

Bifactor Invariance Analysis of Student Conceptions of Assessment Inventory

Daniel Abud Seabra Matos¹
Gavin Thomas Lumsden Brown²
Cristiano Mauro Assis Gomes³

¹Universidade Federal de Ouro Preto, Ouro Preto, MG
²The University of Auckland, Auckland, New Zealand
³Universidade Federal de Minas Gerais, Belo Horizonte, MG

Abstract

Student conceptions of the purposes of assessment are an important aspect of self-regulated learning. This study advances our understanding of the Student Conceptions of Assessment Inventory (SCoA) by examining the generalizability of the factorial structure of the SCoA using bifactor analysis and conducting cross-cultural invariance testing between Brazil and New Zealand. Eight different models were specified and evaluated, with the best model being adopted for invariance testing. This research adds to our understanding of the cross-cultural properties of the SCoA because the introduction of the bifactor model resulted in metric equivalence between countries, which had previously had only partial metric equivalence. Future studies should attempt to create more items around several SCoA constructs.

Keywords: educational assessment; factor analysis; cross-cultural comparison

Análise da Invariância Bifatorial do Inventário de Concepções de Avaliação de Estudantes

Resumo

As concepções de estudantes dos propósitos da avaliação são um aspecto importante da aprendizagem autorregulada. Este estudo avança nossa compreensão do Inventário de Concepções de Avaliação de Estudantes (CAE), pelo exame da generalização da estrutura fatorial do CAE usando análise bifatorial e realizando testes de invariância transcultural entre o Brasil e a Nova Zelândia. Oito modelos diferentes foram especificados e avaliados, com o melhor modelo adotado para o teste de invariância. Esta pesquisa acrescenta à nossa compreensão das propriedades transculturais do CAE porque a introdução do modelo bifatorial resultou em equivalência métrica entre países, que anteriormente tinham apenas equivalência métrica parcial. Estudos futuros devem tentar criar mais itens em torno de vários construtos do CAE.

Palavras-chave: avaliação educacional, análise fatorial, comparação transcultural

Análisis de la Invarianza Bifatorial del Inventario de Concepciones de Evaluación de Estudiantes

Resumen

Las concepciones de los estudiantes sobre los propósitos de evaluación, son un aspecto importante del aprendizaje autorregulado. Este estudio amplía nuestra comprensión sobre el Inventario de Concepciones de Evaluación de Estudiantes (CEE), mediante la investigación de la generalización de la estructura factorial del CEE utilizando análisis bifatorial y realizando tests de invarianza transcultural entre Brasil y Nueva Zelanda. Se especificaron y evaluaron ocho modelos diferentes, con el mejor modelo adoptado para el test de invarianza. Esta investigación aumenta nuestra comprensión de las propiedades transculturales del CEE, ya que la introducción del modelo bifatorial resultó con equivalencia métrica entre países, que anteriormente tenían sólo equivalencia métrica parcial. En el futuro, otros estudios posiblemente tratarán de crear más ítems sobre varios constructos del CEE.

Palabras clave: evaluación educativa; análisis factorial; comparación transcultural

Introduction

Self-regulation theory (Zimmerman, 2008) suggests that greater learning outcomes arise when students (a) activate prior to commencing learning a variety of self-motivation beliefs, (b) control and observe their own performance, and (c) reflect upon and evaluate the self, causes, and outcomes. The self-evaluative phase then iteratively contributes to the activation of various self-motivation beliefs. Hence, self-regulation

of learning requires understanding the purposes and consequences of evaluation, not just controlling learning processes. Self-regulation theory also indicates that certain kinds of cognitions, feelings, and actions lead to increased learning outcomes (Boekaerts & Cascallar, 2006). For example, taking responsibility for one's actions (Zimmerman, 2008), having positive affect in learning (Pekrun, Goetz, Titz, & Perry, 2002), and making use of feedback (Hattie & Timperley, 2007) are adaptive self-regulating responses. In contrast, blaming

external, uncontrollable factors (Weiner, 2000), prioritising emotional well-being (Boekaerts & Corno, 2005), and ignoring learning-related evaluations are examples of maladaptive, non-regulating responses that lead to decreased academic achievement.

Self-regulated learning models incorporate reflection upon performance as an important facet; for higher education students the majority of the performance information is derived from formal assessment events. Assessment processes influence students' behaviors, learning, studying, and achievement (Entwistle, 1991; Peterson & Irving, 2008; Struyven, Dochy, & Janssens, 2005). Hence, student opinions about the nature and purpose of assessment are likely to influence student learning-related behaviours and educational achievement. Thus, an important aspect of self-regulation, often overlooked in learning research, is student conceptions of assessment.

Student Conceptions of Assessment

Literature reviews have demonstrated that students are aware of a number of purposes for assessment (Brown, 2011; Brown & Hirschfeld, 2008; Harris, Harnett, & Brown, 2009; Weekers, Brown, & Veldkamp, 2009). These include awareness that assessment can (a) help improve performance, (b) be negative and ignored, (c) trigger emotional responses, (d) improve classroom climate, (e) evaluate school quality, (f) predict intelligence and future career success, (g) hold students accountable for learning. Further, it would seem as students mature, and especially upon entering secondary schooling with its certification assessment, they tend to become more negative about the function of assessment (Harris, Harnett, & Brown, 2009).

