Is the Greek Version of Teacher’s Evaluation of Student’s Conduct (TESC) a Valid and Reliable Measure?

Artemis Z. Giotsa¹*, Andreas N. Zergiotis², Theodoros Kyriazos³

¹Social Psychology, University of Ioannina, Ioannina, Greece
²Ministry of Education, Athens, Greece
³Department of Psychology, Panteion University, Athens, Greece

Email: *agiota@gmail.com

Abstract

Teacher’s Evaluation of Student’s Conduct Questionnaire (TESC) evaluates students’ conduct as perceived by their teachers. No assumptions for an a pri ori structural model exist. Therefore, this study attempted to identify underlying relationships of the Greek TESC factorial structure initially using Exploratory Factor Analysis, followed by Confirmatory Factor Analysis. A sample of 1201 students was evaluated by 71 teachers. The sample was split in two to perform EFA and CFA in different subsamples. A two-factor structure emerged from the EFA, applying Kaiser > 1 rule, Velicer’s Minimum Average Partial Test and Horn’s Parallel Analysis. This solution confirmed a theoretical classification proposed by Rohner. The CFA supported the bi-dimensional structure developed in EFA, with a two-factor model showing optimal fit. Common Latent Factor (CLF) indicated that no method bias altered the CFA results. Internal Consistency reliability was adequate. This first attempt of factorial analysis of TESC, Greek version offered preliminary evidence of internal consistency and construct validity.

Keywords

Teacher’s Evaluations, Students’ Conduct, Factor Analysis, EFA, CFA, Greece

1. Introduction

Evolution and developmental theorists agree that the need for positive response is a phylogenetic personality trait that humans have developed over time (Rohner, 1975, 1986). Positive response from persons with whom we have a strong emotional bond is even more important (Bjorklund & Pellegrini, 2002; Leary,
This core idea postulates the basic assumption of the Interpersonal Acceptance-Rejection Theory (IPART) of personality development and socialization (Rohner & Rohner, 1980; Rohner, 1986; Rohner & Khaleque, 2010). In childhood parents satisfy this need. Children seek affection, support and acceptance from them (Rohner, 1986). The scope of this search sometimes broadens to include anyone with whom we feel close (Rohner, 2010). This broadening of attachment may include teachers too (Rohner, 2015).

Besides, school–family relationship has been regarded as a significant mesosystem (Bronfenbrenner, 1977), influencing the developmental process and the psychological adjustment of children (Ali, 2011). Furthermore, school performance is a variable that is influenced by physical, social, and psychological and economic factors including students’ relationships with teachers and with their parents (Khan, Haynes, Armstrong, & Rohner, 2010).

The quality of the relationship between adults and children has also been the focus of extensive research. Specifically, the perceived quality of the parent-child relationship is considered an important variable of child-adult relationships (Coleman, 2003; Howes, 1999; Howes & Hamilton, 1992, cited in Erkman, Cancer, Sart, Borkan, & Sahan, 2010). More specifically, scholars suggest that the quality of the child-parent relationship mediates the subsequent relationships with teachers (Ryan, Stiller, & Lynch, 1994). Others argue that the effects of perceived parental acceptance on school performance may be mediated by perceived teacher acceptance (Woolley, Kol, & Bowen, 2009, cited in Khan, Haynes, Armstrong, & Rohner, 2010).

Moreover, research suggests that a positive relationship between parents, children and teachers correlates with high academic performance, richer social skills, improved school performance and better psychological adjustment in both children and youngsters (Rohner, Khaleque, Elias, & Sultana, 2010). Especially, the perceived security by the child is an important dimension in the teacher–child relationship bearing certain similarities to the parent-child relationship (Howes, 1999; Ali, 2011). Other findings suggest that when the teacher-child relationship lacks positivity the child may develop 1) school aversion, school absenteeism and low self-competence (Blankemeyer, Flannery, & Vazsonyi, 2002; Harrison, Clarke, & Ungerer, 2007; Rohner, Parmar, & Ibrahim, 2010) and 2) behavior problems such as frustration intolerance (Rohner, Khaleque, Elias, & Sultana, 2010).

Additionally, results from cross-cultural research on student’s school conduct expose gender and socio-cultural variability (Ali, Khaleque, & Rohner, 2014; Rohner & Khaleque, 2015). Perceived teacher acceptance mediated the relationship between perceived parental acceptance and youth’s psychological adjustment in Bangladesh (Rohner, Khaleque, Elias, & Sultana, 2010) and India (Parmar & Rohner, 2010). Perceived parental acceptance also affected students’ psychological adjustment in Kuwait (Rohner, Parmar, & Ibrahim, 2010) and Estonia (Tulviste & Rohner, 2010). Teacher’s Evaluation of Student’s Conduct (or TESC, Rohner, 2005) is a questionnaire used extensively in the above cross-cultural
studies of parental and teacher’s acceptance and rejection (Ali, Khaleque, & Rohner, 2014; Rohner, 1986; Rohner, 2010; Rohner & Khaleque, 2010; Rohner & Rohner, 1980).

