DOI: 10.1027/2157-3891/a000065
Article category: Article
Manuscript received February 21, 2022
Revision received November 25, 2022
Accepted November 25, 2022
Published online February 28, 2023
Running Head:

Journal of Publication:

International Perspectives in Psychology: Research, Practice, Consultation

Paper Title:

Greek Standardization of the Situational Triggers of Aggressive Responses (STAR^{GR}) Factorial Structure and Methodological Considerations

Kostas Mylonas[1], Claire Lawrence[2], Ioannis Frangistas[1], Mary Bower Russa[3], Sofia Papazoglou[1], Ioanna Papachristou[1], & Anna Zajenkowska[4]

1. Department of Psychology, National and Kapodistrian University of Athens, Greece
2. Lawrence PsychAdvisory, Nottingham, UK
3. Department of Psychology, Grand Valley State University, Allendale, MI, USA
4. Institute of Psychology, Maria Grzegorzewska University, Szczęśliwicka 40, 02-353, Warsaw, Poland

Corresponding author:
Kostas Mylonas
Department of Psychology
National and Kapodistrian University of Athens
15784 Ilissia
Athens, Greece
kmylonas@psych.uoa.gr

Ioannis Frangistas ORCID 0000-0002-1622-184X

IMPORTANT NOTE (*in accordance with the agreement-release form*):
This version of the article may not completely replicate the final authoritative version published in the *International Perspectives in Psychology: Research, Practice, Consultation* journal under [DOI: 10.1027/2157-3891/a000065]. It is not the version of record and is therefore not suitable for citation. To read and cite this paper, please visit DOI: 10.1027/2157-3891/a000065 at Hogrefe Publishing.

Abstract

Aggressive traits and situational factors interact to influence the propensity for aggressive behavior. Individuals' sensitivity to the situational factors provocation and frustration are assessed by the Situational Triggers of Aggressive Responses Scale (STAR) subscales Sensitivity to Provocation (SP) and Sensitivity to Frustration (SF). The aims of this paper were to: i) briefly summarize the basic theory supporting the STAR scale, ii) review previous cross-cultural findings for these STAR constructs, iii) discuss the Greek factorial structure and its cross-cultural validity, iv) describe the methodological-statistical and psychometric properties for a Greek normative sample (*N*=1,094), and v) provide Greek norms and other application aspects regarding the STAR scale. The aforementioned are discussed in the context of both a broader methodological-statistical scope and an applied scope. Finally, and with respect to psychological intervention, a method that can be applied to any sample, independently of the existence of norms, is discussed, for multivariately identifying extreme SP and SF cases (outliers).

Keywords:

STAR scale and Greek norms; Provocation; Frustration; Situationally-triggered aggression; Extreme groups detection method

Impact and Implications.

In line with the UN's Sustainable Development Goals (SDGs) and specifically SDG#11 goal to make cities safer, more inclusive, sustainable, and resilient environments, we investigated the situational triggers of aggressive responses under standardization procedures to gain a better understanding of the nature/character of aggressive cues (sensitivity to provocation or frustration) at a cultural level in the context of urban living (e.g., drivers' aggression). Our research contributes to efforts to tackle violence and lack of respect for humans and the environment. Adopting a prosocial stance by changing the way we think about ourselves and society, we change the way we look at our animate and natural environment, and we promote increased social awareness and responsibility.

Introduction

Individual differences in aggression-related traits. Numerous theories have been proposed as to how and why people become aggressive with models such as the *General Aggression Model* (Anderson & Bushman, 2002) developed as possible frameworks (Lawrence, 2006). Such frameworks include situational triggers for aggression, but individual differences in response to these triggers have been given relatively little attention (Lawrence, 2006). Different triggers will produce different levels of aggression across different people. Thus, in accordance with numerous studies (Crick & Dodge, 1994; Deater-Deckard et al., 2010; Dodge et al., 2006), both aggressive traits and situational parameters produce different levels of aggressive behavior under different conditions across different persons. The delineation of the range of individual difference characteristics that may impact aggressive responding, as well as the mechanisms involved, has been an ongoing source of research attention (e.g., Carver et al, 2008; Robinson et al., 2020). Indeed, recent studies suggest that nearly half of the variance in aggressive responding may essentially be the result of genetic (individualistic) factors (Veroude et al., 2016). Individuals' sensitivity to both provocation and frustration has attracted attention as key predictors of aggressive responding, with characteristics such as trait anger, or emotional impulsivity/dysregulation often mediating the relation between provocation and aggressive responding (e.g., Robinson et al., 2020; Carver et al., 2008; Deater-Deckard et al., 2010). However, attempts to delineate these patterns more clearly or to apply this knowledge in clinical contexts is dependent on the ability to operationally define and assess these risk factors more effectively in research and clinical populations.

Individual differences in sensitivity to aggressive triggers. While levels of aggressive behavior and the relationships between irritation, anger, and frustration have been modeled via the APQ (Aggression Provocation Questionnaire) by O' Connor et al., (2001), the nature of the individual responses to the specific situational triggers of aggression cannot be captured via this instrument (Lawrence, 2006). Attempting to remedy this, the antecedents that make people feel aggressive have been modeled through the Situational Triggers of Aggressive Responses (STAR) Scale (Lawrence, 2006) to "examine whether there are patterns of associations between triggering factors and personality traits linked to aggression and the propensity to act aggressively" (Lawrence, 2006, p. 242).

The STAR scale has been available for research since 2006. This instrument has been examined with respect to its use for specific populations and its methodological and psychometric properties (examples are: Zajenkowska et al., 2014; Zajenkowska, Jankowski, Mylonas, & Rajchert, 2015; Mylonas et al., 2017); one common denominator in these three specific studies is their cross-cultural and/or cultural approach. For example, in 2014, Zajenkowska et al. examined sex differences in situationally triggered aggressive behaviors under a cross-cultural perspective across three countries (Poland, UK, and Greece). Analytics focused on cultural variation in gender-based patterns of aggressive responding to provocation vs. frustration, while accounting for trait aggression and other possible confounds.

The STAR scale properties. The STAR consists of two subscales: *sensitivity to provocations* (SP) and *sensitivity to frustrations* (SF). SF is associated with aggressive feelings in contexts where the individual's ability to achieve a goal is blocked or they find themselves in a situation that is beyond their control. Higher sensitivity to aggressive reactions in situations of frustration is associated with higher trait anger (Lawrence, 2006). However, SF is unlike a personality trait in the sense that personality traits (e.g., neuroticism), are associated with stable and predictable mood, while mood for individuals on SF depends on the situational context (Zajenkowska et al., 2014) such as a stressful exam situation or a car accident. In contrast, SP is the propensity for an individual to perceive other people as aggressive, assaultive or in some way hostile, and to feel aggressive when provoked.

Research supports SP and SF as separate constructs. For example, attentional processes vary as a function of sensitivity to provocation (Zajenkowska & Rajchert, 2020). Low SP individuals gaze longer at non-hostile cues as compared to hostile cues, and those with higher SP do not focus their gaze significantly longer on either hostile or non-hostile cues in ambiguous scenes of social encounters, unless they are also high on trait anger. Also, SP is related to how people interpret other people's behavior (Lawrence & Hutchinson, 2013b) and how they act toward them (level of intention and provocation target). In short, SP is related to events that the individual perceives as directly goaded or provoked by another person. Additionally, both SP and SF are related to depressive symptoms; however, SF is a stronger predictor (Zajenkowska, Zajenkowski, & Jankowski, 2015).

Further insight into SP and SF can be achieved by studying their relation to emotional and cognitive components of trait aggression and to overt physical aggression. In the relative literature (Lawrence, 2006; Lawrence & Hodgkins, 2009) it is evident that high SF is associated with anger and hostility, that is with the emotional and cognitive aspects of aggression; however, SF is not related to overt aggression. In addition, self-concept clarity is less for those with high SF, so a less stable and less coherent self-image is associated with feeling aggressive due to frustrating rather than provocative events. In contrast, high SP is associated with physical aggression and higher levels of narcissism (Lawrence, 2006; Lawrence & Hutchinson, 2013b). Moreover, others will be perceived as more provocative, if SP is high, regardless of other factors such as sex, mood, or general trait aggression.