In accordance with self-regulation frameworks, statistically significant increases in academic performance among high school students in New Zealand have been reported for various adaptive beliefs (Brown & Hirschfeld, 2007, 2008; Brown, Peterson, & Irving, 2009). Increased achievement has been reported when students endorse assessment makes students accountable, assessment is good for me, assessment is valid; assessment makes students accountable; and assessment improves student learning and teacher instruction. In contrast, negative relations were found to performance on standardised tests of reading comprehension and mathematics for the factors assessment was bad, unfair, or irrelevant/ignored. Similarly, factors identifying external attributions (e.g., assessment indicates school quality or predicts student future) had negative

relations to academic performance. Furthermore, factors focused on well-being (e.g., assessment is fun or enjoyable, assessment improves class environment) had negative regressions towards achievement. The proportion of variance in academic performance explained by the conceptions of assessment factors was not trivial, with impact on academic achievement measures reaching, on average, moderate effect sizes (Brown, 2011).

The Student Conceptions of Assessment Inventory

The Student Conceptions of Assessment inventory was developed with New Zealand secondary school students. The SCoA-VI summarises student conceptions of assessment as four inter-correlated constructs (i.e., "Assessment Improves Learning and Teaching [Improvement]", "Assessment Relates to External Factors [External]", "Assessment has Affective Benefit [Affect]", and "Assessment is Irrelevant [Irrelevance]"). Note that details of the SCoA including dictionary of items and New Zealand data files are available at figshare.com (Brown, 2017).

The Improvement conception reflects an adaptive, self-regulating response consisting of two first-order factors (i.e., five items related to students using assessment to evaluate, plan, and improve their learning activities and six items related to teachers interpreting students' assessed performances so as to improve their instruction). The External conception likewise has two first-order factors (i.e., four items in which assessments measure students' future and intelligence and two items in which assessment measures the quality of schooling). These perceptions relate to a lack of personal autonomy or control or external locus of control attributions (i.e., it is about the school and my future) which are maladaptive, non-regulating beliefs. The Affect conception also has two first-order factors (i.e., two items in which assessment is a personally enjoyable experience and six items in which assessment benefits the class environment). These aspects of assessment relate to a sense of 'well-being' and are notionally maladaptive (Boekaerts & Corno, 2005). The Irrelevance conception, consisting of three items on assessment being ignored and a first-order factor in which five items capture students' tendency see assessment as bad or unfair, expresses a maladaptive response since rejecting the validity of assessment lessens a growth-oriented response to being evaluated.

Two validity studies with university students showed that the SCoA factors related to motivational constructs in a manner consistent with self-regulation

theory. Hirschfeld and von Brachel (2008) used a German translation of the SCoA-II (Brown & Hirschfeld, 2008) with undergraduate psychology students to examine their learning behaviours for assessment. In a good fitting model, they found that three of the SCoA factors predicted individualised learning strategies (e.g., mind mapping or summary writing). The paths from student and university accountability predicted increased self-reported usage of these strategies, while the enjoyment affective response acted as a negative predictor of individualised learning strategies. This suggests that agreement with the evaluative purpose of assessment acts adaptively to increase personal responsibility in learning behaviour, while emphasis on the affective domain appears inimical to the growth-related pathway.

The full SCoA version 6 was used with students at one American university which annually administers a low-stakes system evaluation test (Wise & Cotten, 2009). Meaningful relations between SCoA and two measures of motivation (i.e., time taken to respond to a computer administered test—response time effort and attendance at the low-stakes testing day) were found. Less guessing (i.e., longer response times) was associated with greater belief that assessment leads to improvement, while more guessing was predicted by lower Affective benefit and greater Irrelevance of assessment. Attendance on the day of the low-stakes test was considerably higher for those who endorsed improvement and affect and rejected irrelevance.

The Students' Conceptions of Assessment version 6 (SCoA-VI) uses 33 self-report items in which participants rate their level of agreement using an ordinal agreement, six-point, positively-packed rating scale (Lam & Klockars, 1982), with two negative options (strongly disagree, disagree) and four positive options (slightly, moderately, mostly, and strongly agree).

Previous cross-cultural studies of the SCoA

A cross-cultural study with higher education students in Hong Kong, China, New Zealand, and Brazil was reported recently (Brown, 2013). The SCoA inventory was broken into two halves to reduce fatigue among Hong Kong and China university students who were also given new experimental items. Additionally, a previous study with the SCoA in Brazil had eliminated one item related to assessment telling parents about student performance from the Student Future factor (Matos, Cirino, Brown, & Leite, 2013). This meant that comparisons between Brazil and New Zealand

university student responses was done in two parts: Part A consisted of two items for assessment predicts student future and the complete assessment is Irrelevant factor of eight items, while Part B had two items for School Quality, 11 items for Improvement in two factors, and eight items for Social and Affective Benefit in two factors. Four group invariance testing, using maximum likelihood (ML) estimation, found that Part A only had configural invariance, while Part B was completely invariant. Pair-wise comparison among the four samples showed in Part A that the Brazil group differed from all others, suggesting systematic differences may exist in Brazil. An alternative explanation could lie in the use of maximum likelihood (ML) estimation which is intended for continuous variables.