Focusing further on TESC, it is a self-administered questionnaire. It was developed (Rohner, 2005) to assess students’ misbehavior in school environment, as perceived by their teachers. It consists of 18 items of misconduct in the school setting. All items are generally evaluating two broad categories of misconduct during schooling: 1) openly disruptive behaviors and 2) subtly disruptive behaviors (Rohner, 2015; Khan, Haynes, Armstrong, & Rohner, 2010).

On one hand, openly disruptive behaviors include mainly any expression of aggression. In IPART theory (Rohner & Khaleque, 2005) aggression is any behavior that spurs from the intention of hurting someone, something or oneself physically or emotionally (Rohner, Khaleque, & Cournoyer, 2012). Thus, it can be expressed physically, verbally or even non-verbally (Rohner & Khaleque, 2015; Rohner, Khaleque, & Cournoyer, 2012). So, behaviors of open disruption may include first physical expressions of hostility like fighting with peers or damaging foreign property. Secondly, they encompass the communication of verbal aggression with adults and peers, and finally any symbolical expression of aggression and hostility like defiance of teachers’ authority or refusal to do assigned schoolwork (Rohner, 2015; Khan, Haynes, Armstrong, & Rohner, 2010).

Generally, aggression and hostility are parts of the acceptance-rejection syndrome (Rohner, 2004). Some examples from items in this category are “Shoves and hits other people”, “Is quarrelsome”, “Disrupts classroom routine” or “Creates troubles in school”.

On the other hand, behaviors of subtle disruption comprise any action that can cause problems in the educational process in more indirect ways. These behaviors include cheating, stealing or lying (Khan, Haynes, Armstrong, & Rohner, 2010; Rohner, 2015). In other words, these are non-aggressive behaviors, considered unacceptable, being unethical, dishonest or even illegal. Examples of items in this domain are “Lies to get out of trouble”, “Cheats” or “Steals”.

TESC has been proven a valuable tool both in cross-cultural (Khan, Haynes, Armstrong, & Rohner, 2010; Parmar & Rohner, 2010; Rohner, 2010; Rohner, Khaleque, Elias, & Sultana, 2010; Rohner, Parmar, & Ibrahim, 2010; Tulviste & Rohner, 2010) and in multicultural samples (Khaleque, 2014). Specifically, more than nine studies in at least twelve nations used TESC. TESC respondents came from countries like Bangladesh, Colombia, Estonia, Finland, India, Japan, Korea, Kuwait, Pakistan, Puerto-Rico, Turkey, and the United States (Ali, Khaleque, & Rohner, 2014: p. 13). Consequently, some languages that TESC has been translated include Estonian (Tulviste & Rohner, 2010), Bangladeshi (Rohner, Khaleque, Elias, & Sultana, 2010), and Arabic (Ahmed, Rohner, Khaleque, & Gielen, 2010).

In the course of this rich cross-cultural research, TESC has demonstrated acceptable reliability (Rohner, 2005). Cronbach’s alpha ranges from .93 to .97 in recently published cross cultural works. Cronbach’s alphas per country are as
follows: for an Indian sample .96 (Parmar & Rohner, 2010), for a sample in Bangladesh .93 (Rohner, Khaleque, Elias, & Sultana, 2010), for a sample in Tunisia .96 (Tulviste & Rohner, 2010), for a sample in Kuwait also .96 (Rohner, Parmar, & Ibrahim, 2010) and finally for a U.S. sample .97 (Khan, Haynes, Armstrong, & Rohner, 2010).

Regarding factorial structure, TESC is reported to be a valid measure of school conduct assessment both by Melton (2000) and Rohner (1987), who provide relevant information (Rohner, Khaleque, Elias, & Sultana, 2010; Rohner, Parmar, & Ibrahim, 2010).

The current study attempts to identify underlying relationships between measured variables of TESC, Greek version using factorial analysis. Therefore, the current study has the following three objectives: 1) To explore the factorial structure of TESC by means of Exploratory Factor Analysis (EFA), in absence of an a priori model establishing construct validity; 2) To confirm the factorial structure that emerged from EFA by means of Confirmatory Factor Analysis (CFA); 3) To examine the internal consistency of the optimal model emerging from the CFA. Finally, two secondary objectives of this research were as follows: 1) reliability analysis and 2) evaluation of alternative CFA models, as an additional evidence of construct validity. Three research questions emerge from the above goals: 1) Can we identify underlying relationships between measured variables of TESC, Greek version using Exploratory Factor Analysis? 2) Can we confirm the structure that emerged from Exploratory Factor Analysis with Confirmatory Factor Analysis, evidencing construct validity? 3) What is the internal consistency reliability of TESC?