Cultural, cross-cultural, and psychometric aspects of the STAR scale. Much of the work using STAR has had a cross-cultural emphasis. For example, Zajenkowska, Jankowski, Mylonas, & Rajchert (2015) examined correlates of STAR in a Polish and a Greek sample. The study's emphasis on coffee consumption and propensity to experience aggressive feelings under situational triggers bolstered evidence of validity and allowed theoretical comparisons across the two countries with respect to possible antecedents of situationally triggered aggression. In addition, several studies have provided evidence of convergent validity for STAR subscales (Lawrence, 2006; Lawrence & Hodgkins, 2009; Lawrence & Hutchinson, 2013a; Lawrence & Hutchinson, 2013b; Mylonas et al., 2017; Zajenkowska et al., 2014).

Most recently, Mylonas et al. (2017) focused on assessing the methodological, statistical and psychometric properties of the scale using samples from five countries (UK, Poland, Greece, the US, and S. Korea) and an overall *N* of 1,219 participants. Several psychometric strengths of this instrument (such as limited statistical item noise and the absence of erratic behavior in the item pool) were identified in that research. Factor equivalence was demonstrated on two levels of analysis, namely individually compared cultures and clustered cultures (per the Georgas and Berry, 1995, paradigm). While minor, culturally interesting variations were identified, a final "clusters of countries" methodology illustrated that cultural specificities on the STAR dimensions could be minimized (if desired) by forming theoretically meaningful sets of countries sharing common psychological and other characteristics.

Some other methodological considerations related to the STAR scale have already been discussed (Mylonas et al., 2017), such as a) the *Kendall's Tau-b* alternative to *Pearson's r* for the matrices to be analyzed through CFA, and b) the "hit" matrix approach, using Tucker's φ indices as entities, analyzed via MDS-T (a constrained on the circle or the sphere ALSCAL solution). The first consideration is of statistical and metric nature and has been discussed for the purposes of improving the clarity and validity of the factor structure; the second one is of a theoretical nature as it allows for a deeper understanding of the culture-specific scale facets and is capable of revealing clusters of countries under a cross-cultural scope. Following these considerations, cultural and cross-cultural properties of the STAR scale should be further examined, modeled, and described for cultural groups, such as countries and/or even clusters of countries. Specifically, a further description of intraculture statistical and psychometric properties of the STAR scale was the main aim, but several secondary ones will also be addressed under various statistical approaches to our normative data. The emphasis should be placed on several aspects, from the initial adaptation procedures to derived norms; in addition, when a single culture is studied, identifying the psychometric-statistical properties of the STAR scale, such as standard errors of the estimate and confidence limits of raw scores are of specific importance.

Aims of the current study. The evidence so far and the confirmation of the factorial structure, clearly supports the construct validity of the scale, allowing for a standardization study as existing data on the reliability and validity of the STAR scale have increased confidence in its statistical robustness and its theoretical strength. Its application beyond English-speaking populations would allow wider cultures and communities to assess individual differences in aggressive triggers more robustly, and potentially produce a standardized assessment of the STAR scale for these populations. The current paper aimed to do that within a Greek population as the STAR research to date has demonstrated its utility in broad cross-cultural research designs, as well as in Greek populations. Thus, the purpose of the present study was to explicitly standardize the STAR instrument for use in a Greek population, also in line with Barbara Byrne's suggestions (2016). This required translating the instrument in a culturally sensitive way, and delineating STAR statistical and psychometric properties in the Greek population, including focusing on critical details such as standard errors of the estimate and confidence limits of raw scores, along with alternative methods for extreme-scoring identification.

In this standardization process, it should be noted that cultural factors may have strong influence on participants' questionnaire responses. As a member of Southern Europe, Greece has family and collectivist expectations for individuals and groups that may differ from those in other the UK and other areas of Europe. This could create issues around the construct of provocation (as most of STAR's provocation items are about provocation to the self and not the group or family). For example, higher interdependence in a Greek sample (Hareli et al., 2015) was related to lower attention to negative emotions and anger expressions, as Greeks are concerned with collectivistic and harmony rules (Kafetsios & Hess, 2013). Therefore, there may be subtle but important distinctions in the way that the STAR scale is responded to in other cultures beyond those that it was originally designed and tested in.

Method

Sample

The initial sample (*N* = 1,840) was recruited either through an online survey module or via traditional questionnaire administration. Informed consent was obtained, and all participants were free to terminate the procedure at any time. The participants did not receive any compensation. Some parts of the sample were less abiding to random sampling methods than other parts, as some questionnaires were administered to clusters of participants (mostly university students in their classrooms); we excluded these parts from the study sample. In addition, respondents who were part of clinical samples or who were not Greek citizens were also excluded from the final sample. The resulting sample over-represented females and university students. To achieve as equal sex groups as possible, we randomly selected about 50% of the females and cross-validated this random selection through comparisons of demographic characteristics across the random sample and the full female sample. We did not amend for the over-representation of university students (*n*=655) as the non-student sample (*n*=439) appeared to be represented in an acceptable way in the final sample of 1,094 participants.

Description of the study sample (*N*=1,094). Demographic information gathered from the 1,094 respondents is presented in Figure 1. For several characteristics, only a part of the participants provided information, but the percentages derived are indicative of the sample composition and reflect the target population.

----- Insert Figure 1 here -----

Measures: The Situational Triggers of Aggressive Responses Scale: STAR and STAR^{GR}

The focus of this study was the Situational Triggers of Aggressive Responses Scale (STAR), originally created by Lawrence (2006). STAR is a self-report measure comprised of 22 items that are scored by the respondents on a 5-point Likert scale (5="very accurate") with regard to the tendency for each situation to make them feel aggressive. The instrument includes two subscales: Sensitivity to Frustration (SF, 10 items) and Sensitivity to Provocation (SP, 12 items). These two scales assess sensitivity to different types of cues with regard to the elicitation of aggressive responding: "SF is a proneness to feel particularly aggressive in response to having one's goals blocked and in response to uncontrollable negative events. SP relates to feeling aggressive in reaction to goading and provocation from others." (Zajenkowska et al., 2014, p. 355). The STAR is intended for ages 16 or higher and was initially tested through a series of studies on UK samples arriving at acceptable reliability estimates, that is .82 for SF and .80 for SP (Lawrence, 2006; Lawrence & Hodgkins, 2009; Lawrence & Hutchinson, 2013a; Lawrence & Hutchinson, 2013b). More recent research has demonstrated the instrument's psychometric stability along with cross-cultural factor equivalence across five countries (Mylonas et al., 2017), as has been described in the Introduction section, with reliability estimates ranging from .78 to .83 for the SP dimension and from .77 to .80 for the SF dimension.

Adaptation procedures. In order to arrive at the Greek version of the STAR ($STAR^{GR}$), we followed standard translation and back-translation procedures, and amendments were carried out accordingly to match the Greek cultural setting. Specifically, the questionnaire was translated into Greek by two proficient in English collaborators and was then back-translated to English by a bilingual associate. Language and content amendments that followed were minimal, and the overall procedure was approved by the author of the original scale. However, some adaptation was necessary, as a few words and phrases might not prove culturally-fair. In some cases, such adaptations were easy (such as in the case of the item "I am the subject of a practical joke" for which the Greek version "φάρσα" [farssa] as derived from the word "farce" was employed); in some other cases, adaptation proved more difficult, as in the case described in the following paragraph. In general, we adapted the instrument to avoid sex-discriminating language (in Greek, different articles and other parts of speech are employed to denote the sexes) in order to convey the real cultural meaning of items to the best possible extent. A related issue is the Reading Level of the instrument; the STAR scale has a Flesch-Kincaid Reading Grade level of 5/6 (thus, US children in Grade 5 or 6, age 10 or 11, should be able to read and understand the scale without difficulty). It has a Flesch Reading Score of 64.3, meaning it is of standard ease to read. We estimate the Greek instrument's reading level to be approximately 6th grade (11-12 years of age).

One important language problem which became evident during the translation and adaptation procedures, and also later during the pilot and other studies was that the frustration construct itself (mentioned twice in the scale) did not seem to be easily comprehensible by a number of Greek respondents. This raised a cultural-effect question with respect to one of the two dimensions and with respect to the associated two items which specifically refer to "frustration" in their phrasing. These items, in their Greek translation, contained the word "ματαίωση [mαtεosse]" which is the exact translation of "frustration". However, this Greek word is closer to a psychological term than its English equivalent, which is more likely to be used in everyday conversation. With the two words showing different frequency rates of use in each population, it is easier for those using it more frequently to better understand its meaning. In everyday speech, Greeks use other words, such as "αγανάκτηση [αḡαnαktesse]" which is a feeling close to frustration, but it translates back to "indignation". Another possible Greek word is "εκνευρισμός [εknεvresmoss]", which translates back to "irritation" or "annoyance".