In a two-country comparison (Brazil vs. New Zealand) of the full SCoA inventory (Matos & Brown, 2015) a different measurement approach was used. The weighted least squares mean and variance estimation (WLSMV) procedure was used to account for the ordinal nature of the response format and all higher order factors were removed to test an eight-factor inter-correlated model. The fit of the revised model for each sample was acceptable, but the two-group invariance test indicated that the model lacked configural and metric invariance. About half of the items had large differences in item regression weights, as did half of the factor inter-correlations. Large mean score effect sizes ($d > .60$) were seen in favour of New Zealand students for Teacher Improvement, Class Environment, and Student Future, while Bad was in favour of Brazil students.

Hence, while the SCoA seems to have some promising characteristics in terms of cross-cultural invariance, perhaps related to the similarity of assessment cultures in universities world-wide, there are simultaneously differences related to local contexts.

Higher Education Contexts

New Zealand. Until the 11th year of schooling there are no high-stakes assessments in New Zealand. There is much assessment, including the use of standardised testing, but this is school-controlled, done largely for formative and reporting purposes, and there are no negative consequences for schools, teachers, or students as a result of poor performance. All students meeting standards are eligible for publicly funded higher education.

There are 8 public universities in New Zealand and no private universities, although there are a plethora

of private trades and vocations-oriented providers of post-schooling training. University education is highly subsidised by the government, with students contributing about 10-20% of full tuition cost in fees. Entry is via completion of recognised secondary school qualifications, which for most students consists of both internally and externally-administered assessments. However, entry is open for all adults aged 20 plus provided foundation courses are passed by those not having normal secondary school qualifications. Faculties and programmes within universities may set higher entry standards; usually in the most competitive subject areas such as medicine, engineering, or commerce.

Brazil. Relative to the size of its economy, Brazil does not spend much on education. For instance, in higher education, the amount spent per student in Brazil is US\$ 11.8 thousand, while the OECD average is US\$ 16.1 thousand (OECD, 2018). Nonetheless, the number of students enrolled in tertiary level has increased from about 1,500,000 students in 1991 to over 7,000,000 in 2013. There are 2391 universities (301 public and 2090 private institutions), with only about 2,000,000 enrolled in public and 5.300,000 in private universities. Hence, tertiary level education in Brazil is characterised greatly by students in private institutions. However, recently the government has created several scholarship programs for students in private institutions. Additionally, during the last years, quota spaces have been set aside in public universities.

Brazil is a largely examination driven culture in which assessment is used as a student accountability mechanism. Students are evaluated at the end of the elementary, middle, and high school education stages with a standardized test. Brazil has a National System of Higher Education Assessment (SINAES), which includes assessment of student performance (National Exam of Student Performance - ENADE), institutional evaluation, and evaluation of courses.

Method

Because previous studies have demonstrated non-invariance, this study adds to our understanding of whether differences in methods of analysis might have contributed to the lack of invariance. For example, different model structures (i.e., hierarchical vs. first-order only) have resulted in different results. The lack of invariance in the original New Zealand model, other than the ecological argument that contextual differences in how assessment is implemented and consequently

experienced cause non-invariance, may be resolved by using a bifactor method of analysis.

Bifactor models specify a general and domain-specific group factors. The general factor loads on all items and explains the common variance between items across different factors, and explains the item inter-correlations of all items. The group factors are additional to the general factor, and measure the shared variance between items of the same factor after partialling out the general factor. The group factors, thus, measure what is left of the different factors, after controlling for the general factor. A previous attempt at bifactor analysis of the SCoA (Weekers, Brown, & Veldkamp, 2009) used only a four-factor model (i.e., Irrelevance, Improvement, Affect, & External) and only used New Zealand high school data. That study found that the bifactor approach was plausible since a majority of the 33 items had loadings $\geq .35$ from the general factor.

It is also worth noting that most of the published studies with the SCoA have used the maximum likelihood estimator. It can be argued that this is a valid approach because the response scale has more than the minimum five options shown to make an ordinal response scale equivalent to continuous (Finney & DiStefano, 2006). However, the response options are ordered categories and it may prove superior to use an estimator better designed for ordinal options. The Weighted Least Squares Means and Variance (WLSMV) estimator uses an item response theory approach to determining the probability value of each score response threshold, thus placing each response option on a continuous latent scale (Muthén & Muthén, 2012). The WLSMV estimator certainly takes a more conservative approach to determining the fit of a model than maximum likelihood estimation procedures, suggesting it might be more resistant to Type I errors of accepting that the model fits the data when it does not (Li, 2016).

Thus, this study advances our understanding of the SCoA by examining all eight factors of the SCoA using bifactor analysis with the WLSMV estimator and conducting cross-cultural invariance testing also using the WLSMV estimator. We compare the invariance models through the recommendations of Cheung and Rensvold (2002) that a CFI difference between two models higher than .01 indicates that the more constrained model does not fit the data as well as the less constrained model.

