Modeling the Economic Domain of the Export Performance Construct

Autoria: Jorge Carneiro, Angela da Rocha

Abstract

Several constructs in the social sciences exhibit a complex nature, which poses important challenges, both substantive and methodological, to their conceptualization and operational representation. Moreover, it has been recognized that developing appropriate representations of concepts and "specifying the relationship between concepts and operational indicators is equally important to social research as the substantive theory linking concepts to one another" (Carmines and Zeller, 1979:11). In this paper we advance and test a new measurement model of the economic domain of export performance, a construct that has not yet reached agreement in the literature as to the appropriate representation of its complex nature. To accomplish this task, we assembled a comprehensive and integrated set of validation procedures to assess the satisfactoriness of measurement models of multifaceted constructs. 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. The generic nature of the set of validation procedures makes it a readily available instrument for the validation of other complex constructs in several areas of research.

INTRODUCTION

Two stages are involved in theory building: the first is the specification of "relationships between theoretical constructs"; the second, the description of "relationships between constructs and measures" (Edwards and Bagozzi, 2000:155). These two stages are critical, since "theories cannot develop unless there is a high degree of correspondence between abstract constructs and the procedures used to operationalized them." (Peter, 1981:133).

A construct of critical importance to research on exporting is export performance. Yet, although the construct has received the attention of several scholars, no agreed upon measurement model of the construct has yet been reached. This lack of agreement makes it difficult to compare research findings and to develop a shared body of knowledge. In order to contribute to a better understanding of the construct, this paper presents the development and test of a new measurement model so as to offer new insights about the conceptual nature and the operational representation of the construct. Such objective was accomplished by:

i) advancement of a rather comprehensive set of procedures for validating measurement models of complex and multifaceted constructs;

ii) empirical application of these validation procedures for the development of a new measurement model of the export performance construct; and

iii) discussion of the nature and structure of the construct based on the interplay between conceptual reasoning and empirical results.

In our quest for a comprehensive set of procedures to assess and validate measurement models of complex and multifaceted constructs, we critically reviewed and contrasted several works, drawn from quite diverse areas of study, including psychology, education, organizational studies, statistics, strategic management, marketing, and international business. We then organized various guidelines for validation – which, to our best knowledge, can only be found scattered through several independent pieces of conceptual, methodological, empirical, and meta-analytical work – into a comprehensive, internally consistent and a ready-to-employ set of methodological procedures. By putting together and operationalizing several perspectives and criteria by which to judge the adequacy of measurement models, we believe to have produced a better instrument for construct validation.

This paper is organized as follows. We first propose and comparatively test five competing models of the construct, by discussing content validity issues, psychometric properties, concurrent and predictive validity, overall adequacy and stability. New highlights that can be reached from our model are then addressed. Final remarks and some suggestions for future studies close the paper.

PROPOSITION AND VALIDATION OF A NEW MEASUREMENT MODEL OF EXPORT PERFORMANCE

Determining the Conceptual Domain of the Construct

There is a consensus among theoreticians that the first step in the process of developing a new measure for a construct is to carefully specify its boundaries (Churchill, 1979; Spector, 1992; DeVellis, 2003). This conceptual task involves the identification of "what is and what is not included in the domain" (Churchill, 1979:67). As we started the development of our measurement model, we had to decide on the appropriate conceptual domain for which our measurement model would be developed and strive for content validity or at least provide evidence of content adequacy (Schriesheim *et al*., 1993).

Therefore, a first step in the process of arriving to a new measurement model of export performance was to identify the different facets of the construct. Fortunately, several efforts have been described in the literature, which aimed at advancing a generic framework for the delimitation of the conceptual domain of the export performance construct (e.g. Matthyssens and Pauwels, 1996; Katsikeas *et al.,* 2000; Carneiro *et al.*, 2005). From these studies, export performance can be conceptualized as a multi-dimensional construct that includes several classes of measures (economic, market, behavioral, strategic, and overall), two alternative frames of reference (absolute vs. relative) and two perspectives of temporal orientation (static vs. dynamic), as depicted 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: Matthyssens and Pauwels (1996), Katsikeas *et al.* (2000) and Carneiro *et al*. (2005)

In spite of the various classes of measures that can be found in the literature, we decided to concentrated only on the economic aspect of the export performance construct – a narrow, but nonetheless important sub-domain of the phenomenon– in order to get a more accurate representation. We then chose three dimensions to represent the economic domain of export performance: export revenues, growth in export revenues and export profitability (Figure 2).