This raised the issue of whether these two specific items were conceptualized the same way by all respondents. Notably, a similar issue can arise with the English term, as the word 'frustrated' can also refer to annoyance or being aggrieved. However, analyzing a set of independent Greek data (research in progress), for items 5 and 18 which contain the word "frustration", there were no statistically significant differences between a) a group of 63 who explained the concept correctly and full, b) a group of 147 who explained it but not in full, and c) a group of 36 who did not explain it correctly or stated they did not know the term (*Kruskal-Wallis H* criteria were calculated and the respective .607 and .168 χ^2 criteria with 2 degrees of freedom were non-significant). However, as this caveat refers to two out of the 10 SF items (that is 20%), we need to be cautious with respect to the psychometric properties of these two items and consequences regarding the overall SF score they might produce.

Psychometric considerations. Despite caveats, the instrument behaves in a relatively satisfactory manner in many ways, as previous studies have shown. First, it may yield data that satisfy the parametric conditions (e.g., Zajenkowska et al., 2014) allowing for the use of *Pearson's r* throughout the factor analytic approach. Still, this is not a given, as results may be enhanced or even rectified when non-parametric indices, such as *Kendall's Tau-b*, are used (see Mylonas al., 2017). Second, the factor structure that emerges under different, and also cross-cultural, datasets reveal acceptable levels of factor equivalence and construct similarity, at least for a number of cultures. Third, and probably most important, the structures that emerge may include interesting and meaningful cultural variations and specificities, but these are not sufficient to invalidate the original STAR structure shown by Lawrence (2006). Cultural variations with respect to the STAR scale are usually smoothed out (Zajenkowska et al., 2014) by target (Procrustean) rotation and other methods (such as multilevel covariance structure analysis), reproducing a clear two-factor structure of SF and SP and indicating a quite similar although not tautologous structure to the one described by Lawrence in 2006. This should allow for cultural specificities to be represented in the final Greek structure.

Procedure

The data were collected before the SARS-COV2 outbreak, from Spring 2018 to Autumn 2019 via i) [*an internet site -The details here have been removed for blind review reasons*], using the Concerto platform. The study was announced and advertised through social media along with the aid of undergraduate students who advertised the study via their social networks. Data were also collected via ii) group administration to students during lecture hours, and iii) individual administration in occupational settings[1].

Results and Discussion

Item-Level Descriptive Indices and Item Considerations

The average of all items' means was 3.07 (min = 2.35, max = 3.88) with an average standard deviation of 1.16 (min = 1.05, max = 1.30), the average median of the items was 3.18 (min = 2, max = 4) and the observed range for all items was 4 (1 to 5) (Table 1). To get some insight with respect to the "frustration" translation concern relating to the two respective items in the SF subscale, raised earlier, we computed a series of dependent-samples *t*-tests to check whether statistically significant differences across SF items might reveal systematically lower or higher scores for these two items. Although several statistically significant differences were indeed observed, these differences were both positive and negative, showing no particular unidirectional bias for these two items. In addition, we examined the correlations of these two items with all other SF items and found that the coefficients resembled those calculated for any other remaining SF item in the subscale. One note though is that the correlation between the two items including the word "frustration", "ματαίωση [mɑtεosse]" was .24, which is marginally weaker than the average intercorrelation of SF items (.29). Another observation is that the standard deviations of these two items was a little smaller (1.17 and 1.12 for items SF 5 and SF 18 respectively) than the average SF item (1.20) and smaller (by 0.14 to .01) from the other eight SF items. Although these differences were small, these observations reminded us to be cautious with these two items during the factor analytic stage. It should be noted, however, that threat levels

[1] Our true thanks to V. Marneli, L. Sakari, J. Tzalides, and Ch. Vassou for their help during data collection in various settings. About 50% of the data were collected via recruitment type (i), 30% via (ii), and 20% via (iii).

regarding reliability were not strong, possibly reflecting only some minor variation in the data.

----- Insert Table 1 here -----

For some of the items, standard deviations were somewhat elevated (e.g., items 9 and 16), and so are these items' skewness levels. For item 2, the standard deviation was much lower than the average. In addition, for half of the items, kurtosis was also somewhat elevated, and for all items –as one would expect– the *K-S z* (*Lilliefors* correction) and the *Shapiro-Wilk W* criteria were statistically significant. Univariate outliers were detected for six of the 22 items, all assessing SP. All these called for caution if *Pearson's r* indices were to be selected in subsequent analyses, and underlined the need for the possibly preferable alternative approach of using *Kendall's Tau-b* –at least at a safeguard level.

Factorial Structure

A consideration of alternative correlation coefficients. One of the most important aims of the current study was to describe the factorial structure for the Greek population and to test the initial theoretical structure as proposed by Lawrence (2006) and as shown in cross-cultural studies (Zajenkowska et al., 2014, Mylonas et al., 2017). Of particular note, the approach used for the two studies reported in Mylonas et al. (2017) employed an alternative approach to the correlation tables analyzed. Specifically, those studies employed *Kendall's Tau-b* indices, and not *Pearson's r*, to avoid confounds in CFA modeling caused by skewness in the data (which had an effect on the correlation indices when these were compared using *Fisher's* transformation). Although this statistically sound approach served the purposes of those specific studies well, *Pearson's r* indices are preferable (due to their parametric nature) if such an option is acceptable (i.e., when some violation of statistical assumptions does not heavily affect the coefficients to be analyzed).

For the factor structure testing stage, our main statistical tool was Confirmatory Factor Analysis, but we initially employed Exploratory Factor Analysis as well, to gain some preliminary insight by modeling the data theoretically, first on an exploratory manner. During the factorial structure exploration stage, it was evident that when analyzing *Pearson's r* indices, the solutions produced were different and less informative than when analyzing non-parametric correlations instead, namely *Kendall's Tau-b*. A decision on which of the two correlation indices to use was fundamentally important for this study, capitalizing on previous attempts with STAR (Mylonas et al., 2017) and with other psychometric tools (such as the EVS, European Value Survey; Georgas et al., 2004; Georgas & Mylonas, 2006) and as has repeatedly been suggested in the literature (Graziano & Raulin, 1989; Guilford, 1956; Kline, 1993; Tabachnick & Fidell, 2001; Thurstone, 1947). Despite the fact that the correlation matrices in the initial data (*N*=1,094) were not different (3.4% different only, as compared through *Fisher's z* transformation and under a two-tailed H_A), the EFA solutions under the *Pearson's r* option were unstable and the structure was much better when *Kendall's Tau-b* was employed. The method to compare correlation indices through *Fisher's z* transformation follows Winer (1971) (see also Hinkle et al., 1988; Mylonas et al. 2012; Papazoglou & Mylonas, 2017; Zajenkowska et al., 2014) and although an expected family-wise error might indeed inflate the amount of correlation differences, this was not the case (difference levels at 3.4% only), allowing us to explore both correlation alternatives.

Confirmatory Factor Analysis modeling. In our main analyses we tested for three successive factor models and a modification model, namely the Independence model, the Unifactorial model, the two-factor model and a modified for error-covariances two-factor model. We applied these models to the 1,094 cases to compute the CFA solutions under two competing conditions (which is viewed as a "*correlation hypothesis*" in this study): using *Pearson* correlations vs. using *Kendall's Tau-b* correlations as the input for the analyses through maximum likelihood estimation via LiSRel 8.30. Although this was not directly and imperatively dictated by apparent differences in the correlation matrices (in contrast to the Mylonas et al., 2017 study), we had considered this comparison as one of our methodological concerns for the current study. Specifically, we tested for possible improvement in the solution from one condition to the other, even of minor magnitude, to be able to select the best fitting model describing the $STAR^{GR}$'s structure.

