Available online at
http://www.anpad.org.br/bar

BAR, Curitiba, v. 6, n. 4, art. 4,
pp. 331-353, Oct./Dec. 2009

Proposal of a Validation Framework for a New Measurement Model and its Application to the Export Performance Construct

Jorge Carneiro *
E-mail address: jorgemtc@iag.puc-rio.br
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, RJ, Brazil.

Angela da Rocha
E-mail address: amc.darocha@gmail.com
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, RJ, Brazil.

Jorge Ferreira da Silva
E-mail address: shopshop@iag.puc-rio.br
Pontifícia Universidade Católica do Rio de Janeiro
Rio de Janeiro, RJ, Brazil.

ABSTRACT

This paper presents a comprehensive and integrated set of validation procedures to assess the satisfactoriness of measurement models of multifaceted constructs. The validation framework is then applied to a new measurement model of the economic domain of export performance, a construct for which agreement has not yet been reached in the literature concerning the appropriate representation of its complex nature. A sample of 414 large Brazilian exporters of manufactured products was collected, and five competing measurement models of the construct were proposed and comparatively assessed. Insights into the nature and structure of the construct are drawn. In the best fitting of the five tested models, export (venture) performance is represented as a two-dimensional construct – past export revenues and their growth, and past export profitability, with four and two operational indicators, respectively. The set of final indicators provides a reasonable coverage of several conceptual aspects of the phenomenon, namely absolute and relative measures as well as static and dynamic orientations. There is a thorough discussion of validation steps. The validation framework advanced here is generic and comprehensive enough to be employed for modeling other multifaceted constructs in the social sciences.

Key words: construct measurement; validity; export performance; structural equation modeling.

Received 02 February 2009; received in revised form 01 April 2009.

* Corresponding author: Jorge Carneiro
Rua Dezenove de Fevereiro, 127 / 201, Botafogo, Rio de Janeiro, RJ, 22280-030, Brazil.

INTRODUCTION

Several phenomena in the social sciences exhibit a complex and abstract nature, which poses important challenges, both substantive and methodological, to their conceptualization and operational representation. The latent nature of a complex construct means that it cannot be directly observed, but needs to be inferred from its manifestations (Netemeyer, Bearden, & Sharma, 2003). Moreover, it has been recognized that "specifying the relationship between concepts and operational indicators is equally important to social research as the substantive theory linking concepts to one another" (Carmines & Zeller, 1979, p. 11).

A construct of critical importance to research on exporting is export performance. However, although the construct has received the attention of several scholars, none of the already proposed measurement models has reached consensual acceptance. This lack of agreement makes it difficult to compare research findings and develop a shared body of knowledge. As a result, the empirical literature has reached mutually inconsistent results about the effects of determinants of export performance (Zou, Taylor, & Osland, 1998). And although a multidimensional approach to represent the complex nature of the export performance construct has been advocated (Diamantopoulos, 1999; Katsikeas, Leonidou, & Morgan, 2000; Leonidou, Katsikeas, & Samiee, 2002; Madsen, 1987, 1998), most researchers have employed unidimensional models that do not adequately capture the multifaceted nature of the phenomenon.

In addition, even those scholars who have advanced quite elaborate measurement models of export performance (e.g., Cavusgil & Zou, 1994; Lages & Lages, 2004; Lages, Lages, & Lages, 2005; Shoham, 1998, 1999; Styles, 1998; Zou *et al.*, 1998) have not provided a thorough validation screening of the proposed operational models. It is this methodological issue that we address in this paper. The study has the following objectives:

. to offer a rather comprehensive and integrated set of procedures based on structural equation models [SEM] for validating measurement models of complex and multifaceted constructs, which is rooted both in conceptual reasoning and empirical screening;

. to empirically apply these validation procedures to the development of a new measurement model of the export performance construct and the assessment of the degree of satisfactoriness of such a model; and

. to discuss the nature and structure of the construct based on the interplay between conceptual reasoning and empirical results.

In fact, two stages are involved in theory building: the first is the specification of "relationships between theoretical constructs", and the second is the description of "relationships between constructs and measures" (Edwards & Bagozzi, 2000, p. 155). These two stages are critical, since theory building requires "a high degree of correspondence between abstract constructs and the procedures used to operationalize them" (Peter, 1981, p. 133). In this paper, we address mainly the second aspect of theory building.

This paper is organized as follows. After this introduction, we describe the validation framework and present data collection and data treatment procedures. We then apply the validation framework, step by step, to the development of a new measurement model of export performance, and new highlights into the nature of the construct are addressed. Final remarks and some suggestions for future studies close the paper. Although the paper is rather methodological, we also discuss relevant theoretical implications that can be drawn from the analyses.

THE VALIDATION FRAMEWORK

We reviewed and contrasted several works, drawn from quite diverse areas of study, including psychology, education, organizational studies, statistics, strategic management, marketing, and international business. By putting together and operationalizing several perspectives and criteria by which to judge the adequacy of measurement models, we believe we have assembled a useful framework for construct validation.

Our validation framework covers the following eight steps:

1) Conceptualization of the construct and pursuit of content validity
2) Exploratory empirical verification of the dimensionality and content of the construct
3) Advancement of theoretically plausible competing models
4) Assessment of psychometric properties
5) Assessment of concurrent and predictive validity
6) Verification of overall adequacy of the measurement model
7) Verification of (measurement parameters) stability
8) Selection of the most likely model

Table 1 presents the steps and procedures involved in the validation process.

Table 1: The Validation Framework

Step		Procedure	
1	Conceptualization of the construct and pursuit of content validity	(1-a)	Map the conceptual domain of the construct
		(1-b)	Define the breadth of coverage of the phenomenon
		(1-c)	Conjecture on dimensionality (number and content of the dimensions)
		(1-d)	Define hierarchical complexity (number of levels)
		(1-e)	Choose appropriate measurement perspective (reflective vs. formative)
2	Exploratory empirical verification of the dimensionality and content of the construct	(2-a)	Verify whether the same number of pre-specified dimensions emerges from the data
		(2-b)	Verify whether indicators cluster together as theoretically expected
		(2-c)	Verify whether signs of loadings are compatible with theory and, in each factor, have the same direction
		(2-d)	Verify whether indicators do not cross-load
3	Advancement of theoretically plausible competing models	(3-a)	Select relationships with potential to be modeled
		(3-b)	Specify competing measurement models
4	Assessment of psychometric properties	(4-a)	Assess internal consistency of the measurement scale (or of each dimension of the scale)
		(4-b)	Assess unidimensionality of the measurement scale (or of each dimension of the scale)
		(4-c)	Assess reliability of each scale and of each individual indicator
5	Assessment of concurrent and predictive validity	(5-a)	Assess concurrent validity
		(5-b)	Assess predictive validity

Table 1 (continued): The Validation Framework

Step		Procedure	
6	Verification of overall adequacy of the measurement model	(6-a)	Assess parsimony
		(6-b)	Assess goodness-of-fit indices
7	Verification of (measurement parameters) stability	(7-a)	Check for change in the magnitude of indicator loadings in the transition to a larger measurement model and to a corresponding structural model
		(7-b)	Check for change in the state of statistical significance in the transition to a larger measurement model and to a corresponding structural model
8	Selection of the most likely model		Select the model that exhibits overall better properties

METHODS

Population and Sample

A survey was conducted of the largest Brazilian exporters of manufactured products selected from a list provided by FUNCEX, a private foundation supported by Brazilian exporters. Firms controlled by foreign capital were excluded because of potential different objectives and possible transfer pricing mechanisms. Service firms, exporters of commodities and trading companies were also removed in order to make the sample more homogenous, and thus avoid possible confounding effects, resulting in a population of 3,057 exporters of manufactured goods. The unit of analysis was the export venture, i.e., the exporting of a given product line to a given country (Matthyssens & Pauwels, 1996). A sample of 448 exporters was obtained resulting in a response rate of 15.5%, after correcting for non-eligibles. No systematic bias was observed between respondents vs. non-respondents or between early versus late respondents.

Data Collection and Data Treatment Methods

A four-page structured questionnaire covered not only indicators of export performance but also several variables related to determinants of export performance; only the export performance variables are reported here. Firms were mailed a questionnaire with a pre-paid return envelope.