2. Method

2.1. Participants

Our sampling frame comprised a complete list of public schools in the largest metropolitan area of Greece (Athens). Specifically, data were collected from 15 schools selected by simple cluster sampling from all public educational organizations in Athens. If clusters (schools) are randomly selected, then cluster elements (students) also bear similarity to randomly selected cases (Kalton, 1983, cited in Gracia et al., 2012). Reasons for this course of action are 1) the vast majority of students in Greece attend a public school 2) nearly 40% of the total population in Greece lives in Athens.

A total of 1201 students (605 girls and 596 boys), from 1st grade to 12th grade was being evaluated for misbehavior frequency as perceived by their teachers. Students ranged in age from 6 through 17 years ($M = 10.74$ years, $SD = .6$). Regarding grade levels, 4% of the students attended 1st grade, 8% of the students attended 2nd grade, 10% of the students attended 3rd grade, 12% of attended 4th grade, 15% attended 5th grade, 16% attended 6th grade, 5% of the students attended 7th grade, 12% of the students attended 8th grade, 8% attended 9th grade, 3% attended 10th grade, 3% attended 11th grade and 3% of the students
attended 12th grade. In total, 780 students attended all primary school levels (1 - 6), 303 students were attending all middle school levels (7 - 9) and 114 students were attending all high school levels (10 - 12). The grade level was missing in four students. Teachers that participated \( N = 71 \) were from 24 to 58 years, \( M = 36.86, SD = 10.15 \) (69% females).

2.2. Materials

The researchers administered the Teacher’s Evaluation of Student’s Conduct or TESC (Rohner, 2005). TESC is a self-report questionnaire designed to assess teacher’s evaluations of student’s conduct. TESC is composed of 18 (e.g. “My student is impudent”, “Creates troubles in school”, “Shouts at or insults adults”, “Destroys property of others”). Items are scored on a five-point Likert scale from 5 (Very Often) through 1 (Almost Never).

Scores on TESC spread from a low of 18 (no or infrequent conduct problems) to a high of 90 (very often conduct problems). Scores at or above the midpoint of 54 indicate frequent students’ conduct problems as perceived by the teacher (Rohner, 2005). All scores between 18 and 32 indicate a student that never or almost never misbehaves in school. On the contrary, scores between 77 and the maximum value (90) indicate a very frequent behavior problem. In short, the higher the total score, the more frequent the conduct problems are as perceived by the teacher who is taking the test (Khan, Haynes, Armstrong, & Rohner, 2010).

Translation procedure

The instrument was translated in Greek using the translation and back-translation method (Brislin, 1970). More specifically, TESC (Rohner, 2005) was first translated in Greek by Author. Back-translation to English followed by a team member proficient in English, not familiar with the English version. All items of the original English version and the back-translated English version went through a cross-check, item by item to track any ambiguities. Following cross-check and refinements, all ambiguities were resolved, leading to the final Greek versions of the instrument.

2.3. Procedure

School principals from the randomly selected public schools were contacted by team members. They informed them about the study, inviting the school to participate voluntarily. Teachers from each school were also recruited on a volunteer basis. Permission from school authorities was obtained before data were collected. Team members administered the questionnaire to the teachers, explaining to them the purpose of the study. Each teacher filled a maximum number of 25 questionnaires as this is the maximum allowed number of students per classroom.

2.4. Design of the Research

The objectives of the study were pursued in two phases. First, in phase one we
attempted to establish a factorial structure for the Greek Version through Exploratory Factor Analysis, in absence of an a priori model (EFA). Secondly, in phase two we carried out a Confirmatory Factor Analysis (CFA) in the model established in phase one, thus confirming construct validity (Chan, 2014). In this phase we also test alternative structures. Finally, we evaluated the reliability of the instrument.

3. Results

The initial sample had 1245 cases. We carried out a missing value analysis in SPSS 20 (IBM, 2011). Little’s MCAR test results were significant with $\chi^2 (207) = 673.86, p < .001$, suggesting that values were not missing randomly. Missing data (2.5%) were estimated through the expectation-maximization algorithm (EM). The final sample had 1201 cases.

Final sample ($N = 1201$) was split in one-fifth and four-fifths in order to implement analyses in different samples, based on a methodology recommended both by Guadagnoli and Velicer (1988) and MacCallum, MacCallum, Browne, Sugawara (1996). On the one-fifth subsample (201 cases), Exploratory Factor Analysis (EFA) was applied whereas on the four-fifths subsample (1000 cases) Confirmatory Factor Analysis (CFA) was applied. Cases were assigned into the two groups by random number generation.

The one-fifth of the sample (201 cases) was considered appropriate for EFA for the following reasons. To begin with the sample-to-variable ratio in this subsample was 11:1 (201 cases/18 items = 11). That is there are 11 cases for each item in TESC (Hair et al., 1995; Nunnally, 1978; Velicer & Fava, 1998). A ratio of 11:1 is above minimum ratios (5:1 and 10:1) suggested in literature (Comrey & Lee, 1992; Gorsuch, 1983). Furthermore, the four-fifths of the sample (1000 cases) used in CFA was also large enough. The sample-to-variable ratio in this subsample was 56:1 (1000/18 = 56). Comrey and Lee (1992) created a classification scale for sample size where 500 observations were considered as a very good sample and 1000 observations as an excellent sample.