By employing standard criteria in regard to the acceptable levels of the various fit indices (Bolen et al., 2014; Hu & Bentler, 1999; Schumacker & Lomax, 2010), we first computed the outcomes based on the *Pearson* correlation matrix for the 22 STAR items (Table 2). The Independence model was expectedly rejected, and the Unifactorial solution was not acceptable (RMSEA > .06; $\chi^2 \div df = 9.11$) although its Goodness of Fit indices were close to acceptable levels and there was a large improvement compared to the Independence model (TLI ≈ .89). Still, the theoretically driven two-factor model was a better fit, with the AIC and TLI indices showing further improvement along with much better Comparative Fit Index (CFI) exceeding the .90 level, but with GFI and AGFI still remaining below that. For this model, RMSEA improved (.073) but still did not reach acceptable levels (≤.06; Hu & Bentler, 1999). Thus, we attempted a fourth model, including non-zero error covariances between specific items following the impact on model rule and also using the transitivity rule, i.e., if an error covariance was estimated between items 2 and 4 and then another between items 4 and 7, a third estimate was also computed between items 2 and 7 without this being indicated in the first place. This model, despite the unavoidable χ^2 statistical significance, was clearly better in terms of RMSEA, RMR, GFI, AGFI, AIC and TLI indices than the previous one (Table 2, *Pearson's r model c*) and could be accepted as a verification of the theoretically driven STAR factor structure (12 provocation and 10 frustration items). It should also be noted that error covariances were allowed to exist only within each of the two factors; that is, no cross-loadings were allowed as the two factors are assumed to be independent. However, one would argue that this final modified model ("*d*") did not reach acceptable levels in all respects, so there was still room for our proposed use of *Kendall's Tau-b* coefficients (the "correlation hypothesis") to produce further improvement.

----- Insert Table 2 here -----

The CFA computations on the *Kendall's Tau-b* coefficients (Table 3) followed the same route: three successive models and a modified one. The Independence model was rejected for the same reasons; the unifactorial model showed much better fit (TLI ≈ .90). Still, the theoretically driven two-factor model was even better with the AIC and TLI indices showing further improvement along with Goodness of Fit indices

(CFI, GFI and AGFI) exceeding the desired .90 level. In addition, RMSEA and RMR had improved and had reached acceptable levels (Table 3). To test for an error-covariance hypothesis (to examine the impact of adding error-covariances to this 3rd model as we did for the *Pearson's r*-based solution), we computed a fourth, modified model, allowing for non-zero error covariances between specific items (strictly within each of the two factors only). There was some improvement observed, but it was minimal (TLI ≈ .10) and AIC did change but to a small extent. Goodness of Fit indices, RMSEA and RMR improved but only to a limited extent (Table 3). In all, when using the *Kendall's Tau-b* correlation indices to test for the theoretically driven two-factor STAR structure, there was no benefit to employing non-zero error covariances to verify the existence of these two factors, whereas when using *Pearson's r* indices instead, the non-modified model less clearly supports the two theoretical factors. With respect to the current study, the finally accepted solution for our data is *Model c* (using the *Kendall's Tau-b* solution) which is graphically presented in Figure 2.

----- Insert Table 3 here -----

----- Insert Figure 2 here -----

The initially observed difference between the two correlation matrices is an important issue at this point; it seems that even minor, statistically non-significant changes in the correlation indices (that is, even if $|z|$ criteria for transformed correlations following Fisher do not exceed the critical z; Mylonas et al., 2012) can make a vast difference with respect to the way the data behave and may eliminate discrepancies (such as the ones calling for modifications of a close-to-acceptance model). When computing the mean difference between the 231 pairs of correlations (*Pearson's r* vs. *Kendall's Tau-b*), their approximate mean difference was .048, which is a non-trivial difference with regards to correlations. A possible implication stemming from this outcome is that when comparing two correlation matrices one might consider the absolute average of the paired differences, as this might be a less statistical but still informative indication.

Aggregate scores and further statistical considerations. Having reached the CFA outcomes on the factor structure, we could now compute the factor aggregates by adding the items corresponding to each factor (averaging them by the number of items involved, 12 and 10 items, respectively, for SP and SF). These final aggregate scores were tested for their distribution characteristics and are presented in Table 4.

----- Insert Table 4 here -----

The basic descriptives showed larger variability for the SF dimension and a smaller mean value, with the difference being statistically significant (t=24.624, df=1,093, p<.001). The *Kolmogorov-Smirnov* and *Shapiro-Wilk* tests showed a departure from normality. Although this deviation may be attributed to the large N, it requires further investigation with respect to the nature and behavior of the factor aggregate scores. To better understand this departure from normality and to gain further insight with respect to the factor structure shown to hold through CFA modeling, several methods were employed. The first involved the calculation of the regression factor scores for each of the two dimensions. To do so, we first estimated the necessary factor coefficients through our CFA modeling and then calculated the respective factor scores via *z*-scores for each item and each participant. Thus, two sets of scores were now available, one consisting of the factor scores computed as just described and the other consisting of the two factor aggregate scores (simple sums over *k* items). If the two sets exhibit high correlation levels (≥.95), this would verify that important information hidden in the computed factor structure would not have been lost due to the standard method of averaging item responses within each dimension to arrive to factor aggregate scores. Indeed, both correlations exceeded .97. This, along with the strong reliability estimates (*Cronbach's alpha,* and *McDonald's omega* as suggested by Hayes & Coutts, 2020) assured us that the factor aggregate scores could be further employed in the analyses with safety.

Tucker's Phi and Target Rotation. The UK and Greek factor structures were compared using several methods. A first observation was that the UK and Greek factor structures were almost identical, albeit with somewhat lower loadings in the Greek version; especially for items 2, 16 and 21. For the next step, target rotation was applied between the Greek and the UK factor structures. This was calculated to explore cultural specificities and identify whether these could or should be smoothed out. The results are presented in Table 5.

----- Insert Table 5 here -----

The Procrustean solution closely followed the initial UK factor structure (Lawrence, 2006) and the differences in the loadings' magnitude observed even from the CFA stage do not pose a serious threat to the structure as their error loading is close to null; however, we should further discuss these three 'irregularities'. In detail, items 2, 16 and 21 are the ones which may reflect different cultural aspects across the two countries (Greece and the UK). Practical jokes (item 2) may be interpreted in different ways in the two cultures. In Greece, such an act might be considered offensive, even on culturally sanctioned days such as "April-fool's day". In the UK, the tradition of playing practical jokes is stronger, thus it is perceived as less personal and more game-like. The offense of reckless driving (item 21) may also be associated with the extent to which the drivers –especially male ones– are expected to respond to such provocation; this extent may depend on the masculinity levels being threatened (O'Dea et al., 2018) as a part of an honor belief system, possibly different in the two cultures. According to the same research, "men are expected to defend themselves, but only following a threat, insult, or other form of provocation against one's masculinity" (p. 131), as social rewards are expected if such a reaction is successful. Sex differences may be involved in such a process, along with the possibly different weights assigned by different cultures to this provocation. With respect to the inconsiderate behavior of drunk people (item 16), drinking alcohol to the point of getting drunk is more common amongst British individuals (according to the Global Drug Survey, 33.7% reported getting drunk in 2020, with data collected before the SARS-COV2 outbreak) than Greeks (14.3%), although it is not clear if these percentages reflect drunk drivers too. Thus, drinking behavior is not as similar in the two cultures as one might initially expect. As a result, alcohol-related aggression may simply be less experienced or less visible in Greece and this may result in different weights assigned to it. In all, such contained item irregularities were expected and pose minor threat, but they should be further studied with regard to the interpretation levels and provocation targets as perceived by the provoked. Both sex and culture may play an important role.

Apart from the Procrustean rotation, for the next step, we computed a series of congruence coefficients comparing our current and previous (Mylonas et al., 2017) study and its alternative solutions with the original STAR factor structure as supported by Lawrence (2006). *Tucker's φ* indices were calculated to assess factor congruence between previous versions of the STAR questionnaire with the original 2006 structure. All *φ* indices ranging from .89 to .92 indicated factors being identical (or nearly identical) between the structure revealed in the current study and the factor structures described in Mylonas et al. (2017) (across countries and across clusters of countries). Factor indices between the Greek and the UK 2006 versions ranged from .91 to .92 for the SP factor and from .89 to .90 for the SF factor. These indices indicated similar (nearly identical) structures between the two solutions. The current study's structure was shown being identical to the original UK 2006 one ($\varphi \geq .90$). Thus, through this approach, we demonstrated consistent findings with the original STAR structure being supported (12 SP items and 10 SF items), with only small discrepancies (all congruence coefficients are presented in Table 6). The marginal *φ* values (.89) pointing to the Greek data presented in the 2017 study and the second overall cluster of countries (Poland, UK, and S. Korea) might indicate cultural specificities possibly not accounted for by the adaptation procedures. Of course, one might model these specificities instead of smoothing them out, depending on their extent (in this case, it might be argued that it is important to model those rather than "eliminating" them).

