The Task and Ego Orientation in Sport Questionnaire: Testing for Measurement Invariance and Latent Mean Differences in Spanish and Portuguese Adolescents

Isabel Castillo, Inés Tomás, and Isabel Balaguer
University of Valencia, Valencia, Spain

António M. Fonseca and Cláudia Dias
Porto University, Porto, Portugal

Joan L. Duda
University of Birmingham, Birmingham, United Kingdom

Within the theoretical framework of achievement goals (Nicholls, 1989), Duda and Nicholls (see Duda, 1989; Duda & Whitehead, 1998) developed the Task and Ego Orientation in Sport Questionnaire (TEOSQ) to assess individual differences in achievement goal orientations. This study searches for validity evidence of the TEOSQ in the case of Spanish (2473) and Portuguese (2486) junior high school. The results showed that factor structure, factor loadings, and intercepts could be considered invariant across groups. Between group differences in latent means showed slightly higher mean scores for the Spanish group for both factors. The results also provide high support for the internal consistency of both the Spanish and Portuguese versions of this scale.

Keywords: achievement goal theory, motivational orientations, measurement invariance, multigroup confirmatory factor analysis, adolescents, TEOSQ

Within the theoretical framework of achievement goals (Nicholls, 1989), Duda and Nicholls (see Duda, 1989; Duda & Whitehead, 1998) developed the Task and Ego Orientation in Sport Questionnaire (TEOSQ) to assess individual differences
in proneness for emphasizing task and ego involving criteria for defining success in athletic settings. The operational definitions of task and ego orientation stem specifically from the work of Nicholls (1989). According to Nicholls (1989), there are individual differences in the means by which people strive to demonstrate competence; namely, task or ego goal orientations. Task orientation involves the purposes of gaining skill or knowledge and performing one’s best. Thus, perceptions of competence tend to be self-referenced if the individual is task oriented. On the other hand, when an ego orientation prevails, people tend to be preoccupied with their ability and see the personal demonstration of superior competence as fundamental to success and demonstrated ability. Ego-oriented individuals tend to judge their own competence by comparison to others.

Psychometric tests on the scores gathered by TEOSQ across various populations have shown the instrument is supported by evidence of reliability and validity and characterized by a two-dimensional factor structure that replicated previous work with the English version of this instrument (see Duda & Whitehead, 1998). As far as we know, this pattern has held across samples from the United States (e.g., Duda, 1989; Duda, Chi, Newton, Walling, & Catley, 1995), Korea (e.g., Kim & Gill, 1995), Greece (e.g., Papaioannou & Digglelidis, 1996), China (Cindi & Koenraad, 2005), Japan (Hayashi & Weiss, 1994), Spain (Balaguer, Castillo, & Tomás, 1996; Guivernau & Duda, 1994), the United Kingdom (e.g., Spray & Biddle, 1997), Portugal (Fonseca & de Paula Brito, 2005), and Croatia (Bariae & Horga, 2006).

Previous research with the Spanish version of this scale using exploratory factor analysis (Balaguer et al., 1996) and confirmatory factor analysis (Castillo, Balaguer, & Duda, 2002) provided evidence of adequate psychometric properties in the case of Spanish adolescents. The exploratory factor analysis showed a two-factor solution that accounted for an acceptable percentage of variance (49.2%). The confirmatory factor analysis supported the predominant two-dimensional structure and showed a good fit for the model. Previous research with the Portuguese version of this scale using exploratory and confirmatory factor analyses has also provided evidence of adequate psychometric properties in an adolescent sample (Fonseca & de Paula Brito, 2005). The exploratory factor analysis showed a two-factor solution that accounted for an acceptable percentage of variance (49.8%). The confirmatory factor analysis also supported the hypothesized two-dimensional structure (Fonseca & de Paula Brito, 2005).

