The STEP Scale: A Measure of Short-Term Export Performance Improvement

This article is a direct response to a recent observation in the literature that managers appear to be short-term oriented in their assessment of the performance of an export venture (Madsen 1998). On the basis of a cross-national survey of exporting firms, the authors present a three-dimensional scale for assessing managerial judgment of short-term export performance (i.e., the STEP scale). The three dimensions are (1) satisfaction with short-term performance improvement, (2) short-term exporting intensity improvement, and (3) expected short-term performance improvement. The scale presents evidence of reliability as well as convergent, discriminant, and nomological validity, and it reveals factorial similarity and factorial equivalence across both samples. The authors outline managerial and public policy implications that stem from the scale and identify avenues for further export marketing research.

Managers’ maps of export performance are often very static, narrow, and short-term oriented.

—Madsen (1998)

The measurement of export performance is a topic of current interest for managers, public policymakers, and marketing researchers (Katsikeas, Leonidou, and Morgan 2000). Research that analyzes export performance is of managerial interest because it may provide guidelines for firms to reduce production costs and dependency on the domestic market while stabilizing cyclical demand. From the point of view of governments, a better understanding of export performance is also crucial because it allows for the accumulation of foreign exchange reserves; enhances societal prosperity; and helps national industries develop, improve productivity, and create new jobs (Czinkota 1994). These increasing micro and macro policy concerns about exporting activity are leading to a growing interest in export-performance research (Zou, Taylor, and Osland 1998).

Beginning with the pioneering work of Tookey (1964), researchers have been examining export performance for nearly four decades. Because of the innumerable difficulties associated with export-performance conceptualization, operationalization, and measurement, it had not been possible to find consensus on a unique export-performance construct.

Submitted October 2002
Accepted June 2003

© Journal of International Marketing
Vol. 12, No. 1, 2004, pp. 36–56
ISSN 1069-031X
during this period. The existing diversity of measures affects the reliability of existing findings, because researchers question whether existing performance findings are a consequence of either independent variables or export-performance operationalization (Zou, Taylor, and Osland 1998). To overcome this state of affairs, a few notable efforts to advance export theory have been made in recent years through the development of broad export-performance scales that can be used to compare findings across different national settings (e.g., Styles 1998; Zou, Taylor, and Osland 1998). Nevertheless, and despite Madsen’s statement (1998, p. 91) that “managers’ maps of export performance are often very static, narrow, and short-term oriented,” none of the studies were particularly concerned with the assessment of short-term export performance.

We believe that this lack of concern is surprising, given that this is a topic of both practitioner and academic interest. Recently, the Marketing Science Institute (2002) announced that the need to develop global and international measures to assess short-term performance was among its “Top Tier Priority Topics” for 2002–2004. This is a particularly important topic at both practitioner and academic levels because it is not possible to make future decisions without the evaluation of export ventures in the short run (Madsen 1998). When assessing performance, both managers and public policymakers emphasize short-term factors and are particularly worried about setting annual targets and assessing annual export-performance improvement (Lages and Montgomery 2003). Although long-term export performance is crucial, if the exporting activities of the firm are unsuccessful in the short run, it will be extremely difficult for both managers and public policymakers to focus on the future. Altogether, this provides support for Madsen’s view (1998, p. 82) that “the real world is not as simple as the economic theory, according to which the long-term profitability or maximization of rent-earning abilities are the ultimate goals for the firm.” Managers appear to be short-term oriented in their definition of their strategies and assessment of the performance of an export venture.

At the research level, an integrated approach to short-term performance measurement is lacking. It is our aim to develop a scale (i.e., the STEP scale) that researchers can use, in order that the quality of empirical research on export performance is improved. Madsen’s (1998) research indicates that managers have a difficult time directly assessing the quality of export performance. To do so, they often seek out benchmarks and vary their assessments by venture. Although the STEP scale does not directly address these issues, it provides a research path that can assist in clarifying the actual processes underlying managerial assessments and future
planning directions. Furthermore, the STEP scale allows for the incorporation of firms’ strategic goals and objectives without an exhaustive specification for each venture.

In summary, this article aims to develop a short-term export performance measure. Through the STEP scale, we expect to contribute to theory development by providing a unified measure for capturing short-term performance, which will contribute to a greater consensus in future literature that explores the topic of export performance. The STEP scale may also be instrumental in the evaluation of export activities. It can be applied to multiple countries, with the final objective of helping managers and public policymakers assess the performance of short-term exporting ventures, which will help in the development of future tactical and strategic decisions.