----- Insert Table 6 here -----

Psychometric Properties

At this stage, our aim was to describe some of the basic psychometric properties of the $STAR^{GR}$ scale (Table 7).

----- Insert Table 7 here -----

First, the satisfactory reliability estimates (both *Cronbach's alphas* and *McDonald's omegas*), as evident for both dimensions (SP and SF), deserve merit; this was a supportive outcome with respect to the overall Greek version of the STAR scale. Second, the arithmetic mean, the trimmed mean (5%) and the Huber-estimator within each dimension are similar, although not identical. This might be an indication of outliers under both a univariate and a multivariate perspective, an indication further explored at the final stage of the analysis. Third, the arithmetic differences among the Mean, the Median and the Mode for each dimension, which indicate skewness were further tested via *K-S z* and *Shapiro-Wilk W.* These were statistically significant for both STAR dimensions. However, the stem-and-leaf plots did not suggest large departures from normality. Finally, the extent of the standard error of the estimate and its reflection on the estimated range of confidence around each raw score was relatively limited (for both dimensions) but not less than one unit. We should note that this final score (for both SP and SF) is an aggregate score, computed by averaging all respective factor items so as to express each dimension on the original rating scale (1 to 5). Thus, for such a raw SP score to differ from another, an absolute difference of 1.13 is necessary; for the SF dimension, the necessary algebraic difference is somewhat larger, that is 1.27. So, for example, if an individual scores 3.27 on the SF-aggregate, then this person differs from another person who scored 2.00, but does not differ from another person who scored 2.43. In all, these range limits around the raw aggregate scores for each dimension may be considered a little wider than expected (i.e., it would be better if this range was closer to 1.00), but they are acceptable in terms of score-discrimination power.

Standard Scores and Norms Table

Finally, we computed the standard scores corresponding to the averaged aggregates for each dimension and for simple sums for these scores. The correspondence between raw and standard scores for each STAR^{GR} dimension (SP and SF) is presented in Table 8 for the two extremes of the distribution (the rest of the Table is omitted for brevity).

For the norms provided in Table 8, the factor sums (item raw scores, summed) for the first factor ranged from a score of 12 to 60 to the corresponding 20 to 80 T-scores while for the second factor sums ranged from 10 to 50 with corresponding T-scores of 20 to 80. The averaged aggregates are more directly comparable. Thus, for the *z*-score (standard score) to +1.50 (which is a T-score of 65), one has to reach a raw score (averaged aggregate) of 3.90 on SF, whereas for the same T-score one has to score 4.25 on SP. In short, the two dimensions are not distributed the same way which is indicative of their different nature as well. To discuss this finding further, a shift seems to exist (under frustration one can reach a T-score of 80, whereas under provocation, the same T-score is not reached and vice versa) and it shows primarily that the SF peak is reached earlier, even with smaller raw scores. This difference may reflect a cultural specificity stemming from the characteristics of the Greek sample and/or the situational setting of everyday life in Greece.

----- Insert Table 8 here -----

A Method for Identifying Multivariate Extreme Scores

How can we identify a person or a subset within a larger set (sample) who are extremely sensitive to provocation and to frustration? One option might be to use the normative scores but there are two problems with that: first, we do not have the norms for other than the Greek population yet; and second, we can univariately identify such extreme scores, but a multivariate approach would be better and much more effective in identifying special cases (Penny & Jollife, 2001). To achieve this goal, we need to use multivariate distances and, in this case we may employ *Mahalanobis's* D^2 as our multivariate indicator. The method has been supported (Mardia et al., 1989; Tabachnick & Fidell, 2001) and has been employed to identify gifted students (using intelligence and creativity scores) among a large set of more than 1,000 children (Gari et al., 2000). This identification method first capitalized on the computation of multivariate clusters, and for the cluster with the highest scores, D^2 was computed and evaluated for its statistical significance under the χ^2 distribution. Following this method, we first clustered our data on the concurrent multivariate basis of the SP and SF scores, and we then computed the D^2 indices for each cluster separately. By calculating statistical significance levels for each of them as χ^2 criteria with 2 degrees of freedom, we depicted 46 multivariate outliers for the high-scores cluster and 21 for the low-scores cluster. This was achieved under a lenient critical χ^2 ($p<.05$). When we used a more stringent *a* level ($p<.01$), the multivariate outliers dropped to just 10 for the high-score cluster and just 4 for the low-score cluster. Under the strict *a* of .001, just one multivariate outlier (high-scores cluster) was identified. However, this very strict *a* level should not be applied under the circumstances and as the method used is not addressing both ends of the multivariate distribution concurrently; that is, the method addresses each of the two multivariate clusters separately, one at a time, seeking multivariate outliers first within the high-scores cluster and then, separately, within the low-scores cluster. Also, the meaning of outliers will differ in terms of whether the person is scoring very low or very high on STAR dimensions. A person who scores high might be easily triggered and may have a problem in everyday life with angry emotions and the perception of others behaving in a provocative or frustrating manner towards them. Those who score extremely low on the STAR dimensions may experience flattened affect or may find themselves being too flexible to the needs and behavior of others, as seen in dependent personality disorder for example. Thus, it becomes evident that either an *a* level of .01 should be used or, to avoid excluding marginal cases, a more lenient *a* of .05 should also be considered while evaluating the χ^2 criteria under this specifically tailored method.

Conclusion, Limitations, and Future Research

Through this multifold study, apart from the standardization of the STAR scale in its Greek version and the description of the psychometric properties involved, we attempted several other aims, mostly methodological and statistical. First, it is clear to us that the alternative use of correlation coefficients (*Kendall's Tau-b*) is a necessity when analyzing any STAR scale data, as was evident in our previous studies as well; this still holds true despite our current efforts to reinstate the standard *Pearson's r* approach. Although the initial correlation matrices regarding the two alternative methodologies did not really differ, even small such differences seem to have an impact on the final factor structure describing the scale. Thus, we believe that any future correlational approach using the STAR scale should consider employing *Kendall's Tau-b* instead of *Pearson's r* in the statistical analysis.

Through an alternative methodology, we created standardized scores and considered multivariate approaches (e.g., *Mahalanobis's* D^2; Gari et al., 2000; Mardia et al., 1989) to identifying extreme scores. This approach may be beneficial to use of the STAR scale in Greek clinical and research settings (for example, as a screening tool or to identify high or low-risk groups in experimental designs). This methodology can serve as a front-line alternative to identifying extreme groups when culture standardization and norms are not available, enabling the person-oriented potential of the STAR scale, in order to identify and examine individuals' specific situational sensitivities for aggression. Thus, our method for identifying extreme scores in a multivariate sense may be an important step ahead with respect to the clinical use of the scale. Even if no standardization exists for a population, this method can reveal at least the extreme cases at both ends of the continuum, so that action can be taken at the individual level. Even more, clustering a non-normative set of cases using the STAR scale data in any population of interest and in any culture should yield appropriate information for the high-end cluster to be explored further through the multivariate outliers detection method described in the Results and Discussion section. By identifying such a high-risk STAR cluster (or by knowing that an individual is part of such a high-risk cluster), and then considering other correlates of these high-risk scores for individuals in such a cluster, the practitioner can create a potentially useful profile for these persons-at-risk. This may allow better intervention planning for these individuals and/or populations.

The current study did not attempt to suggest a possible network of related constructs, linking SP and SF to e.g. trait aggression, anxiety, violence, or even psychopathology, as such efforts would appear premature. However, future research should capitalize on these findings to explore such networks. Research focused on better understanding of the factors that may contribute to aggression is particularly relevant in a context in which aggression and domestic and street violence, as well as deadly attacks inside schools and other institutions have become all too common: *Education Week* has reported 27 school shootings so far in 2022 and 119 since 2018, (Education Week, 2022, January 5), and an earlier study (Agnich, 2015) had reported 282 mass-murder school attacks across 38 nations. Similar aggressive patterns hold true in Greece as well, where recent violent crimes –with predominantly female victims– have become common news. The SF and SP patterns in the behavior of the perpetrators in these acts of violence are often a common denominator. A successful attempt to delineate such patterns more clearly or to apply this knowledge in clinical contexts to support practical management of aggression is dependent on the ability to assess these risk factors more effectively in research and clinical populations. To study the mechanisms involved in aggressive acts within a culture, we should capitalize on the parameters reflected in the available psychometric properties of the constructs assessed via a valid instrument. For example, and especially when norms are not available, it is crucial for the practitioner to be sufficiently informed of the necessary raw-score absolute distance between two persons, for these to be considered different with respect to their SP and SF levels; in the present standardization, a consultant might accept a raw-score distance of 1.3 being the safest way to discriminate between two people, for both STAR dimensions. Further on –and from a practical standpoint, proper adaptation, standardization, and availability of norms are necessary for using any instrument in a clinical or applied research setting (Byrne, 2016). But when instruments are to be used in cultural samples that differ from that in which the instrument was normed, attention to translational issues and nuanced aspects of the culture are as important as psychometrics and normative sampling in demonstrating the utility of these instruments in the new cultural context. Finally, when an instrument such as STAR is normed and standardized for use across an array of cultures, it opens the door for a more complex exploration of ways in which culture may interplay with other factors to shape aggressive responding.