Although the Spanish and Portuguese studies of the factorial validity of the TEOSQ have provided evidence for the use of this instrument with adolescents, this does not necessarily indicate the absence of cultural sensitivity. In general, most cross-cultural research on goal orientations in sport has involved an examination of group means and has not considered the factorial equivalence of the TEOSQ’s hypothesized structure across cultures (e.g., Hayashi & Weiss, 1994; Williams, Kim, & Gill, 1995). Several researchers have addressed the issue of equivalence
in measurement in cross-cultural research (see e.g., Gangyan & Hing-chu, 2007; Duda & Hayashi, 1998). Measures with sound psychometric properties within and across countries will facilitate cross-cultural comparative research and should help in understanding diverse cultural variations in the motivation underlying sport participation around the world. Although researchers have called for an examination of the applicability of sport psychology theories in other countries and cultures (Gangyan & Hing-chu, 2007; Duda & Hayashi, 1998), to date few studies have been done.

In sum, the major purpose of this research is to determine the extent to which the TEOSQ is equivalent (i.e., invariant) across cultures, specifically across Spanish and Portuguese adolescents. The second purpose is to test the differences in the factor latent means across the two countries. The establishment of measurement invariance would lead us to a major contribution that would allow us to make appropriate comparisons between the two different cultural groups studied. There are some similarities (e.g., in values, educational systems, etc.) between Spain and Portugal due to the fact that both are countries marked by a collectivist culture (Hofstede, 1980). Thus, it was hypothesized that the two dimensions of the TEOSQ will be equivalent across both countries and, furthermore, there will not be differences in the factor latent means across countries.

**METHOD**

**Participants and Procedure**

*The Spanish sample.* Responses to the TEOSQ were obtained from 2473 adolescents (49.1% males, 50.9% females). Ages ranged from 13 to 18 years ($M = 15.4$ years; $SD = 1.41$). The age distribution was as follows: 13 years old (17%), 14 years old (2.8%), 15 years old (28.6%), 16 years old (28.3%), 17 years old (18%), and 18 years old (5.3%). The students in the Spanish sample were from different types of coeducational establishment (public, semiprivate and private high schools), all of which were located in the same eastern region of Spain. Students reflected a wide range of socioeconomic levels but similar ethnic backgrounds. The Spanish version of the TEOSQ was completed by this sample.

*The Portuguese sample.* The Portuguese sample was specifically selected so that it matched the Spanish sample as closely as possible in terms of demographic variables such as gender, age, and grade level. Responses to the TEOSQ were obtained from 2486 high school students (45.9% males, 54.1% females). Ages ranged from 13 to 18 ($M = 15.4$ years; $SD = 1.43$). The age distribution was similar to the one obtained in the Spanish sample: 13 years old (17%), 14 years old (2.8%), 15 years old (28.4%), 16 years old (27.9%), 17 years old (18.1%), and 18 years old (5.8%). The students in the Portuguese sample were from different types of coeducational establishment (public, semiprivate and private high schools),
all of which were located in the same northwestern region of Portugal. Students reflected a wide range of socioeconomic levels but similar ethnic backgrounds. The Portuguese version of the TEOSQ was completed by this sample.

In both samples of adolescents, questionnaires were administered by trained research assistants to classroom groups during school hours.

Instrumentation

The TEOSQ is a 13-item questionnaire designed to assess task (7 items) and ego (6 items) orientations as proposed by Nicholls’ (1989) model of achievement motivation (Duda, 1989, 2001). When completing the instrument, the participants were requested to think of when they felt most successful in a particular sport and then indicate their agreement with items reflecting task-oriented (e.g., “I feel successful in sport when I work really hard”) or ego-oriented (e.g., “I feel successful in sport when the others can’t do as well as me”). Responses are indicated on a 5-point Likert-type scale with 1 = strongly disagree and 5 = strongly agree.

Translations of the TEOSQ into Spanish and into Portuguese were conducted. The source English version of the questionnaire was translated to Spanish (Baloguer et al., 1996) and to Portuguese (Fonseca & de Paula Brito, 2005) following the back-translation procedure widely described in the literature (e.g., Hambleton & Kanjee, 1995).