In the next part of this article, we develop the STEP scale, which we test using a survey of Portuguese and British exporting managers. We then present the results and discuss their implications for theory. We conclude with the implications of the results for public policymaking and managerial practice, and we consider fruitful directions for further research.

In the export-marketing literature, researchers have used many different measures to assess export performance (Diamantopoulos and Schlegelmilch 1994) because no consensus exists about its conceptual and operational definitions (Shoham 1998; Zou, Taylor, and Osland 1998). This may explain why the task of addressing performance is so complex. Indeed, as Bonoma and Clark (1988, p. 1) comment, “perhaps no other concept in marketing’s short history has prove[d] as stubbornly resistant to conceptualization, definition, or application.”

Although some researchers might argue that it is possible to measure export performance with a single variable, we argue that to capture the complexity of export success, it is advisable to construct a scale based on a set of different variables (Bijmolt and Zwart 1994). The aggregation of various performance measures into a single measure of export performance partially overcomes the difficulty of performance measurement (Katsikeas, Leonidou, and Morgan 2000). This is the approach that we incorporate herein.

In terms of the mode of performance assessment, research might use objective and subjective indicators (Venkatraman and Ramanujam 1986). The measure we develop in this article is subjective, which is an approach used by nearly half the studies in export marketing (Katsikeas, Leonidou, and Morgan 2000). More specifically, in this study, we rely on the manager’s perception of export performance to develop sub-
jective measures. Because the unit of analysis is a specific export venture, responding executives have detailed knowledge of the venture’s export performance.

There are several factors that support the use of a subjective approach to performance measurement. For research purposes, it is often impossible to establish a common definition or fixed reference points across firms. Financial success for one company may constitute failure for another. This may be because performance itself is a complex construct in the view of the firm (Greve 1998). Managers tend to use their own perceptions of performance, rather than objective values, to formulate their own decisions (Bourgeois 1980). By measuring perceived performance, instead of performance per se, we are able to capture the degree to which performance has matched the aspiration levels of the firm from one year to the next. It is as if we established an imaginary boundary line that can be used as a reference point for success and failure. A firm realizes success if performance is better than or equal to expected performance, and it realizes failure if performance is lower than expected. This line can also be a useful starting point for future decision making. Depending on the firm’s location relative to the line, it will take different actions. For example, an improvement from a good position in the previous year may require much more effort than an improvement from a bad position. In broad terms, it can be assumed that an individual export venture will be successful when it meets or exceeds its targets (Lages and Jap 2003).

Although objective assessments for the measurement of actual performance can be considered trustworthy, this type of approach may raise different measurement problems. First, as Styles (1998) indicates, because samples are often drawn from a heterogeneous population of exporting firms, the researcher, not the manager, usually fixes the imaginary boundary line between success and failure. A second measurement problem is that some measures (e.g., profitability, return on investment, cash flow) might raise comparability problems because of different accounting practices across firms. A third practical concern is that company reports and financial statements rarely distinguish between domestic- and export-market operations and even more rarely provide specific information on the different export ventures. Moreover, the obtaining of accurate financial data on export performance is a difficulty researchers acknowledge, because export managers might be unwilling to respond openly and effectively to absolute values (Katsikeas, Piercy, and Ionnidis 1996; Yang, Leone, and Alden 1992).

Finally, both stakeholders and managers may have different opinions about which operational measures to use when setting targets, and thus it becomes difficult to find agreement
on how to use financial measures to assess export performance (Madsen 1998). Performance assessment is often idiosyncratic to the type of firm and its setting. For example, financial export-performance measures do not indicate whether a firm has adequately exploited existing export opportunities (Cavusgil 1984). Moreover, the existing differences in terms of market characteristics, competition, and technology intensity may lead to a comparison of financial measures that does not have the same meaning across the various firms (Katsikeas, Piercy, and Ionnidis 1996).

In summary, it might be extremely difficult for managers to provide accurate exporting information through financial data. All these reasons might explain why subjective measures have proved more valid in measuring export performance and in determining the way performance is associated with managerial decisions (see Katsikeas, Leonidou, and Morgan 2000, p. 505). Nevertheless, because we do not include objective, financial measures in this study (and because of the unavailability of this type of data for export operations), the scale we present herein should be regarded as suggestive rather than conclusive. Consequently, when using the STEP scale, researchers, managers, and public policymakers alike must be aware that the scale presents unfinished measures of export performance because it cannot completely capture the complex domain of the export-performance construct.