Taking advantage of the present data, psychological traits such as impulsivity and interdependence may mediate/moderate SP to aggression, but the extent of these effects will certainly vary across cultures. For example, since the trait of interdependence is higher for Greeks than for other nationalities or cultures (Hareli et al. 2015), its effect on SP may be active only from some SP threshold score upwards (the SP raw score distribution indicates this threshold being 47/60 or so), possibly suppressing the manifestation of aggression due to this sensitivity dimension. To gain further insight into such a possibility, we might attempt to "read" the norms table for possible departures from the standard curve –in terms of non-perfect correspondence. In our standardization, very high raw SP scores correspond to lower than expected norms (as evident in the absence of raw scores in the highest norm area of 78 to 80) and this might be an indication of a "suppressive" effect, resembling a moderation-type of interference (as this effect is not present in the lower levels of the normed data). Thus, it takes more high SP-score answers for a Greek adult to correspond to an extreme manifestation of SP in this specific culture. This is not true for SF though; for that, even the lowest raw SF score of 10 does not correspond to the lowest T-score of 20, meaning that even with very low raw scores, Greeks are already experiencing some amount of SF. Thus, with fewer than expected high SF responses, Greeks reach the standardized average amount of SF in their culture. While the pattern with low SF is more complicated, one can easily identify people at the very high levels of SF; for practitioners, such identification of people at the high end of risk is often most critical. Of course, several psychological traits and other factors should be considered, with the SP patterns possibly attributed for example, to interdependence (or a possible social desirability effect?) and the SF's larger dispersion attributed to all the everyday Greek life cumbersome situations one has to deal with; all these certainly require further research with the norms and the rest of the available psychometric characteristics always to be taken into account.

Limitations, Further Future Research Considerations, and Concluding Remarks

A central limitation in the present study is the overrepresentation of the university students population. We did not amend for this as by doing so we would limit the sample's normative *power* and as this overrepresentation does not pose a serious threat to the normative *nature* of the sample. In more detail, one would expect 60% of our sample to have no university training but we have only 40% non-university students (ages 18 to 40 years). Of course, this limitation has to be kept in mind with respect to representativeness and norm use based on this paper. Still, even though our normative sample is not ideal, we do have quite a bit of diversity with respect to sex, urban/rural locations, educational level, family status, and socioeconomic status.

The Procrustean solution showed that three SP items might be associated with cultural specificities or differences which should be modeled in future research. These items should be examined with regard to sex to model possible cultural invariance at least at the item level or at higher levels as well (metric, scalar). Previous cross-cultural STAR modeling (2014, 2017) has shown cross-cultural equivalence across a number of countries but invariance testing taking a sex by culture interaction into consideration has not been attempted yet.

With respect to CFA modeling, although goodness of fit indices and model-improvement indices are acceptable, a word of caution is warranted: the $\chi^2 \div df$ ratios are high and remain at non-acceptable levels even for the finally accepted model. This pattern may reflect some cultural specificities or adaptation issues to be pursued further in future research. Future validation studies and small-scale studies on specific Greek target populations might add to our knowledge with respect to norms verification and adjustment.

Considering all, the standardization itself may reveal interesting and useful clinical insight to the practitioner, who will also combine other assessed psychological traits of the individual to arrive at a complete picture. Through the use of norms, available discriminatory power, identification of extreme cases within a larger group, and other information on the scale's characteristics, we believe we aid the researcher and the consultant to better understand the nature of situationally triggered aggression and to be able to avoid missing signals from individuals at risk. Despite the minor caveats, the findings here demonstrate that the STAR scale retains its robust psychometric properties within a Greek cultural context and future work can focus on the extent to which it can be used effectively within clinical, forensic, and educational interventions.

References

Agnich, L. E. (2015). A comparative analysis of attempted and completed school-based mass murder attacks. *American Journal of Criminal Justice, 40*(1), 1-22 https://doi.org/10.1007/s12103-014-9239-5

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology*, *53*, 27-51. https://doi.org/10.1146/annurev.psych.53.100901.135231

Bollen, K. A., Harden, J. J., Ray, S., & Zavisca, J. (2014). BIC and alternative Bayesian Information Criteria in the selection of structural equation models. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(1), 1-19. https://doi.org/10.1080/10705511.2014.856691

Byrne, B. M. (2016). Adaptation of assessment scales in cross-national research: Issues, guidelines, and caveats. *International Perspectives in Psychology: Research, Practice, Consultation, 5*(1), 51-65. https://doi.org/10.1037/ipp0000042

Carver, C. S., Johnson, S. L., & Joormann, J. (2008). Serotonin function, two-mode models of self- regulation, and vulnerability to depression: What depression has in common with impulsive aggression. *Psychological Bulletin, 134*(6), 912-943. https://doi.org/10.1037/a0013740

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social-information processing mechanisms in children's development. *Psychological Bulletin, 115*(1), 74-101. http://doi.org/10.1037/0033-2909.115.1.74

Deater-Deckard, K., Beekman, C., Wang, Z., Kim, J., Petrill, S. A., Thompson, L. A. & DeThorne, L. S. (2010). Approach/positive anticipation, frustration/anger, and overt aggression in childhood. *Journal of Personality, 78*(3), 991-1010. https://doi.org/10.1111/j.1467-6494.2010.00640.x

Dodge, K. A., Coie, J. D. & Lynam, D. (2006). Aggression and antisocial behaviour in youth. In W. Damon, & N. Eisenberg, (Eds.) *Handbook of child psychology, Vol. 3: Social, emotional and personality development* (pp. 719-788). Wiley. ISBN-10: 0471272906

Education Week (2022, January 5). *School shootings this year: How many and where*. https://www.edweek.org/leadership/school-shootings-this-year-how-many-and-where/2022/01

Gari, A., Kalantzi-Azizi, A. & Mylonas, K. (2000). Adaptation and motivation of Greek gifted pupils: exploring some influences of primary schooling. *High Ability Studies, 11*(1), 55-68. https://doi.org/10.1080/713669173

Georgas, J., & Berry, J. W. (1995). An ecocultural taxonomy for cross-cultural psychology. *Cross Cultural Research, 29*(2), 121-157. https://doi.org/10.1177/106939719502900202

Georgas, J., & Mylonas, K. (2006). Cultures are Like All Other Cultures, Like Some Other Cultures, Like No Other Culture. In U. Kim, K.-S. Yang, & K.-K. Hwang (Eds.), *Indigenous and cultural psychology: Understanding people in context* (pp. 197–221). Springer Science and Business Media. https://doi.org/10.1007/0-387-28662-4_9

Georgas, J., Mylonas, K., Gari, A. & Panagiotopoulou, P. (2004). Families and Values in Europe. In W. Arts & L. Halman (Eds.), *European values at the end of the millennium*, (pp. 167-204). Leiden: Brill Academic Publishers. ISBN: 90-75001-46-0.

Global Drug Survey. https://www.globaldrugsurvey.com/wp-content/uploads/2021/01/GDS2020-Executive-Summary.pdf

Graziano, A. M., & Raulin, M. L. (1989). *Research methods: A process of inquiry*. Harper and Row Publishers, Inc. ISBN-10: 0060424796

Guilford, J. P. (1956). *Psychometric methods.* McGraw-Hill.