3.1. Exploratory Factor Analysis

Although a CFA was the primary goal of this research, such an analysis would be unfeasible without a preexisting EFA model (Finch & West, 1997; Timm, 2002; Williams et al., 2010). Since an EFA model was absent, both for the Greek Version, and for the original TESC (Rohner, Khaleque, Elias, & Sultana, 2010; Rohner, Parmar, & Ibrahim, 2010), EFA had to be the first step of this research. In order to explore the factor structure of TESC, IBM SPSS AMOS Version 20 (IBM, 2011) was used. EFA subsample ($N = 201$) was examined for factorability. EFA participants were 132 boys and 69 girls aged 6 to 17 years ($M = 10.69$, $SD = .73$).

Criteria of factorability used in this procedure are as follows: 1) Data Suitability Verification, 2) Extraction Method Selection, 3) Factor Extraction Criteria, 4)
Rotational Method Selection, 5) Factor Interpretation (Williams et al., 2010). Kaiser-Meyer-Olkin Measure of Sampling Adequacy (KMO = .92) suggested that 92% of the variance can be explained with Exploratory Factor Analysis. Bartlett’s Test of Sphericity was significant at $\chi^2 (153, N = 201) = 2520.63, p < .001$. Given these overall factorability indicators, exploratory factor analysis was conducted with all 18 items of TESC scale.

Principal Axis Factoring method was employed for extraction as the assumption of normality (DeCarlo, 1997) had been violated (Costello & Osborne, 2005; Fabrigar, Wegener, MacCallum, & Strahan, 1999; Fabrigar & Wegener, 2012). In fact, all skewness values were found significant ($p < .05$). Regarding kurtosis, 12 out of 18 values were found significant ($p < .05$). Additionally, Mardia’s coefficient (Mardia, 1970) was 538.58 (critical ratio 47.18).

A key issue of EFA is the optimal estimate of the number of factors to retain (Courtney, 2013). The Eigenvalue > 1 Rule (Kaiser, 1960) has received criticism by EFA researchers as unreliable (Costello & Osborne, 2005; Courtney, 2013; Ledesma & Valero-Mora, 2007; Velicer & Jackson, 1990). Consequently, two additional factor extraction criteria were employed, which are generally considered more statistically robust (Courtney, 2013; Gorsuch, 1983; Hayton, Allen, & Scarpello, 2004; Ruscio & Roche, 2012; Zwick & Velicer, 1986). First Velicer’s Minimum Average Partial Test (MAP) was used both in its original (Velicer, 1976) and revised versions (Velicer, Eaton, & Fava, 2000). The revised MAP test (based on the work of O’Connor, 2000) with the partial correlations raised to the 4th power rather than squared is considered an alternative procedure for determining both factors and components to retain (Velicer, Eaton, & Fava, 2000). Secondly an alternative version of Horn’s (1965) Parallel Analysis was used, suitable for PAF method (O’Connor, 2000), with raw data permutations (Buja & Eyuboglu, 1992).

More specifically, the Eigenvalue > 1 Rule (Kaiser, 1960) suggested three factors (Table 1), with the third marginally above one (1.08). However, according to the original Minimum Average Partial Test (Velicer, 1976) the minimum average squared partial correlation value (.0273) is attained for a two-factor solution. In line with original MAP, the revised MAP (Velicer, Eaton, & Fava, 2000) showed that the minimum average partial correlation value (0.0019) is also attained for a two-factor solution. Regarding Parallel Analysis (Horn, 1965), only for the first two factors, the eigenvalues from the raw data permutations were noticeably lower than those from the original data. Therefore, Parallel Analysis also supported a two-factor solution.

Total variance explained by the two factors was 55% (see Table 1). Eigenvalues suggested that the first factor explained 49% of the variance and the second factor 6% of the variance.

Factors had an acceptable correlation of .37 so Oblique Rotation (Oblimin) method was considered more suitable.

All communalities were within a range of .41 to .73, except item “Steal” and “Refuses to do school work” with a value of .28 and .17 respectively. Note that
these items were also difficult to handle during translation process. Table 2 presents communalities per item along with the pattern matrix derived. All items of the pattern matrix had tolerable to acceptable primary loadings (.38 to .90). More specifically, the first factor had 13 items (items 1 - 13) with

**Table 1.** Total variance explained.

<table>
<thead>
<tr>
<th>Factors</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% Variance</td>
</tr>
<tr>
<td>1</td>
<td>9.23</td>
<td>51</td>
</tr>
<tr>
<td>2</td>
<td>1.60</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>1.08</td>
<td>6</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Axis Factoring.

**Table 2.** Factor loadings (Pattern Matrix) and communalities based on principal axis factoring with oblimin rotation for the 18 items of the tesc (N = 201).