Participants

Since the two samples being compared had quite different research agendas there were few common

demographic characteristics available. However, both groups consisted completely of undergraduate students. Consistently, nearly twice as many females participated as males in each sample; no doubt consistent with the greater tendency for voluntary participation among women (Table 1). Only the Brazil sample met the conventional expectation of large sample size with more than 400 participants (Boomsma & Hooglund, 2001). The Brazilian sample was older on average but with much less university education experience than the New Zealand sample. It is also worth noting that in New Zealand all students were enrolled in one publicly funded institution; whereas, a mixture of public and private enrolments were seen in Brazil. The Brazilian sample reflects the contextual reality, since the majority of students in Brazil are engaged in private institutions. These various experience and institutional factors may contribute to patterns of equivalence or non-equivalence.

Instrument

In an adaptation and validation of the Students' Conceptions of Assessment (SCoA) - version VI for the Brazilian context, Matos et al. (2013) translated the inventory into Portuguese. Afterwards, three independent researchers evaluated the translation quality via back translation. Additionally, cognitive interviews were made with 12 undergraduate students from public and private universities. Only one item was eliminated from the Student Future factor (i.e., item

33- Assessment tells my parents how much I've learnt) on the basis that Brazilian tertiary students believed this item only made sense for younger students (Matos et al., 2013). Hence, for comparison purposes in this paper, item 33 from the New Zealand data has been excluded. Supplementary Appendix 1 provides the items by factor in both languages.

Analysis

The motive of the study was to find a model that retained as much of the original SCoA structure as possible while maximizing the probability that the model would fit the data from both samples equally well. The models analyzed in this study were derived initially from the structure of the original SCoA-VI model as published in two different studies (Models 1-3). A combination of these models was used to introduce the bifactor approach (Models 4-6). Then, because of relatively poor fit of the models to both data sets, improved fit was sought by introducing pairs of covarying item residuals identified by Lagrange modification indices and by exploratory factor analysis of the SCoA with the Brazilian data (Models 7-8). The eight different models specified and evaluated were:

- a) Eight correlated 1st-order factors representing the specific dimensions of the SCoA-VI inventory as specified in Matos et al. (2013);
- b) Hierarchical model #1 in which four correlated 2nd-order factors predicted eight 1st-order factors as

Table 1.
Participant Characteristics by Jurisdiction

Demographic Characteristic	Jurisdiction	
	Brazil	New Zealand
Sex		
Female	495	212
Male	204	101
Missing	3	
Age <i>M (SD)</i>	24.49 (5.42)	20.80 (4.58)
Current semester of study	70.6% 3rd or less	58.2% 3rd or less
Type of Higher Education Institution		
Public	297	313
Private	405	
Total sample	702	313

- specified in Brown, Irving, Peterson, & Hirschfeld (2009);
- c) Hierarchical model #2 in which four correlated 2nd-order factors predicted seven 1st-order factors as specified in Brown, Peterson, and Irving (2009);
 - d) Bifactor model #1 consisting of a general factor predicting all items plus four 2nd-order factors from Models 2 and 3, with no correlations between factors, as specified in Weekers, Brown, & Veldkamp (2009);
 - e) Bifactor model #2 consisting of a general factor predicting all items plus the eight uncorrelated factors from Model 1;
 - f) Bifactor model #3 consisting of a general factor predicting all items plus hierarchical Model #1 (i.e., four 2nd-order factors predicting eight 1st-order factors);
 - g) Bifactor model #4 consisting of a general factor predicting all items plus five correlated factors with one pair of residuals covarying;
 - h) Bifactor model #5 consisting of a general factor predicting all items plus three correlated factors with three pairs of residuals covarying.

Data analysis first evaluated the fit of each model, using the mean and variance-adjusted weighted least squares estimator (WLSMV). Then, configural (unconstrained model) and metric invariance (equivalent regression weights), also using the WLSMV estimator, between the two samples was evaluated for models that had converged with adequate fit. Strong invariance (i.e., equivalent regression weights, intercepts, factor means, thresholds, and residuals) was tested for the best fitting model. All analyses were performed using Mplus version 7.0.

The following fit indexes were used: the comparative factor index (CFI), the root mean square error of approximation (RMSEA), gamma hat, the weighted root mean residual (WRMR). A good data fit occurs when gamma hat and CFI are ≥ 0.95 and RMSEA $< .06$ (Hu & Bentler, 1999; Schumacker & Lomax, 2004). CFI and gamma hat values between 0.90 and 0.95 suggest an acceptable data fit, as well as RMSEA values between 0.06 and 0.09. Values outside these ranges suggest the model does not fit the data sufficiently to be accepted. Adequate fit is indicated when the weighted root-mean-square residual (WRMR) is close to 1.00, though this is an experimental fit index and little is known concerning values that indicate rejection (Yu,

2002). We use the recommendations of Cheung and Rensvold (2002) that $\Delta CFI \geq 0.01$ indicates that the more constrained model produces a worse data fit than the less constrained model. In this case, the lack of invariance of the model indicates the less constrained model is preferred.

Results

The eight models were evaluated for the Brazilian and New Zealand samples separately and fit values inspected (Table 2). For Models 1-3, based on the original SCoA-VI specification, fit was generally acceptable for both groups.