Hareli, S., Kafetsios, K., & Hess, U. (2015). A cross cultural study on emotional expression and the learning of social norms. *Frontiers in Psychology, 6*, Article 1501. https://doi.org/10.3389/fpsychg.2015.01501

Hayes, A. F. & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But… *Communication Methods and Measures, 14*(1), 1-24, https://doi.org/10.1080/19312458.2020.1718629

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1988). *Applied statistics for the behavioral sciences* (2nd ed.). Houghton Mifflin Company. ISBN: 0395369118.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. https://doi.org/10.1080/10705519909540118

Kafetsios, K., & Hess, U. (2013). Effects of activated and dispositional self-construal on emotion decoding accuracy. *Journal of Nonverbal Behavior, 37*(3), 191–205. https://doi.org/10.1007/s10919-013-0149-x

Kline, P. (1993). *The handbook of psychological testing*. Routledge. ISBN: 0415054818.

Lawrence, C. (2006). Measuring individual responses to aggression-triggering events: Development of the situational triggers of aggressive responses (STAR) scale. *Aggressive Behavior*, *32*, 241-252. https://doi.org/10.1002/ab.20122

Lawrence, C., & Hodgkins, E. (2009). Personality influences on interpretations of aggressive behavior: The role of provocation sensitivity and trait aggression. *Personality and Individual Differences, 46*(3), 319-324. https://doi.org/10.1016/j.paid.2008.10.022

Lawrence, C., & Hutchinson, L. (2013a). The influence of individual differences in sensitivity to provocations on provoked aggression. *Aggressive Behavior, 39*(3), 212-221. https://doi.org/10.1002/ab.21473

Lawrence, C., & Hutchinson, L. (2013b). The impact of non-aggressive behaviour early in aggressive interactions: Sex differences in direct and indirect aggression in response to provocation. *British Journal of Psychology, 105*(1), 127-144. https://doi.org/10.1111/bjop.12020

Mardia, K. V., Kent, J. T., & Bibby, J. M. (1989). *Multivariate Statistics*. Academic Press. ISBN: 9780124712522

Mylonas, K., Lawrence, C., Zajenkowska, A., & Bower Russa, M. (2017). The Situational Triggers of Aggressive Responses scale in five countries: Factor structure and country clustering solutions. *Personality and Individual Differences, 104*(1), 172-179. https://doi.org/10.1016/j.paid.2016.07.030

Mylonas, K., Veligekas, P., Gari, A., & Kontaxopoulou, D. (2012). Development and Psychometric Properties of the Scale for Self-Consciousness Assessment. *Psychological Reports: Measures and Statistics, 111*(1), 233-252. https://doi.org/10.2466/08.02.07.PR0.111.4.233-252

O'Connor, D. B., Archer, J., Wu, F. W. C. (2001). Measuring aggression: Self-reports, partner reports and responses to provoking scenarios. *Aggressive Behavior 27*(2), 79-101. https://doi.org/10.1002/ab.2

O'Dea, C. J., Chalman, S. T., Castro Bueno, A. M., & Saucier, D. A. (2018). Conditional aggression: Perceptions of male violence in response to threat and provocation. *Personality and Individual Differences, 131,* 132-141. https://doi.org/10.1016/j.paid.2018.04.029

Papazoglou, S., & Mylonas, K. (2017). An examination of Alternative Multidimensional Scaling Techniques. *Educational and Psychological Measurement, 77*(3), 429-448. https://doi.org/10.1177/0013164416661823

Penny, K. I. & Jollife, I. T. (2001). A comparison of multivariate outlier detection methods for clinical laboratory safety data. *Journal of the Royal Statistical Society. Series D (The Statistician), 50*(3), 295-308. http://www.jstor.org/stable/2680933

Robinson, M. D., Traurig, E. & Klein, R. J. (2020). On looking versus leaping: A situation multilevel approach to trait anger and the anger-aggression relationship. *Personality and Individual Differences, 164*, Article 110130. https://doi.org/10.1016/j.paid.2020.110130

Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed.). Routledge. ISBN: 978-1-84169-890-8

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (4th ed.). Allyn & Bacon. ISBN-10: 0321056779

Thurstone, L. L. (1947). *Multiple factor analysis: A development and expansion of the Vectors of the mind.* The University of Chicago Press.

Veroude, K., Zhang-James, Y., Fernandez-Castillo, N., Bakker, M. J., Cormand, B., Faraone, S. V. (2016). Genetics of aggressive behavior: An overview. *American Journal of Medical Genetics. Part B, Neuropsychiatry and Genetics. 171B*, 3-43. https://doi.org/10.1002/ajmg.b.32364

Winer, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). McGraw-Hill. ISBN-10: 0070709815

Zajenkowska, A. & Rajchert, J. (2020). How sensitivity to provocation shapes encoding and interpretation of ambivalent scenes in an eye tracking study. *Journal of Cognitive Psychology, 32*(2), 1-19. https://doi.org/10.1080/20445911.2020.1717498

Zajenkowska, A., Zajenkowski, M., & Jankowski, K. (2015). The relationship between mood experienced during an exam, proneness to frustration and neuroticism. *Learning and Individual Differences. 37*, 237-240. https://doi.org/10.1016/j.lindif.2014.11.014

Zajenkowska, A., Jankowski, K., Mylonas, K., & Rajchert, J. (2015). Coffee consumption and propensity to experience aggressive feelings in provoking and frustrating situations. *Current Issues in Personality Psychology*, *3*(2), 105-111. https://doi.org/10.5114/cipp.2015.52890

Zajenkowska, A., Mylonas, K., Lawrence, C., Konopka, K., & Rajchert, J. (2014). Cross-cultural sex differences in situational triggers of aggressive responses. *International Journal of Psychology, 49(*5), 355-363. https://doi.org/10.1002/ijop.12052

Table 1. Item-level basic descriptive statistics in the STAR^{GR} scale

N = 1,094	mean	s.d.	median	*IQR*
1. A friend betrays me	3.43	1.15	4	1
2. I am the subject of a practical joke	2.35	1.05	2	1
3. People I live with show a lack of consideration	3.37	1.18	4	1
4. Someone steals from me	3.88	1.10	4	2
5. I feel frustrated	2.87	1.17	3	2
6. Someone insults me	3.76	1.09	4	2
7. I have academic problems	2.94	1.19	3	2
8. I experience family dispute	3.17	1.22	3	2
9. I feel hot and crowded	2.60	1.26	3	3
10. Someone ignores me	3.03	1.16	3	2
11. Someone behaves in an inconsiderate manner towards me	3.60	1.13	4	1
12. I am in pain	2.73	1.23	3	2
13. I am goaded or provoked by someone	3.33	1.08	3	1
14. I have been let down by someone	2.82	1.18	3	2
15. I feel stressed	2.99	1.21	3	2
16. Someone is drunk and behaves inconsiderately towards me	2.62	1.30	3	3
17. I hear a noise I cannot control	2.49	1.19	2	2
18. I am frustrated with services	2.81	1.12	3	2
19. Other around me are becoming aggressive	3.09	1.14	3	2
20. Someone makes offensive remarks to me	3.70	1.11	4	1
21. Another driver commits a traffic violation	3.15	1.15	3	2
22. I argue with a friend	2.90	1.08	3	2