Semantic-differential scales of perceptual measures were employed instead of asking firms to provide objective information (Matthyssens & Pauwels, 1996; Shoham, 1998). This was deemed necessary to improve the response rate and minimize missing values since most firms do not keep objective public data for each export venture, segregated from the firm's other ventures. Moreover, it has been reported that subjective measures correlate highly with objective measures of performance as well as with overall assessments of performance (Dess & Robinson, 1984; Venkatraman & Ramanujam, 1987) and that managers' assessments are as reliable as data from objective sources (Wong & Saunders, 1993). Also, managerial decisions tend to be driven by perceptions rather than solely by "cold" data (Bourgeois, 1980; Matthyssens & Pauwels, 1996).

Variables and cases with more than 15% missing values were removed (Hair, Black, Babin, Anderson, & Tatham, 2006), which led to the exclusion of one indicator of export venture performance (past export venture's volume *vs*. other Brazilian firms exporting to the same country) and 34 cases. Since missing data exhibited an missing completely at random [MCAR] pattern at the 10% significance level, it was possible to estimate the missing values. Given that three estimation methods (mean substitution pairwise, regression imputation and EM approach) provided very similar estimates, a simple average of these three methods was used (*cf*. Hair *et al*., 2006). The resulting sample (414 cases) showed no indication of the presence of multivariate outliers as far as the 10

remaining operational indicators of export venture performance were considered. Parameters were estimated by an asymptotic distribution-free method [ADF] because variables did not follow a normal distributional pattern. SPSS 15 and AMOS 7.0 were employed to run the statistical analyses.

Application of the Validation Framework to a New Measurement Model of Export Performance

The validation framework consisted of eight steps.

Step 1: Conceptualization of the Construct and Pursuit of Content Validity

First of all, it is necessary to **(1-a)** map the conceptual domain of the construct (DeVellis, 2003; Spector, 1992). This task involves the identification of "what is and what is not included in the domain" (Churchill, 1979, p. 67). Therefore, we had to decide on the appropriate conceptual domain of the export performance construct for which our measurement model would be developed and strive for content validity, or at least provide evidence of content adequacy (Schriesheim, Powers, Scandura, Gardiner, & Lankau, 1993).

Content validity was sought by means of an extensive review of the literature on the focal construct as well as theoretical reflection and consultation with academic experts. We searched the most prominent journals on International Business (Dubois & Reeb, 2000) for a 30-year period (1976-2005) in order to identify studies (conceptual, empirical, meta-analytical, and consolidation works) which seemed to represent the best efforts to characterize the multifaceted nature of the export performance phenomenon. We also reviewed the proceedings of two leading conferences in the field: the Academy of International Business [AIB] and European International Business Academy [EIBA]. Building on these studies, export performance is conceptualized as a multi-dimensional construct that includes several classes of measures (economic, market, behavioral, strategic and overall), two alternative frames of reference (absolute and relative), and two perspectives of temporal orientation (static and dynamic), as shown in Figure 1.

Figure 1: Generic Analytical Framework for the Characterization of Export Performance

Classes of measures	Frame of reference	Temporal orientation
✓economic ✓market ✓behavioral ✓strategic ✓overall ✓other measures	✓absolute ✓relative - to competitors - to a benchmark - to domestic operations - to other international ventures of the firm - to pre-set goals	✓static - recent past - future expectations ✓dynamic - change in recent past - expected future change

Source: Carneiro, Hemais, Rocha and Silva (2005), Katsikeas *et al.* (2000), Matthyssens and Pauwels (1996).

After mapping the domain of the construct, it is necessary to **(1-b)** define the breadth of coverage of the phenomenon. From the various classes of measures in the literature, we chose to concentrate on only one, the economic aspect of export venture performance. This was deemed necessary in order to avoid having too many indicators, which might lead to fatigue bias when eliciting information from respondents. We used the two frames of reference (absolute and relative) and both types of temporal orientation (static and dynamic measures, with a past and a future view in each case). In addition, it was necessary to conjecture on the **(1-c)** dimensionality (number and content of the dimensions), **(1-d)** hierarchical complexity (no. of levels), and **(1-e)** measurement perspective (reflective vs. formative).

Whether a construct ought to be viewed as unidimensional or multidimensional depends on the level

of abstraction used to define it (Jarvis, MacKenzie, & Podsakoff, 2003). One can look at each facet as a separate construct, but at a more abstract level all facets are integral parts of the overall construct. If the construct is deemed to be multidimensional, the researcher has to speculate about the relationships among the dimensions (Law, Wong, & Mobley, 1998). Are they just correlated? Does there seem to be a higher-order underlying factor that ties them together? Or do they concur to define a higher-level representation of the construct? If the construct is deemed to be better represented by an arrangement of a higher-order level together with its lower-order counterparts, then the researcher has to decide whether either a reflective or a formative perspective of measurement (Bollen & Lennox, 1991) would better characterize the relationship between higher and lower levels. A reflective (*vis-à-vis* formative) perspective would make sense if, among other aspects, the dimensions are expected to co-vary, to be affected by the same antecedents and to have the same consequences in a given nomological network where the researcher expects the construct to be used (Jarvis *et al*., 2003). Therefore, the choice of the appropriate dimensional and hierarchical arrangement (single vs. multiple dimensions and single- vs. higher-order structure among the dimensions), and of the measurement perspective (reflective vs. formative) may depend on the conceptual breadth chosen to represent the phenomenon.

In order to identify indicators that could operationally represent the construct, we reviewed 62 empirical studies, ranging from 1985 to 2005, and uncovered 116 distinct indicators of export performance, of which 35 were related to economic aspects of the export venture. We then grouped the economic-related indicators in categories according to their similarity of content. We initially modeled the economic domain of the construct as multi-dimensional because it is composed of distinct, albeit related, aspects, i.e., revenues, revenues growth, and profitability (Figure 2). For each dimension a reflective perspective was considered appropriate to represent its relationship with the respective indicators. Many of the 35 indicators were either redundant or very similar in content. Therefore, we initially selected 11 operational indicators (Table 2) to measure the three dimensions. As a set, these indicators cover reasonably well the diversity of conceptual aspects of the phenomenon, while maintaining a reasonable degree of parsimony.

Figure 2: Economic Domain of the Export Performance Construct

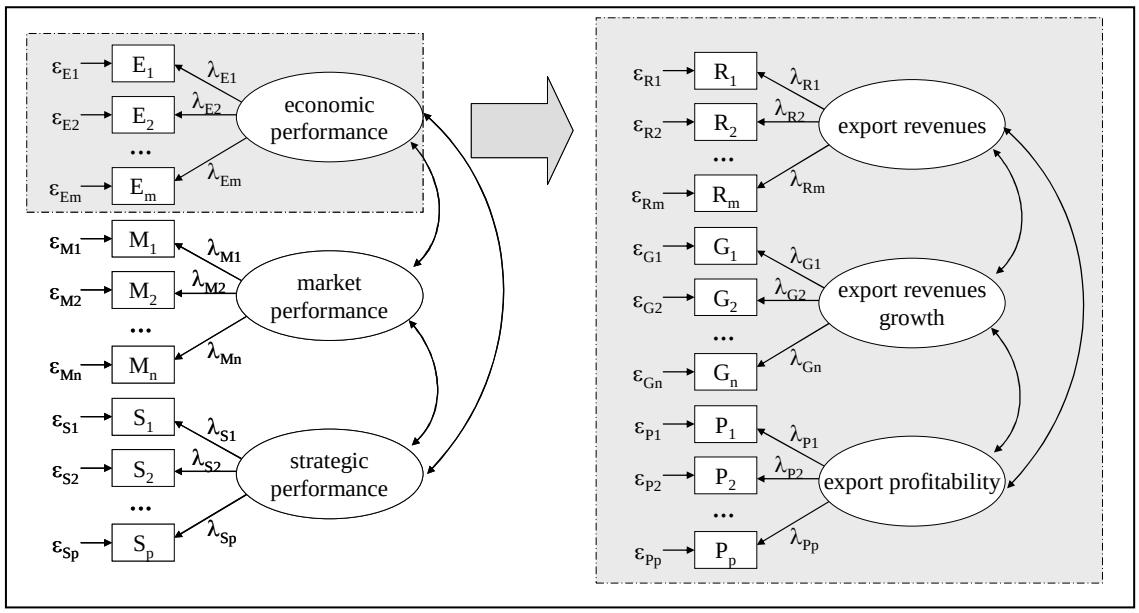


Table 2: *A priori* Dimensions and Indicators of the Economic Domain of the Export Performance Construct