Only Bifactor Model #4, using the four main factors, reached convergence, but with unacceptable data fit in New Zealand sample (see CFI index in Table 2). Hence, the development of Models 7 and 8 was necessary to determine if an underlying bifactor model was present. Although Model 7 fit the Brazilian data well, it was non-converging for the New Zealand data. Model 8 (i.e., bifactor plus three factors and three covarying pairs of residuals) had good data fit in the Brazilian sample and acceptable fit to the New Zealand data.

Based on these results, two-group invariance tests were run with the WLSMV for all Models except #5 to 7.

Table 3 shows configural invariance fit of the tested models. Model 2 did not converge and Model 4 had an unacceptable data fit (see CFI in Table 3). Models 1, 3, and 8 all had acceptable CFI and RMSEA values, but only Model 8 also had gamma hat > 0.90 and an RMSEA value close to the 0.06 threshold.

The four proper solution models (i.e., 1, 3, 4, & 8) were analyzed for metric equivalence (Table 4). Model 1 presented a difference in CFI $> .01$ between the metric equivalent model and the configural model, indicating that model 1 is not invariant in terms of regression weights. Model 3 showed $\Delta CFI = .01$, indicating that this model is not invariant in terms of regression weights. On the other hand, model 4 showed an unacceptable data fit (see CFI in Table 4). The only metric invariant model was Model 8 (i.e., Bifactor Model #5 with three unique factors and three pairs of correlated residuals). This model was subsequently tested strong invariance (i.e., equivalent regression weights, intercepts, factor means, thresholds, and residuals) and was found to have unacceptable fit ($\chi^2 [1092] = 4671.48$; $\chi^2/df = 4.28$; RMSEA = .080; CFI = .87; gamma hat = .82) and was non-equivalent between groups ($\Delta CFI > .01$). Hence,

Table 2.
WLSMV Model Fit for Brazil and New Zealand Separately

Model	Brazil										New Zealand						
	<i>k</i>	χ^2	df	RMSEA	CI lo 90%	CI hi 90%	CFI	WRMR	gamma hat	χ^2	df	RMSEA	CI lo 90%	CI hi 90%	CFI	WRMR	gamma hat
Model 1: 8 correlated 1st-order factors																	
	32	2043.94	436	0.072	0.069	0.076	0.92	1.67	0.88	1072.19	436	0.068	0.063	0.074	0.92	1.27	0.89
Model 2: Hierarchical Model #1																	
	32	2167.64	450	0.074	0.071	0.077	0.91	1.79	0.87	1166.10	450	0.071	0.066	0.076	0.91	1.39	0.88
Model 3. Hierarchical Model #2																	
	32	2193.10	452	0.074	0.071	0.077	0.91	1.83	0.87	1150.35	452	0.070	0.065	0.075	0.91	1.40	0.88
Model 4: Bifactor #1 plus 4 broad Factors																	
	32	1885.90	432	0.069	0.066	0.072	0.93	1.58	0.89	1400.46	432	0.085	0.080	0.090	0.88	1.53	0.84
Model 5. Bifactor #2 plus 8 specific factors																	
	32																no convergence
Model 6. Bifactor #3 plus 4 broad factors plus 8 specific factors																	
	32																no convergence
Model 7. Bifactor #4 plus 5 specific factors plus one covariance																	
	32	1376.60	439	0.055	0.052	0.058	0.95	1.34	0.92								no convergence
Model 8: Bifactor #5 plus 3 factors plus three covariances																	
	32	1372.35	436	0.055	0.052	0.059	0.95	1.31	0.92	1118.70	436	0.071	0.066	0.076	0.91	1.32	0.88

Note. *k* = number of items; RMSEA= root mean square error of approximation; CI lo=lower bound of confidence interval; CI hi=upper bound of confidence interval; CFI=comparative fit index; WRMR=Weighted root mean residual; WLSMV=Weighted Least Squares Mean and Variance.

we conclude that there is only metric equivalence in the best fitting two-group model between New Zealand and Brazil student responses to the SCoA.

Discussion

The best model discovered in this study (i.e., Model 8: Bifactor #5 containing a general factor predicting all items plus three factors and three pairs of covarying residuals) adds to our understanding of previous data analyses of the SCoA inventory. The bifactor model seems to have identified correctly, as did previous research (Weekers, Brown, & Veldkamp, 2009), that there is a general latent trait accounting for substantial covariance among the SCoA items. Clearly, when thinking about the purposes of the items, there

is a common latent trait driving responses, perhaps the reason why the Improvement, External, and Affect factors are positively correlated and all inversely correlated with Irrelevance. The presence of three domain-specific factors reinforces the claim that the purposes do have additional meaning to the general function of assessment, strengthening the claim that the SCoA is multi-dimensional. Hence, this study advances our understanding of the SCoA dimensionality.

While the introduction of correlated residuals has been given warrant (Byrne, 2001) this is a step that ought to be taken cautiously since it rests on the presumption that the unexplained variance of one item systematically covaries with the unexplained variance of another item, but not with other items. A more cautious approach considers that unexplained variance

Table 3.

