two second-order factors (general verbal self-concept and general mathematics self-concept; Figure 1b); and (3) the incomplete bifactor model (Figure 1e).

Taking into account that Brunner et al.'s (2008) model is an incomplete bifactor model, in this paper, two more models will be analysed: an ESEM incomplete bifactor model allowing cross-loadings between specific factors (Figure 1f) and a complete bifactor model with three specific and uncorrelated factors, and general academic self-concept as the general factor (complete BIF, Figure 1g).

Cross-cultural research on academic self-concept

Although self-concept has been broadly studied, cross-cultural differences have not been commonly analysed. Markus and Kitayama (1991) expressed that "people in different cultures have strikingly different construal of the self, of others, and of the interdependence of the two. This construal can determine the very nature of individual experience, including cognition, emotion and motivation" (p. 224). Therefore, culture might determine, at least in part, self-representation in terms of collectivism and individualism (Chen et al., 2004; Markus & Kitayama, 1991). It is argued that self-conceptions in individualistic societies are predominantly abstract and decontextualized, while in collectivistic societies, they tend to be more influenced by the social context (Shweder & Bourne, 1982).

Some authors have claimed that individuals from Western societies have greater self-concept clarity than their counterparts from Eastern societies (Campbell et al., 1996). Studies have shown that students of collectivistic cultures, like China, get lower scores on *non-academic* areas of self-concept, but stand out in *academic* ones when compared with students from individualistic cultures, such as Australians and Americans (Wästlund et al., 2001). In adolescence, Japanese students had a lower self-concept than Swedish students in *verbal* ($\eta^2 = .21$) and *general school* ($\eta^2 = .42$), but non-significance differences were found in *mathematic* self-concept (Nishikawa et al., 2007). Likewise, Inglés et al. (2009) showed that Spanish and Portuguese students had higher scores than Chinese students in *verbal* ($d = .25$ for Spain and $d = .20$ for Portugal) and *general school* ($d = .15$ for Spain and $d = .25$ for Portugal), while Chinese students presented higher scores in *math* than Spanish ($d = .47$) and Portuguese ($d = .31$) students. On the other hand, Portuguese students had higher scores than Spanish ones in *math* ($d = .14$), but as can be seen most effect sizes were small. However, it is important to emphasize that these studies did not analyse the measurement invariance of the questionnaire before comparing the means, so the results should be used with caution because it is a prerequisite to comparing group means (Putnick & Bornstein, 2016).

Nevertheless, there are some studies that have analysed and confirmed the measurement invariance. Leung et al. (2016) examined the responses of Australian and Chinese high school students finding small but systematic differences in favour of Australians compared with their Chinese counterparts (*math* = −.54, *verbal* = −.67, and *general school* = −.87). These authors pinpointed that those results could be explained by taking into account that collectivism and humbleness are emphasized in Chinese culture under the influence of Confucian values. In contrast, individualism and values of outperforming others are emphasized in Western culture (Kurman, 2003; Li et al., 2006). Hence, Chinese students may downgrade themselves in relation to others in comparison with American students (Kurman, 2003), and consequently, group means from these two countries (Australia and China) could not be compared to generate meaningful interpretations. Marsh et al. (2013) also revealed that students from Arab countries (Saudi Arabia, Jordan, Oman, and Egypt) showed higher self-concepts than students from Anglo countries (the United States, Australia, England, and Scotland), since Arab students have been socialized to be less critical of themselves in school and family (Abu-Hilal, 2001), and thus, they have a higher perception of themselves than students from Anglo countries. Considering that the cultural norms regarding modesty and self-assertion are different, Marsh et al. (2006) also argued that it may not be appropriate to analyse the mean-level differences across cultures, since cultural differences in modesty and self-enhancement might affect the mean responses.

Therefore, an important question is whether the structure of self-concept is the same or different according to cultural context. It seems that individuals from Anglo-Saxon and European countries perceive different but related domains of themselves (multi-dimensional self-concept), although these main findings should be tested more exhaustively in other cultural contexts. Specifically, China has received special attention due to its particular cultural traits based on Confucianism and the pursuit of the supreme Chinese virtue, *jen* (Chen et al., 2004). In this regard, Chen et al. (2020) found that despite some cross-cultural variations in self-concept among individuals from Western and non-Western societies, both European (Spanish) and Eastern (Chinese) individuals perceived themselves by a multi-dimensional related structure of self-concept. On the other hand, English and Chen (2007) found that Asian Americans were less consistent in their self-descriptions across relationships contexts than were

European Americans, but Asian Americans' self-descriptions showed high consistency within this context over time. It seems that, whereas Westerners tend to emphasize global conceptions of the self-concept that remain consistent across contexts, East Asians tend to elaborate context-specific selves. However, more research in Eastern countries is clearly needed to clear up all these issues.

The present study

As shown above, the conceptual and methodological differences regarding the second-order confirmatory model and the bifactor model of ASC have generated considerable debate due to its important practical applications: score interpretation cannot be separated from its internal structure and bifactor models are particularly useful when working with domain-specific subscores (Reise et al., 2010). The latest findings about bifactor models in the various psychological areas show a better fit for this type of models compared to second-order confirmatory models (Rodríguez et al., 2016). Thus, from the set of most representative ASC models, we expect the bifactor with specific math and verbal domains to be the model that will best fit the data (hypothesis 1). The valuable contribution generated by Brunner et al.'s (2009) nested model in the understanding of ASC structure is undeniable. However, instead of using a full questionnaire, the three best items from each of the 10-item self-concept scales for verbal, math, and ASC were selected (Marsh & O'Niell, 1984), which was precisely the measure conducted in the PISA study (Organisation for Economic Co-operation and Development [OECD], 2001). Likewise, even if the bifactor model was supported across 26 countries (Brunner et al., 2009) and in Spain (Esnaola et al., 2018), it has not been analysed in a Chinese sample. Finally, the measurement invariance of the bifactor model was not supported by Brunner et al. (2009) in the 26 countries analysed.

Therefore, it is intended to investigate more precisely the ASC structure by analysing the measurement invariance of the best model between countries. In fact, although there are some cross-cultural studies on self-esteem, there appears to be few cross-cultural studies on ASC. This is a key issue because if the measurement invariance of the questionnaire is not supported, mean comparisons are not allowed. It is mandatory to test the extent to which both the instrument and the psychological meaning and structure of its underlying constructs are group equivalent (Putnick & Bornstein, 2016). Hence, it is a relevant goal to analyse whether the structure of ASC is invariant across countries and to compare the differences in latent means. Taking into account that previous studies have used structural equation modelling (SEM) to analyse the measurement invariance, in this study we want to go further and use also a new approach, network psychometrics (NP) conducting an exploratory graph analysis (EGA). To the best of our knowledge, this is the first study that has analysed the internal structure of ASC with this approach. Indeed, there is empirical evidence that EGA approach is superior to traditional dimension reduction techniques (Golino et al., 2021). According to previous research, we do not expect invariance across countries in the structure of ASC (hypothesis 2).

Therefore, the aim of this study is twofold: to analyse (1) the internal structure of ASC through the SDQII-short questionnaire (the most validated self-concept measure available for use with adolescents) in a Spanish and a Chinese sample using two approaches, SEM and NP, conducting an EGA; and (2) the measurement invariance of the best model across countries and investigate the cross-cultural differences in ASC.

METHOD

Participants

The sampling process was carried out on the basis of convenience or incidental sampling (in accordance with schools' accessibility), and the sample was composed of 651 adolescents from middle-class families who attended urban schools (public in China and public and semi-private in Spain). Participants where from two countries, Spain [$n = 262$, 118 males (45%) and 144 females (55%), $M_{age} = 16.50$, SD = .87, range 15 to 18] and China [$n = 389$, 160 males (41.1%) and 229 females (58.9%), $M_{age} = 16.55$, SD = .64, range 15.10 to 17.92]. No statistical differences were found by gender ($\chi^2 = .82$, $p = .364$) and age ($t_{446.71} = .94$, $p = .349$) between the two subsamples. Likewise, in both countries, the amount of time of math classes/per week is the same time as language classes; so, although it can be believed that in China schools put stronger emphasis in math, both subjects are two of significant subjects in the biggest examination, National Higher Education Entrance Examination. On the other hand, the student sample can represent the average student population because in both countries, participants are from a medium-size city and middle-class families. To sum up, this information may indicate with some certainty that both samples are comparable.