Dimension	Indicator	Description of the meaning of the indicator	Conceptual aspects covered *
Export revenues (ExpRev)	SPasRev [1]	Satisfaction with past export venture revenues	absolute, past (static)
	PasReOt [2]	Export venture past revenues *vs.* average revenues of other export ventures of the firm	relative (to other export ventures), past (static)
	PasVoCo [2 +]	Past export venture volume *vs*. other Brazilian firms exporting to the same country	relative (to competitors), past (static)
	FutVoOt [2]	Expected future export venture volume *vs.* expected average volume of other export ventures of the firm	relative (to other export ventures), future (static)
Export revenues growth (ExpRevG)	SPaReGr [1]	Satisfaction with past growth of export venture revenues	absolute, past (dynamic)
	PasVGOt [2]	Past growth of export venture volume *vs.* average volume growth of other export ventures of the firm	relative (to other export ventures), past (dynamic)
	FutVGOt [2]	Expected future growth of export venture volume *vs.* expected average volume growth of other export ventures of the firm	relative (to other export ventures), future (dynamic)
Export profitability (ExpProf)	SPasPro [1]	Satisfaction with past profit margin of export venture	absolute, past (static)
	PasPrOt [2]	Past export venture profitability *vs.* average profitability of other export ventures of the firm	relative (to other export ventures), past (static)
	FutProf [2]	Expected future export venture profitability	absolute, future (static)
	FutPrOt [2]	Expected future export venture profitability *vs.* expected average profitability of other export ventures of the firm	relative (to other export ventures), future (static)

Note: temporal bracket explicitly stated in the questions was "last three years" or "next three years"
* besides the economic aspect, which is, by design choice, covered by all of the indicators
[+] this indicator was subsequently dropped due to the high incidence of missing data
[1] rated on five-point semantic differential scales with anchor words "very dissatisfied" … "very satisfied"; e.g., for SPasRev:

Total value of your exports of this product to that country in the period 2004 thru 2006 — very dissatisfied | 1 | 2 | 3 | 4 | 5 | very satisfied

[2] rated on five-point semantic differential scales with anchor words "much lower" … "much higher"; e.g., for FutPrOt:

Compared with the average profitability of your exports to other countries, the export profitability of this product to that country in the next three years will probably be — much lower | 1 | 2 | 3 | 4 | 5 | much higher

Step 2: Exploratory Empirical Verification of the Dimensionality and Content of the Construct

In order to avoid capitalization on chance, i.e., overfitting to the idiosyncrasies of a particular sample (MacCallum, Roznowski, & Necowitz, 1992), we split the original sample into a calibration sub-sample and a validation sub-sample (around 1/3 and 2/3 of the total cases, respectively) – the former for an exploratory and the latter for a confirmatory factor analysis.

An exploratory factor analysis [EFA] was thus run on the calibration sample in order to check whether the factorial structure that emerged from empirical data replicated what was expected from theoretical considerations. Although it might at first seem odd to run an exploratory factor analysis [EFA] instead of moving directly to a confirmatory factor analysis [CFA] of the proposed operational model, the use of EFA permits us to verify whether **(2-a)** the same number of pre-specified dimensions emerges from the data; **(2-b)** indicators cluster together as theoretically expected; **(2-c)** signs of the loadings are compatible with theory and, in each factor, have the same direction; and **(2-d)** indicators do not cross-load. This procedure provides a preliminary empirical account of the dimensionality and content of the construct (Hinkin, 1998).

In the orthogonal solution (Table 3), the first factor can be interpreted as "expected future (absolute and relative) export venture performance". If one accepts that, on substantive argumentation, indicator PasPrOt can be associated with the third factor, on which it loads high (.38), and not with second factor, on which it loads highest (.50), then the second factor could be interpreted as "satisfaction with past absolute export venture performance" and the third factor would be "past relative export venture performance". The oblique solution produced a similar pattern of associations as long as one adopts the same reasoning for associating PasPrOt with the third factor. As desired, signs of the indicators in each factor have the same direction.

Table 3: Pattern of Associations between Indicators and Factors Uncovered by EFA

	Factor		
	1	2	3
SPasRev	-.04	**.73**	.00
PasReOt	.03	.12	**.68**
VFutOt	**.79**	.08	.09
SPaReGr	.22	**.54**	.19
PasVGOt	.19	.19	**.83**
FutVGOt	**.86**	.02	.08
SPasPro	.22	**.62**	.22
PasPrOt	.24	.50	**.38**
FutProf	**.60**	.44	.16
FutPrOt	**.73**	.30	.14

Extraction method: principal axis factoring.
Rotation method: Varimax with Kaiser normalization

	Factor		
	1	2	3
SPasRev	.14	**.80**	.10
PasReOt	.08	-.02	**-.72**
VFutOt	**-.83**	-.07	.01
SPaReGr	-.14	**.51**	-.10
PasVGOt	-.05	-.01	**-.86**
FutVGOt	**-.91**	-.14	.03
SPasPro	-.12	**.59**	-.12
PasPrOt	-.14	.42	**-.31**
FutProf	**-.57**	.33	-.03
FutPrOt	**-.73**	.17	-.01

Extraction method: principal axis factoring.
Rotation method: Oblimin with Kaiser normalization

Notes: High loadings (absolute value ≥ .30) are shadowed
The suggested association of indicators with factors is shown in bold-face type

Although the rotated factors can be neatly interpreted, the factor structure that emerged from empirical data was different from what was expected. Correlation analysis was then used to shed some more light on the relationships among factors; 44 out of 45 correlations among the 10 indicators are statistically significant at the .01 level, and 17 are reasonably high (over .40); such correlations would suggest that a single-factor model should not be immediately ruled out.

Step 3: Advancement of Theoretically-plausible Competing Models

Previous analyses, based on the interplay of conceptual reasoning with empirical results, should be used for **(3-a)** the selection of relationships with potential to be modeled, and **(3-b)** the specification of competing measurement models. Given the fact that preliminary empirical evidence did not confirm theoretical expectations and that the factor structure was not conclusive – EFA suggested three factors while correlation analysis suggested one factor, we decided to keep five models for further comparative assessment (Figures 3-a through 3-e).

Figure 3-a - Model # P1 of export performance

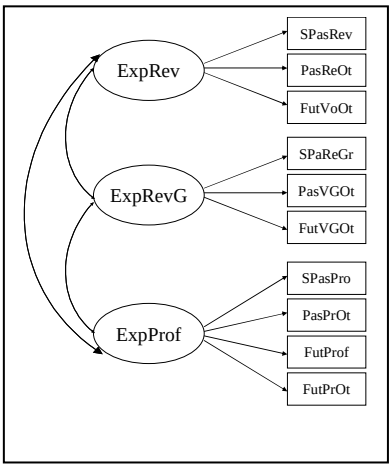


Figure 3-b - Model # P2 of Export performance

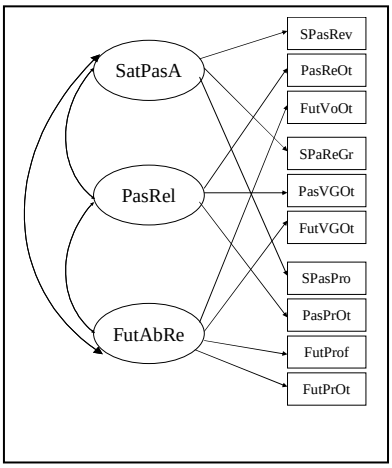


Figure 3-c - Model # P3 of Export performance

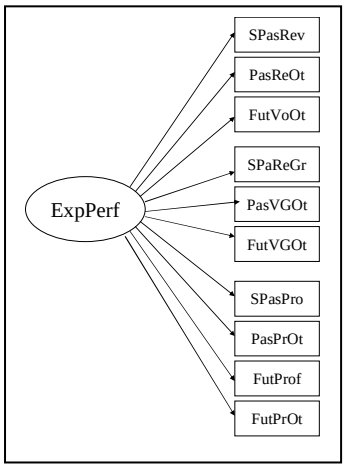


Figure 3-d - model # P4 of export performance

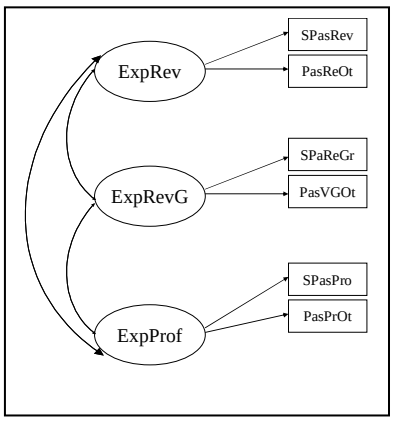
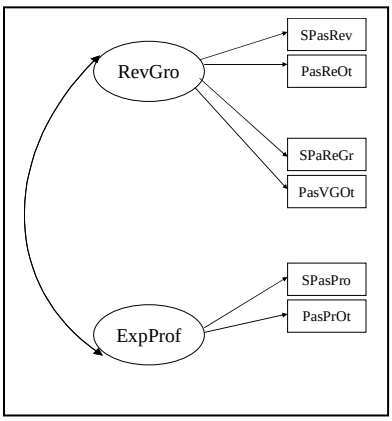