Whether a construct ought to be viewed as unidimensional or multidimensional depends on the level of abstraction used to define it (Jarvis *et al*., 2003). One can look at each facet as a separate construct, but at 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 and Mobley, 1998). Are they just correlated or 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? In case of a higher-level representation, a reflective (*vis-à-vis* a formative) perspective of measurement (Bollen and Lennox, 1991) would make sense if, among other aspects, the dimensions are expected to co-vary together and 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, it should be recognized that 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.

Figure 2 – Economic Domain of Export Performance

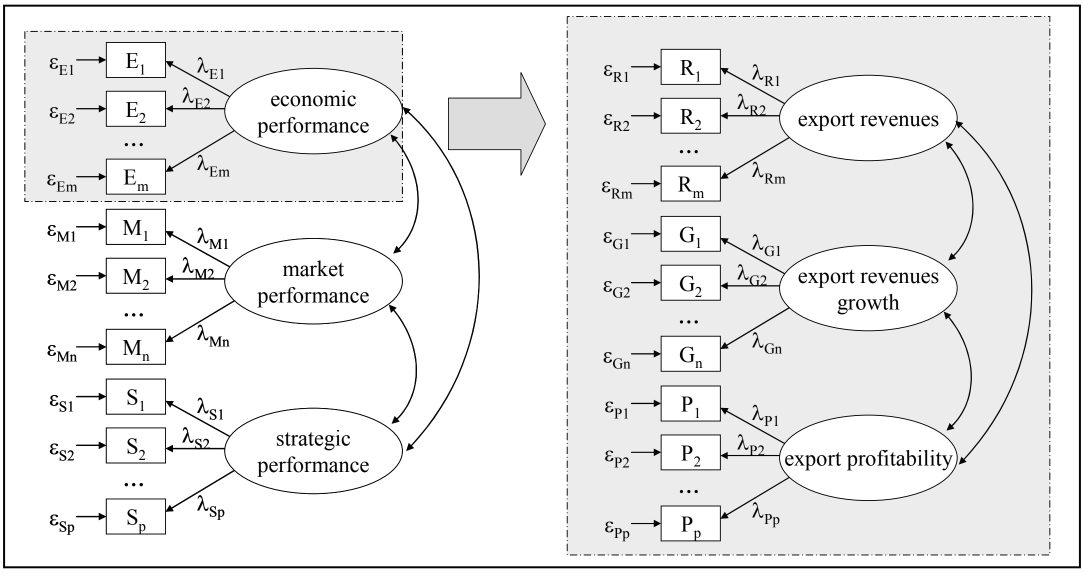


Figure 3 – *A Priori* Dimensions and Indicators of the Economic Domain of the Export Performance Construct

Dimension	Indicator	Description of the meaning of the indicator	Conceptual aspects covered *
Export revenues	SPasRev [1]	Satisfaction with past export venture's revenues	absolute, past (static)
	PasReOt [2]	Export venture's past revenues *vs.* average revenues of other export ventures of the firm	relative (to other export ventures), past (static)
	PasVoCo [2 +]	Past export venture's volume *vs.* other Brazilian firms exporting to the same country	relative (to competitors), past (static)
	FutVoOt [2]	Expected future export venture's volume *vis-à-vis* expected average volume of other export ventures of the firm	relative (to other export ventures), future (static)
Export revenues growth	SPaReGr [1]	Satisfaction with past growth of export venture's revenues	absolute, past (static)
	PasVGOt [2]	Past growth of export venture's volume *vis-à-vis* 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's volume *vis-à-vis* expected average volume growth of other export ventures of the firm	relative (to other export ventures), future (dynamic)
Export profitability	SPasPro [1]	Satisfaction with export venture's past profit margin	absolute, past (static)
	PasPrOt [2]	Past export venture's profitability *vis-à-vis* average profitability of other export ventures of the firm	relative (to other export ventures), past (static)
	FutProf [2]	Expected future export venture's profitability	absolute, future (static)
	FutPrOt [2]	Expected future export venture's profitability *vis-à-vis* 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 later on dropped due to high incidence of missing data
[1] rated on five-point semantic differential scales with anchor words "very dissatisfied" … "very satisfied"
[2] rated on five-point semantic differential scales with anchor words "much lower" … "much higher"