Statistical Analyses

A series of multi-sample Confirmatory Factor Analysis (CFAs) with mean and covariance structure (MACS) were conducted with LISREL 8.80 (Jöreskog & Sörbom, 2006). Considering the ordinal nature of the TEOSQ items, weighted least square (WLS) was used as the estimation procedure, and the polychoric covariances matrix, the estimated asymptotic covariance matrix of the polychoric variances and covariances, and the means vector of the observed variables were used as input for the analysis (for further discussion of SEM, see Bollen, 1989; Byrne, 1998). As far as distributional aspects are concerned, the tests of multivariate normality offered by PRELIS (which is part of LISREL) as well as the univariate distributions of the items indicated that both the Portuguese and the Spanish scores were non-normally distributed. As the analyzed data were ordinal and non-normally distributed, the Satorra and Bentler (1994) scaled corrected chi-square statistic was reported. The $SB\chi^2$ was designed to be robust against non-normal distributed variables. Nevertheless, other additional fit measures were also used as it is explained later on in the Goodness-of-fit criteria section.

Analyses were based on 4952 respondents (2466 Spanish adolescents and 2486 Portuguese adolescents) who had reasonably complete data for the TEOSQ in
that they had missing values for no more than two of the 13 TEOSQ items. The inevitable missing data are a potentially important problem, particularly when the amount of missing data exceeds 5% (e.g., Graham & Hofer, 2000). In our study, however, the percentage of missing data was small (.6% in the Spanish sample and .7% in the Portuguese sample) and assumed to be completely at random so that it was unlikely to be a serious problem. Nevertheless, for LISREL analysis purposes, within-group covariance matrices were constructed using the Expectation Maximization Algorithm, a widely recommended approach for imputation of missing data, as operationalized using missing value analysis in SPSS.

Test for item parameter invariance. Testing for TEOSQ equivalence encompassed a series of hierarchically ordered steps addressed to test the invariance of the factor structure and the invariance of item parameters of the TEOSQ across the groups. First, the a priori factor structure was fitted separately for each group to determine the extent to which the baseline model fitted the data for each group individually. Thus, Model 0a (M0a) tested the baseline model for the Portuguese group, and Model 0b (M0b) tested the baseline model for the Spanish group. Then, a set of three nested models (Models 1 to 3) using multi-sample CFAs tested for increasingly stringent levels of constrained equivalence across the groups. Model 1 (M1: Structural Invariance Model) addressed the equality of number and pattern of factor loadings across groups. Thus, M1 tested whether a 2-common-factor model held in the two samples, but no invariance constraints were imposed. This model was used as a baseline for fit comparisons against the later, more restricted models. Model 2 (M2) addressed the invariance of factor loadings across groups. And in Model 3 (M3), factor loadings and intercepts were constrained to be invariant across groups.

Test for latent mean differences. First, Model 4 (M4: Factor Loading, Intercept and Latent Mean Invariance Model) addressed the equality of latent means across groups. Then, between-group latent mean differences across groups were estimated in Model 5 (M5). The test of latent mean differences requires that the hypothesized CFA model shows a good fit in both compared groups and that all factor loadings and intercepts be equivalent across groups. In this study, the Portuguese group was arbitrarily selected as the reference group; its latent means were fixed to 0. Latent means for the Spanish sample were freely estimated. Statistical significance associated with differences between the latent means for the groups was determined on the basis of the z-statistic. In all the models, for identification purposes and to establish the scale of measurement, one measured variable for each of the 2 TEOSQ factors was selected to be a reference indicator, and its factor loading was fixed to be 1.