Figure 3-e - Model # P5 of export performance



Latent variables ("dimensions")	
ExpRev	export revenues
ExpRevG	export revenues growth
ExpProf	export profitability
SatPasA	satisfaction with past absolute export perf.
PasRel	past relative export performance
FutAbRe	future export perform.
RevGro	export revenues and their growth
ExpPerf	export performance

. Model # P1: three factors as suggested by the initial conceptual discussion, i.e., export venture revenues (past and future, absolute and relative), export venture revenues growth (past and future, absolute and relative), and export venture profitability (past and future, absolute and relative), composed of, respectively, three, three and four indicators;

. Model # P2: three factors as suggested by the exploratory factor analysis, i.e., satisfaction with past absolute export venture performance, past relative export venture performance, and future (absolute and relative) export venture performance, composed of, respectively, three, three and four indicators; and

. Model # P3: one single factor (export venture performance) incorporating all ten indicators as suggested by correlation analysis; given its relative simplicity (i.e., only one dimension), this is an important baseline model to use for comparison with more complex models such as # P1 and # P2 (three dimensions each).

Considering that at the time of the survey Brazilian exporters expected that the real (Brazilian currency) would be strongly valued against the dollar, it is possible that this expectation might have affected differently the indicators of past performance *vis-à-vis* those of future performance. Consequently, we decided to use two other models that would not incorporate indicators of future performance:

. Model # P4: three factors involving only past indicators, i.e., past export venture revenues (absolute

and relative), past export venture revenues growth (absolute and relative), and past export venture profitability (absolute and relative), each composed of two indicators (this model is the counterpart of # P1, considering only past indicators); and

. Model # P5: two factors involving only past indicators, i.e., past export venture revenues (absolute and relative) and their growth, and past export venture profitability (absolute and relative), composed respectively of four and two indicators. This model explicitly assumes that absolute value and growth of the absolute value over time may reflect the same aspect of performance instead of two distinct, albeit complementary, dimensions.

Step 4: Assessment of Psychometric Properties

For each dimension of each model, it is necessary to verify whether it conforms to desirable psychometric properties, specifically: **(4-a)** internal consistency, **(4-b)** unidimensionality, and **(4-c)** reliability (Anderson & Gerbing, 1988, 1991, 1992; Bollen, 1989; Carmines & Zeller, 1979).

A confirmatory factor analysis [CFA] was run on the validation sub-sample, for each pre-specified competing measurement model. This analysis was conducted in both an absolute mode (i.e., compliance with minimum adequacy thresholds) and a comparative mode (i.e., verification of whether some of the models attended the desirable properties better than others).

Before running CFA, some decisions regarding the statistical identification of the model had to be made. In order to make a model identifiable, it is necessary to define a measurement scale for each latent construct (Anderson & Gerbing, 1988; MacCallum & Browne, 1993), which can be accomplished by constraining one of the paths from one of the indicators to some nonzero value or by fixing the latent variable's variance at 1.0 (i.e., standardizing it). We chose the latter option since we were more interested in estimating loadings of indicators. As for the error terms, we fixed their loadings onto their respective indicators at 1.0 because we were interested in their variance. For all models, ADF estimation converged and there were no improper solutions.

Assessment of internal consistency (4-a) – Internal consistency relates to the homogeneity of the items within a scale (DeVellis, 2003) or, in the case of multidimensional constructs, within each latent variable (dimension) of the measurement model. Four tests were used to help determine the degree of internal consistency: compatibility of signs of the indicators with theoretical expectations; magnitude and level of statistical significance of inter-item within-construct correlations (both observed correlations and model-implied correlations); magnitude and level of statistical significance of item-to-total within-construct correlations; magnitude and statistical significance of standardized loadings.

As for the signs of loadings, in all the estimated models they were compatible with theoretical expectations and, in each latent variable (dimension), had the same direction. As for the magnitude and level of statistical significance of inter-item within-construct correlations, as implied by the measurement model, they were all adequately high (each $\geq .20$; and on average $\geq .30$, *cf*. Kim & Mueller, 1978, except for the latent variable 'export revenues' in model # P1) and statistically significant at the 5% level. So the set of indicators *a priori* associated with each latent variable seemed to offer a good joint representation of it.

We also checked whether item-to-total within-construct correlation was high enough (each $\geq .50$; on average $\geq .70$) and statistically significant. The original theoretically-derived model # P1 fails these quality criteria in half of the correlations. EFA-derived model # P2 performs best and models # P3 and # P5 are acceptable. In model # P4, this test is redundant with respect to the previous test since each dimension has only two indicators. As for the magnitude of the average item-to-total within construct correlations, EFA-derived model # P2 performed better and the original theoretically-derived model # P1 performed worse.

It is also desirable that standardized loadings are statistically significant and adequately high ($\geq .50$ and, ideally, $\geq .707$), showing that more than 50% ($.707^2$) of an indicator's variance is explained by its

respective construct, while the rest is (random or indicator-specific) measurement error (Hair *et al*., 2006). All models satisfied this rule, but the loading for indicator SPasRev was lower than .707 (although still higher than .50 and statistically significant at the 1% level) in some of the models.

On the whole, all models satisfactorily met the internal consistency requirement, except, to a certain extent, the theoretically-conceived model # P1.

Assessment of unidimensionality (4b) - Unidimensionality refers to the extent to which a set of items reflects one single underlying trait (Anderson & Gerbing, 1988; Hair *et al*., 2006; Hattie, 1985; McDonald, 1981). If the construct is conceptualized as multidimensional, unidimensionality should be assessed for each separate subscale (latent variable), each one measuring one of the distinct dimensions of the construct (Peter, 1981). We employed five tests to verify the degree of unidimensionality: inter-item within-construct correlations vs. inter-item between-construct correlations; pattern of standardized residuals; completely standardized expected parameter changes; modification indices; and discriminant validity.

In order for unidimensionality to be established, inter-item within-construct correlations should be statistically higher than inter-item between-construct correlations (Campbell & Fiske, 1959). A visual inspection of the correlations raises serious questions about the validity of model # P1 and some suspicion regarding the validity of model # P5, since it seems that some indicators are more related to other latent variables than to the one they had been *a priori* assigned to. Model # P2 performs well. As for model # P3, this test does not apply since it has only one dimension.

Residuals, which are differences between empirically observed covariances and model-derived covariances, also provide evidence of the degree of association between indicators and latent variables. It is desirable that an indicator does not show large negative standardized residuals (absolute value above 2.58, *cf*. Diamantopoulos & Siguaw, 2006) with indicators in its assigned dimension because a large negative residual would suggest that the two indicators seem not to be related to the same dimension (Steenkamp & Trijp, 1991). Models # P2 and # P4 satisfy this condition, suggesting that their indicators in fact "cluster together" as specified. However, given that model # P2 proposes a rather different (compared with model # P4) grouping of indicators around dimensions, these results make it difficult to clearly discern the dimensionality and the content of the dimensions of the construct. The pattern of residuals of the other models suggests that indicators related to the future would not seem to represent the same dimensions as indicators related to the past, while past indicators seem to cluster well together in the way suggested by the models.

A complementary test is whether an item does not show large positive standardized residuals with another item associated with a different latent variable: if items assigned *a priori* to distinct dimensions show a large positive standardized residual, this would imply cross-loading (Steenkamp & Trijp, 1991) – violating the desirable unidimensionality of each latent variable – or would imply that the two items should in fact be assigned to the same dimension, (and not to distinct dimensions). In model # P1 there are four (12%) between-construct standardized residuals that are high (greater than 2.58), suggesting that revenues, growth and profitability might somehow represent the same facet, rather than three distinct dimensions, of the construct. There are also some large standardized residuals between satisfaction-related indicators, but this may be due to a possible method bias – i.e., respondents may have, inadvertently or not, tried to provide similar answers to questions about **satisfaction** with performance (indicators SPasRev, SPaReGr, and SPasPro). On the whole, the pattern of inter-item between-construct residuals neither invalidates any of the five competing models nor places any of them as indisputably better than the others.