Fit for Unconstrained Configural Invariance of Two-Group (Brazil-New Zealand) Data by Model

Model	k	χ^2	df	RMSEA	CI lo 90%	CI hi 90%	CFI	gamma hat	WRMR
Model 1: 8 correlated factors	64	3002.72	880	0.069	0.066	0.072	0.92	0.88	2.17
Model 2: Hierarchical Model #1					no convergence				
Model 3: Hierarchical Model #2	64	3308.56	914	0.072	0.069	0.075	0.91	0.87	2.42
Model 4: Bifactor #1 plus 4 broad Factors	64	5107.87	920	0.095	0.092	0.097	0.85	0.79	3.09
Model 8: Bifactor #5	64	2503.85	876	0.061	0.058	0.063	0.94	0.91	1.90

Note. All models calculated with WLSMV estimator; k =number of items; RMSEA=root mean square error of approximation; CI lo=lower bound of confidence interval; CI hi=upper bound of confidence interval; CFI=comparative fit index; WRMR=Weighted root mean residual; values in bold indicate acceptable fit quality.

Table 4.

Metric Equivalence Tests of Two-Group (Brazil-New Zealand) Data by Model

Model	Unconstrained Configural Equivalence			Metric Equivalence Constrained			Fit Difference after Equivalence Constraint
	χ^2	df	CFI	χ^2	df	CFI	Δ CFI
Model 1: 8 correlated factors	3002.72	880	0.92	3641.38	904	0.90	>.01
Model 3: Hierarchical Model #2	3308.56	914	0.91	3706.25	942	0.90	.01
Model 4: Bifactor #1 plus 4 broad Factors	5107.87	920	0.85	5129.30	948	0.85	<.01
Model 8: Bifactor #5	2503.85	876	0.94	2504.01	932	0.94	<.01

Note. All models calculated with WLSMV estimator; CFI=comparative fit index; values in bold indicate equivalence.

has a zero relationship with all other error variances. Nonetheless, the three pairs of error covariances in the preferred model do not appear completely random. The three pairs were:

- Pair 1. Items 8 and 9 from Assessment Improves Teaching;
- Pair 2. Items 6 and 31 from Assessment is Enjoyable; and
- Pair 3. Items 24 and 11 from Assessment Evaluates School Quality.

It is clear that the pairs of items came from matching SCoA factors and had very similar wording. This suggests that either there were insufficient items to detect the intended factor or else the items function as ‘bloated specifics’ artificially creating a scale because of repeated wording (Kline, 1994). Since, Model 1 with eight specific factors had acceptable fit for each group separately, it is likely that the factors do exist and have simply been insufficiently operationalized with the introduction of the bifactor approach. This suggests future studies should attempt to create more items around these three constructs to ensure that the specific contribution the factors make can be detected even after the shared general factor is introduced. It also probable that greater specificity in these constructs would improve invariance analysis results.

This research adds to our understanding of the cross-cultural properties of the SCoA because a previous two-country comparison of the same data sets (i.e., Brazil and New Zealand) (Matos & Brown, 2015) showed lack of configural and metric invariance. The additional introduction of the bifactor combined with the restructuring of the unique factors in the SCoA and the introduction of three pairs of correlated residuals resulted in metric equivalence between countries. This indicates that, while starting values are different for Brazilian and New Zealand students, the regression slopes between the latent traits and the items differ only by chance. This suggests that the SCoA inventory may have cross-cultural validity between countries with quite different higher education arrangements, perhaps because of the fundamentally similar relationship assessment plays in higher education (i.e., it evaluates student learning).

It seems reasonably safe to conclude that any difference in factor means and inter-factor correlations between the New Zealand and Brazil samples is a function of differences in populations and environments

rather than deficiencies in estimation method or model specification. The common model across samples fits equally well and is partially invariant; hence, the differences are best explained by reference to different ecologies rather than deficient measurement.

References

- Boekaerts, M., & Cascallar, E. (2006). How far have we moved towards the integration of theory and practice in self regulation? *Educational Psychology Review*, 18(3), 199-210. doi:10.1007/s10648-006-9013-4
- Boekaerts, M., & Corno, L. (2005). Self-regulation in the classroom: A perspective on assessment and intervention. *Applied Psychology: An international review*, 54(2), 199-231.
- Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. Du Toit, & D. Sorbom (Eds.), *Structural equation modeling: Present and future* (pp. 139-168). Lincolnwood, IL: Scientific Software International.
- Brown, G. T. L. (2013). Student conceptions of assessment across cultural and contextual differences: University student perspectives of assessment from Brazil, China, Hong Kong, and New Zealand. In G.A.D. Liem & A. B. I. Bernardo (Eds.), *Advancing Cross-cultural Perspectives on Educational Psychology: A Festschrift for Dennis McInerney* (pp. 143-167). Charlotte, NC: Information Age Publishing.
- Brown, G.. (2017, February 23). Students Conceptions of Assessment (Version 2). figshare. <https://doi.org/10.17608/k6.auckland.c.3694258.v2>
- Brown, G. T. L., & Hirschfeld, G. H. F. (2007). Students' conceptions of assessment and mathematics: Self-regulation raises achievement. *Australian Journal of Educational & Developmental Psychology*, 7, 63-74.
- Brown, G. T. L., & Hirschfeld, G. H. F. (2008). Students' conceptions of assessment: Links to outcomes. *Assessment in Education: Principles, Policy & Practice*, 15(1), 3-17. doi: 10.1080/09695940701876003
- Brown, G. T. L., Irving, S. E., Peterson, E. R., & Hirschfeld, G. H. F. (2009). Use of interactive-informal assessment practices: New Zealand secondary students' conceptions of assessment. *Learning & Instruction*, 19(2), 97-111. doi: 10.1016/j.learninstruc.2008.02.003