We initially modeled the construct as multi-dimensional because we believed that it is composed of distinct, albeit related, aspects (that is, revenues, revenues growth, and profitability).

Content validity was sought by means of an extensive review of past conceptual and empirical literature on the focal construct and other related ones, as well as theoretical reflection and consultation with academic experts. We reviewed 62 empirical studies, covering a period from 1985 to 2005, and uncovered 116 distinct indicators of the construct. Initially, we selected 11 operational indicators (Figure 3) to measure the three dimensions. As a set, these indicators reasonably cover the diversity of conceptual aspects of the phenomenon – specifically, absolute vs. relative (either to competitors or to other export ventures of the firm), static vs. dynamic operational, and past vs. future measures.

Data Collection

A survey was conducted with the largest Brazilian exporters of manufactured products. The survey was part of a larger research project and was conveyed through a four-page structured questionnaire, which covered not only indicators of export performance but also several variables related to determinants of export performance, but only the export performance variables are reported in the present paper. Firms were selected from a list of the 5,000 largest Brazilian exporters provided by FUNCEX, a private foundation supported by Brazilian exporters. Multinationals, service firms, trading companies and exporters of commodities were removed from the list, resulting in a population of 3,057 Brazilian exporters of manufactured goods.The unit of analysis was the export venture, that is, the exporting of a given (line of) product to a given country since the alternative of measuring export performance at the corporate level would provide information on aggregated results of several export ventures, but almost no insights on the individual contribution of each export venture or product, be it “success” or “failure” (Matthyssens and Pauwels, 1996).

Semantic-differential scales were employed instead of asking firms to provide objective information (*cf*. Matthyssens and Pauwels, 1996; Shoham, 1998). This was deemed necessary to improve response rate and minimize missing values since most firms do not keep “objective” public data for each export venture, segregated from other ventures or other projects in the firm. Firms were mailed a questionnaire with a pre-paid return envelope.

A sample of 448 exporters was obtained resulting in an effective response rate of 15.5%. No systematic bias was observed, which suggests that the sample can be considered a reasonable representation of the population of the largest Brazilian exporters of manufactured products. Variables and cases with more than 15% missing values were removed, which led to the exclusion of one indicator of export performance (past export venture’s volume *vs*. other Brazilian firms exporting to the same country). The resulting sample size was 414 cases. Since missing data exhibited an MCAR (missing completely at random) pattern at the 10% significance level, it was possible to estimate the missing values. The resulting sample showed no indication of the presence of multivariate outliers as far as the 10 remaining operational indicators of export performance were considered. Parameters were estimated by an asymptotic distribution-free method (ADF) because variables did not follow a normal distributional pattern.

Purifying the Measure

In order to avoid capitalization on chance, i.e., overfitting to the idiosyncrasies of a given sample (MacCallum, Roznowski and 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 run on the calibration sample in order to check whether the factorial structure (number of distinct factors and the particular association of indicators to factors) that emerged from empirical data replicated what was expected from theoretical considerations. This would provide a preliminary account of the dimensionality and structure of the construct.

Figure 4 – Pattern of Associations between Indicators and Factors Uncovered by EFA

	Factor		
	1	2	3
SPasRev	-.04	**.73**	.00
PasReOt	.03	.12	**.68**
VFutOut	**.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**
VFutOut	**-.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 to factors is shown in bold-face type

In the orthogonal solution (Figure 4), the first factor can be interpreted as "expected future (absolute and relative) export 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 performance" and the third factor would be "past relative export performance". The oblique solution produced a similar patter 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.