We subsequently develop a measurement scale for short-term export performance that incorporates (1) satisfaction with short-term performance improvement, (2) short-term exporting intensity improvement, and (3) expected short-term performance improvement over a one-year period. We use the one-year time frame to assess short-term export performance because it is already established in the literature (e.g., Cooper and Kleinschmidt 1985; Kaynak and Kuan 1993; Lages 2000; Lages and Montgomery 2003).

We define the first dimension, satisfaction with short-term performance improvement, as a compound psychological variable (an affective state) that assesses the effectiveness of a marketing program in terms of its sales, profitability, and market share, as well as its overall performance (see Bonoma and Clark 1988) from one year to the next. This measure is crucial in the assessment of export performance, because it covers managers’ beliefs about how well they are meeting their exporting goals, and it addresses some objective measures of export performance (e.g., sales volume, profitability) (Wang and Olsen 2002). Satisfaction in the assessment of export performance is also one of the most-studied outcome variables in the marketing literature (for a review, see Geyskens, Steenkamp, and Kumar 1999), and it is well estab-
lished in the export marketing field (e.g., Shoham 1998; Zou, Taylor, and Osland 1998).

We used the second dimension, short-term exporting intensity improvement, to assess managers' perceived change in the importance of the exporting activity (in terms of sales volume and profitability) to a firm's overall performance from one year to the next. Historically, export performance has been measured by export sales intensity, because this variable is associated with most aspects of export behavior (Bilkey and Tesar 1977; Cavusgil and Nevin 1981; Zhao and Zou 2002). Katsikeas, Leonidou, and Morgan's (2000) meta-analysis of export-performance studies indicates that export sales intensity is the most common measure in the literature (61% of the reviewed studies in export marketing use this measure). In the current study, as in previous empirical studies (e.g., Beamish, Craig, and McLellan 1993; Bilkey 1982; Kotabe and Czinkota 1992), we used profit intensity and export sales intensity. Some researchers might argue that the use of export intensity has some shortcomings, such as that a higher export-performance intensity does not necessarily imply higher export performance. Nevertheless, this export-performance measure is already established in the export-performance literature (Leonidou, Katsikeas, and Samiee 2002). Furthermore, previous research (e.g., Lages and Jap 2003; Shoham 1998) has already demonstrated the existence of a strong association between export intensity and other export-performance measures.

The third dimension of the STEP scale is expected short-term performance improvement. As Katsikeas, Leonidou, and Morgan's (2000) review of the literature reveals, only two studies have used measures of anticipated future export performance. They suggest that this is somewhat surprising because even with the use of cross-sectional studies, this approach helps gain insight into the export-performance phenomenon while taking into consideration the time-horizon dimension.

In terms of expected short-term performance improvement, managers can report on their expectations of improvement from one year to the next while taking into consideration their own perception of their firm's reference groups. As with export-performance intensity, some researchers might argue that the use of expected export-performance improvement does not necessarily imply higher export performance in the current year. However, previous research reveals that current strategy is determined as a function of its expected effects on short-term performance and that performance levels tend to reinforce one another from period to period (Lages and Jap 2003). This works in two ways. First, when the firm performs well, internal publics (e.g., employees, union repre-
sentatives) and external publics (e.g., clients/customers, suppliers, investors, credit institutions) are more likely to react favorably to the firm, thus facilitating continued performance improvement (Isen and Baron 1991). Second, poor performance may negatively influence performance in the next period, because the reputation of both the firm and top management are harmed by poor performance (Sutton and Callahan 1987). The perception of failure on the part of the different entities that interact with the company, enhanced by the firm’s internal instability, will lead the organization into vicious cycles of “unsuccess” (Masuch 1985), which explains why previous empirical research has shown a strong correlation between expected performance improvement and other export-performance measures (e.g., Wang and Olsen 2002).

Our research setting is Portugal and the United Kingdom, two developed countries that are members of the European Union (EU). Research in this arena is particularly pertinent, because the EU is the world’s largest exporter of goods, maintaining a stable share of approximately one-fifth of total world exports (intra-EU trade excluded) since 1990 (European Commission 2000).

In line with previous research in international marketing (e.g., Styles 1998; Zou, Taylor, and Osland 1998), we used two countries to test our scales. Although we selected both countries for convenience, “they are similar enough for the same dimensions of performance to be relevant, but different enough for the possibility of substantial variation” (Styles 1998, p. 19). A key similarity is that national firms are dominated by short-term export performance in both countries (for Portuguese firms, see Lages and Jap 2003; for British firms, see Styles 1998). Another similarity is that both Portuguese and British economic growth depends heavily on the exporting success of national firms. Moreover, the majority of the countries’ trade is with other EU countries. The main differences across both countries are the language, culture, and values.