Measure

Self-Description Questionnaire II-Short (SDQII-S) (Marsh et al., 2005). A shortened version (51 items) of the original 102-item SDQII (Marsh, 1992; Chinese translation: Hau et al., 2003; Kong, 2000; Spanish translation: Inglés et al., 2012) was used. This questionnaire is designed to assess adolescent's self-concept and retains all eleven factors of SDQII with items scored on a 6-point Likert scale (1 = *False*, 6 = *True*). As the aim of this research involves the ASC, only three scales were used, *math* (four items), *verbal* (five items), and *general school* (four items). Reliability indices will be shown later as a product of the instrument's latent structure analytic strategy.

Procedure

After obtaining Ethical permission from the Committee on Ethics of Research and Teaching (CEID) from the University of the Basque Country (UPV/EHU) to conduct the study, schools' participation was requested. Subsequently, the parents were asked for written informed consent to authorize their children to participate. The questionnaires were voluntary and anonymously completed during school hours. Research assistants were at classes during questionnaires administration to provide help to participants.

Data analysis

A descriptive data analysis was previously performed to assess the quality of data and to check the fulfilment of univariate (Anderson-Darling test) and multivariate normality (Mardia's skewness and kurtosis coefficients) as statistical assumptions. There were no missing values and, consequently, no imputation methods were implemented. The first step was to estimate a one-factor confirmatory model for all ASC dimensions (*math*, *verbal*, and *general school*) before the models' comparison process on the latent structure of the construct. In this way, we tried identifying possible items with inadequate psychometric functioning within each of the three factors and considering the two subsamples. After that, we used a model comparison approach to test the fit of a set of seven latent models with the items of the SDQII-S (Figure 1): (a) the Shavelson et al.'s (1976) second-order factor model (*math* and *verbal* as first-order factors and *general school* as a second-order factor); (b) Marsh/Shavelson model (Marsh, 1990), three-factor CFA model with two second-order factors (Mac = general mathematic self-concept; Vac = general verbal self-concept) and three first-order factors (*math*, *verbal*, and *general school*); (c) three-factor confirmatory model (CFA 3F); (d) three-factor model that allows the cross-loadings estimation (exploratory structural equation modelling, ESEM 3F); (e) Brunner et al.'s (2008) model, incomplete bifactor model with two specific and correlated factors (*math* and *verbal*; incomplete BIF); (f) the e model but allowing cross-loadings between specific factors (ESEM incomplete BIF); and (g) complete bifactor model with three specific and uncorrelated factors (complete BIF). Once the seven models were estimated, we selected the one that obtained a better fit for each subsample.

After selecting the best-fitted model, we implemented a complete invariance study, including partial invariance, through the Spanish and Chinese subsamples. All SEM models were estimated using the maximum likelihood method with robust standard errors and a mean-adjusted chi-square statistic (MLM), given that the items have six-point graded scale and deviations from normality were found (Viladrich et al., 2017). The direct Schmid–Leiman method with theta parameterization and target rotation for bifactor models was also used, which has been recently identified as one of the best methods to recover non-hierarchical and hierarchical bifactor structures (Giordano & Waller, 2019). All analyses were performed using R program (R Core Team, 2022) and the packages 'Lavaan' (Rosseel, 2012) and 'semTools' (Jorgensen et al., 2022).

With regard to the overall fit indices used, the recommendations of Hu and Bentler (1999) were followed. Thus, a non-significant chi-square, a Comparative Fit Index (CFI) and a Tucker–Lewis Index (TLI) $\geq .95$, the root mean square error of approximation (RMSEA) $\leq .05$ and its 90% confidence interval, and the standardized root mean squared residuals (SRMR) $\leq .08$ are indicative of adequate fit. The power *post-hoc* analysis for RMSEA index was calculated using the 'semPower' R package (Moshagen & Erdfelder, 2016). This procedure estimates how large the probability (power) of falsifying a model is if it is wrong, at least to the extent defined by the chosen effect (in this case, we consider .50 as the minimum value/effect for a loading factor). In general, a power around or above .80 is adequate (Cohen, 1992).

The models' fit comparison takes into account whether or not they are nested or non-nested models. In general, nested models are those for which the parameter space associated with one model is a subset of the parameter space associated with the other model; every set of parameters from the less complex model can be translated into an equivalent set of parameters from the more complex model (Merkle et al., 2016). In this sense, the seven competing models of this study can be considered nested since they all include the same set of observed variables (the same 13 items); what differs between models is the number of latent variables. According to this scenario, we applied the Satorra–Bentler scaled chi-square test (Satorra & Bentler, 2010). Regarding invariance, also a CFI cut-off criterion is considered: a model comparison is non-significant when the ΔCFI value is equal to or less than .01 (Cheung & Rensvold, 2002).

Once the model comparison and the selection of the model with the best fit were carried out, at the factor level, we computed the Cronbach's alpha (α) and Omega coefficient (ω) as reliability indicators of the latent structure.

To obtain more latent structure validity evidence from another perspective, we used NP and conducted an EGA (Golino & Epskamp, 2016) using the package 'EGAnet' (Hudson & Christensen, 2022). There is empirical evidence that EGA approach is superior to traditional dimension reduction techniques such as exploratory factor analysis (EFA) and even simple or complex confirmatory factor analysis (CFA; Golino et al., 2021). We implemented the Gaussian graphical model using graphical LASSO with extended Bayesian information criterion ("glasso") with the "walktrap" algorithm to select optimal regularization parameter for the network estimation. We evaluated the fit of the EGA model using the total correlation of the dataset which high values indicate high interdependence and the Entropy Fit Index (EFI) with the Miller-Madow correction. The model with a lower EFI value is indicative of best fit. Unlike the traditional SEM indices (RMSEA, CFI, TLI, and SRMR), we cannot use EFI to assess the absolute fit of a model but rather as a measure of relative fit between models (Golino et al., 2021). Also, the small-world statistic serves as indicator of the network as random (values near 1), lattice (values near −1), and small-world (values between −.5 and .5). Small-world networks have two primary characteristics: a short average path length between nodes (items) and high clustering. If the items of the network are fully connected, the clustering coefficient is 1 and a value close to 0 means that there are hardly any connections in the network.

Following the network inference analysis, we conducted a non-parametric bootstrap confidence interval ($n = 1000$), with the estimation of the correlation stability coefficient, and bootstrapped difference tests to study the significance of the estimates of the EGA glasso network obtained using 'bootnet' R package for each subsample (Epskamp et al., 2018). The correlation stability (CS) coefficient provides a measure to determine whether centrality indices vary significantly between each incremental case-dropping subset, with CS >.5 indicating appropriate stability. Structural consistency was computed along with item stability under reliability analysis instead of internal consistency using the *EGAnet* package. A graphical comparison between the EGA models estimated for the two subsamples is composed and displayed for visual inspection of the relationships among all items and latent factors. Finally, we tested the metric invariance (equal loadings) of the EGA structure obtained across the two subsamples. Checking more restrictive invariance levels is not yet available using NP. Finally, the non-invariant parameters detected by the two approaches were compared. The correlation matrix of the items is displayed in Table 1 to facilitate secondary analyses. Data from this study are publicly shared at https://figshare.com/articles/dataset/ASC_invariance_SPAIN_CHINA_dataset/23742129.

RESULTS

Preliminary analysis and model comparison approach

Table 1 includes the correlation matrix, the main univariate descriptive statistics of the items of the three dimensions of ASC, plus the Anderson-Darling normality test. None of the items fitted the normal distribution in the Spanish or Chinese subsamples ($p < .001$). We rejected the hypothesis for multivariate normality using Mardia skewness and kurtosis coefficients: 995.74 ($p < .001$) and 16.99 ($p < .001$) for the Spanish subsample and 1672.92 ($p < .001$) and 44.06 ($p < .001$) for the Chinese subsample, respectively. Failure to comply with multi-variate normality led us to use the MLM method to estimate the hypothesized latent models.

The first foray into the study of the latent structure of the ASC construct was to carry out a one-factor independent CFA for each of the three dimensions (*math*, *verbal*, and *general school*). Table 2 shows the overall fit of all one-factor models and the estimated factor loadings by country. Results indicate an adequate fit for the three dimensions, with non-significant chi-square tests in all cases ($p \geq .05$), except for the *verbal* dimension in the Chinese subsample ($\chi_{\text{scaled}}{}^2 = 12.607$, $df = 5$, $p = .027$); anyway, the model obtained acceptable results attending the rest of the fit indices (RMSEA = .063, CI90% RMSEA = [.032; .095], p[RMSEA ≤ .05] = .223, CFI = .992, TLI = .984, SRMR = .024). Accordingly to these fit, the factor loadings were all statistically significant and ranging from .692 (*general school*2) to .939 (*math*3) in the Spanish subsample and from .525 (*general school*1) to .891 (*verbal*3) in the Chinese subsample.