Although the rotated factors can be neatly interpreted, they may not be useful for some studies where the interest might lie on understanding the effects of other variables on distinct aspects of economic performance, e.g., revenues vs. profitability, and not on past vs. future performance. Besides a method bias may have played a role here, since questions in the questionnaire were placed together in the same way they "came out" associated in the rotated factor solutions. Given the fact that empirical evidence on the factor structure was not conclusive, we decided to keep five models for further comparative assessment (Figures 5-a through 5-e).:

- Model # P1: three factors as suggested by the initial conceptual discussion, that is, export revenues (past and future, absolute and relative), export revenues growth (past and future, absolute and relative), and export 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, that is, satisfaction with past absolute export performance, past relative export performance, and future (absolute and relative) export performance, composed of, respectively, three, three and four indicators; and
- Model # P3: one single factor (export performance) incorporating all ten indicators.

In addition, 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 indicators of past performance *vis-à-vis* those of future performance. So, we decided to use two other models that would not incorporate indicators of future performance:

- Model # P4: three factors involving only past indicators, that is, past export revenues (absolute and relative), past export revenues growth (absolute and relative), and past export profitability (absolute and relative), each composed of two indicators; and
- Model # P5: two factors involving only past indicators, that is, past export revenues (absolute and relative) and their growth thereof, and past export profitability (absolute and relative), composed respectively of four and two indicators.

Figure 5-a – Model # P1 of export performance

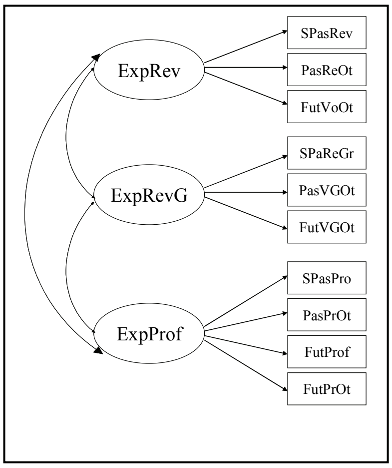


Figure 5-b – Model # P2 of export performance

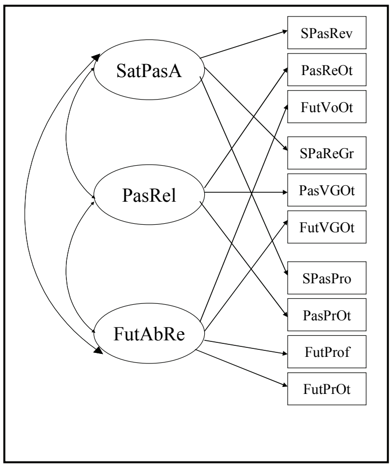


Figure 5-c – Model # P3 of export performance

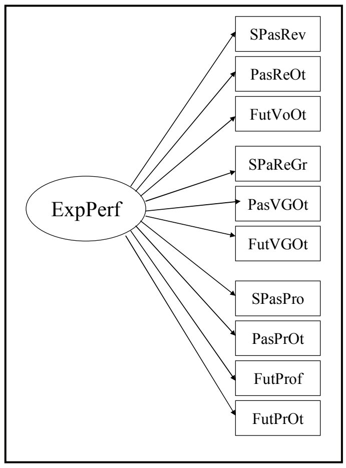


Figure 5-d – Model # P4 of export performance

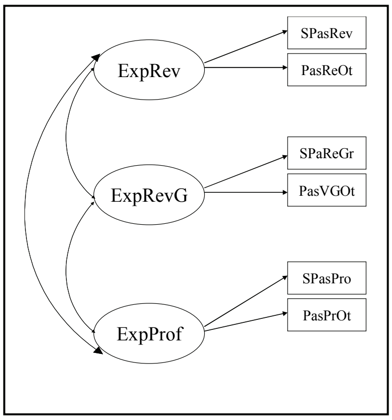
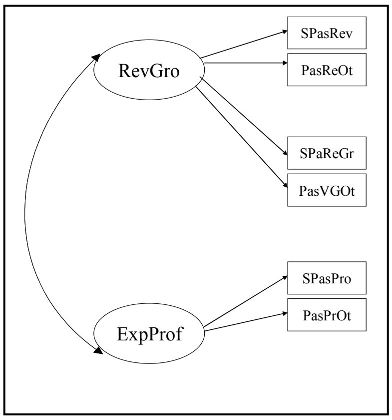