As with Cavusgil and Zou (1994), our focus is on a single export venture, because this approach of a single product or product line exported to a single foreign market will enable future researchers who use the measures to associate export performance more precisely with its antecedents and outcomes.

For both countries, we developed a questionnaire that incorporates a variety of multi-item measures and indicators of the conceptual framework. We also included additional indicators derived from exploratory interviews in the research context.
The questionnaire was initially developed in English and then translated into Portuguese. Four judges (university lecturers in marketing) assessed the content and face validity of the items; each judge was asked to assess how representative each item was of the final construct. The survey was revised according to their comments. It was then given to a pretest sample of 15 managers involved in export operations. We used the pretest results to refine the questionnaire further. To avoid translation errors, a different researcher translated the questionnaire into English. A full listing of the final items and their scale reliabilities for Portugal and the United Kingdom is presented in Table 1. The average internal reliability (Cronbach’s alpha) for the Portuguese and British samples were .93 and .87, respectively.

**Portugal.** We randomly generated a sample of 2500 firms from the database of Instituto do Comércio Externo de Portugal (1997), a government agency. The database of 4765 Portuguese exporters was the most comprehensive and up-to-date database available in the Portuguese market at the time of data collection (i.e., 1999). The pretest results indicated a strong need for an incentive to motivate the respondents to participate. In the cover letter, we stated that in return for a completed survey, we would provide respondents with a list of contacts for potential overseas importers or clients and a report of the final results. In addition, we assured them of confidentiality.

In the first mailing, a cover letter, a questionnaire, and an international postage-paid business reply envelope were sent to the person responsible for exporting in each of the 2500 Portuguese firms. This missive was followed by a second mailing that included a reminder letter and a reply envelope. Of the sample of 2500 Portuguese managers, 29 stated that they no longer exported, and the postal service returned 119 questionnaires from firms that had either closed or moved without leaving a forwarding address. Thus, the sample size was reduced to 2352. Of these, 519 valid questionnaires were returned, for a 22% response rate.

**United Kingdom.** We randomly generated a sample of 1564 enterprises from the British Chamber of Commerce’s British Exports 2000 database (Reed Business Information 2000). In the cover letter, we stated that in return for a completed questionnaire, the findings would be available after the completion of the study. We also assured respondents of confidentiality.

As with the Portuguese survey, a cover letter, a questionnaire, and a postage-paid business reply envelope were sent to the person responsible for exporting at each firm in the sample. Unfortunately, in contrast to the Portuguese survey,
### Table 1. Confirmatory Factor Analysis Results for Portuguese and British Samples

<table>
<thead>
<tr>
<th>Dimensions and Items of Short-Term Export Performance</th>
<th>Portugal(^a)</th>
<th>United Kingdom(^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>SAT: Satisfaction with Short-Term Performance Improvement</strong></td>
<td>α/ρ(_{\text{vco}})/ρ</td>
<td>Item Loading</td>
</tr>
<tr>
<td>How satisfied are you with the results of your export venture from Year 1 to Year 2? (1 = &quot;much less satisfied in Year 2 than in Year 1,&quot; 5 = &quot;much more satisfied in Year 2 than in Year 1&quot;)</td>
<td>.95/.83/.95</td>
<td></td>
</tr>
<tr>
<td>SAT1: Export sales volume</td>
<td>.93</td>
<td>27.85</td>
</tr>
<tr>
<td>SAT2: Export profitability</td>
<td>.89</td>
<td>25.88</td>
</tr>
<tr>
<td>SAT3: Market share in the importing market</td>
<td>.87</td>
<td>24.94</td>
</tr>
<tr>
<td>SAT4: Overall export performance</td>
<td>.95</td>
<td>28.96</td>
</tr>
<tr>
<td><strong>INT: Short-Term Exporting Intensity Improvement</strong></td>
<td>.90/.82/.90</td>
<td></td>
</tr>
<tr>
<td>With regard to your export venture, to what extent did the following change from Year 1 to Year 2? (1 = &quot;large decrease from Year 1 to Year 2,&quot; 5 = &quot;large increase from Year 1 to Year 2&quot;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT1: Percentage of exporting venture to total sales volume</td>
<td>.92</td>
<td>25.56</td>
</tr>
<tr>
<td>INT2: Percentage of exporting venture to total profitability</td>
<td>.89</td>
<td>24.46</td>
</tr>
<tr>
<td><strong>EXP: Expected Short-Term Performance Improvement</strong></td>
<td>.94/.80/.94</td>
<td></td>
</tr>
<tr>
<td>What do you anticipate the results of your export venture will be this year compared to the past year? (1 = &quot;worsen significantly,&quot; 5 = &quot;improve significantly&quot;)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXP1: Export sales volume of the export venture</td>
<td>.90</td>
<td>26.15</td>
</tr>
<tr>
<td>EXP2: Export profitability of the export venture</td>
<td>.85</td>
<td>24.00</td>
</tr>
<tr>
<td>EXP3: Achievement of the objectives of the export venture</td>
<td>.92</td>
<td>26.90</td>
</tr>
<tr>
<td>EXP4: Satisfaction with the exporting venture</td>
<td>.91</td>
<td>26.78</td>
</tr>
</tbody>
</table>