We also verified whether the completely standardized expected parameter change – which indicates the probable change that the standardized value of a model parameter that has previously been fixed (usually at zero) during the estimation process would undergo if it were allowed to be freely estimated (Diamantopoulos & Siguaw, 2006) – was reasonably small. In all five models under evaluation, the magnitude of standardized expected changes – of paths linking two indicators directly to one another, of paths between an indicator and a latent variable (dimension) to which it had not previously been

assigned, and of correlations between error terms – are small (less than 25%) compared with the standardized loadings of the respective indicators. This is a desirable property as it means that there is no reason to believe that indicators or error terms would be more related than implied by the respective specified models.

We also checked whether modification indices would suggest a different arrangement of indicators around dimensions. "A modification index indicates the minimum decrease in the model's chi-squared value if a previously fixed parameter is set free and the model re-estimated" (Diamantopoulos & Siguaw, 2006, p. 108). A value of 3.84 or more would mean a significant (at 5% level) improvement in model fit. In models # P3 and # P5 modification indices show that there would be a significant improvement in model fit if some direct paths between indicators were set free to be estimated. Good measurement practices, however, recommend that any given indicator be solely determined by one latent construct, rather than by another indicator. So, pending further evidence, we considered that this might be a spurious and idiosyncratic result of this particular sample. It should be noted that no modification index suggests linking any indicator to another latent variable, meaning that the proposed factorial structures are, in some sense, all compatible with empirical data. Model # P4 performed better, with the smallest number of relevant modification indices, while models # P3 and # P5 performed somewhat worse.

We also performed tests of discriminant validity, i.e., whether dimensions that are expected to be distinct, theoretically, seem to be distinct statistically. Such tests would not apply to model # P3 because it has only one dimension. First we checked whether average variance extracted [AVE] for each latent variable was higher than the square of the correlation between this latent variable and any other latent variable (inter-construct correlation). The logic behind this argument is that a latent construct should explain its items better than it explains another construct (Fornell & Larcker, 1981; Hair *et al*., 2006). None of the dimensions of models # P1, # P4 or # P5 satisfied this requirement, suggesting that they might not be distinct. On the other hand, model # P2 fulfilled this condition.

Second, we tested whether the Δ-χ^2 between a model that constrains the correlation between two latent variables to equal 1.0, and another model that allows this correlation to be freely estimated is statistically significant (Δ-χ^2 is the difference in the degree of fit between two models, taking into account the difference in the degrees of freedom). Hughes, Price and Marrs (1986) stated that if Δ-χ^2 is significant, this would provide evidence of the distinctiveness of the two dimensions. In model # P1, when we fixed the correlation between 'export revenues' and 'export revenues growth', Δ-χ^2 was not significant, suggesting that the two dimensions might not be independent. When correlations between each pair of dimensions in model # P2 were fixed, one at a time, at 1.0, the estimation process either reached an improper solution (negative variances for error terms) or the covariance matrix was not positive definite and thus could not be inverted, therefore preventing the proper estimation of model parameters. This, together with the fact that all three Δ-χ^2 were significant at the .001 level, would lead one to conclude for the relative independence of the dimensions. A similar situation was found when we fixed the correlation between the dimensions in models # P4 and # P5. Results suggest that in each model the proposed dimensions seem to be distinct, except for model # P1.

Third, we tested whether the for the statistical significance of Δ-χ^2 between a model where each dimension keeps its *a priori* assigned indicators and another one where all indicators of any two dimensions are considered indicators of one same dimension. If Δ-χ^2 is not statistically significant, the two dimensions cannot be considered distinct (Hair *et al*., 2006). Results indicate that in models that put together indicators of the past and of the future the dimensions of 'export revenues' and of 'export revenues growth' might not be distinct. However, in models that contain only indicators of the past, these dimensions do, indeed, seem to be distinct. As for 'export profitability', when its indicators are placed under the same dimension as the other indicators, the large Δ-χ^2 suggests that it might be measuring a distinct facet of export performance. In model # P2 we joined the indicators of the two dimensions related to past performance, keeping the independence of the dimension related to the future (given its conceptual distinctiveness). Given that Δ-χ^2 was statistically significant, one would conclude for the distinctiveness of the two dimensions of past performance in model # P2.

Assessment of reliability (4-c) – Reliability refers to accuracy or precision of the measuring instrument or, in other words, absence of random measurement error. One ought to assess both the reliability of each scale (latent variable or dimension) and of each individual indicator.

Coefficient alpha, which is frequently employed in the assessment of the reliability of latent variables (scales), was not used in this study because it assumes identical loadings and equal error variances for all indicators of each latent variable (Cortina, 1993), which is not the case here. Therefore, we assessed the reliability of latent variables by means of:

. composite reliability

. average variance extracted [AVE]

Composite reliability, ρ_c (calculated as (Σ standardized $\lambda_i)^2$ divided by ((Σ standardized $\lambda_i)^2 + \Sigma\delta_i$), where λ_i represents the standardized loading of indicator "i" and δ_i represents the variance of the error term of indicator "i", (*cf*. Bagozzi, 1984; Bagozzi & Yi, 1988) was used instead of coefficient alpha because it takes into account differences among indicators and shows how high loadings as a set are relatively to error terms as a set (Bollen, 1989). All latent variables in all models exhibit composite reliability coefficients higher than .60, indicating good reliability of the indicators as a set. Moreover, unidimensional models show better reliability than two-dimension models and these show better reliability than their correspondent three-dimension models. This is probably a result of the fact that reliability estimates tend to be higher as more items are added to a latent variable (Carmines & Zeller, 1979) and cannot be taken at face value to indicate that unidimensional models would better represent the focal construct.

We also calculated average variance extracted [AVE], which represents the average percentage of variation in the latent variable explained among its indicators (Hair *et al*., 2006). Low (less than .50) values of AVE mean that more error remains in the items, as a set, than variance explained by the latent factor structure imposed on the measures (Hair *et al*., 2006). Except for the latent variable 'export revenues' in model # P1 (AVE = .472, just a little below .50), all other latent variables in all five models fulfill the desired threshold level.

Besides assessing the reliability of latent variables, we also assessed the reliability of each individual item, since the reliability of a scale also depends on the reliability of its indicators (Rossiter, 2002). We employed two tests: item reliability index, and statistical significance of (measurement) error variances.

One can empirically determine the degree of an item's reliability, ρ_i (defined as: $\lambda_i^2 / (\lambda_i^2 + \delta_i)$ *cf*. Bagozzi, 1984). Although we were unable to find in the literature a clear recommended minimum for ρ_i, it seems reasonable to assume a value of .50, meaning that error variance would be less than the respective proportion of variance of the indicator explained by its latent variable. Indicator SPasRev fails to meet this requirement in models # P1, # P4 and # P5. Although these and other results place SPasRev under suspicion, the item was nonetheless kept because of statistical identification needs and also because other tests did not clearly condemn it. In model # P2 all indicators satisfied this condition. In model # P3 indicator SPasPro is just marginally below the threshold.

It is also worth mentioning that measurement error variances (variances in the error terms of each indicator) are statistically significant (at the 1% level) in all the models. This rule may seem counter-intuitive and needs justification. In Diamantopoulos and Siguaw's (2006, p. 89) words, "although one is clearly interested in **minimizing** measurement error, **zero** measurement error is a cause for concern" (emphasis in the original) because, as stated by Bagozzi and Yi (1988, p. 77), "nonsignificant error variances usually suggest specification errors, since it is unreasonable to expect the absence of random error in most managerial and social science contexts."

Step 5: Assessment of Concurrent and Predictive Validity

The fifth step consists of assessing both **(5-a)** concurrent and **(5-b)** predictive validity.

Concurrent validity (5-a) – Concurrent validity refers to the degree to which the focal measure correlates with measures of other constructs that are expected to be correlated with the focal latent variable (Bollen, 1989). To conclude in favor of concurrent validity, "the variables, at a minimum, should demonstrate [statistically significant] covariation above and beyond what can be attributed to shared method variance" (DeVellis, 2003, p. 55). Such a criterion variable has to be an accepted standard against which to compare the newly developed measure (Bollen, 1989) and should at least exhibit some degree of content validity. Also, there should be solid theoretical reasons to justify such expected relationship (Diamantopoulos & Winklhofer, 2001). However, since a generally-accepted measure of export performance has not yet been unequivocally established by previous research, we decided instead to test for concurrent validity by eliciting information about two overall assessments of export performance, which would summarize the construct, rated as separate questions (Diamantopoulos & Winklhofer, 2001; Smith, 1999). The degree of concurrent validity was measured as the pairwise correlation of each indicator of export performance with each of the two overall assessments. All correlations were significant at the 1% level and greater than .40, suggesting good concurrent validity for all indicators.