Figure 5-e – Model # P5 of export performance



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

Evaluating the Psychometric Properties of the Models (Internal Consistency, Unidimensionality, and Reliability)

A confirmatory factor analysis (CFA) was run for each pre-specified competing measurement model in order to verify whether the individual dimensions of the construct (as operationalized in each model) satisfactorily attended the desirable psychometric properties, that is, *internal consistency, unidimensionality, and reliability* (Anderson and Gerbing, 1988, 1991, 1992; Bollen, 1989; Carmines and Zeller, 1979; DeVellis, 2003; Hair *et al*., 2005).

In order to make a model statistically identifiable, it is necessary to define a measurement scale for each latent construct (Anderson and Gerbing, 1988; MacCallum and 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 (standardizing it). We chose the latter option since we were more interested in estimated indicators' loadings. As for the error terms, we fixed their loadings onto their respective indicators at 1.0 because we were interested in their variance (that is, error variance). For all models, ADF estimation converged and no improper solutions were produced.

Assessment of internal consistency – 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. In all estimated models, signs of loadings are compatible with theoretical expectations and, in each construct, have the same direction. Besides, inter-item within-construct correlations (that is, correlations between pairs of indicators associated with the same latent variable), as implied by the measurement model, are adequately high (≥.20). The rationale for this requirement is that if an indicator does not correlate well with others *a priori* associated with the same latent variable, then it would not seem to represent the same underlying factor as the others. Also, correlations are all significant at the 5% level. Moreover, for each latent variable ("dimension") – except for the latent variable 'export revenues' in model # P1 –, average inter-item within-construct correlations is adequately high (≥.30), so the set of indicators *a priori* associated with the same latent variable seems to offer a good joint representation of it. It is worth mentioning that there is a visually large difference between observed correlations and model # P1-implied correlations, the former being usually lower than the latter, suggesting that the indicators may not be so associated with one another as the model implies or that the sample may not be a good representation of the population.

We also checked whether each item-to-total within-construct correlation (correlation of a given indicator with the sum of the scores of all other indicators associated with the same latent variable, excluding itself) was high enough (≥.50). The original theoretically-derived model # P1 fails this quality criteria in half of the correlations. EFA-derived model # P2 performs best and models # P3 and # P5 are reasonably acceptable. In all models such correlations were statistically significant, as desired. As for model # P4, this test is redundant since each dimension has only two indicators. As for, the magnitude of the average item-to-total within construct correlations – the higher, the better –, 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 (≥.50 and, ideally, ≥.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.*, 2005). 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.

Assessment of Unidimensionality - Unidimensionality refers to the extent to which a set of items reflects one single underlying trait (Hair *et al.*, 2005). In order for unidimensionality to be established, inter-item within-construct correlations should be statistically higher than inter-item between-construct correlations. A visual inspection of the correlations seriously questions the validity of model # P1 and raises some suspicion over 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.

Also, it is desirable that an item does not show large negative standardized residuals (absolute value above 2.58, *cf.* Diamantopoulos and Siguaw, 2006) with items in its assigned dimension because, since residuals are differences between empirically observed covariances and model-derived covariances, a large negative residual would suggest that the two items seem not to be related to the same dimension (Steenkamp and Van Trijp, 1991). Models # P2 and # P4 satisfy this condition, suggesting that their indicators may in fact "cluster together". However, given that model # P2 proposes a rather different, from model # P4, grouping of indicators around dimensions, these results make it difficult to clearly discern the dimensionality 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 seems 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 because, if items *a priori* assigned to distinct dimensions show a large positive standardized residual, this would imply cross-loading (Steenkamp and Van 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, not to distinct dimensions. In model # P1 there are four (12%) between-construct standardized residuals that are high (greater than 2.58), suggesting that maybe revenues, growth and profitability might somehow represent the same facet, rather than three distinct dimensions, of the construct. There are some large standardized residuals between satisfaction-related indicators, but this may be due to a possible method bias. On the whole, the pattern of inter-item between-construct residuals does not invalidate 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 it were allowed to be freely estimated (Diamantopoulos and Siguaw, 2006) – was reasonably small. In all five models under evaluation here, the magnitude of standardized expected changes of paths linking two indicators directly to one another or of correlations between error terms are small (less than 25%) compared with the standardized loadings of the respective indicators. This is a desirable feature property as it means that there is no reason to believe that indicators or error terms should 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 and Siguaw, 2006: 108). A value of 3.84 or more would mean a significant (at 5%) improvement in model fit. In models # P3 and # P5 modification indices indicate 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 indices 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.