\(^a\)Portugal (n = 519): χ² = 129.73, 32 d.f., p < .01; normed fit index = .98; nonnormed fit index = .98; comparative fit index = .98; incremental fit index = .98.

\(^b\)United Kingdom (n = 111): χ² = 70.42, 32 d.f., p < .01; normed fit index = .91; nonnormed fit index = .93; comparative fit index = .95; incremental fit index = .95.

it was not possible to obtain governmental funding to conduct the research. Consequently, because of a lack of financial resources, it was not possible to send a reminder mailing.

We conducted the data collection in 2002. Of the 1564 exporters, we received 111 replies, for a raw response rate of 7% (111/1564). To identify the problems associated with this low response rate, we used Menon and colleagues’ (1999) method of contacting 100 randomly chosen respondents to determine nondeliverable and noncompliance rates. We then assessed final response rates. We determined that 34% of the mailings were nondeliverable because of incorrect address, an additional 40% did not reach the person responsible for the export operations in the firm, and 4% of the respondents reported a corporate policy of not responding to academic surveys. In line with Menon and colleagues’ method, the total of 111 usable returned questionnaires represents a 32% effective response rate, which is quite satisfactory, given that average top-management survey response rates are in the range of 15%–20% (Menon, Bharadwaj, and Howell 1996).

We tested nonresponse bias by assessing the differences between the early and late respondents with respect to the means of all the variables for both samples (Armstrong and Overton 1977). We defined early respondents as the first 75% of the returned questionnaires and late respondents as the last 25%. These proportions approximate the actual way respondents returned questionnaires. We found no significant differences between the early and late respondents, suggesting that response bias was not a significant problem in the study.

The entire size range of firms is represented in the sample. Both Portuguese and British exporting industries primarily consist of small and medium-sized enterprises. Of the exporting firms represented in the sample, 5% of Portuguese firms and 6% of British firms had more than 500 employees. For the Portuguese sample, the average annual export sales of the firms ranged from US$350,000–$1.5 million. For the British sample, the average annual export sales of the firms ranged from US$470,000–$1.6 million.

We directed both surveys to employees who were primarily responsible for exporting operations and activities; the job titles of respondents included president, marketing director, managing director, and exporting director. We asked respondents in both countries to indicate their degree of experience in exporting on a scale where 1 = “none” and 5 = “substantial.” The mean response for Portugal was 3.6 (standard deviation [s.d.] = .84, range 1 to 5) and 3.8 (s.d. = .93, range 1 to 5) for the United Kingdom. Collectively, this indicates that...
though the title of the respondents’ positions may be wide ranging, the respondents appear to have considerable knowledge in the specific exporting activities of the firm and are experienced with exporting in general.

We used Churchill’s (1979) traditional approach to scale development. As Churchill notes, to increase reliability and decrease measurement error, it is advisable to use multi-item scales rather than single-item scales. Gerbing and Anderson (1988) have expanded Churchill’s approach to scale development with the use of confirmatory factor analysis (CFA). To assess the measurement properties of the existing scales, we used CFA with full-information maximum likelihood estimation procedures in LISREL 8.3 (Jöreskog and Sörbom 1993); CFA provides a better estimate of reliability than coefficient alpha (Steenkamp and Van Trijp 1991). Although coefficient alpha assumes that different indicators have equal factor loadings ($\lambda$) and error variances ($\delta$), CFA takes into account the differences among the existing indicators (Styles 1998).

In this article, we separate the analysis of Portuguese and British data because data collection occurred in two different years, and it is possible that managerial perceptions of performance are different in the two countries. As is shown in Table 1, convergent validity is evidenced by the large, significant standardized loadings of each item on its intended construct (average loading size was .90 for the Portuguese sample and .84 for the British sample). As is also shown in Table 1, all constructs present the desirable levels of composite reliability (Bagozzi 1980) and discriminant validity (Fornell and Larcker 1981).