The Omega reliability coefficient is over .80 for the three dimensions in the Spanish subsample (*general school* = .892, *math* = .890, and *verbal* = .887), explaining 68%, 67%, and 61% of the variance, respectively. These results were similar in the Chinese subsample (*verbal* = .897, *math* = .886, and *general school* = .761), explaining 64%, 66%, and 45% of the variance, except the *general school* dimension that obtained lower values both for reliability and explained variance.

Once we obtained adequate results for the fit of the measurement model of each of the factors of the ASC construct, we proceeded to estimate the fit of the seven hypothesized latent structures using a SEM model comparison strategy. The overall fit indices obtained after estimating the seven hypothesized models on the latent structure of ASC for the two subsamples are shown in Table 3. The list of models appears in descending order from best to worst overall fit. The best-fitted model for the two subsamples was the three-factor ESEM model (Figure 1d), with better overall fit in the Spanish subsample ($\chi_{\text{scaled}}{}^2 = 47.864$, $df = 42$, $p = .247$, RMSEA = .023, CI90% RMSEA = [.000; .047], CFI = .997, TLI = .994, SRMR = .018) than the Chinese subsample ($\chi_{\text{scaled}}{}^2 = 140.839$, $df = 42$, $p < .001$, RMSEA = .078, CI90% RMSEA = [.067; .089], CFI = .955, TLI = .916, SRMR = .041). The RMSEA *post-hoc* statistical power for the three-factor ESEM model was .79 and .95 for the Spanish and Chinese subsamples, respectively. The second and third best results corresponded to the incomplete bifactor ESEM model (Figure 1f) and the

TABLE 1 Correlation matrix of the items and univariate descriptive statistics by subsamples.

Items	1	2	3	4	5	6	7	8	9	10	11	12	13
Spanish subsample ($n = 262$)													
1. Math1	–												
2. Math2	.575	–											
3. Math3	.683	.793	–										
4. Math4	.617	.652	.723	–									
5. Verbal1	−.084	.006	.009	.020	–								
6. Verbal2	−.022	.075	.067	.083	.600	–							
7. Verbal3	−.187	−.118	−.100	−.100	.440	.585	–						
8. Verbal4	−.109	−.020	.021	.022	.587	.650	.521	–					
9. Verbal5	−.117	−.043	−.014	−.028	.623	.733	.647	.672	–				
10. S1	.130	.205	.307	.208	.392	.399	.160	.509	.376	–			
11. S2	.239	.291	.345	.304	.352	.409	.215	.360	.443	.512	–		
12. S3	.168	.193	.305	.249	.414	.414	.188	.520	.412	.725	.571	–	
13. S4	.185	.207	.320	.250	.428	.518	.268	.546	.489	.733	.662	.762	–
Mean	2.88	3.25	3.11	3.27	4.15	3.86	2.68	3.80	3.81	4.77	4.09	4.16	4.22
Median	3.00	3.00	3.00	3.00	4.00	4.00	3.00	4.00	4.00	5.00	4.00	4.00	4.00
SD	1.69	1.64	1.56	1.73	1.49	1.50	1.58	1.59	1.44	1.42	1.28	1.36	1.34
AD norm.	11.34	7.48	8.64	8.91	9.15	7.37	11.91	9.80	7.71	18.16	7.98	8.54	8.57
Chinese subsample ($n = 389$)													
1. Math1	–												
2. Math2	.551	–											
3. Math3	.699	.611	–										
4. Math4	.675	.642	.790	–									
5. Verbal1	−.043	.150	.006	.055	–								
6. Verbal2	.104	.065	.116	.137	.565	–							
7. Verbal3	.071	.001	.118	.114	.493	.769	–						
8. Verbal4	.104	.062	.158	.208	.438	.678	.754	–					

(Continues)

TABLE 1 (Continued)

Items	1	2	3	4	5	6	7	8	9	10	11	12	13
9. Verbal5	.097	.105	.105	.199	.464	.700	.699	.683	–				
10. S1	.181	.391	.163	.218	.164	.095	.084	.101	.099	–			
11. S2	.338	.382	.440	.508	.072	.169	.118	.260	.310	.327	–		
12. S3	.271	.249	.374	.342	−.013	.243	.210	.295	.240	.390	.390	–	
13. S4	.346	.340	.390	.539	.102	.203	.177	.288	.264	.429	.521	.585	–
Mean	3.70	4.36	3.65	3.81	4.69	3.80	3.48	3.72	3.81	4.31	3.80	4.00	4.07
Median	4.00	5.00	4.00	4.00	5.00	4.00	3.00	4.00	4.00	5.00	4.00	4.00	4.00
SD	1.51	1.38	1.39	1.36	1.47	1.46	1.50	1.36	1.36	1.39	1.16	1.28	1.28
AD norm.	9.45	13.46	9.18	10.05	24.70	9.08	8.42	9.18	9.41	13.64	11.91	11.05	11.83

Note: All correlations and AD values are statistically significant ($p < .001$).

Abbreviations: AD norm. = Anderson-Darling normality test statistic; S = general school; SD = standard deviation.

TABLE 2 Overall fit indices of one-factor CFA models and estimated factor loadings by dimensions and subsamples.

Models	χ^2_{Scaled}	Scaling factor	*df*	*p*	CFI	TLI	RMSEA	$CI_{90\%}$ RMSEA	*p* (RMSEA ≤.05)	SRMR
Spain										
Math	3.650	1.78	2	.161	.998	.993	.056	.000–.123	.361	.019
Verbal	5.389	1.49	5	.370	.999	.999	.017	.000–.077	.108	.020
General school	4.727	2.32	2	.094	.993	.978	.072	.013–.128	.209	.023
China										
Math	2.129	1.48	2	.345	1.00	.999	.013	.000–.088	.698	.009
Verbal	12.607	1.89	5	.027*	.992	.984	.063	.032–.095	.223	.024
General school	2.106	1.14	2	.349	1.00	.999	.012	.000–.097	.654	.023

	Factor loadings							
	Spain	China		Spain	China		Spain	China
Math1	.728**	.774**	Verbal1	.710**	.581**	General school1	.814**	.525**
Math2	.837**	.706**	Verbal2	.835**	.861**	General school2	.692**	.605**
Math3	.939**	.889**	Verbal3	.700**	.891**	General school3	.851**	.689**
Math4	.780**	.888**	Verbal4	.772**	.823**	General school4	.908**	.847**
			Verbal5	.883**	.804**			

$^*p < .05$; $^{**}p < .01$.

TABLE 3 Summary of the SEM model comparison strategy with the seven hypothesized latent structures overall fit indices (from the best to the worst).

Models	χ^2_{Scaled}	Scaling factor	Df	p	CFI	TLI	RMSEA	$CI_{90\%}$ RMSEA	p(RMSEA ≤.05)	RMSEA power	SRMR
Spain											
ESEM 3F (1d)	47.864	1.265	42	.247	.997	.994	.023	.000–.047	.971	.79	.018
ESEM incomplete BIF (1f)	62.158	1.267	48	.082	.992	.987	.033	.000–.053	.913	.83	.031
Brunner et al.'s model (1e)	75.268	1.223	55	.036	.989	.984	.037	.014–.055	.868	.87	.038
Complete BIF (1g)	93.967	1.083	52	<.001	.974	.961	.055	.038–.072	.288	.85	.082
Shavelson et al.'s model (1a)	111.109	1.224	62	<.001	.973	.966	.055	.040–.070	.282	.90	.059
CFA 3F corr. (1c)	119.978	1.133	62	<.001	.964	.955	.060	.044–.075	.141	.90	.056
Marsh/Shavelson model (1b)	128.339	1.215	64	<.001	.965	.957	.062	.048–.076	.083	.91	.096
China											
ESEM 3F (1d)	140.839	1.745	42	<.001	.955	.916	.078	.067–.089	<.001	.95	.041
ESEM incomplete BIF (1f)	158.533	1.681	48	<.001	.949	.918	.077	.067–.087	<.001	.97	.045
Brunner et al.'s model (1e)	168.818	1.301	55	<.001	.948	.926	.073	.063–.083	<.001	.98	.046
Shavelson et al.'s model (1a)	197.447	1.601	62	<.001	.938	.922	.075	.066–.084	<.001	.99	.055
Marsh/Shavelson model (1b)	208.849	1.615	64	<.001	.934	.919	.076	.067–.086	<.001	.99	.082
Complete BIF (1g)	210.504	1.155	52	<.001	.909	.864	.089	.077–.100	<.001	.98	.047
CFA 3F corr. (1c)	230.337	1.372	62	<.001	.904	.879	.084	.074–.094	<.001	.99	.052

incomplete bifactor model (Figure 1e), but in both cases, the fit was worse than that of the three-factor ESEM model (Figure 1d). Subsequent models are already far from being able to compete for adequately representing the construct's latent structure.