We also performed tests of discriminant validity, that is, of whether dimensions that are theoretically supposed to be distinct do seem to be statistically distinct. 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 that 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 and Larcker, 1981; Hair *et al*., 2005). 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 restricts the correlation between two latent variables to be 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, thus 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. All in all, results suggest that in each model the proposed dimensions seem to be distinct, except maybe for model # P1.

Third, we tested whether the Δ-χ^2 between a model where each dimension keeps its a *priori* assigned indicators and one where all indicators of any two dimensions are considered indicators of one same dimension was statistically significant. If Δ-χ^2 is not statistically significant, the two dimensions cannot be considered distinct (Hair *et al*., 2005). 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 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 indicates that it seems to measure 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 – The next step was to assess the reliability of each latent variable and that of each individual indicator. Coefficient alpha was not used because it assumes identical loadings and equal error variances for all indicators of each latent variable, which is not the case. 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 and Yi, 1988) was used instead 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 representatives of their constructs. 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 and 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.*, 2005). Low (less than 0.5) 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.*, 2005). 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 assess the reliability of each individual item that composes the scale, since the reliability of the scale also depends on the reliability of its indicators (Rossiter, 2002). 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 not able 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 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 models. This rule may seem counter-intuitive and needs justification. In Diamantopoulos and Siguaw's (2006: 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: 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."

Assessing Concurrent and Predictive Validity

In order to test for concurrent validity, we elicited information on two overall assessments of export performance, which would somehow summarize the construct, rated as separate questions (as suggested by Diamantopoulos and Winklhofer, 2001, and 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. We also tested for predictive validity, that is, the ability of the new measure to predict intentions of future behavior (Nunnally, 1978; Smith, 1999). The question was phrased "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 ten indicators of export performance. Except for PasReOt, all indicators had a correlation significant at the 1% level, but four of them were relatively low (in the range of .16 through .29) thereby suggesting just reasonable predictive validity.

Assessing Overall Adequacy

We assessed overall adequacy in terms of parsimony and of fit indices. Models should be as parsimonious as possible. The competing models advanced here contain at most ten indicators, which would not violate parsimony requirements. Besides, 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, that is, 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 and Cheek, 1986). Besides checking the psychometric properties of indicators and latent variables and judging the level of concurrent and predictive validity, we also assessed the overall adequacy of each model as an integrated set. However, the decision about what is 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 chose to comparatively assess goodness-of-fit by absolute and relative indices (Figure 6). The significant χ^2, suggesting models do not fit data, is not conclusive because this test is over sensitive to sample size. 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*. (2005) 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.

Figure 6 – Goodness-of-fit indices of the measurement models of export performance

		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

Note: shadowed cells emphasize the three best models in each given fit criterion

Given overall fit indices – which, except for GFI, are worse than desired –, one would be tempted to abandon the unidimensional model (# P3). Model # P2 performs best, but, again, we decided not to recommend it because of a probable method bias that might have played a role. Overall, the bi-dimensional model (# P5) seems to fit data better than three-dimensional models (# P1 and # P4). Given conceptual and empirical considerations, 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.

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. 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*., 2005). 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 (Figure 7).

Figure 7 – Goodness-of-fit indices (integrated measurement model and 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

Assessing Stability

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, along 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.

Selecting the "Winner" 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, but only half of them are above .707.