Discriminant validity is also evidenced by the correlation estimates between any two constructs (Jöreskog and Sörbom 1993). No correlation includes a value of one (Anderson and Gerbing 1988), and the highest correlation is for satisfaction and intensity improvement (Portugal: .70; United Kingdom: .67).

To assess cross-cultural validation, we tested (1) factorial similarity, (2) factorial equivalence, and (3) measurement equivalence (see Mullen 1995; Singh 1995), which is a procedure frequently used in the international business literature (e.g., Brady and Robertson 2001; Cadogan, Diamantopoulos, and Mortanges 1999; Styles 1998).

**Factorial Similarity.** The first step was for us to test factorial similarity across the Portuguese and British samples using LISREL 8.3 (Jöreskog and Sörbom 1993). We developed two measurement models for each sample (see Table 1). Each item is restricted to load on its prespecified factor, and we allowed the error variances and factor loadings to correlate.
freely. The chi-square for both models is significant (Portu-
gal: $\chi^2 = 129.73, 32$ degrees of freedom [d.f.], $p < .01$; United
Kingdom: $\chi^2 = 70.42, 32$ d.f., $p < .01$). Because the chi-square
statistic is sensitive to sample size, we also assessed addi-
tional fit indexes: the normed fit index (NFI), the nonnormed
fit index (NNFI), the comparative fit index (CFI), and the
incremental fit index (IFI). All the fit indexes for the Por-
tuguese sample present satisfactory values (NFI = .98,
NNFI = .98, CFI = .98, and IFI = .98). Similarly, for the British
sample, the NFI, NNFI, CFI, and IFI are .91, .93, .95, and .95,
respectively, which suggests that the model fits the data well.

**Factorial Equivalence and Measurement Equivalence.** After
testing for factorial similarity, we assessed factorial equiva-
ence. The baseline model we estimated involved a three-
construct multi-indicator specification, whereby we allowed
the factor loadings, error variances, and factor correlations to
differ across Portuguese and British samples (see Table 2,
Model A). Then, we estimated a more constrained model (see
Table 2, Model B), whereby we allowed the error variances to
differ and we set the factor loadings and factor correlations to
be invariant across both samples.

We compared Model B with the baseline model (see Table 3).
The chi-square difference test between Model A and Model B
is nonsignificant ($\Delta \chi^2 = 12.82; \Delta$ d.f. = 13; $p > .10$). This
reveals that the STEP scale presents factorial equivalence,
because the factor loadings across the two samples are the
same within the statistical bounds we set.

The final step was to test full metric equivalence; the chi-
square difference test between Model A and the most con-
strained model (Model C, whereby factor loadings, error vari-
ances, and factor correlations were invariant across both
samples) is significant ($\Delta \chi^2 = 221.89; \Delta$ d.f. = 23; $p < .001$).
Thus, the STEP scale does not present full metric equivalence.

Our validation across the Portuguese and British samples
indicates that the STEP scale reveals factorial similarity and
factorial equivalence, but it does not present full metric
equivalence. We now present some possible explanations. A
possibility is that we collected data in two different years
and used different types of incentives in the two countries,
which might have influenced respondents’ willingness to
answer the questionnaire. Another possibility is the existing
scalar inequivalence due to language differences between
Portuguese and British managers. Nevertheless, we largely
minimized this situation with the pretest (see Douglas and
Craig 1983) and the straightforward questions (Styles 1998).

Another possible explanation is the interpretation of the con-
textual variables (Craig and Douglas 2000; Douglas and Craig