<table>
<thead>
<tr>
<th>Item #</th>
<th>ITEM</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Communalities</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Disrupts classroom routine</td>
<td>.90</td>
<td>.73</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Is disobedient</td>
<td>.88</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Is unruly in school</td>
<td>.86</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Is disrespectful</td>
<td>.81</td>
<td>.65</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Is impudent</td>
<td>.77</td>
<td>.59</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Creates troubles in school</td>
<td>.74</td>
<td>.66</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Shoves and hits other people</td>
<td>.71</td>
<td>.71</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Is rebellious or defiant of teacher’s wishes</td>
<td>.70</td>
<td>.50</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Is cruel and makes fun of others</td>
<td>.70</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Shouts at or insults adults</td>
<td>.69</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Takes revenge on other children</td>
<td>.60</td>
<td>.33 .61</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Is quarrelsome</td>
<td>.60</td>
<td>.55</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Refuses to do school work</td>
<td>.57</td>
<td>.28</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Lies to get out of trouble</td>
<td>.47</td>
<td>.45</td>
<td>.58</td>
</tr>
<tr>
<td>15</td>
<td>Destroys property of others</td>
<td>.42</td>
<td>.40</td>
<td>.46</td>
</tr>
<tr>
<td>16</td>
<td>Is abusive to younger or smaller children</td>
<td>.39</td>
<td>.38 .41</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>Cheats</td>
<td>.66</td>
<td>.44</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>Steals</td>
<td>.40</td>
<td>.17</td>
<td></td>
</tr>
</tbody>
</table>

satisfactory loadings, from .57 to .90. The second factor contained 2 items (items 17 - 18) with acceptable loadings from .40 to .66. Moreover, three cross-loading items emerged (items 14, 15, 16) with loadings from .38 to .47. They had marginally higher loadings on the first factor. However, this difference was as low as 0.1 to 0.2. Therefore, we went into CFA in an attempt examine factor structure further.

Nonetheless the two factors that emerged group items in a theoretically understandable way. Specifically, our labeling followed a naming convention adopted by Rohner (Khan, Haynes, Armstrong, & Rohner, 2010; Rohner, 2015). Factor 1 was labeled “Overtly Disrupting Behaviors” while Factor 2 was called “Subtly Disrupting Behaviors”.

### 3.2. Confirmatory Factor Analysis

In order to confirm the factor structure of TESC establishing construct validity, IBM SPSS AMOS Version 20 (IBM, 2011) was used. Our CFA subsample was N = 1000. CFA participants were 464 boys and 536 girls aged 6 to 17 years (M = 10.78, SD = .47).

Data were non-normally distributed for this dataset too (DeCarlo, 1997). All skewness and all kurtosis values were found significant (p < .05). Likewise, Mardia’s coefficient (Mardia, 1970) was 2477.36 (critical ratio 1247.67). Under these circumstances, Maximum Likelihood Estimation Method (ML) was unsuitable. Given the non-normality of the data, bootstrapping was used (2000 bootstrap samples) to find out parameter estimates; and the Bollen-Stine corrected p value of the chi-square was estimated (Byrne, 2010). Using a conventional significance level of p < .05, the model was rejected because of poor fitting to the data.

As an alternative the Unweighted Least Square (ULS) Estimation Method was employed. Due to this, model fit was evaluated with fit measures compatible with the ULS method, excluding chi-square, which is not (Byrne, 1994). Researchers recommend the use of an extensive array of measures (Marsh, Balla, & Hau, 1996). AMOS evaluates model fit for ULS Method by means of (Arbuckle, 2005) the Standardized Root Mean Square Residual (SRMR), the Goodness-of-fit Index (GFI), the Adjusted Goodness-of-fit Index (AGFI) and the Normed Fit Index (NFI). As far as acceptability criteria of the above measures, SRMR has a cut-off value close to .08 or below (Hu & Bentler, 1999). GFI and AGFI values beyond .90 show tolerable model fit (Byrne, 1994), while a value above.95 underpins a significant model fit (Kelloway, 1998; Hu & Bentler, 1999). Finally, NFI values that exceed .90 (Byrne, 1994) or .95 (Schumacker & Lomax, 2004) are similarly suggesting a sustainable fit. Goodness-of-fit indexes applicable to this data set are summarized in Table 3 along with their results for the optimal model fit found.

More specifically about optimal solution found, model had 2 latent factors. The first factor (also called Overtly Disrupting Behaviors) was equivalent to Factor 1 of EFA containing 15 manifest variables, i.e., items 1-13 of EFA plus 2
Table 3. Fit indexes for optimal two-factor model (with 15 items in factor 1 and 3 items in factor 2).