The next step was to test the invariance of the ESEM three-factor model through the two subsamples. At the first level, configural invariance obtained an adequate overall fit ($\chi_{scaled}{}^2 = 164.25$, $df = 84$, $p < .001$, RMSEA = .054, CI90% RMSEA = [.044; .064], CFI = .980, TLI = .962, SRMR = .026), with a RMSEA *post-hoc* power of 1.00. The second level, metric invariance, that constrains the factor loadings to be equal across the two subsamples, obtained an adequate overall fit but slightly worse ($\chi_{scaled}{}^2 = 228.73$, $df = 114$, $p < .001$, RMSEA = .056, CI90% RMSEA = [.047; .064], CFI = .971, TLI = .960, SRMR = .047), with a RMSEA *post-hoc* power of 1.00; the ΔCFI (.009) was very near to the non-significant difference (.010), but the $\Delta\chi^2$ test showed significant statistical differences ($\Delta\chi^2 = 65.954$, $\Delta df = 30$, $p = .00016$). Consequently, mean comparisons between countries are not allowed and we tried to explore those non-invariant factor loadings that were impeding reaching metric invariance.

The partial invariance analysis detected as non-invariant the direct loadings of the items *math*2, *verbal*3, *general school*1, *general school*2, and *general school*4, and the cross-loadings of the *verbal*1, *verbal*3, and *general school*1 on the *math* dimension; the *math*2 item on the *verbal* dimension, and the items *verbal*1 and *verbal*3 on the *general school* dimension. A metric partial model releasing those parameters was re-estimated and obtained a good fit ($\chi_{scaled}{}^2 = 189.81$, $df = 103$, $p < .001$, RMSEA = .051, CI90% RMSEA = [.042; .060], CFI = .978, TLI = .967, SRMR = .033), with a RMSEA *post-hoc* power of 1.00. The chi-square test between the configural and this partial metric model was statistically non-significant ($\Delta\chi^2 = 22.933$, $\Delta df = 19$, $p = .240$), and ΔCFI (.002) was clearly below .01.

To complete the partial invariance study, we tested the next level, scalar invariance, that constrains the items intercepts to be equal, but keeping non-invariant factor loadings unconstrained. Scalar invariance model obtained a relatively adequate fit ($\chi_{scaled}{}^2 = 257.07$, $df = 113$, $p < .001$, RMSEA = .063, CI90% RMSEA = [.054; .071], CFI = .963, TLI = .950, SRMR = .040), with a RMSEA *post-hoc* power of 1.00. The chi-square difference test between this scalar and the partial metric model was statistically significant ($\Delta\chi^2 = 100.41$, $\Delta df = 10$, $p < .001$), and ΔCFI (.015) was clearly above .01. In a similar way to the exploration carried out with the factor loadings, we proceeded to detect the invariant and non-invariant intercepts. The non-invariant intercepts detected corresponded to the items *math*2, *math*3, *verbal*1, *verbal*2, *verbal*3, *general school*1, and *general school*4. Once again, the scalar model was re-estimated by keeping the non-invariant loadings unconstrained and releasing the equality constraint on the non-invariant intercepts. The fit of the partial scalar model significantly improved ($\chi_{scaled}{}^2 = 194.73$, $df = 107$, $p < .001$, RMSEA = .050, CI90% RMSEA = [.047; .074], CFI = .978, TLI = .968, SRMR = .033), with a RMSEA *post-hoc* power of 1.00. The chi-square difference test between this partial scalar and the partial metric model was statistically non-significant ($\Delta\chi^2 = 9.754$, $\Delta df = 4$, $p = .05$) and ΔCFI (.000) was clearly below .01.

The invariance study process stopped at this point as five direct and six non-invariant cross factor loadings were detected, as well as seven non-invariant intercepts. These non-invariant parameters invalidate the possibility of testing the following possible levels of invariance, that is, the equality of the variances and covariances of the latent factors, the latent means, and the residuals. At this point, we resorted to NP in order to further explore and understand these partial invariance results.

An alternative way using network psychometrics

Since the fit of a complete metric invariance was not adequate, we used a strategy based on the analysis of the NP to try to unravel the network of relationships between the items of the three dimensions composing the ASC construct. The NP algorithms induce a sequence of partitions into communities (in our case, latent factors), using distance metrics based on the strength (or degree) of the association between nodes (in this case, items). We conducted an EGA analysis in the two subsamples using all items of the three ASC dimensions.

The estimated EGA model perfectly identifies the hypothesized three-factor structure in which items ascription is adequate in both subsamples. Respecting the model fit, the Miller-Madow correction for the total correlation of the dataset was .52 for the two subsamples and the clustering coefficient was .625 and .580 for the Spanish and Chinese subsamples, respectively. The Miller-Madow correction of the Entropy Fit Index (EFI) obtained very low values for the Spanish (−3.40) and Chinese (−3.59) subsamples. The R-squared Fit for Scale-free Network, that is, a basal uncorrelated model, was .004 and .002, respectively. The small-world index was .139 for the Spanish and .184 for the Chinese subsample, indicating the networks are small-world (far from lattice or random networks). All these indices indicate a good fit of the estimated EGA models.

Although these results were adequate, the next step was to implement a bootstrap EGA that allows for the consistency of factors and items to be evaluated across bootstrapped EGA results. This procedure provides information about whether data are consistently organized in coherent factors or fluctuate between dimensional configurations. We simulated 1000 bootstrap samples; results indicated that the three-factor model appeared 100% of the time for the Spanish subsample and all items were 100% consistently related to the correct factor. In the Chinese subsample, the three-factor model emerged the 96.4%, a two-factor model the 3%, and a four-factor model the remaining 0.6%. This dimensional fluctuation can be explained by the relationship of *general school* items to *math* dimension (*general school*2 = 4.3%, *general school*1 = 4.2%, *general school*4 = 3.1%, and *general school*3 = 2.9% of 1000 samples). Despite this, the structural consistency of the *verbal* dimension was .971, .940 for *general school*, and .929 for *math*. The average stability of the items was 1.000 for the *math* dimension, .982 for *verbal*, and .962 for *general school*.

Dimensional consistency and items stability for the three-factor EGA model in the two subsamples were high, and a configural invariance of the latent structure can be interpreted. But it is important to compare visually and analytically the relationships between the factors inside the network. The *EGAnet* package includes a graph function for comparing networks that locate the items in the same position in the space. This way, it facilitates the comparison of the two networks' configurations. Figure 2 shows a configural difference between the Spanish and Chinese networks; the *general school* dimension appears more related to the *verbal* factor in the Spanish subsample and to the *math* dimension in the Chinese subsample. In addition, the relationship between *verbal* and *math* factors in Spanish students does not reach enough magnitude to be represented in EGA, but this connection is more relevant for Chinese students. These two differences may be behind the difficulty in finding complete metric invariance

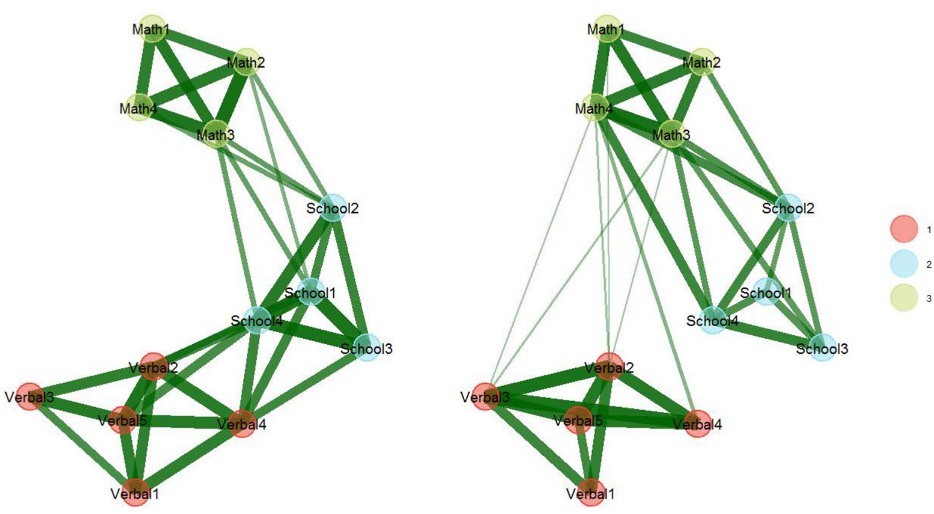

FIGURE 2 Visual comparison of the EGA models from Spanish (left) and Chinese (right) subsamples.

using SEM models. It is a question of the construct's nature, less related to analytical phenomena and deserves deeper discussion later.