Goodness-of-fit criteria. Model fit was assessed using different criteria. As previously stated, the Satorra and Bentler scaled corrected chi-square statistic was reported. This statistic, which is the same as the conventionally reported $\chi^2$ statistic, assesses whether the hypothesized model holds exactly in the population.
This is an unrealistic assessment that even makes models with a reasonable fit be rejected if the sample size is large enough. Therefore, the use of other fit measures is strongly recommended. Following Marsh, Hau, and Grayson (2005; also see Marsh, Hau, & Wen, 2004), we considered the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMSR), the non-normed fit index (NNFI), and the comparative fit index (CFI) to evaluate goodness of fit as well as an evaluation of parameter estimates. In order to test invariance hypotheses using CFA, a modeling rationale was used (Little, 1997, 2000; Marsh et al., 2005). With a modeling rationale, practical fit indices are used to assess the difference in goodness of fit between nested models. From this viewpoint, invariance hypotheses are evaluated comparing the practical fit indices of two nested models. Even though there are no generally accepted standards, some criteria have been proposed in the literature to interpret differences in practical fit indices. Thus, for example, Widaman (1985) considered differences not larger than .01 between NNFI values ($\Delta$NNFI) as an indication of negligible practical differences. Regarding differences between CFI values ($\Delta$CFI), which are based on the results of a simulation work, Cheung and Rensvold (2002) suggested that decreases in fit greater than .01 might be important. From a modeling rationale, in the present study we will consider that a tested invariance hypothesis is tenable if: (a) the model that posits the hypothesis has an acceptable practical fit, and (b) the difference in the NNFI and CFI values provided by this model and a model without constraints can be considered as negligible (equal to or less than .01).

RESULTS

Preliminary Analyses

The descriptive statistics and coefficient alpha are presented in Table 1. The estimates of internal consistency for the scores gathered with the two TEOSQ scales were satisfactory in the Spanish and the Portuguese samples.

| Scale | Portuguese Group | | Spanish Group |
|-------|------------------|------------------|
|       | N = 2486         | N = 2466         |
|       | Mean  | SD   | $\alpha$ | Mean  | SD   | $\alpha$ |
| 1. Task | 3.93  | .61  | .78     | 4.14  | .56  | .80     |
| 2. Ego | 2.49  | .91  | .85     | 2.76  | .87  | .85     |

Note: $\alpha$ = Coefficient alpha.
### TABLE 2
Goodness of Fit Indices for Tested Models

<table>
<thead>
<tr>
<th>Model Description</th>
<th>df</th>
<th>$SB_\chi^2$</th>
<th>RMSEA</th>
<th>(90% CI)</th>
<th>SRMSR</th>
<th>NNFI</th>
<th>CFI</th>
<th>$\Delta$NNFI</th>
<th>$\Delta$CFI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>M0a</strong> Baseline Model Portugal</td>
<td>64</td>
<td>718.1**</td>
<td>.07</td>
<td>(.063–.072)</td>
<td>.07</td>
<td>.95</td>
<td>.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M0b</strong> Baseline Model Spain</td>
<td>64</td>
<td>597.4**</td>
<td>.06</td>
<td>(.056–.065)</td>
<td>0.5</td>
<td>.96</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M1</strong> Structural Invariance (Baseline Model)</td>
<td>128</td>
<td>1326.8**</td>
<td>.06</td>
<td>(.061–.067)</td>
<td>.07</td>
<td>.96</td>
<td>.96</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>M2</strong> FL Invariance</td>
<td>139</td>
<td>1427.2**</td>
<td>.06</td>
<td>(.061–.067)</td>
<td>.07</td>
<td>.96</td>
<td>.96</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td><strong>M3</strong> FL + INT Invariance</td>
<td>150</td>
<td>1895.7**</td>
<td>.07</td>
<td>(.068–.074)</td>
<td>.07</td>
<td>.95</td>
<td>.95</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td><strong>M4</strong> FL + INT + LM Invariance</td>
<td>152</td>
<td>2072.1**</td>
<td>.07</td>
<td>(.071–.077)</td>
<td>.07</td>
<td>.94</td>
<td>.94</td>
<td>.02</td>
<td>.02</td>
</tr>
<tr>
<td><strong>M5</strong> FL + INT + estimation of between group differences in LM</td>
<td>150</td>
<td>1895.7**</td>
<td>.07</td>
<td>(.068–.074)</td>
<td>.07</td>
<td>.95</td>
<td>.95</td>
<td>.01</td>
<td>.01</td>
</tr>
</tbody>
</table>