Predictive validity (5-b) – We also tested for predictive validity, i.e., the ability of the new measure to predict intentions of future behavior (Nunnally, 1978; Smith, 1999). Respondents were asked to answer the question "If this decision was up to you, would you recommend that exporting this product to that country should" and a 5-point scale was offered, ranging from 1 = **be interrupted** to 5 = **be increased substantially**. Predictive validity was measured as the pairwise correlation of this intention of future behavior with each of the indicators of export performance. Except for PasReOt, all indicators had a significant correlation at the 1% level, but four of them were relatively low (in the range of .16 through .29) thereby suggesting reasonable predictive validity.

Readers may ask why not also assess convergent validity – the degree to which multiple attempts to measure the same concept with maximally dissimilar methods are in agreement, i.e., their correlations are "significantly different from zero and sufficiently large" (Campbell & Fiske, 1959, p. 82). The different methods selected should not share similar sources of method variance (Walker, Olson, & Chow, 1992). Dissimilar methods and sources of data – such as objective vs. subjective data, interviews, questionnaires, archival data, participant observation, multiple managers in different key functions, published secondary data, expert opinion, and use of different types of scales (Venkatraman & Grant, 1986) or different respondents outside the firm – reduce the chance that high correlations are a result of shared methods variation. Therefore, if there were already established and accepted measures of the construct, they should be collected in order to evaluate the relationship between the newly-proposed and previously validated measures (Hinkin, 1998). However, the very assumption of this paper is that there is no **good** measure of export performance known and "[t]o show convergent validity, the new measure would have to show that it is highly correlated with the very measure it claims to be superior to" (Rossiter, 2002, p. 326). Although the multitrati-multimethod (MTMM) matrix (Campbell & Fiske, 1959) has been used for the assessment of convergent validity, Rossiter (2002) criticizes it because high (low) correlation would not provide information as to which of the two measures is more (less) valid. Moreover, correlations can be low due to several reasons, some of which have nothing to do with the indicators being poor representations of the focal construct (Bollen, 1989). Given this argument, and the fact that objective data about export performance results are usually not available and that there were financial and methodological limitations to eliciting information from additional sources, we considered that assessment of convergent validity of the proposed measurement models of export performance was not feasible.

Step 6: Verification of Overall Adequacy of the Measurement Model

Besides checking the psychometric properties of indicators and latent variables and judging the level

of concurrent and predictive validity, it is necessary to verify whether the model, as an integrated set, seems to convey an overall satisfactory picture of the construct. However, the decision about what constitutes a **satisfactory picture** usually depends on the researcher's discretionary judgment, since some signs may suggest good adequacy while others may paint a poor picture. We assessed overall adequacy in terms of (**6-a**) parsimony and of (**6-b**) goodness-of-fit indices.

Parsimony (6-a) – Parsimony was assessed by means of: number of indicators; average inter-item within construct correlation; and item-to-total within construct correlation. The competing models advanced here contain at most ten indicators, which is not a very large number and would not thereby violate parsimony requirements. Furthermore, for each indicator in all five models, average inter-item within-construct correlation and all item-to-total within-construct correlations are not too high, i.e., they are each lower than .90. The rationale here is that too high correlations would mean that the indicators would jointly be too redundant to efficiently sample the domain of the construct (Briggs & Cheek, 1986).

Goodness-of-fit (6-b) – We assessed fit using both absolute and relative indices for:

. each of the stand-alone measurement models of the construct;

. each measurement model of the construct inserted into a larger measurement model with other constructs (that were part of the larger study);

. each measurement model of the construct inserted into a larger structural model (nomological network) with other constructs to which it is expected to be related (besides the dimensions of export venture performance, the structural models were composed of five other constructs: psychic distance, business distance, legal barriers in the target country, status of the export activity, and systematization of export planning; but these will not be further detailed here).

The fit indices for the stand-alone models are shown in Table 4. The significant χ^2, suggesting models do not fit data, is not conclusive because this test is oversensitive to sample size (Hair *et al*., 2006). Normed χ^2 (χ^2 / *df*) should ideally be less than 5.0. For a model with 10 or fewer observed variables estimated with a sample of more than 250 cases, Hair *et al*. (2006) recommend Tucker-Lewis fit index [TLI] and comparative fit index [CFI] both higher than .95 and root mean square error of approximation [RMSEA] below .07. Also, goodness-of-fit index [GFI] higher than .90 is usually considered the minimum threshold for acceptable fit.

Table 4: Goodness-of-fit Indices of the Stand-alone Measurement Models

		absolute			Relative	
	χ^2	χ^2 / *df*	GFI	RMSEA **	TLI	CFI
Model # P1	203.8*	6.4	.835	.138 (.120; .156)	.363	.547
Model # P2	109.9*	3.4	.911	.093 (.074; .112)	.711	.794
Model # P3	244.4*	7.0	.802	.146 (.129; .158)	.290	.448
Model # P4	63.9*	10.7	.917	.185 (.189; .241)	.257	.703
Model # P5	72.0*	9.0	.906	.168 (.134; .205)	.384	.672

* $p < .001$ ** Confidence interval shown in parentheses *df* = degrees of freedom.
Note: shadowed cells emphasize models that comply with desirable values in each fit criterion.

Model # P2 performs best, but we decided not to take this at face value because of a possible method bias. Overall, the bi-dimensional model (# P5) seems to fit data better than three-dimensional models (# P1 and # P4). Unidimensional model # P3 performs worst. Given conceptual considerations and empirical results, we decided to keep three models for further evaluation: # P1, # P4 and # P5.

We then inserted each of these three remaining measurement models into integrated measurement models with other constructs that were part of the larger research project and assessed them with the previously presented steps. Given that the integrated models contained more than 12 observed

variables, Hair *et al*. (2006) recommend GFI above .90, CFI and TLI above .92, and RMSEA below .07. The integrated measurement models were then transformed into corresponding structural models, by removal of correlational paths between dependent and independent constructs and addition of structural (causal) paths with variables that were part of the larger model of the full study. We estimated the structural model with the same sample used to assess and purify the measurement model. This practice is acceptable as long as few (less that 5%) of the parameters are changed (added, moved or dropped) along the purification phase (Hair *et al*., 2006). The process of parameter estimation and assessment of the psychometric properties of the structural models showed that one of the models was clearly superior to the others. In this model, the operationalization of the export performance construct corresponded to that of model # P5. The goodness-of-fit indices of the best-fitting structural model and the corresponding measurement model were quite reasonable (Table 5).

Table 5: Goodness-of-fit Indices (of the Integrated Measurement Model and the Structural Model)

	χ^2	χ^2 / *df*	GFI	RMSEA **	TLI	CFI
Integrated measurement model	470.3*	3.2	.957	.074 (.066; .081)	.907	.927
Structural model	519.3*	3.9	.946	.086 (.078; .094)	.873	.901

* $p < .001$ ** Confidence interval shown in parentheses *df* = degrees of freedom.

Step 7: Verification of (Measurement Parameters) Stability

Measurement parameters should present good stability – in terms of **(7-a)** change in the magnitude of indicator loadings and of **(7-b)** change in the state of statistical significance – in the transition to a larger (integrated with other constructs) measurement model and the transition to a corresponding structural model. As desired, estimated parameter values did not change substantially (in magnitude or significance level) when each of the measurement models of export performance was inserted into the larger integrated measurement models. In the three remaining integrated models, average absolute variation of export performance indicators was 7.4%, 2.9% and 2.4%, respectively. However, throughout the transition to the structural model, although all loadings remained statistically significant, they did not show good stability. Average absolute change in value for the export performance indicators was 15.4% and one of the loadings changed as much as -31.9%. These results question whether the indicators arranged in those dimensions represent the focal construct well.

Step 8: Selecting the Most Likely Model

Given the overall empirical results coupled with theoretical considerations, model # P5 was picked up as the best fitting model. Composite reliability for the two dimensions – past export revenues and their growth, and past export profitability – was fine (.84 and .71), but average variance extracted was not – although it was just marginally below the minimum threshold of .50; standardized loadings are all above .50 as desired, but only half of them are above .707 (Table 6).