Finally, Table 4 shows the EGA factor loadings by subsamples and the metric invariance test item by item. The *EGAnet* package computes the null distribution of loading differences from randomized samples. There is evidence that the multi-group CFA fails to reject the null hypothesis when it is false to a greater extent than EGA (Jamison et al., 2022). Results indicate that non-invariant loadings with higher differences between subsamples corresponded to items *verbal*3 (−.183), *general school*1 (.149), *verbal*5 (.136), *math*2 (.124), and *math*3 (.121). The classic metric invariance and the EGA metric invariance studies coincide in pointing out the items *math*2 and *general school*1 as non-invariant, but they differ regarding the items *math*3 (EGA), *verbal*2 (SEM), *verbal*3 (EGA), *verbal*5 (EGA), *general school*2 (SEM), and *general school*4 (SEM). In short, taking into account the evidence provided by both approaches, eight of the 13 items are under suspicion of non-invariance.

DISCUSSION

Several studies have consistently shown the importance of ASC for learning and other educational outcomes, such as academic achievement (Marsh & Craven, 2006), interest and satisfaction in school (Marsh et al., 2005), persistence and long-term attainment (Guo et al., 2015), etc. For example, the *self-enhancement model* claims that self-concept is a primary determinant of academic achievement. For these reasons, it is a relevant goal to analyse the ASC structure across countries, to allow for a fuller conceptualization of it. Most of the studies have analysed the internal structure of ASC through SEM, but to the best of our knowledge, nobody has used a different approach as NP. The novelty of this research is that we have used both approaches, SEM and NP, conducting an EGA. In fact, its use has been of particular importance, since it has made possible the representation of the differences between China and Spain with respect to the relationships established between ASC dimensions (*verbal* and *math*) and *general school* self-concept. Some authors (Golino et al., 2021) claim that EGA approach is superior to traditional dimension reduction techniques such as EFA and even simple or complex CFA. Therefore, the aims of this study have been (1) to analyse the internal structure of ASC in Spanish and Chinese adolescents' samples using two approaches, SEM and NP conducting an EGA; and (2) to analyse the measurement invariance of the best model across countries and investigate the cross-cultural differences

TABLE 4 EGA loadings and metric invariance.

Items	Factor	Spanish EGA loadings	Chinese EGA loadings	Loadings differences	*p*
Math1	Math	.321	.352	−.031	.858
Math2		.397	.273	.124	.008**
Math3		.591	.470	.121	.036*
Math4		.378	.458	−.080	.124
Verbal1	Verbal	.282	.236	.046	.270
Verbal2		.399	.477	−.078	.050
Verbal3		.294	.477	−.183	.016*
Verbal4		.315	.373	−.058	.292
Verbal5		.505	.369	.136	.016*
General school1	General school	.361	.212	.149	.002**
General school2		.220	.184	.036	.966
General school3		.403	.325	.078	.314
General school4		.515	.432	.083	.592

$^*p < .05$; $^{**}p < .01$.

in ASC. It is important to highlight that some studies that have analysed cross-cultural differences in ASC do not have analysed the measurement invariance of the questionnaire before comparing the means (Inglés et al., 2009; Nishikawa et al., 2007); other studies have analysed the measurement invariance, but the equivalence has not been supported (Brunner et al., 2009); and finally, other studies have been successful finding the measurement invariance and have had the possibility to compare the means (Leung et al., 2016). However, it is well known that it is not easy to find the measurement invariance because people in different countries have different construal of the self. Therefore, it is an interesting goal to analyse how Spanish and Chinese adolescents conceptualized ASC.

As we have said, the first goal of the study was to analyse the internal structure of ASC through the SDQII-S questionnaire. Considering that empirical findings about self-concept have been mainly obtained from studies conducted in Western societies, cross-cultural research questions if those findings can be extended to Eastern societies (Spencer-Rodgers et al., 2009). Thus, to analyse the internal structure of ASC in a Chinese sample is relevant bearing in mind that it still has not been performed. In order to fulfil this objective, seven models were tested. This might be considered a strength of our study, as previous studies have not compared as many models as we have. Marsh et al. (2006) claimed that the first-order factor model was reasonably invariant across 25 countries, but they did not compare different models. On the other hand, Brunner et al. (2009) compared two models and Esnaola et al. (2018) evaluated three. Among all of the models, the differences between second-order and bifactor models become specifically relevant because bifactor models have some advantages over the second-order models: (a) they allow a differentiated interpretation of academic-related domain self-concept profiles; (b) invariance can be assessed for general and group factors; and (c) the relationship between group factors and external criteria can be analysed (Chen et al., 2006).

Our results have failed to confirm the first hypothesis, which claimed that the bifactor model would be the best taking into account Brunner et al. (2009) and Esnaola et al. (2018) studies; the first one found a good fit to the data of all 26 countries analysed and the later in Spain (Esnaola et al., 2018). However, it can be helpful to remind that Brunner et al. (2009) did not used the SDQII-S questionnaire subscales, but they selected the three best items for *verbal*, *math*, and ASC used in the PISA study (OECD, 2001). In our study, although Brunner et al. (2008) model has shown a good fit, the best-fitted model for the two subsamples has been the three-factor ESEM model (Figure 1d), which allows the cross-loadings estimation. This result cannot be compared with previous studies, where the first-order model (Figure 1c) has shown a good fit, like in Marsh et al.'s (2006) study in 25 countries or Brunner et al.'s (2009) study in 26 countries. It can be important to highlight that Brunner et al. (2009) found that both models (first-order factor model and nested-factor model) provided a good fit, in the total sample and in each country-specific data, so both models can be considered an appropriate structural conception of ASC, although these authors claim that the nested-factor model offered a more differentiated perspective than the popular first-order factor model on how domain-specific ASCs are associated with students' characteristics (Brunner et al., 2009). However, the first-order model analysed in those two studies did not allow cross-loadings estimations, and as we will see later, the items of the different ASC factors are sometimes related.

Our results mean that the multi-dimensionality of ASC has been confirmed in both countries being consistent with previous studies (Chen et al., 2020), but the hierarchy has not. However, our results are in line with those who have failed to find a well-fitting structural model with general ASC at its apex (Yeung et al., 2000); models including general ASC at the top of the hierarchy have rarely provided a good fit to the empirical data.

The second purpose was to analyse the measurement invariance of the best model (the three-factor ESEM model) across countries and to investigate the cross-cultural differences in ASC. Results have shown that only configural invariance is supported. Thus, taking into account that scalar invariance has not been supported, it is not possible to compare cross-cultural differences, and consequently, hypothesis 2 has been supported. This result is not a surprise, because as Byrne (2002) claimed, "responses to self-concept questionnaires will be influenced by a cultural bias that ultimately leads to differential perceptions of self" (p. 903). This error has been categorized into two types: *conceptual non-equivalence* (self-concept has a different meaning) and *measurement non-equivalence* (bias caused by the instrument, e.g., because of culturally different response styles; Byrne et al., 2009). In order to control this type of bias, Byrne (2002) suggested using interviews alongside the established self-concept measures. However, the use of qualitative data in self-concept remains extremely rare, although in a recent work Rüschenpöhler and Markic (2019) have used a mixed method approach in order to evaluate the culture-sensitive of ASC. Concretely, they analysed the chemistry self-concept with secondary school students using both interviews and a questionnaire. It is also significant to highlight that Brunner et al. (2009) did not find the measurement invariance in either of the two models analysed in their study (the first-order model and the nested-factor model) across the 26 countries.

Taking into account that we have not found scalar invariance, partial invariance analysis has been performed. Results have shown that using SEM and EGA, items *math*2 ("I do badly in tests of mathematics") and *general school*1 ("I get bad marks in most school subjects") are non-invariant; SEM analysis shows that *verbal*2 ("Work in English classes is easy for me"), *general school*2 ("I learn things quickly in most school subjects"), and *general school*4 ("I am good at most school subjects") are non-invariant; and EGA analysis shows that non-invariant items are *math*3 ("I get good marks in mathematics"), *verbal*3 ("English is one of my best subjects"), and *verbal*5 ("I learn things quickly in English classes").