*Note:* df = degrees of freedom; RMSEA = root mean square error of approximation; 90% CI = 90% confidence interval for the RMSEA; SRMSR = standardized root mean square residual; NNFI = non-normed fit index; CFI = comparative fit index; FL = factor loadings; INT = intercepts; LM = latent mean. All the $\Delta$ index comparisons are made with respect to the baseline model (M1). ** $p < .01$.

### Item Parameter Invariance

Previously to carrying out the multi-sample CFAs, the hypothesized factorial structure for the TEOSQ was fitted separately in both groups. As can be seen in Table 2 (Models M0a and M0b), the proposed factorial structure adequately fitted the data for the Portuguese (RMSEA = .07; SRMSR = .07; NNFI = .95; CFI = .96) and for the Spanish (RMSEA = .06; SRMSR = .05; NNFI = .96; CFI = .97) groups considered separately. For these models, all parameter estimates were statistically significant.

The goodness-of-fit indices for the models used to test item parameter invariance are also presented in Table 2. Three multi-sample models (M1 to M3) were tested across Portuguese and Spanish adolescents, each representing an increasingly more restricted parameterization than its predecessor.
With regard to the multi-sample baseline model, in which no equality constraints were imposed, results showed that the fit was acceptable (RMSEA = .06; SRMSR = .07; NNFI = .96; CFI = .96). Thus, it could be concluded that the same factor model was able to fit the data from each group. Consequently, the freely estimated two-factor model (Model 1) was used as the baseline against which all remaining models were compared in the process of determining evidence of invariance.

The practical fit index values obtained for the model that posits factor loading invariance (M2) showed that this model had an acceptable fit (see Table 2, RMSEA = .06; SRMSR = .07; NNFI = .96; CFI = .96). Furthermore, the NNFI and CFI values obtained for M1 and M2 were the same, indicating that there were no differences in practical fit between Model 2 and the baseline model. Thus, it could be concluded that no important differences were found in the factor loadings across the two samples.

Model 3 (M3) hypothesized the invariance of factor loadings and intercepts. The fit of this model was acceptable (RMSEA = .07; SRMSR = .07; NNFI = .95; CFI = .95). And when compared with the baseline model (M1), the ΔNNFI and ΔCFI values obtained did not exceed the criterion value .01. Therefore, it could be concluded that no important differences in the intercepts across the two samples were found.

Latent Mean Invariance and Between-group Differences

As the lack of invariance of factor loadings and intercepts was not supported across the Portuguese and Spanish groups, we carried out two additional multi-sample CFAs in order to test for latent mean invariance across groups and, if required, to test for between-group differences. Thus, Model 4 (M4) imposed invariance constraints on factor loadings, intercepts, and factor means. As shown in Table 2, this model presented an acceptable practical fit (RMSEA = .07; SRMSR = .07; NNFI = .94; CFI = .94), but the difference in the NNFI and CFI values between M1 and M4 was larger than .01 for both indices. Hence, based on an evaluation of practical incremental fit indexes, there was no support for latent mean invariance.

Finally, Model 5 (in which factor loadings and intercepts were constrained to be invariant across groups) estimated between-group differences in latent means. The Portuguese sample was used as the reference group, so its latent means were fixed to 0. Then, latent means for the Spanish group were freely estimated. Thus, the estimated values represent the estimated mean differences between the two groups. The mean score for the Spanish group for both factors was higher than the mean score for the Portuguese group (.22 points higher for Task factor, and .29 points higher for Ego factor). These differences were statistically significant (p < .01), although the sizes of the differences were small. To assess effect size, d statistic (Cohen, 1988) was estimated. For Task and Ego factors (both with
significant differences in the latent means across groups), d statistic value was .27 and .30, respectively. Those values most likely indicate that the effects were in the range of small to medium.