Table 6: Association between Indicators and Constructs in the Most Likely Model

		standardized loading	critical coefficient	composite reliability	average variance extracted
Export revenues				*.84*	*.47*
SPasRev	Satisfaction with past export venture revenues	.714	n.a.		
PasReOt	Export venture past revenues *vs.* average revenues of other export ventures of the firm	.535	10.541*		
SPaReGr	Satisfaction with past growth of export venture revenues	.770	18.498*		
PasVGOt	Past growth of export venture volume *vs.* average volume growth of other export ventures of the firm	.694	12.334*		
Export profitability				*.71*	*.49*
SPasPro	Satisfaction with past profit margin of export venture	.834	n.a.		
PasPrOt	Past export venture profitability *vs.* average profitability of other export ventures of the firm	.528	7.991*		

n.a. = not applicable because the unstandardized parameter value was fixed at 1.0 in order to provide a measurement scale for the latent variable.
* $p < .001$

Discussion and Final Considerations

The results of the assessment process indicate that none of the models is indisputably better. Although model # P5 seems to have presented a more desirable set of measurement characteristics, it has, nonetheless, failed to achieve good signs of adequacy in some aspects. Far from being a weakness of the best fitting model, this apparent inconsistency highlights the importance of using a competing-models approach and a comprehensive set of validation procedures. In fact, the natural consequence of achieving mutually contradictory results may serve as an explicit recognition of the strengths and weaknesses of the model, which might have been missed if simpler validation procedures had been employed. We now turn to the strong vs. weak points of the best fitting measurement model and discuss some insights from this research.

Strong and Weak Aspects of the Validation Framework and of the Proposed Model

The set of validation procedures presents some strong points that ought to be stressed:

. use of a structural equation modeling approach, whereby the complex nature of the phenomenon is recognized and represented;

. use of a competing models approach; and

. employment of an extensive set of validation procedures, including assessment of content validity, psychometric properties (internal consistency, unidimensionality, reliability), concurrent and predictive validity, overall adequacy, and stability.

The best fitting model has some strong points in its favor: good overall compliance with desirable measurement properties; good stability of the parameters when inserted into a larger measurement model with other constructs (supposed antecedents of the phenomenon); and parsimony. However, the best fitting model also has some weak points:

. some signs of inadequacy, such as worse than desirable fit indices when the model was estimated in isolation of other constructs, although they did improve when jointly estimated;

. poor stability of the parameters after the transition from the larger measurement model to the corresponding structural model;

. no undisputable establishment of its dimensional structure (number and content of dimensions, and hierarchical complexity), although the conjoint discussion of empirical results and substantive arguments would suggest two dimensions;

. only two indicators remaining for the export venture profitability dimension, which may not be enough to capture the complex nature of export venture performance; in addition, the loading on one of these, although significant, is the lowest in the entire model (.528);

. no measures of performance *vis-à-vis* competitors (due to high incidence of missing data);

. no use of **objective**, but only of perceptual measures.

Insights on the Conceptual Nature and the Operational Structure of the Construct

Some interesting insights about the nature of the export performance phenomenon and the methodological procedures to measure it can be gained from this research. While other researchers (Cavusgil & Zou, 1994; Katsikeas *et al*., 2000; Lages & Lages, 2004; Lages *et al*., 2005; Leonidou *et al*., 2002; Matthyssens & Pauwels, 1996; Shoham, 1998, 1999; Styles, 1998; Zou *et al*., 1998) have also advocated and empirically determined that export performance would be a multidimensional phenomenon, our study showed how difficult it may be to unequivocally establish the number of dimensions, their specific content and the level of hierarchical complexity. Moreover, no study has tested a second-order arrangement of the construct.

The development of our model suggests only two dimensions: (i) export revenues and their growth, and (ii) export profitability. In terms of content, such dimensions would be composed of a single class of measure (economic, by virtue of the narrow conceptual domain we chose), measures of the past only, both static and dynamic orientations, and both absolute and relative references (only *vis-à-vis* other export ventures of the firm). The fact that several managers did not provide information on performance relative to competitors prevented a more thorough representation of the construct and might in fact suggest that practitioners employ a narrower (as compared to theoreticians) conceptualization of the construct. A single-order reflectively-measured structure seems to portray the nature of the construct adequately. Nevertheless, a higher-order structure could not be tested (although the relatively high correlation between the first-order dimensions indicates that it should not be hastily ruled out) because more than three first-order dimensions would be needed (Marsh & Hocevar, 1985) to make the second-order model statistically identifiable. A formative perspective was not tried because the relatively narrow conceptual frontier we chose would seem to suggest a reflective arrangement. It should be noted that the procedures for validation used here are appropriate only when the construct is measured in a reflective perspective (Bollen & Lennox, 1991; Diamantopoulos, 1999; Jarvis *et al*., 2003); a formative perspective of measurement would call for a distinct set of validation procedures.

Furthermore, empirical results of the structural relationships of export venture performance with other constructs (not reported here, but available from the authors upon request) and also substantive reasoning indicate that each dimension of export venture performance may suffer distinct influence from other constructs and may also have distinct impacts on other constructs. This suggests keeping the dimensions distinct when inserting the construct into a nomological network and in fact argues against the use of a higher-order construct or an aggregate (single) measure of the construct.

Furthermore, our findings show that measures of past performance and of future performance may represent distinct aspects of the phenomenon. This may indicate the temporal volatility of the

antecedent variables of the phenomenon, which would diminish correlation between past and (expected) future results. As a suggestion, more indicators of future performance could be collected, in order to permit the use of locally identifiable latent variables representing this facet.

Past export venture revenues and growth in past export venture revenues seem to represent one same facet of export venture performance. Similar results were reached by Zou *et al*. (1998). On the other hand, some researchers (e.g., Madsen, 1987; Shoham, 1996) have argued that change would be a distinct dimension and Shoham's (1998) empirical results indicate that revenues and growth in revenues would be distinct dimensions. It might be conjectured that measures of revenues and their growth might be poorly correlated at low levels of export intensity (low revenues and possible high growth due to the small basis for comparison) and high export intensity (where export revenues would be high, but there might be little space to grow further, due to internal or to market limitations, and the fact that the basis for comparison would be large, thereby diminishing the growth index), but might correlate well at mid-levels. On the other hand, profitability was revealed as a distinct dimension, confirming arguments and empirical results of several past studies (Cavusgil & Zou, 1994; Shoham, 1998; Zou *et al*., 1998).

Shoham (1998, p. 62) argued that "satisfaction-based measures provide richer assessments of each sub-dimension, rather than additional, independent sub-dimensions". We followed his advice – as have others (e.g., Shoham, 1999; Zou *et al*., 1998), although some have not (e.g. Lages & Lages, 2004; Lages *et al*., 2005) – and employed measures of satisfaction as additional indicators of each corresponding dimension. Internal consistency and unidimensionality tests indicate that this seems to be fine, but results from the exploratory factor analysis indicate otherwise. These measures should be submitted to further screening in future studies.

Final Remarks

From an academic perspective, the results of this paper are important because measurement soundness enhances substantive reasoning and permits to avoid inconsistent and conflicting research results. Moreover, this particular application of the validation framework reinforces the fact that the establishment of the degree of adequacy of a model is highly dependent on the set of tests it is submitted to. Therefore, one might question whether some elaborate models of export performance that have been advanced in the literature might have been rejected, or at least placed under suspicion, had additional validation checks been employed. Moreover, comparability across studies needs to be enhanced in order to achieve consistency in construct measurement.

From a managerial standpoint, this endeavor is relevant because better measures of constructs are expected to lead to sounder normative orientation by way of better theory development.

By recognizing that (a) phenomena in the social sciences are usually of a complex nature and, consequently, that (b) the content and nature of such constructs ought to be represented as latent variables, measured by multiple observed variables (indicators), whereby measurement error is explicitly incorporated, this paper has: (1) presented an integrative and stringent set of procedures for validating competing operationalizations of multifaceted constructs, and (2) illustrated the application of the validation procedures with the particular case of the export performance construct. Although drawing heavily on previously published works on construct measurement, this framework has moved further by integrating complementary issues that have been scattered around distinct pieces of conceptual, methodological and empirical research and also by better organizing the sequence of methodological and procedural steps involved in the empirical validation of measures of complex constructs.

Limitations

The study suffered from several limitations. First, the conceptualization of the export performance phenomenon was circumscribed to its economic domain and to a single export venture. Although the

choice of a narrower domain is acceptable, it limits the scope of the investigation on the nature of the construct. Second, responses might have been affected by a method bias, due to the particular sequencing of the questions in the questionnaire. Results concerning Model #2, in particular, might have been affected by this potential bias, since the association between indicators and factors mirrored the grouping of questions in the questionnaire. Third, this study was run in a single country (Brazil), therefore possibly being affected by country characteristics. Fourth, for each export venture, only one single respondent provided data. Fifth, although the companies studied were among the largest Brazilian exporters of manufactured goods, they were not necessarily as large as comparable samples from developed countries. Therefore, the results should not be generalized to other countries, industries or types of firms. Further research is needed to test these models in different research settings.