---

**CONCLUSION**

**Discussion**
## Table 2: Confirmatory Factor Analysis Results for Two-Group Model

<table>
<thead>
<tr>
<th>Dimensions of Short-Term Export Performance</th>
<th>Model A&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Model B&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Model C&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Portugal</td>
<td>United Kingdom</td>
<td>Portugal</td>
</tr>
<tr>
<td></td>
<td>Standardized Item Loading</td>
<td>Standardized Item Loading</td>
<td>Standardized Item Loading</td>
</tr>
<tr>
<td></td>
<td>Error Variance</td>
<td>Error Variance</td>
<td>Error Variance</td>
</tr>
<tr>
<td>SAT</td>
<td>.93</td>
<td>.13</td>
<td>.84</td>
</tr>
<tr>
<td>SAT1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SAT2</td>
<td>.89</td>
<td>.20</td>
<td>.70</td>
</tr>
<tr>
<td>SAT3</td>
<td>.87</td>
<td>.24</td>
<td>.82</td>
</tr>
<tr>
<td>SAT4</td>
<td>.95</td>
<td>.10</td>
<td>.88</td>
</tr>
<tr>
<td>INT</td>
<td>.90</td>
<td>.16</td>
<td>.95</td>
</tr>
<tr>
<td>INT1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INT2</td>
<td>.89</td>
<td>.21</td>
<td>.74</td>
</tr>
<tr>
<td>EXP</td>
<td>.92</td>
<td>.20</td>
<td>.75</td>
</tr>
<tr>
<td>EXP1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EXP2</td>
<td>.88</td>
<td>.28</td>
<td>.74</td>
</tr>
<tr>
<td>EXP3</td>
<td>.96</td>
<td>.18</td>
<td>.67</td>
</tr>
<tr>
<td>EXP4</td>
<td>.95</td>
<td>.18</td>
<td>.69</td>
</tr>
</tbody>
</table>

**Fit Indexes**
- NFI = .97; NNFI = .97; CFI = .98; IFI = .98
- NFI = .96; NNFI = .97; CFI = .98; IFI = .98
- NFI = .94; NNFI = .95; CFI = .95; IFI = .95

<sup>a</sup>Factor loadings, error variances, and factor correlations are variant.
<sup>b</sup>Factor loadings are invariant, error variances are variant, and factor correlations are invariant.
<sup>c</sup>Factor loadings, error variances, and factor correlations are invariant.
As Craig and Douglas (2000, p. 271) note, “scales that attempt to capture firm or organizational characteristics use managers or employees as respondents.... The results of the analysis are [then] used to make inferences about firm behavior.” Thus, it is possible that as a consequence of the different contexts, when Portuguese and British respondents evaluated the performance of the selected export venture, they chose different types of products and exporting markets. For example, although Portuguese exporters might rely more on traditional (e.g., textiles, shoes) products and less on high-technology products, this reliance is much less likely to occur for British exporters. Similarly, although British firms are typically international firms, Portuguese exporting firms typically depend on a neighboring country (e.g., Spain). Styles (1998) also suggests that firms’ different levels of familiarity and experience with export operations are another problem. However, this does not seem to be a concern in this study. As we previously discussed, when managers rated their degree of experience in exporting, the mean responses for Portugal (mean = 3.6; s.d. = .84, range 1 to 5) and the United Kingdom (mean = 3.8; s.d. = .93, range 1 to 5) were similar.

The first limitation of this research is that, as with other studies in international business, our findings may be biased as a result of self-report and perceptual data (Skarmeas, Katsikeas, and Schlegelmilch 2002), particularly if we consider that aspirations and goals may be conflicting inside the firm and that we collected data in two different years and used different types of incentives in the two countries.

A second limitation is that our research instrument (i.e., the questionnaire) may have created common method variance. This could be particularly threatening if the respondents were aware of the conceptual framework of interest. However, they were not told the specific purpose of the study, and we separated and mixed all the construct items. Further-
more, we guaranteed confidentiality to all survey participants, which also helps reduce the possibility of bias in performance reports for self-presentation reasons (Singh 2000). In addition, if common method bias exists, a CFA containing all constructs should produce a single method factor (Podsakoff and Organ 1986). The goodness-of-fit indexes (Portugal: NFI = .57, NNFI = .46, CFI = .58, IFI = .58; United Kingdom: NFI = .53, NNFI = .41, CFI = .54, IFI = .55) indicate a poor fit for both models, which suggests that bias from common method variance is unlikely.

The final limitation is related to the small size of the British sample and the inexistence of full metric equivalence. Consequently, the results should be regarded as suggestive rather than conclusive.

Managerial and Public Policy Implications

From a practitioner’s perspective, several reasons justify the need for a sound evaluation of short-term export performance. The STEP scale might help managers monitor performance in the short run, thereby enabling them to take short-term performance as a reference point when defining future actions and when allocating resources to specific export ventures. Recent empirical research by Lages and Montgomery (2002, in press) and Lages and Jap (2003) indicates that poor/successful performance in export operations has an immediate impact on strategic decisions. When performance decreases in any given year, both internal (e.g., top management, employees, union representatives) and external (e.g., shareholders/investors, suppliers, credit institutions) publics consider it a potential threat to the entire organization, demanding in turn immediate improvements in short-term export performance. Moreover, many firms depend on short-term performance for survival. This is particularly true for firms that lack financial resources and for those that operate in markets with low margins (as a result of a high level of competition or market saturation).