The reasons for these differences can be multiple, but we do not believe that one of them is about the differences in the general characteristics between the two subsamples. All participants are mid-range schools in mid-size cities, so they may represent the average student population in both countries. The amount of math class time/week is the same time as language classes, so schools put equal weight on both subjects, and by socioeconomic level, both subsamples include middle-class families. Therefore, we believe that both samples present an adequate degree of comparability so as not to have biased the invariance procedure. Likewise, the interpretation of the differential comprehensions of the individual items is not easy because we have not found previous studies where partial invariance analyses have been carried out on the items of the SDQII-S questionnaire comparing different countries. Therefore, we cannot compare our results with previous studies; however, we are going to point out some hypotheses.

Some authors claim that the cultural differences in means level in self-concept can be related to collectivism and humbleness emphasized in Chinese culture (Leung et al., 2016), the cultural norms regarding modesty and self-assertion (Marsh et al., 2006), or to be less critical of themselves in the case of Arab students (Abu-Hilal, 2001). These reasons could be also the key point to explain why people from different cultures might understand self-concept items in a different way.

If we focus on the two items that both approaches (SEM and EGA) have identified as non-invariant, we could say that both of them are reverse items. As the author of the SDQ-II himself (Marsh, 1994) stated, sometimes this kind of items is problematic. Likewise, although SDQII is one of the most reliable questionnaire to measure self-concept, it is also true that both items could be interpreted in different ways. That is, what does mean "to do badly" or "get bad marks"? For some people "to do badly" or "get bad marks" can be understood as to fail or do not pass the examinations, whereas for other ones can be to pass the examinations but with low marks; or for others, it may be when grades are simply lower than expected (a student may get good marks but not as high as he/she would like considering the time and energy invested in studying). Likewise, there are another two items that have been identified as non-equivalent: *general school2*, "I learn things quickly in most school subjects" (through SEM) and *verbal5*, "I learn things quickly in English classes" (through EGA) that refer to the same content. Here again arises the doubt about what means "to learn quickly," as it is also a subjective statement. Thus, it seems that another reason of the results about non-invariant items could be related to bias associated with test content or its interpretation, that is, conceptual non-equivalence, which is defined as a construct having not the same meaning across groups.

These possible different interpretations of the meaning of some items could be related with the *Big Fish Lite Pond Effect* (BFLP; Marsh, 1984), which means that students in selective schools would have lower ASC compared to those with comparable ability that attend regular or non-selective schools. That is, this effect states that the ASC of an individual student is based on the student's own academic performance levels and also on the average of the performance levels of students in the same school he or she attends. Marsh and Hau (2003) probed the BFLPE's cross-cultural generalizability in a very large cross-cultural study with nationally representative samples of 15-year-old adolescents from 26 countries, and Marsh et al. (2019) found the same results in 68 countries. In our case, we are not talking about mean differences, but the understanding of what means "to do badly" or "to get bad marks"; evidently, this understanding can be also influenced by BFLP effect. Likewise, extending the BFLPE theory, Marsh et al. (2019) highlighted the *paradoxical cross-cultural self-concept effect*, which means that country average achievement has a negative effect on academic self-concept. These two paradoxical effects could be behind the different understanding about what means "to do badly" or "get bad marks." In our case, two countries have been compared, Spain and China.

Related to *paradoxical cross-cultural self-concept effect*, it can be said that the educational system in China is more exigent than the Spanish one. As Davey et al. (2007) stated, Chinese university education markedly increases life chances in China, as society is very competitive and the number of university applicants exceeds the number of available places. Competition is fierce, especially for entry into prestigious universities, through the National Entrance Examination for Higher Education. In consequence, teachers and parents (who have usually high expectations in terms of education) place considerable pressure on their children to succeed in school (parents' pressure may influence how individuals interpret the items of SDQII-S), and as a result, preparation for the entrance examination begins at an early age. For this reason, Chinese children and teenagers spend most of their time studying (in class and finishing overloaded homework out of school) and report heightened fear of failure. Moreover, this situation is aggravated by the fact that school failure is traditionally associated with individual, family, and national shame. Maybe this is the reason that in the last PISA study results (2012, 2015, and 2018), China always has been clearly in front of Spain. To sum up, as Marsh et al. (2019) claim, contextual effects matter, resulting in significant and meaningful effects on self-beliefs by local school level (BFLPE) and also at the macro-contextual country level. Hence, the average of the students' performance level in the same school, the country average achievement, and the requirements of educational systems, as well as the personal expectations of individuals (related to conceptual non-equivalence) may influence the differential understandings of the items.

Since the fit of the metric invariance was not adequate, we used a new strategy based on the analysis of the NP to see the relationships between the items. The estimated EGA model perfectly identifies the hypothesized three-factor structure in which items ascription is adequate in both subsamples. However, the bootstrap EGA has shown that the three-factor model appears 100% of the time in the Spanish subsample, but in the Chinese one this model emerges the 96.4%. If we compare visually and analytically the relationships between the factors inside the network (Figure 2), the results have shown a configural difference between the two subsample networks. The *general school* dimension appears more related to the *verbal* factor in the Spanish adolescents, but in the Chinese subsample, it is more related to *math* dimension. Likewise, the relationships between *verbal* and *math* factors in Spanish students do not reach enough magnitude to be represented in EGA figure, being consistent with some studies (Marsh, 1986, 1990; Möller et al., 2009), but in Chinese adolescents' this connection is more relevant. Therefore, the results found in Spanish adolescents are in line with the idea that people think of themselves either as mathematics people or as verbal ones, but not as both (Marsh & Hau, 2004). However, in the Chinese subsample, it seems that the same idea is not true.

These differences may be behind the difficulty in finding invariance using SEM models. To the best of our knowledge, this is the first study which has used NP approach to analyse the internal structure of academic self-concept. As we have said in the beginning of this paper, although the multidimensionality of ASC is well known, it has been a challenge to find a structural model of ASC capable of accounting the hierarchy and/or the relations between *math* and *verbal* self-concepts; and it seems that these relationships can be different depending on the culture.

This study has some limitations. On the one hand, although the sample size is considered big enough as the power estimation proofed, it is not as large as could be expected, so it would be interesting to increase the number of students in the following studies and to use a stratified sample looking for a better representativeness. On the other hand, cross-cultural studies are scarce. In this study, the ASC has been analysed in two countries, one Asian Eastern country (China) and another Western country (Spain). Hence, in future studies, it would be interesting to add more countries with different cultural backgrounds, analysing for example one country for each of the eleven cultural clusters that Ronen and Shenkar (2013) proposed. Finally, cross-cultural measurement invariance has not been supported, so the need for further research in this area is mandatory; or maybe, researchers simply could reanalyse their data using NP (conducting an EGA) to verify whether the results found in this paper can be confirmed in other countries. Another option, as Byrne (2002) and Rüschenpöhler and Markic (2019) have suggested and developed, would be that self-concept measure and cross-cultural comparison might need a mixed method (quantitative and qualitative) approach.

Despite these limitations, it could be said that this study have some strengths and its relevance has both a methodological and a substantive aspect: (1) the methodological aspect consists of presenting a rigorous analysis strategy on the latent structure (comparing seven models, much more than previous studies) of a test that combines more classic procedures such as SEM, with cutting-edge procedures such as NP; (2) the applied one refers to the need not to overlook cultural differences that can vary the definition of the constructs to be measured; that is, to analyse invariance, but not as a purely mathematical–statistical procedure, but rather as a way of obtaining information that allows us to understand whether the behaviour of the areas of interest of a construct and its internal relationship is invariant or not between cultures.

As we have said before, it is not easy to know why some items are non-invariant and we have tried to explain it taking into account the cross-cultural self-concept theory and its relationships with the highest requirement of the Chinese education system and the BFLPE. These explanations could be behind the differences in understanding of what means "to do badly," "get bad marks," "learn quickly," etc., for each subsample. One practical implication of this study would be the need to follow the recommendations of Byrne (2002) and Rüschenpöhler and Markic (2019) that mentioned the convenience of using mixed-methods (quantitative and qualitative) measures.

This approach will give us the possibility to ask participants qualitatively (in an open question, for example) what they understand when, for example, they have to answer a question about "to do badly" or "get bad marks." This is related with *validity evidence based on response processes*, which was first considered explicitly as a source of validity evidence in the *Standards for Educational and Psychological Testing* (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 1999); that is, we would need more information about how individuals cope or understand the meaning of the items and answer them (in our case what they understand about the meaning of "to do badly" or what means "to get bad marks") to interpret the differences that we have found in those items. One option could be to use methods that allow us directly access to the psychological processes or cognitive operations (think aloud, focus group, and interviews) to identify elements (words, expressions, response format, etc.) which may be problematic for the test or questionnaire respondents, or if the researcher intends to identify how people refer to the object, content, or specific aspects included in the items (Padilla & Benítez, 2014). That information would be very relevant to analyse the differential understandings or interpretations of items content that participants do, and try to understand the conceptual non-equivalence.