REFERENCES

Anderson, J., & Gerbing, D. (1988). Structural equation modeling in practice: a review and recommended two-step approach. *Psychological Bulletin, 103*(3), 411-423.

Anderson, J., & Gerbing, D. (1991). Predicting the performance of measures in a confirmatory factor analysis with a pretest assessment of their substantive validities. *Journal of Applied Psychology, 76*(5), 732-740.

Anderson, J., & Gerbing, D. (1992). Assumptions and comparative strengths of the two-step approach: comment on Fornell and Yi. *Sociological Methods and Research, 20*(3), 321-333.

Bagozzi, R. (1984). A prospectus for theory construction in marketing. *Journal of Marketing, 48*(1), 11-29.

Bagozzi, R., & Yi, Y. (1988). On the evaluation of structural equation models. *Journal of the Academy of Marketing Science, 16*(1), 74-94.

Bollen, K. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: a structural equation perspective. *Psychological Bulletin, 110*(2), 305-314.

Bourgeois, L. (1980). Performance and consensus. *Strategic Management Journal, 1*(3), 217-248.

Briggs, S., & Cheeks, J. (1986). The role of factor analysis in the evaluation of personality scales. *Journal of Personality, 54*(1), 106-148.

Campbell, D., & Fiske, D. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81-105.

Carmines, E., & Zeller, R. A. (1979). Reliability and validity assessment. In J. Sullivan & R. Niemi (Eds.). *Quantitative applications in the social sciences* (Vol. 7, Chap. 17, pp. 1-71). Beverly Hills: Sage Publications.

Carneiro, J. M. T., Hemais, C. A., Rocha, A. da, & Silva, J. F. da (2005). Conceptualisation and measurement of export performance: proposition of a generic analytical framework and critical assessment of the empirical literature 1999-2004. *Proceedings of the Annual Conference European International Business Academy,* Oslo, Norway, 31.

Cavusgil, S. T., & Zou, S. (1994). Marketing strategy-performance relationship: an investigation of the empirical link in export market ventures. *Journal of Marketing, 58*(1), 1-21.

Churchill, G. (1979). A paradigm for developing better measures of marketing constructs. *Journal of*

Marketing Research, 16, 64-73.

Cortina, J. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98-104.

DeVellis, R. (2003). *Scale development: theory and applications* (Vol. 26, 2nd ed.). (Applied Social Research Methods Series). Thousand Oaks, CA: Sage.

Dess, G., & Robinson, R. (1984). Measuring organizational performance in the absence of objective measures: the case of the privately held firm and the conglomerate business unit. *Strategic Management Journal, 5*(3), 265-273.

Diamantopoulos, A. (1999). Export performance measurement: reflective versus formative indicators. *International Marketing Review, 16*(6), 444-457.

Diamantopoulos, A., & Siguaw, J. (2006). *Introducing LISREL*. London: SAGE Publications.

Diamantopoulos, A., & Winklhofer, H. (2001). Index construction with formative indicators: an alternative to scale development. *Journal of Marketing Research, 38*(2), 269-277.

DuBois, F., & Reeb, D. (2000). Ranking the international business journals. *Journal of International Business Studies, 31*(4), 689-704.

Edwards, J., & Bagozzi, R. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155-174.

Fornell, C., & Larcker, D. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39-50.

Hair, J., Jr., Black, W., Babin, B., Anderson, R., & Tatham, R. (2006). *Multivariate data analysis* (6th ed.). Upper Saddle River, NJ: Prentice Hall.

Hattie, J. (1985). Methodology review: assessing unidimensionality of tests and items. *Applied Psychological Measurement, 9*(2), 139-164.

Hinkin, T. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104-121.

Hughes, M., Price, R. L., & Marrs, H. (1986). Linking theory construction and theory testing: models with multiples indicators of latent variables. *Academy of Management Review, 11*(1), 128-144

Jarvis, C., MacKenzie, S., & Podsakoff, P. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research, 30*(2), 199-218.

Katsikeas, C., Leonidou, L., & Morgan, N. (2000). Firm-level export performance assessment: review, evaluation and development. *Academy of Marketing Science, 28*(4), 493-511.

Kim, J-O., & Mueller, C. (1978). Introduction to factor analysis. In J. Sullivan (Ed.). *Quantitative applications in the social sciences* (Vol. 7, Chap. 13, pp. 1-80). Beverly Hills: Sage Publications.

Lages, L., & Lages, C. (2004). The STEP scale: a measure of short-term export performance improvement. *Journal of International Marketing, 12*(1), 36-56.

Lages, L., Lages, C., & Lages, C. R. (2005). Bringing export performance metrics into annual reports: the APEV scale and the PERFEX scale. *Journal of International Marketing, 13*(3), 79-104.

Law, K., Wong, S-C., & Mobley, W. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review, 23*(4), 741-755.

Leonidou, L., Katsikeas, C., & Samiee, S. (2002). Marketing strategy determinants of export performance: a meta-analysis. *Journal of Business Research, 55*(1), 51-67.

MacCallum, R., & Browne, M. (1993). The use of causal indicators in covariance structure models: some practical issues. *Psychological Bulletin, 114*(3), 533-541.

MacCallum, R., Roznowski, M., & Necowitz, L. (1992). Model modifications in covariance structure analysis: the problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490-504.

Madsen, T. (1987). Empirical export performance studies: a review of conceptualizations and findings. In S. T. Cavusgil & C. Axinn (Eds.). *Advances in international marketing* (Vol. 2, pp. 177-198). Greenwich: JAI Press.

Madsen, T. (1998). Managerial judgment of export performance. *Journal of International Marketing, 6*(3), 82-93.

Marsh, H., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: first and higher-order factor models and their invariance across groups. *Psychological Bulletin, 97*(3), 562-582.

Matthyssens, P., & Pauwels, P. (1996). Assessing export performance measurement. In S. T. Cavusgil & C. Axinn (Eds.). *Advances in international marketing* (Vol. 8, pp. 85-114). Greenwich: JAI Press.

McDonald, R. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Social Psychology, 34*, 100-117.

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures: issues and applications*. Thousand Oaks: Sage.

Nunnally, J. (1978). *Psychometric theory* (2nd ed.). New York: McGraw-Hill.

Peter, J. P. (1981). Construct validity: a review of basic issues and marketing practices. *Journal of Marketing Research, 18*(2), 133-145.

Rossiter, J. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing, 19*(4), 305-335.

Schriesheim, C., Powers, K., Scandura, T., Gardiner, C., & Lankau, M. (1993). Improving construct measurement in management research: comments and a quantitative approach for assessing the theoretical content adequacy of paper-and-pencil survey-type instruments. *Journal of Management, 19*(2), 385-417.

Shoham, A. (1996). Effectiveness of standardized and adapted television advertising: an international field study approach. *Journal of International Consumer Marketing, 9*(1), 5-23.

Shoham, A. (1998). Export performance: a conceptualization and empirical assessment. *Journal of International Marketing, 6*(3), 59-81.

Shoham, A. (1999). Bounded rationality, planning, standardization of international strategy, and export performance: a structural model examination. *Journal of International Marketing, 7*(2), 24-50.

Smith, A. (1999). Some problems when adopting churchill's paradigm for the development of service quality measurement scales. *Journal of Business Research, 46*(2), 109-120.

Spector, P. (1992). *Summated ratings scales construction*. Newbury Park, CA: Sage.

Steenkamp, J-B., & Trijp, H. van (1991). The use of LISREL in validating marketing constructs. *International Journal of Research in Marketing, 8*(3), 283-299.

Styles, C. (1998). Cross-cultural examination of export performance. *Journal of International Marketing, 6*(3), 5-31.

Venkatraman, N., & Grant, J. (1986). Construct measurement in organizational strategy research: a critique and proposal. *Academy of Management Review, 11*(1), 71-87.

Venkatraman, N., & Ramanujam, V. (1987). Measurement of business economic performance: an examination of method convergence. *Journal of Management, 13*(1), 109-122.

Walker, B., Olson, J., & Chow, S. (1992). Is construct validity a problem of measurement or theoretical generalization? A reply to malhotra. *Journal of Business Research, 25*(2), 187-195.

Wong, V., & Saunders, J. (1993). Business orientation and corporate success. *Journal or Strategic Management, 30*(1), 20-40.

Zou, S., Taylor, C., & Osland, G. (1998). The EXPERF scale: a cross-national generalized export performance measure. *Journal of International Marketing, 6*(3), 37-58.