- Brown, G. T. L., Peterson, E. R., & Irving, S. E. (2009). Beliefs that make a difference: Adaptive and maladaptive self-regulation in students' conceptions of assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 159-186). Charlotte, NC: Information Age Publishing.
- Byrne, B. M. (2001). *Structural Equation Modeling with AMOS: Basic Concepts, Applications, and Programming*. Mahwah, NJ: LEA.
- 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. DOI: 10.1207/S15328007SEM0902_5
- Entwistle, N. J. (1991). Approaches to learning and perceptions of the learning environment: Introduction to the special issue. *Higher Education*, 22, 201-204.
- Finney, S. J., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G. R. Hancock & R. D. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269-314). Greenwich, CT: Information Age Publishing.
- Hattie, J., & Timperley, H. (2007). The power of feedback. *Review of Educational Research*, 77(1), 81-112.
- Hirschfeld, G. H. F., & von Brachel, R. (2008, July). *Students' conceptions of assessment predict learning strategy-use in higher education*. Paper presented at the Biannual Conference of the International Test Commission (ITC), Liverpool, UK.
- Hu, L.T. & 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.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.
- Lam, T. C. M., & Klockars, A. J. (1982). Anchor point effects on the equivalence of questionnaire items. *Journal of Educational Measurement*, 19(4), 317-322.
- Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods*, 48(3), 936-949. doi:10.3758/s13428-015-0619-7
- Matos, D. A. S., & Brown, G. T. L. (2015). Comparing university student conceptions of assessment: Brazilian and New Zealand beliefs. In C. Carvalho & J. Conboy (Eds.), *School Feedback, Identity and Trajectories: Dynamics and Consequences* (pp. 177-194). Lisbon, Portugal: Universidade de Lisboa, Instituto de Educação.
- Matos, D. A. S., Cirino, S. D., Brown, G. T. L., & Leite, W. L. (2013). A avaliação no ensino superior: Concepções múltiplas de estudantes Brasileiros [Assessment in higher education: Multiple conceptions of Brazilian students]. *Estudos em avaliação educacional*, 24(54), 172-193. doi: 10.18222/eae245420131907
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide* (7th ed.). Los Angeles, CA: Muthén & Muthén.
- OECD (2018). *Education at a Glance 2018: OECD Indicators*, OECD Publishing, Paris.
- Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic emotions in students' self-regulated learning and achievement: A program of qualitative and quantitative research. *Educational Psychologist*, 37(2), 91-105.
- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18(3), 238-250.
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling*. London: Lawrence Erlbaum Associates.
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: A review. *Assessment & Evaluation in Higher Education*, 30(4), 325-341.
- Weekers, A. M., Brown, G. T. L., & Veldkamp, B. P. (2009). Analyzing the dimensionality of the Students' Conceptions of Assessment (SCoA) inventory. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning*. (pp. 133-157). Charlotte, NC US: Information Age Publishing.
- Weiner, B. (2000). Intrapersonal and interpersonal theories of motivation from an attributional perspective. *Educational Psychology Review*, 12, 1-14.

- Wise, S. L., & Cotten, M. R. (2009). Test-taking effort and score validity: The influence of student conceptions of assessment. In D. M. McInerney, G. T. L. Brown, & G. A. D. Liem (Eds.), *Student perspectives on assessment: What students can tell us about assessment for learning* (pp. 187-205). Charlotte, NC: Information Age Publishing.
- Yu, C.-Y. (2002). *Evaluating cutoff criteria of model fit indices for latent variable models with binary and continuous outcomes* (unpublished doctoral dissertation). University of California, Los Angeles, Los Angeles, CA.
- Zimmerman, B. J. (2008). Investigating self-regulation and motivation: Historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45(1), 166-183.
- Recebido em: 02/06/2018
Reformulado em: 10/11/2018
Aprovado em: 07/12/2018