DISCUSSION AND FINAL CONSIDERATIONS

The results of the assessment process indicate that there is no indisputably better model. Although all in all model # P5 seems to have presented a more desirable set of measurement characteristics, it has, nonetheless, failed to have achieved good signs of adequacy in some aspects. Far from being a weakness of the "winner" 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. Now we turn to the strong vs. weak points of the winner measurement model and discuss some insights from this research.

Strong and Weak Aspects of the Proposed Measurement 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 construct is recognized and represented;
- use of a competing models approach; and
- employment of an extensive and well-grounded set of validation procedures, including assessment of content validity, psychometric properties (internal consistency, unidimensionality, reliability), concurrent and predictive validity, overall adequacy, and stability.

The winner 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 winner model also has some weak points that should not go unnoticed:

- some signs of inadequacy, such as worse than desirable fit indices when the model was estimated in isolation of other constructs, although they improved 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, hierarchical complexity), although the conjoint discussion of empirical results

and substantive arguments would suggest two dimensions;

- no discussion of possible formative perspective of measurement (although this can be justified given the rather narrow frontier we chose to operationalize the phenomenon);
- 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. Some of them confirm empirical results of past works, while some question past results. While other researchers (Cavusgil and Zou, 1994; Matthyssens and Pauwels, 1996; Katsikeas *et al*., 2000; Lages and Lages, 2004; Lages *et al*., 2005; Leonidou *et al*., 2002; Shoham, 1998, 1999; Styles, 1998; and Zou *et al*., 1998) had 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 (types of operational indicators) and the level of hierarchical complexity. Since there has not been agreement in the literature on the conceptual frontier of the construct and on which operational indicators to collect, it is not surprising that conflicting results have been reached. Besides, no study that we know of has tested a second-order arrangement of the construct.

The development of our model suggests only two dimensions: export revenues and their growth, and export profitability; in terms of content, such dimensions would be composed of a single class of measure (economic, by consequence of the narrow conceptual domain we chose), measures of the past only, both static and dynamic measures, and both relative (only *vis-à-vis* other export ventures of the firm) and absolute measures; a single-order, reflectively-measured, structure seems to adequately portray the nature of the construct, but 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 and Hocevar, 1985) in order to make the second-order model statistically identifiable. A formative perspective was not tried because the relatively narrow, albeit relevant, conceptual frontier would seem to suggest a reflective arrangement. It should be noted that the procedures for validation used here are appropriate only when the construct in measured in a reflective (*vs*. formative) perspective (Bollen and Lennox, 1991; Diamantopoulos, 1999; Diamantopoulos and Winklhofer, 2001; Jarvis *et al*., 2003).

Furthermore, empirical results of the structural relationships of export performance with other constructs (not reported here) indicate that each dimension of export performance may suffer distinct influence from other constructs. This suggests keeping the dimensions distinct and may in fact argue against the use of a higher-order construct or an aggregated (single) measure of the construct.

Also, our results show that measures of past performance and of future performance may represent distinct aspects of the phenomenon and not just complementary aspects of the same facets. One might conjecture that 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, so as to permit the use of locally identifiable (three or more operational indicators, *cf*. Hair *et al*., 2005) latent variables representing this facet.

It is interesting to notice that (past) revenues and growth in (past) revenues seem to represent the same facet of export 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 argued that revenues measures 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 demand limitations, and the fact that the basis for comparison would be large thus diminishing the growth index), but might correlate well at mid-levels.

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 and Lages (2004) and Lages *et al*. (2005) – and employed measures of satisfaction as additional indicators of the other corresponding dimensions. Internal consistency and unidimensionality tests indicate that this seems to be fine.

By recognizing that (a) constructs in the social sciences are usually of a complex nature – they are multifaceted, usually not directly observable, and cannot be measured with full precision – and, as a consequence, (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 set of procedures for empirically 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, 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, procedural and instrumental steps involved in the empirical validation of measures of complex constructs. The generic nature of the set of validation procedures makes it a readily available instrument for the validation of other complex constructs in several areas of research.

References

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.

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. (1988). Structural Equation Modeling in Practice: A Review and Recommended Two-Step Approach. *Psychological Bulletin*, 103: 411-423.