<table>
<thead>
<tr>
<th>Index</th>
<th>Value in this study</th>
<th>Acceptability Value</th>
<th>Commenting</th>
</tr>
</thead>
<tbody>
<tr>
<td>GFI</td>
<td>.988</td>
<td>&gt;.95</td>
<td>Acceptable</td>
</tr>
<tr>
<td>AGFI</td>
<td>.985</td>
<td>&gt;.95</td>
<td>Acceptable</td>
</tr>
<tr>
<td>NFI</td>
<td>.979</td>
<td>&gt;.95</td>
<td>Acceptable</td>
</tr>
<tr>
<td>SRMR</td>
<td>.062</td>
<td>&lt;.08</td>
<td>Acceptable</td>
</tr>
</tbody>
</table>

GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; NFI = normed fit index; SRMR = standardized root mean square residual.

out of 3 cross-loaders. The second factor (Subtly Disrupting Behaviors) paralleled to Factor 2 of EFA containing 3 variables. Namely, items “Cheats” and “Steals” like Factor 2 of the EFA plus item “Lies to get out of trouble” which was one of the cross-loaders in EFA.

This model confirmed EFA structure and while verifying the theoretical categorization proposed by Rohner (Khan, Haynes, Armstrong, & Rohner, 2010; Rohner, 2015). Additionally, model fit measures resulted in values according to Table 3. For GFI, AGFI, and NFI, typically a value that is greater than .90 suggests adequate model fit whereas a value that is greater than .95 suggests significant model fit (Hu & Bentler, 1999). Similarly, when SRMR stays below .08 model fit is considered adequate whereas even lower SRMR values (below .05) show significant fit (Hu & Bentler, 1999). SRMR value in optimal model was marginally above this level (.062).

By convention, a single factor might also represent the data adequately, exposing maximum parsimony (Crawford & Henry, 2004). Consequently, a single-factor model including all 18 items of TESC was also examined (MODEL 1).

The maximum parsimony hypothesis was not confirmed (GFI = .988, AGFI = .985, NFI = .978, and SRMR = .065) as MODEL 2 remained the optimal model. This finding suggested that the current data set was best represented by more than one factor. This assumption led to an effort to test an alternative two-factor model (MODEL 3). Table 4 compares all models tested. GFI and AGFI values were identical. NFI and SRMR made the difference in favor of MODEL 2 with a value of .979 and .062 respectively.

As to item allocation per factor (or more precisely per latent variable) in the alternative two-factor model, we tested Factor 1 with 14 items Factor 2 with 4 items. This alternative two-factor model had 1 item less in Factor 1 in comparison to the optimal model. On the contrary, in Factor 2 it had item “Takes revenge on other children” plus items “Cheats”, “Steals”, “Lies to get out of trouble”, as the optimal model.

As far as loadings and correlations for the optimal model are concerned (Figure 1), Standardized Regression Weights were ranging for Factor 1 from .42 (“Refuses to do Schoolwork”) to .76 (“Is disobedient”, “Disrupts classroom routine”, “Is unruly in school”). In Factor 2 they had weights from .33 (“Steals”)
Table 4. Fit indexes for alternative models tested in comparison with optimal model.

<table>
<thead>
<tr>
<th>Model #</th>
<th>Description</th>
<th>GFI</th>
<th>AGFI</th>
<th>NFI</th>
<th>SRMR</th>
</tr>
</thead>
<tbody>
<tr>
<td>MODEL 1</td>
<td>1 Factor with all 18 items</td>
<td>.988</td>
<td>.985</td>
<td>.978</td>
<td>.065</td>
</tr>
<tr>
<td>MODEL 2</td>
<td>2 Factors with 15 &amp; 3 items (Optimal)</td>
<td>.988</td>
<td>.985</td>
<td>.979</td>
<td>.062</td>
</tr>
<tr>
<td>MODEL 3</td>
<td>2 Factors with 14 and 4 items</td>
<td>.988</td>
<td>.985</td>
<td>.978</td>
<td>.063</td>
</tr>
</tbody>
</table>

GFI = goodness-of-fit index; AGFI = adjusted goodness-of-fit index; NFI = normed fit index; SRMR = standardized root mean square residual.

Figure 1. The graphic representation of the standardized solution for optimal Model (Model 2). Conventionally, latent factors are represented by large circles, errors as smaller circles and manifest variables as rectangles. Single-headed arrows connecting the variables represent a causal path while double-headed arrows denote correlation between variables.

to .80 ("Lies to get out of trouble"). The two latent variables (factors) “Overtly Disrupting Behaviors” and “Subtly Disrupting Behaviors” were correlated with a value of .81.

Finally, we carried out a common method bias test in the optimal model to find out if a method bias was altering the results of our measurement model. We used the “unmeasured latent factor” method (Podsakoff et al., 2003) suitable for
studies that do not explicitly evaluate a common factor as here. The comparison of the standardized regression weights before and after the addition of the Common Latent Factor (CLF) indicated that none of the regression weights were affected by the CLF. Deltas were less than .200 and both Composite Reliability (CR) and Average Variance Extracted (AVE) for each construct still met minimum thresholds.