As Model 3 turned out to be the more parsimonious and plausible model, the item parameter (except for intercepts) and the factor parameter estimates yielded by M3 are displayed in Figure 1. As can be seen, the differences in latent mean values estimated in M3 are congruent with the absolute mean differences between groups estimated in M5.

Because achievement goal theory (Nicholls, 1989) assumes that the two goal orientations (task and ego) are independent, an empirical examination of this relationship is of theoretical interest. The results showed that the relationship was found to be positive and statistically significant for the Spanish ($\phi = .12; p < .01$) and Portuguese ($\phi = .9; < .01$) sample. It appears that in our study the two latent variables were correlated, but to a small degree. Additionally, we also tested for

![Figure 1](image)

**FIGURE 1**
Factor loading, uniqueness, factor correlation and latent mean estimates in the Portuguese and Spanish groups for Model 3.

*Note.* LM = latent mean. For uniquenesses, factor correlations and latent means, the first value is for the Portuguese group. All parameter estimates are standardized in relation to a common group metric, and all are statistically significant ($p < .01$).
the difference between the correlation coefficients for the Spanish and Portuguese groups, finding a statistically significant difference ($z = 6.3; p < .01$).

**DISCUSSION**

The goal of the present study was to analyze the measurement equivalence of the TEOSQ across Spanish and Portuguese adolescents, using multi-sample Confirmatory Factor Analysis with Mean and Covariance Structure. As a prelude to the main analyses, support was provided for the reliability of the scores gathered with the Spanish and Portuguese language versions of the TEOSQ.

Drawing from the results of the CFA for each sample separately, support for a hypothetical two-factor structure was found across the two groups that were compared, stemming from an examination of multiple fit indexes. The multi-sample CFA reveals that the hypothetical two factor structure of the TEOSQ is identical across samples. Moreover, on the basis of overall adequacy of model fit together with $\Delta$CFI and $\Delta$NNFI values no larger than .01, it was concluded that no important differences were found on the item parameters (factor loadings and intercepts) across Spanish and Portuguese adolescents.

In light of our findings regarding factor loadings and intercept invariance, we feel confident that the TEOSQ can be considered to operate equivalently across Spanish and Portuguese adolescents. Given the known psychometric soundness of the TEOS, together with the popularity of its use, those results have the potential to add importantly to the construct validity of the instrument with respect to its Spanish and Portuguese adaptations.

Afterwards, we proceeded to test for latent mean invariance, but it was not supported. Finally, we tested for latent means differences across groups. These tests revealed statistically significant mean differences between groups.

The results revealed factor mean scores for Spanish adolescents to be slightly higher than those for their Portuguese counterparts on the two TEOSQ factors (task and ego orientation). However, taking into account the absolute size of these differences and the large sample size (Cohen, 1988), the practical significance of these findings is questionable. From this result, it seems likely that our findings of a discrepancy in the means of the two factor dimensions between Spanish and Portuguese adolescents need to be further verified. Additionally, while it is hypothesized that task and ego are independent, we found a small interfactor correlation. The factor correlation also needs to be further verified.

It is important to note at least some limitations of our reported results. First, although the samples were matched in terms of age, education, and gender breakdown, they cannot be considered totally representative of the populations of the two countries compared. Thus, there may be some risk that we have overgeneralized our findings. Finally, we did not measure the cultural orientations of the
individual students. We assumed that our two samples, which were drawn from different nations, each shared a common culture (collectivistic), but it is clearly a generalization that ignores variation within a nation.

In conclusion, the results of this study suggest that the factorial structure of the TEOSQ represents a valid and reliable theoretical construct for studying sport goal orientations among adolescents in Spain and Portugal.

REFERENCES