From the manager’s point of view, there is also a common practice of focusing on annual performance results because it is much easier to establish and quantify results annually than it is in the long run. Furthermore, managers consider short-term performance vital because it relates to their own personal interests. In recent years, there has been increasing mobility of managers across firms, and top managers spend fewer years in the same organization. Moreover, performance improvement at the end of the year might have an immediate effect in terms of personal income (e.g., salary bonus), which may lead managers to place more importance on short-term performance.

It might be argued that in particular situations, public policymakers have more short-term interest than do managers, because the latter are responsible for the firm’s long-term
development. However, at the public policy level, the evaluation of short-term performance is equally important. All the benefits provided by the exporting activity encourage public policymakers to implement export-promotion programs with the objective of helping firms improve their competitive advantage and ultimately enhancing their performance in the international arena (Lages and Montgomery 2003).

Particularly in times of recession, some countries (e.g., Portugal) explore the export activity for short-term solutions, such as decreasing the nation’s budget deficit in a short period (see, e.g., Financial Times 2002). A decrease in national firms’ export performance might put pressure on public policymakers to support the export activity and to demand a better allocation of the assistance received from managers. As with managers, public policy actions are constantly evaluated by several publics and, consequently, have (dis)incentives (e.g., reelection). Naturally, if public policymakers want to remain active, they need to be concerned with national firms’ short-term export performance, because this will have a direct impact on a country’s economic health. Moreover, given that long-term success in export allocation is also a result of short-term actions, public policymakers will favorably view a new scale with which to assess short-term export performance. A proper allocation of export assistance will enable public policymakers to save resources that either can be used to generate reserves or can be allocated to other activities.

Major advances in export marketing will only be made possible by means of a more integrated approach to the conceptualization and measurement of export performance (Cavusgil and Zou 1994; Shoham 1998). In this article, we develop the STEP scale, which assesses short-term export performance on three dimensions: (1) satisfaction with short-term performance improvement, (2) short-term exporting intensity improvement, and (3) expected short-term performance improvement. We expect that this scale will enhance the quality of empirical research on export performance and make it more comparable, while also helping researchers capture performance in the context of actual business time (the business year).

It is also our aim to help managers and public policymakers assess short-term export performance. As Katsikeas, Leonidou, and Morgan (2000) indicate, although some practitioners are focused on monitoring short-term export performance, others judge the long-term existence of organizations. It is our aim to provide a scale that both managers and public policymakers can use to support the development of tactical and strategic decisions. Regardless of firms’ level of interest in short-term performance, it is important that the STEP scale not be misused to achieve short-term goals, without consideration of organizations’ long-term orientations.

Directions for Further Research
Instead of treating the STEP scale as a unidimensional construct, we present various measurement units for each of the three dimensions. To refine the STEP scale presented herein, we suggest that researchers replicate this study across various industries and particularly across different countries.

In the assessment of export performance, researchers should be concerned with exploring the short time horizon that managers and public policymakers frequently use to assess performance (Madsen 1998). Given the long-term failures and successes of a firm as a function of its short-term actions, research that attempts to understand short-term performance and its antecedents and effects can yield valuable insights into long-term performance improvement. Organizations and people constantly set goals and adjust their behavior in response to favorable and unfavorable feedback (Cyert and March 1963; March and Simon 1958). The scale we present helps assess this feedback. International marketing researchers should simultaneously consider how previous strategy affects short-term performance and how short-term performance affects strategy, because strategic decisions are motivated by a combination of short-term proactive and reactive behaviors (March and Sutton 1997). By better understanding the one-year relationship between strategy and performance, researchers might help managers avoid becoming caught in a vicious cycle of successive unsatisfactory results.

Finally, we encourage future researchers to develop a performance scale for the other side of the dyad: the importer (see Skarmeas and Katsikeas 2001). This approach might be particularly difficult if the other side is an individual consumer rather than an organizational customer (Lages and Jap 2003).

In summary, international business scholars have been challenged to state their managerial and public policy implications more clearly (Czinkota 1994; Madsen 1998). In trying to meet this call, our research attempts to capture practical knowledge in a theoretical framework. We believe that the combined perspectives of theory and practice focus our research agenda. Moreover, we suggest that in conceptual argument and methodological development, the use of a short-term time frame (e.g., one year) might assess export performance in a novel way. More specifically, this research creates the STEP scale that measures improvement in short-term export performance, which may align real-world time constraints with methodological soundness.

1. We generated the contact list using online information mainly from information available on the Web sites of several chambers of commerce, which list importers by sector.