3.3. Reliability Analysis

Internal reliability was estimated for the latent variables emerged in CFA by three different methods: 1) Cronbach’s alpha coefficient (Cronbach, 1951), 2) Spearman-Brown Coefficient, 3) Guttman Split-Half Coefficient. Cronbach’s alpha coefficient for the whole scale was .89 (N = 18). Cronbach’s alpha for the subscale “Overtly Disrupting Behaviors” was .88 (N = 15). For subscale “Subtly Disrupting Behaviors” Cronbach’s alpha was .56 (N = 3). Guttman Split-half Coefficient was .87 for the total scale while Spearman-Brown for the total scale was .90. All reliability coefficients and descriptive statistics are presented in Table 5.

4. Discussion

The purpose of this study was to explore the factorial structure of TESC, Greek version. Exploratory Factor Analysis (EFA) was carried out first in the lack of assumptions for an a priori model (Finch & West, 1997; Timm, 2002; Williams et al., 2010). Only unpublished similar studies are quoted in literature (Melton, 2000; Rohner, 1987 quoted in Rohner, Khaleque, Elias, & Sultana, 2010 and Rohner, Parmar, & Ibrahim, 2010). CFA followed.

The sample was split randomly and both factor analyses were carried out in different subsamples, following a methodology proposed by Guadagnoli and Velicer (1988) and by MacCallum, Browne and Sugawara (1996). In this study CFA subsample was four times larger than the EFA subsample. Scholars are in debate about sample splitting, when a researcher wishes to carry out both EFA and CFA. Sample splitting was carried out because the sample was large enough to allow this procedure without sacrificing the reliability of the results (Comrey & Lee, 1992; Gorsuch, 1983; Guadagnoli & Velicer, 1988).

Table 5. Descriptive statistics and reliability coefficients of scales that emerged in CFA for the teacher’s evaluation of student’s conduct (N = 1000).

<table>
<thead>
<tr>
<th>Subscales</th>
<th>M</th>
<th>SD</th>
<th>Cronbach’s alpha</th>
<th>Spearman-Brown</th>
<th>Guttman Split-Half</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overtly Disrupting Behaviors</td>
<td>16.73</td>
<td>3.80</td>
<td>.88</td>
<td>.90</td>
<td>.82</td>
</tr>
<tr>
<td>Subtly Disrupting Behaviors</td>
<td>3.15</td>
<td>.63</td>
<td>.56</td>
<td>.63</td>
<td>.59</td>
</tr>
<tr>
<td>Total TESC Score</td>
<td>19.88</td>
<td>4.23</td>
<td>.89</td>
<td>.90</td>
<td>.87</td>
</tr>
</tbody>
</table>
Regarding sample power, literature is rich in suggestions (Williams, Onsman, & Brown, 2010) extending from 3:1 to 6:1 (Cattell, 1978), 5:1 (Gorsuch, 1983), 10:1 (Hair et al., 1995; Velicer & Fava, 1998; Nunnally, 1978), even 20:1 (Tabachnick & Fidell, 2007). However, no more than 15.4% of the studies fall within the range of greater than 10:1 and less than or equal to 20:1 (Costello & Osborne, 2005). Here, caution was taken to keep both subsamples powerful enough. To this end for the EFA, being a necessary step of this study in the absence of an a priori model (Finch & West, 1997; Timm, 2002), the sample-to-variable ratio was kept just above 10:1. A Minimum of 5 to 10 cases per variable have been suggested (Comrey & Lee, 1992; Gorsuch, 1983). Simultaneously, any sample size above 200 cases is supposed to provide sufficient statistical power (Hoe, 2008). The Confirmatory Factor Analysis (CFA) subsample was on the threshold of 1000 cases. Comrey and Lee (1992) suggested that 1000 cases are an excellent sample size for factor analysis.

More concisely, key findings that emerged from this attempt were the following:

1) Our data were positively skewed, suggesting that perceived students’ misbehavior as reported by Greek teachers was assessed overusing the low end of the Likert scale.

2) Our EFA analysis suggested a two-factor structure for TESC. However, a CFA was necessary to offer a clearer picture of the factorial structure of the Greek TESC due to: a) three cross-loading items and b) a third factor with an eigenvalue marginally greater than c) The third factor had also questionable stability as it contained only two items. Anyhow, the CFA that followed corroborated the two-factor structure of EFA revealing a model with fifteen items on the first factor and three items on the second factor.

3) The CFA that followed EFA, confirmed the construct validity of the instrument because the model found reached acceptable values. At the same time and additional models tested to further support the construct validity of the optimal model found.

4) Yet another evidence of construct validity is that these two factors grouped the 18 items of TESC in a theoretically understandable way, since Rohner in literature (Khan, Haynes, Armstrong, & Rohner, 2010; Rohner, 2015) has already classified TESC behaviors contained in our EFA and CFA factors as Overtly Disrupting and Subtly Disrupting.