In summary, the results of this research show that: (1) Results have supported the multidimensional nature of the ASC as well as the reliability of SDQII-S academic subscales; (2) The three-factor ESEM model has obtained the best fit in both countries; (3) Only configural invariance has been supported across countries, therefore it is not possible to compare cross-cultural differences; (4) Partial invariance results showed that eight of the thirteen items are under suspicion of non-invariance; (5) The estimated EGA model perfectly identifies the hypothesized three-factor structure. The graph function shows that the *general school* dimension appears more related to the *verbal* factor in the Spanish subsample and to the *math* dimension in the Chinese subsample. Likewise, the relationships between *verbal* and *math* factors in Spanish students does not reach a high enough magnitude to be represented in EGA, but this connection is more relevant for Chinese students. These differences may be behind the difficulty in finding invariance using SEM models. Therefore, it seems that instead of striving to analyse the measurement invariance, it could be more interesting to understand how the ASC is conceptualized in different cultures; so maybe the present findings could open a new way to analyse the invariance of the ASC items. The classical way to analyse the internal structure of ASC has been the SEM approach and as we have seen in the findings of this study, SEM and EGA approach have shown diverging results about the invariance of the items. Hence, our results lead to the need for a more comprehensive exploration of data about psychometric latent structure evidence. Therefore, it would be interesting to triangulate the results found through SEM with the new approach used in this study. We humbly invite all researchers that are interested in ASC to reanalyse their data to confirm or not our findings in other countries using NP (conducting an EGA) approach. We should try to clear up, if the differences in the understanding of the items are related with the content or if they are because we use different techniques or statistical approaches to analyse that.

AUTHOR CONTRIBUTIONS

Igor Esnaola: Conceptualization; supervision; investigation; writing – original draft; writing – review and editing. **Albert Sesé:** Formal analysis; methodology; writing – review and editing; writing – original draft. **Lorea Azpiazu:** Conceptualization; investigation; writing – original draft; writing – review and editing; supervision. **Yina Wang:** Supervision; writing – review and editing.

CONFLICT OF INTEREST STATEMENT

All authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are openly available in Figshare at https://figshare.com/articles/dataset/ASC_invariance_SPAIN_CHINA_dataset/23742129.

ORCID

Igor Esnaola https://orcid.org/0000-0002-4159-3565

REFERENCES

Abu-Hilal, M. M. (2001). Correlates of achievement in The United Arab Emirates: A sociocultural study. In D. M. McInerney & S. van Etten (Eds.), *Research on sociocultural influences on motivation and learning* (Vol. *1*, pp. 205–230). Information Age.

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. American Educational Research Association.

Arens, K. A., Yeung, A. S., Craven, R. G., & Hasselhorn, M. (2011). The twofold multidimensionality of academic self-concept: Domain specificity and separation between competence and affect components. *Journal of Educational Psychology*, *103*(4), 970–981. https://doi.org/10.1037/a0025047

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*(4), 387–403. https://doi.org/10.1016/j.lindif.2008.11.008

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*(1), 137–172. https://doi.org/10.1080/00273170701836737

Byrne, B. M. (2002). Validating the measurement and structure of self-concept: Snapshots of past, present, and future research. *American Psychology*, *57*(11), 897–909. https://doi.org/10.1037/0003-066X.57.11.897

Byrne, B. M., Oakland, T., Leong, F. T. L., van de Vijver, F. J. R., Hambleton, R. K., Cheung, F. M., & Bartram, D. (2009). A critical analysis of cross-cultural research and testing practices. *Training and Education in Professional Psychology*, *3*(2), 94–105. https://doi.org/10.1037/a0014516

Campbell, J., Trapnell, P., Heine, S., Katz, I., Lavallee, L., & Lehman, D. (1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology*, *70*(1), 141–156. https://doi.org/10.1037/0022-3514.70.1.141

Chen, F., García, O. F., Fuentes, M. C., García-Ros, R., & García, F. (2020). Self-concept in China: Validation of the Chinese version of the Five-Factor Self-Concept (AF5) questionnaire. *Symmetry*, *12*(5), Article 798. https://doi.org/10.3390/sym12050798

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*, 189–225. https://doi.org/10.1207/s15327906mbr4102

Chen, X., Zappulla, C., Lo Coco, A., Schneider, B., Kaspar, V., De Oliveira, A., He, Y., Li, D., Li, B., Bergeron, N., Chi-Hang Tse, H., & DeSouza, A. (2004). Self-perceptions of competence in Brazilian, Canadian, Chinese and Italian children: Relations with social and school adjustment. *International Journal of Behavioral Development*, *28*(2), 129–138. https://doi.org/10.1080/01650250344000334

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155–159. https://doi.org/10.1037/0033-2909.112.1.155

Davey, G., De Lian, C., & Higgins, L. (2007). The university entrance examination system in China. *Journal of Further and Higher Education*, *31*(4), 385–396. https://doi.org/10.1080/03098770701625761

English, T., & Chen, S. (2007). Culture and self-concept stability: Consistency across and within contexts among Asian Americans and European Americans. *Journal of Personality and Social Psychology*, *93*(3), 478–490. https://doi.org/10.1037/0022-3514.93.3.478

Epskamp, S., Borsboom, D., & Fried, E. I. (2018). Estimating psychological networks and their accuracy: A tutorial paper. *Behavior Research Methods*, *50*(1), 195–212. https://doi.org/10.3758/s13428-017-0862-1

Esnaola, I., Elosua, P., & Freeman, J. (2018). Internal structure of academic self-concept. *Learning and Individual Differences*, *62*, 174–179. https://doi.org/10.1016/j.lindif.2018.02.006

Giordano, C., & Waller, N. G. (2019). Recovering bifactor models: A comparison of seven methods. *Psychological Methods*, *25*(2), 143–156. https://doi.org/10.1037/met0000227

Golino, H., Moulder, R., Shi, D., Christensen, A. P., Garrido, L. E., Nieto, M. D., Nesselroade, J., Sadana, R., Thiyagarajan, J. A., & Boker, S. M. (2021). Entropy fit indices: New fit measures for assessing the structure and dimensionality of multiple latent variables. *Multivariate Behavioral Research*, *56*(6), 874–902. https://doi.org/10.1080/00273171.2020.1779642

Golino, H. F., & Epskamp, S. (2016). Exploratory graph analysis: A new approach for estimating the number of dimensions in psychological research. *PLoS One*, *12*(6), Article e0174035. https://doi.org/10.1371/journal.pone.0174035

Guo, J., Marsh, H. W., Morin, A. J., Parker, P. D., & Kaur, G. (2015). Directionality of the associations of high school expectancy-value, aspirations, and attainment: A longitudinal study. *American Educational Research Journal*, *52*(2), 371–402. https://doi.org/10.3102/0002831214565786

Hau, K. T., Kong, C. K., & Marsh, H. W. (2003). Chinese Self-Description Questionnaire. Cross-cultural validation and extension of theoretical self-concept models. In H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *International advances in self-research* (pp. 49–65). Information Age Publishing.

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), 1–55. https://doi.org/10.1080/10705519909540118

Hudson, G., & Christensen, A. P. (2022). *EGAnet: Exploratory graph analysis – A framework for estimating the number of dimensions in multivariate data using network psychometrics. R package version 1.1.1.* https://cran.r-project.org/web/packages/EGAnet/

Inglés, C., Torregrosa, M. S., & García-Fernández, J. M. (2009). *Self-description questionnaire II (SDQ-II): A pilot study conducted with Spanish, Portuguese, and Chinese adolescent samples* [Conference]. 2nd International Conference of Education, Research and Innovation, Madrid, Spain.

Inglés, C., Torregrosa, M. S., Hidalgo, M. D., Nuñez, J. C., Castejón, J. L., García-Fernández, J. M., & Valle, A. (2012). Validity evidence based on internal structure of scores on the Spanish version of the Self-Description Questionnaire-II. *Spanish Journal of Psychology*, *15*(1), 388–398. https://doi.org/10.5209/rev_SJOP.2012.v15.n1.37345

Jamison, L., Golino, H., & Christensen, A. P. (2022). *Metric invariance in exploratory graph analysis via permutation testing*. PsyArXiv, PPR498583. https://doi.org/10.31234/osf.io/j4rx9

Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *SemTools: Useful tools for structural equation modeling. R package version 0.5-6*. https://CRAN.R-project.org/package=semTools

Kong, C. K. (2000). *Chinese students' self-concept: Structure, frame of reference, and relation with academic achievement*. [doctoral dissertation]. The Chinese University of Hong Kong.