Bagozzi, R. (1994). Structural Equation Models in Marketing Research: Basic Principles. In R. Bagozzi, (ed.). *Principles of Marketing Research*, p.317-385, Oxford: Blackwell.

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*. John Wiley & Sons.

Bollen, K. & Lennox, R. (1991). Conventional Wisdom on Measurement: A Structural Equation Perspective. *Psychological Bulletin*, 110(2): 305-314.

Briggs, S. & Cheeks, J. (1986). The Role of Factor Analysis in the Evaluation of Personality Scales. *Journal of Personality*, 54(1): 106-148.

Carmines, E. & Zeller, R. (1979). Reliability and Validity Assessment. In Sage University Paper Series on *Quantitative Applications in the Social Sciences*, 07-017, 6th printing, Beverly Hills: Sage Publications.

Carneiro, J., Hemais, C., Da Rocha, A. & Da Silva, J. (2005). Conceptualisation and Measurement of Export Performance: Proposition of a Generic Analytical Framework and Critical Assessment of the Empirical Literature 1999-2004. Oslo, Norway: *Proceedings* of *the European International Business Academy, Annual Conference*, 31p.

Cavusgil, S. & Zou, S. (1994). Marketing Strategy-Performance Relationship: An Investigation of the Empirical Link in Export Market Ventures. *Journal of Marketing*, 58, 1-21.

Churchill, G. (1979). A Paradigm for Developing Better Measures of Marketing Constructs. *Journal of Marketing Research*, 16(Feb), 64-73.

DeVellis, R. (2003). *Scale Development: Theory and Applications*. 2nd ed., Applied Social Research Methods Series, vol.26, Thousand Oaks, CA: Sage Publications.

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 Ltd.

Diamantopoulos, A. & Winklhofer, H. (2001). Index Construction with Formative Indicators: An Alternative to Scale Development. *Journal of Marketing Research*, 38(2): 269-277.

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*, 17: 39-50.

Hair, Jr., J., Black, B., Babin, B. & Anderson, R. (2005). *Multivariate Data Analysis*. 6th ed., Upper Saddle River, NJ: Prentice Hall.

Hughes, M., Price, R. & Marrs, D. (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.

Lages, L., Lages, C. & Lages, C. (2005). Bringing Export Performance Metrics into Annual Reports: The APEV Scale and the PERFEX Scale. *Journal of International Marketing*, 13(3), 79-104.

Lages, L. & Lages, C. (2004). The STEP Scale: A Measure of Short-Term Export Performance Improvement. *Journal of International Marketing*, 12(1), 36-56.

Law, K., Wong, C.-S. & 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, 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: *Advances in International Marketing*, vol.2, pp.177-198, Greenwich, CT: JAI Press Inc.

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 Cavusgil, S. T. & Axinn, C. (eds.). *Advances in International Marketing*, Greenwich, CT: JAI Press Inc., vol.8, pp.85-114.

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

Peter, J. (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: 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: 385-417.

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.

Shoham, A. (1998). Export Performance: A Conceptualization and Empirical Assessment. Journal of International Marketing, 6(3), 59-81.

Shoham, A. (1996). Effectiveness of Standardized and Adapted Television Advertising: An International Field Study Approach. *Journal of International Consumer Marketing*, 9(1), 5-23.

Smith, A. (1999). Some Problems When Adopting Churchill's Paradigm for the Development of Service Quality Measurement Scales. *Journal of Business Research*, 46: 109-120.

Spector, P. (1992). *Summated Ratings Scales Construction*. Newbury Park, CA: Sage Publications.

Steenkamp, J.-B. & Van Trijp, H. (1991). The Use of LISREL in Validating Marketing Constructs. *International Journal of Research in Marketing*, 8: 283-299.

Styles, C. (1998). Cross-Cultural Examination of Export Performance. *Journal of International Marketing*, 6(3), 5-31.

Zou, S., Taylor, C. R. & Osland, G. E. (1998). The EXPERF Scale: A Cross-National Generalized Export Performance Measure. *Journal of International Marketing*, 6(3), 37-58.