5) Measures of reliability reached acceptable values suggesting that items of TESC measure misbehavior consistently. As the second scale had three items, Cronbach’s alpha coefficient would not be enlightening, being dependent on the number of items under evaluation (Cortina, 1993; Green, Lissitz, & Mulaik, 1977; Nunnally & Bernstein, 1994). So, except Cronbach’s alpha, two more coefficients, were employed. All reliability coefficients had very comparable values (Table 5). Finally, the value of Cronbach’s alpha for the whole scale was consistent with the ones cited in literature further endorsing reliability. More crucially,
Moving away from key findings into a more detailed view on EFA, data was suitable for an Exploratory Factor Analysis for the following reasons. To begin with, Kaiser-Meyer-Olkin measure of sampling adequacy was far above the recommended value of .6. For the Kaiser-Meyer-Olkin index a value that is greater than .50 is adequate for factor analysis (Kaiser & Rice, 1974). Moreover, Bartlett’s test of sphericity showed high significance. Given these overall indicators, Exploratory Factor Analysis with all 18 items of the scale was carried out. Principal Axis Factoring was used because of the violation of the normality assumption (Costello & Osborne, 2005; Fabrigar et al., 1999). Besides, when variables have high reliability the differences between PCA & PAF are almost eliminated (Thompson, 2004 as quoted in Williams et al., 2010). This premise was verified here, as PCA provided very similar results. Oblique rotation produces correlated factors thus more accurate results in humanities research where correlation between variables is more often than not, expected (Costello & Osborne, 2005). Since our factors were correlated, oblique rotation (Oblimin) was employed. Two factors were retained based on the Eigenvalue > 1 Rule. We could have retained a third factor with Eigenvalue marginally above 1 (1.08), but Kaiser’s Criterion is known to overestimate the actual number of factors (Ruscio & Roche, 2012 as quoted in Courtney, 2013). So, additionally two other methods were used for determining the number of factors to retain, as their results are generally considered more accurate and stable (Courtney, 2013; Henson & Roberts, 2006; Zwick & Velicer, 1986). More specifically Horn’s Parallel Analysis (1965) and Minimum Average Partial Test (Velicer, 1976; Velicer, Eaton, & Fava, 2000) were used. Taken into consideration jointly, these results provide support for a two-factor solution. The cumulative percentage of variance explained by the two retained factors was between the acceptable levels of explained variance for humanities (Hair et al., 1995). Moreover, this two-factor model was proposing adequate theoretical support as it had already been proposed by Rohner in literature (Khan, Haynes, Armstrong, & Rohner, 2010; Rohner, 2015).

Next, in an attempt to further establish construct validity, we performed CFA. Summing up the CFA results, optimal model revealed two similar latent factors, also validating both EFA analysis and TESC theory (Khan, Haynes, Armstrong, & Rohner, 2010; Rohner, 2015). More explicitly, first latent variable (or Overtly Disrupting Behaviors) replicated Factor 1 of the EFA, this time containing 15 items instead of 13. The second latent variable (or Subtly Disrupting Behaviors) was similar to Factor 2 of the EFA, containing three items instead of two. Besides items “Cheats” and “Steals”, the item “Lies to get out of trouble” was included which was one of the cross-loaders. The rest of the cross-loaders (2 items or “Destroys property of others” and “Is abusive to younger or smaller children”) were included in Latent variable “Overtly Disrupting Behaviors” (or Factor 1) with acceptable Standardized Regression Weights.

Finally, alternative models were also tested. At first, a single-factor structure was tested based on the assumption that this structure is exposing the maximum
parsimony (Crawford & Henry, 2004). Results suggested that the two-factor model with 15 and 3 items per factor best fitted our data. Further on, another two-factor model was tested with 14 items and 4 per factor. Still the two-factor model with 15 and 3 items per factor, remained the optimal model, as fit indices showed. In particular, the SRMR and the NFI of the optimal fit model were better than the ones in the rest of the models. GFI and AGFI values were identical for all models.

5. Conclusions

Generally, cross-cultural research suggested that behaviors measured by TESC are characterized as culturally and gender specific (Rohner, 2010; Rohner & Khaleque, 2015). Therefore, this solution contains only what the Greek cultural context has categorized as Overtly Disrupting and Subtly Disrupting Behaviors, and this classification is also context specific. However, there are more culturally specific issues. The positive skew of the sample distribution, suggesting an over-use of the lower levels of the Likert scale could equally be culturally dependent. Perhaps a way to overcome this underscoring tendency would be to rephrase certain items slightly, making them more convergent to the language of a Greek teacher.

A limitation of this study is the absence of convergent and discriminant validity, but in the Greek context no choices of long established and reliable measures exist for this purpose. However, preliminary evidence from CFA used for construct validity are encouraging. Future research could further build on these initial findings of construct validity adopting the modern, and holistic view of construct validity (Messick, 1989), also adopted by research standards for education and psychology research (AERA, APA, & NCME, 1999). According this view validity is a unified construct tailored around construct validity (Chan, 2014). A second limitation is the use of EM to fill missing values that are not missing at random. Despite these limitations, the Greek version of TESC is both a valid and a reliable measure for the evaluation of students' conduct, as perceived by their teachers. Finally, the validation of TESC, Greek version may boost the research of school-related conduct problems in Greece in relation to rejection and acceptance theory (Rohner, 1975).

References