Kurman, J. (2003). Why is self-enhancement low in certain collectivist cultures? An investigation of two competing explanations. *Journal of Cross-Cultural Psychology*, *34*(5), 496–510. https://doi.org/10.1177/0022022103256474

Leung, K. C., Marsh, H. W., Craven, R. G., & Abduljabbar, A. S. (2016). Measurement invariance of the Self-Description Questionnaire II in a Chinese sample. *European Journal of Psychological Assessment*, *32*(2), 128–139. https://doi.org/10.1027/1015-5759/a000242

Li, H. Z., Zhang, Z., Bhatt, G., & Yum, Y. O. (2006). Rethinking culture and self-construal: China as a middle land. *Journal of Social Psychology*, *146*(5), 591–610. https://doi.org/10.3200/SOCP.146.5.591-610

Markus, H., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, *98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224

Marsh, H. W. (1984). Self-concept: The application of a frame of reference model to explain paradoxical results. *Australian Journal of Education*, *28*(2), 165–181. https://doi.org/10.1177/000494418402800207

Marsh, H. W. (1986). Verbal and math self-concepts: An internal/external frame of reference model. *American Educational Research Journal*, *23*, 129–149. https://doi.org/10.3102/00028312023001129

Marsh, H. W. (1990). The structure of academic self-concept: The Marsh/Shavelson model. *Journal of Educational Psychology*, *82*(4), 623–636. https://doi.org/10.1037/0022-0663.82.4.623

Marsh, H. W. (1992). *Self-Description Questionnaire (SDQ) II: A theoretical and empirical basis for the measurement of multiple dimensions of adolescent self-concept*. University of Western Sydney, SELF Research Centre.

Marsh, H. W. (1994). Using the national longitudinal study of 1988 to evaluate theoretical models of self-concept: The Self-Description Questionnaire. *Journal of Educational Psychology*, *86*(3), 439–456. https://doi.org/10.1037/0022-0663.86.3.439

Marsh, H. W., Abduljabbar, A. S., Abu-Hilal, M. M., Morin, A. J. S., Abdelfattah, F., Leung, K. C., Xu, M. K., Nagengast, B., & Parker, P. (2013). Factorial, convergent, and discriminant validity of TIMSS math and science motivation measures: A comparison of Arab and Anglo-Saxon countries. *Journal of Educational Psychology*, *105*(1), 108–128. https://doi.org/10.1037/a0029907

Marsh, H. W., & Craven, R. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective. *Perspectives on Psychological Science*, *1*, 133–163. https://doi.org/10.1111/j.1745-6916.2006.00010.x

Marsh, H. W., Ellis, L. A., Parada, R. H., Richards, G., & Heubeck, B. G. (2005). A short version of the Self-Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment*, *17*(1), 81–102. https://doi.org/10.1037/1040-3590.17.1.81

Marsh, H. W., & Hau, K. T. (2003). Big-fish-little-pond effect on academic self-concept. A cross-cultural (26-country) test of the negative effects of academically selective schools. *The American Psychologist*, *58*(5), 364–376. https://doi.org/10.1037/0003-066x.58.5.364

Marsh, H. W., & Hau, K. T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal-external frame of reference predictions across 26 countries. *Journal of Educational Psychology*, *96*(1), 56–67. https://doi.org/10.1037/0022-0663.96.1.56

Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: Cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing*, *6*(4), 311–360. https://doi.org/10.1207/s15327574ijt0604_1

Marsh, H. W., & O'Niell, R. (1984). Self-Description Questionnaire III (SDQ III): The construct validity of multidimensional self-concept ratings by late-adolescents. *Journal of Educational Measurement*, *21*(2), 153–174. https://doi.org/10.1111/j.1745-3984.1984.tb00227.x

Marsh, H. W., Parker, P. D., & Pekrun, R. (2019). Three paradoxical effects on academic self-concept across countries, schools, and students: Frame-of-reference as a unifying theoretical explanation. *European Psychologist*, *24*(3), 231–242. https://doi.org/10.1027/1016-9040/a000332

Merkle, E. C., You, D., & Preacher, K. J. (2016). Testing nonnested structural equation models. *Psychological Methods*, *21*(2), 151–163. https://doi.org/10.1037/met0000038

Möller, J., Pohlmann, B., Köller, O., & Marsh, H. W. (2009). A meta-analytic path analysis of the internal/external frame of reference model of academic achievement and academic self-concept. *Review of Educational Research*, *79*(3), 1129–1167. https://doi.org/10.3102/0034654309337522

Moshagen, M., & Erdfelder, E. (2016). A new strategy for testing structural equation models. *Structural Equation Modeling*, *23*, 54–60. https://doi.org/10.1080/10705511.2014.950896

Nishikawa, S., Norlander, T., Fransson, P., & Sundbom, E. (2007). A cross-cultural validation of adolescent self-concept in two cultures: Japan and Sweden. *Social Behavior and Personality*, *35*(2), 269–286. https://doi.org/10.2224/sbp.2007.35.2.269

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*. https://www.oecd-ilibrary.org/education/knowledge-and-skills-for-life_9789264195905-en

Padilla, J. L., & Benítez, I. (2014). Validity evidence based on response processes. *Psicothema*, *26*(1), 136–144. https://doi.org/10.7334/psicothema2013.259

Parker, P. D., Marsh, H. W., Ciarrochi, J., Marshall, S., & Abduljabbar, A. S. (2014). Juxtaposing mathematics self-efficacy and self-concept as predictors of long-term achievement outcomes. *Educational Psychology*, *34*(1), 29–48. https://doi.org/10.1080/01443410.2013.797339

Putnick, D., & Bornstein, M. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

R Core Team. (2022). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. http://www.R-project.org/

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*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Rodríguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods*, *21*(2), 137–150. https://doi.org/10.1037/met0000045

Ronen, S., & Shenkar, O. (2013). Mapping world cultures: Cluster formation, sources and implications. *Journal of International Business Studies*, *44*(9), 867–897. https://doi.org/10.1057/jibs.2013.42

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1–36. https://doi.org/10.18637/jss.v048.i02

Rüschenpöhler, L., & Markic, S. (2019). A mixed methods approach to culture-sensitive academic self-concept research. *Education Sciences*, *9*(3), Article 240. https://doi.org/10.3390/educsci9030240

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled difference chi-square test statistic. *Psychometrika*, *75*(2), 243–248. https://doi.org/10.1007/s11336-009-9135-y

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research*, *46*(3), 407–441. https://doi.org/10.3102/00346543046003407

Shweder, R. A., & Bourne, E. J. (1982). Does the concept of the person vary cross-culturally? In A. Marsella & G. White (Eds.), *Cultural conceptions of mental health and therapy* (pp. 97–117). Reidel.

Spencer-Rodgers, J., Boucher, H. C., Mori, S. C., Wang, L., & Peng, K. (2009). The dialectical self-concept: Contradiction, change, and holism in East Asian cultures. *Personality & Social Psychology Bulletin*, *35*(1), 29–44. https://doi.org/10.1177/0146167208325772

Viladrich, C., Angulo-Brunet, A., & Doval, E. (2017). A journey around alpha and omega to estimate internal consistency reliability. *Anales de Psicología*, *33*(3), 755–782. https://doi.org/10.6018/analesps.33.3.268401

Wästlund, E., Norlander, T., & Archer, T. (2001). Exploring cross-cultural differences in self-concept: A meta-analysis of the Self-Description Questionnaire-1. *Cross-Cultural Research*, *35*(3), 280–302. https://doi.org/10.1177/106939710103500302

Wigfield, A., & Karpathian, M. (1991). Who am I and what can I do? Children's self-concepts and motivation in achievement situations. *Educational Psychologist*, *26*(3–4), 233–261. https://doi.org/10.1207/s15326985ep2603&4_3

Yeung, A. S., Chui, H.-S., Lau, I. C., McInerney, D. M., Russell-Bowie, D., & Suliman, R. (2000). Where is the hierarchy of academic self-concept? *Journal of Educational Psychology*, *92*, 556–567. https://doi.org/10.1037/0022-0663.92.3.556

How to cite this article: Esnaola, I., Sesé, A., Azpiazu, L., & Wang, Y. (2024). Revisiting the academic self-concept transcultural measurement model: The case of Spain and China. *British Journal of Educational Psychology*, *94*, 89–111. https://doi.org/10.1111/bjep.12635