Supplementary Appendix 1
SCoA-VI items by factor and language

Item	English	Portuguese
<i>Irrelevance: Assessment is Bad</i>		
3	Assessment is unfair to students	A avaliação é injusta com os alunos
13	Assessment interferes with my learning	A avaliação interfere no meu aprendizado
18	Teachers are over-assessing	Os professores avaliam exageradamente
22	Assessment results are not very accurate	Os resultados da avaliação não são muito exatos
26	Assessment is value-less	A avaliação é sem valor
<i>Irrelevance: Assessment is Irrelevant</i>		
7	I ignore assessment information	Eu ignoro as informações da avaliação
29	I ignore or throw away my assessment results	Eu desconsidero os meus resultados de avaliação
32	Assessment has little impact on my learning	A avaliação tem um impacto pequeno no meu aprendizado
<i>Affect/Benefit: Assessment Helps Class Climate</i>		
2	Assessment encourages my class to work together and help each other	A avaliação encoraja a minha turma a trabalhar junta e a ajudar uns aos outros
12	Assessment motivates me and my classmates to help each other	A avaliação me motiva e aos meus colegas a ajudarem uns aos outros
17	Our class becomes more supportive when we are assessed	A nossa turma se dá mais apoio quando nós somos avaliados
21	When we do assessments, there is a good atmosphere in our class	Quando nós fazemos avaliações, existe um bom clima na nossa turma
25	Assessment makes our class cooperate more with each other	A avaliação faz a nossa turma cooperar mais uns com os outros
28	When we are assessed, our class becomes more motivated to learn	Quando nós somos avaliados, a nossa turma se torna mais motivada para aprender
<i>Affect/Benefit: Assessment is Enjoyable</i>		
6	Assessment is an engaging and enjoyable experience for me	A avaliação é uma experiência envolvente e agradável para mim
31	I find myself really enjoying learning when I am assessed	Eu realmente aprecio o aprendizado quando eu sou avaliado
<i>External: Assessment Predicts Student Future</i>		
4	Assessment results show how intelligent I am	Os resultados da avaliação mostram o quanto eu sou inteligente
16	Assessment results predict my future performance	Os resultados da avaliação predizem o meu desempenho futuro
20	Assessment is important for my future career or job	A avaliação é importante para a minha carreira futura ou emprego

(Continue)

Supplementary Appendix 1
SCoA-VI items by factor and language (Continuation)

Item	English	Portuguese
<i>External: Assessment Holds Schools Accountable</i>		
11	Assessment provides information on how well schools are doing	A avaliação fornece informação sobre como as escolas estão indo
24	Assessment measures the worth or quality of schools	A avaliação mede a qualidade das escolas
<i>Improvement: Assessment Improves Student Learning</i>		
1	I pay attention to my assessment results in order to focus on what I could do better next time	Eu presto atenção nos meus resultados de avaliação para me concentrar no que eu posso melhorar da próxima vez
10	I make use of the feedback I get to improve my learning	Eu faço uso do <i>feedback</i> que recebo para melhorar meu aprendizado
14	I look at what I got wrong or did poorly on to guide what I should learn next	Eu observo o que eu fiz de errado ou de maneira insuficiente para guiar o que eu deveria aprender em seguida
15	I use assessments to take responsibility for my next learning steps	Eu uso as avaliações para assumir responsabilidade para as minhas próximas etapas de aprendizagem
19	I use assessments to identify what I need to study next	Eu uso as avaliações para identificar o que eu preciso estudar em seguida
<i>Improvement: Assessment Improves Teaching</i>		
5	Assessment helps teachers track my progress	A avaliação ajuda os professores a acompanhar o meu progresso
8	Assessment is a way to determine how much I have learned from teaching	A avaliação é uma forma de determinar o quanto eu aprendi do ensino
9	Assessment is checking off my progress against achievement objectives or standards	A avaliação averigua o meu progresso em comparação com os objetivos de aprendizagem
23	My teachers use assessment to help me improve	Os meus professores usam a avaliação para me ajudar a melhorar
27	Teachers use my assessment results to see what they need to teach me next	Os professores usam os meus resultados da avaliação para ver o que eles precisam me ensinar em seguida
30	Assessment measures show whether I can analyse and think critically about a topic	A avaliação mostra se eu posso analisar e pensar criticamente sobre um assunto

Acknowledgments:

The third author had a Research Productivity Fellowship granted by Brazilian National Council for Scientific and Technological Development – CNPq.

About the authors:

Daniel Abud Seabra Matos holds a degree in Psychology and a Master's degree in Education from Universidade Federal de Minas Gerais, and a PhD in Education from UFMG, with a Sandwich period at the University of Florida. Postdoctoral at Universidade of Beira Interior, Portugal. Professor of the Department of Education and advisor in the Graduate Program of Education at Universidade Federal de Ouro Preto. Associate Editor of the Journal *Estudos em Avaliação Educacional*.

ORCID: <https://orcid.org/0000-0001-7955-4302>

E-mail: danielmatos@ufop.edu.br

Gavin Thomas Lumsden Brown obtained his Ph.D. in Education at the University of Auckland (UoA), New Zealand. Associate Dean Postgraduate (Research) and Director of the Quantitative Data Analysis and Research Unit in the Faculty of Education and Social Work at UoA. Affiliated Professor, Umea University, Sweden. Honorary Professor, Education University of Hong Kong. Chief Section Editor: *Frontiers in Education: Assessment, Testing, and Applied Measurement*.

ORCID: <http://orcid.org/0000-0002-8352-2351>

E-mail: gt.brown@auckland.ac.nz

Cristiano Mauro Assis Gomes is a coordinator of the Cognitive Architecture Research Laboratory (LAICO) at UFMG. Works mainly on the themes: architecture and cognitive development, construction and validation of psychological and educational measuring instruments, cognitive intervention and development, machine learning and psychometrics. Associate Researcher of the National Science Network for Education (CpE). CNPq Research Fellow, level 2, triennium 2014-2016, and triennium 2017-2019.

ORCID: <https://orcid.org/0000-0003-3939-5807>

E-mail: cristianomaurogomes@gmail.com

Contact:

Departamento de Educação - Instituto de Ciências Humanas e Sociais/ICHS - UFOP.

Rua do Seminário, s/n. Centro.

Mariana-MG, Brasil

CEP: 35420-000