***Note**: IQR=InterQuartile Range*

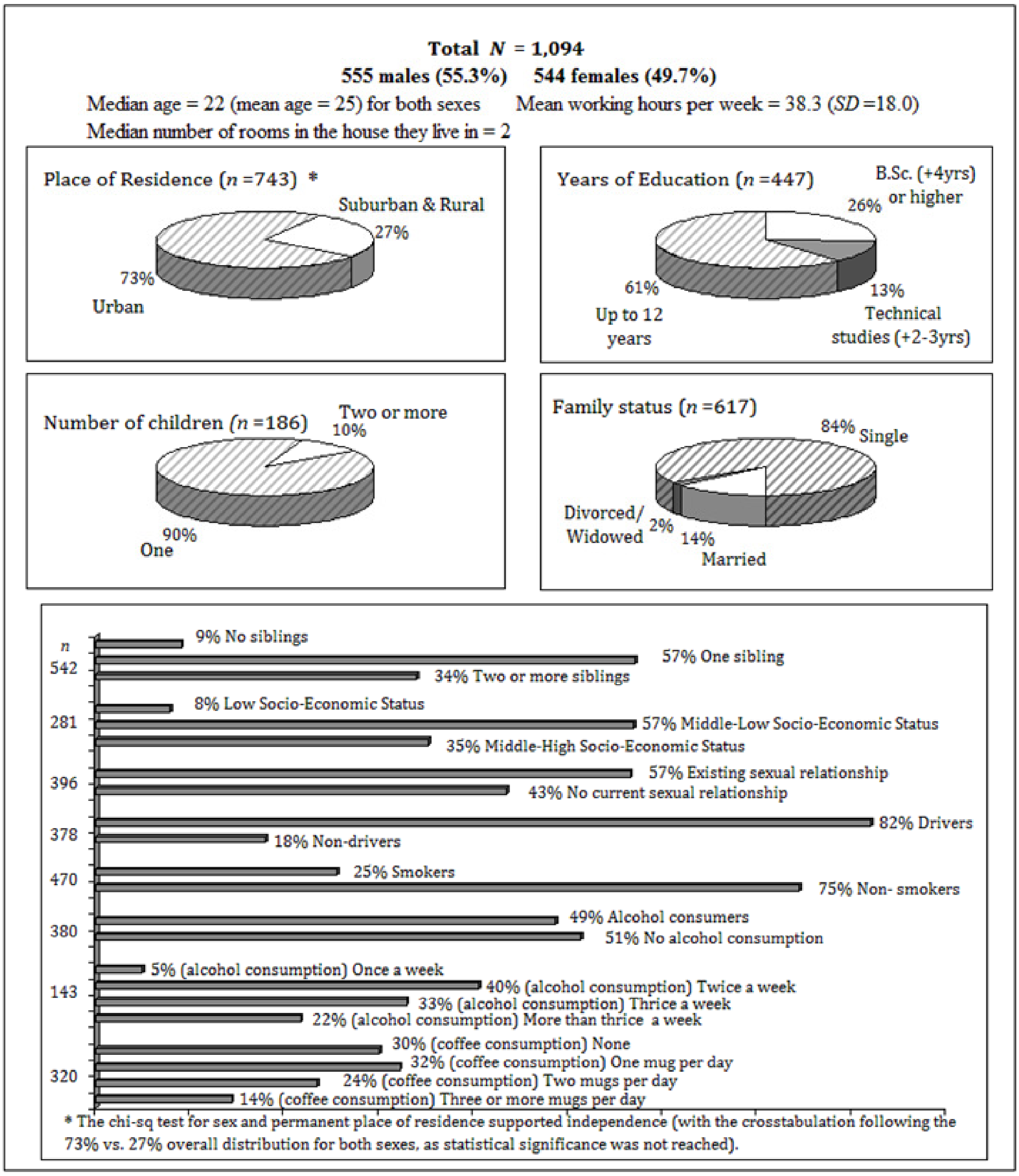

Figure 1. Sample demographics

Table 2. CFA Solutions on *Pearson's r* indices (*N*=1,094)

Models	χ^2	*df*	$\chi^2 \div df$	RMSEA [CI 90 %]	RMR	GFI	AGFI	CFI	AIC	$\Delta\chi^2$	Δdf	TLI
a	17,027.60	231	73.71	-	-	-	-	-	17,071.60	-	-	-
b	1,904.88	209	9.11	.086 [.083 - .090]	.061	.86	.83	.86	1,992.88	[a-b]15,122.72	22	.888
c	1,422.30	208	6.84	.073 [.070 - .077]	.058	.89	.87	.94	1,512.30	[b-c]482.58	1	.281
d	956.56	194	4.93	.062 [.058 - .065]	.050	.92	.90	.96	1,114.56	[b-d]948.32	15	.516

Note: *TLI* for c-d = .327 ($\Delta\chi^2$ = 466.74, *df*=14, p<.01). All χ^2 and $\Delta\chi^2$ values are statistically significant at the .001 level.

KEY: *a* = independence model, *b* = unifactorial solution, *c* = two-factor solution, *d* = two-factor solution with 14 estimated error covariances strictly within factors. This estimation was empirically based and reflects possible cultural fluctuations. The impact on model improvement and the transitivity rules were employed. For all models, Maximum Likelihood estimates were computed via LiSRel 8.30.

Table 3. CFA Solutions on *Kendall's Tau-b* indices (*N*=1,094)

Models	χ^2	*df*	$\chi^2 \div df$	RMSEA [CI 90 %]	RMR	GFI	AGFI	CFI	AIC	$\Delta\chi^2$	Δdf	TLI
a	11,437.08	231	49.51	-	-	-	-	-	11,481.08	-	-	-
b	1,256.22	209	6.01	.068 [.064 - .071]	.053	.91	.89	.93	1,344.22	$^{a\text{-}b}$10,180.86	22	.897
c	944.27	208	4.54	.057 [.053 - .061]	.050	.93	.91	.94	1,034.27	$^{b\text{-}c}$311.95	1	.294
d	857.56	206	4.16	.054 [.050 - .058]	.047	.93	.92	.95	951.56	$^{b\text{-}d}$398.66	3	.369

Note: TLI for a-c = .927 and for c-d .106 ($\Delta\chi^2$ = 86.71, *df*=2, *p*<.01). All χ^2 and $\Delta\chi^2$ values are statistically significant at the .001 level.
KEY: *a* = independence model, *b* = unifactorial solution, *c* = two-factor solution, *d* = two-factor solution with 2 estimated error covariances strictly within factors. This estimation was empirically based and reflects possible cultural fluctuations. The impact on model improvement and the transitivity rules were employed. For all models, Maximum Likelihood estimates were computed via LiSRel 8.30.

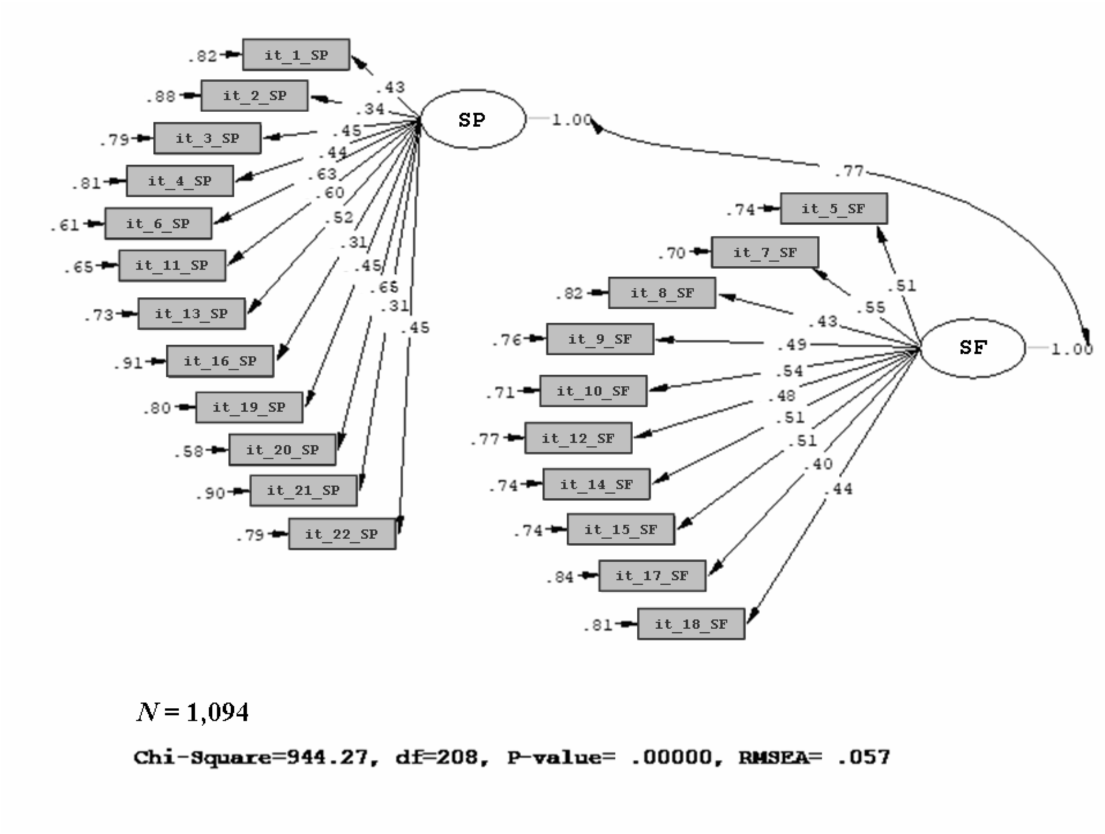

Figure 2. Confirmatory Factor Analysis STAR^{GR} outcomes: *Kendall's Tau-b* indices, two-factor model (unmodified).

Table 4. Descriptive statistics, tests of normality for the two factor aggregate scores, and reliability estimates

N = 1,094	Sensitivity to Provocation (SP)	Sensitivity to Frustration (SF)
Mean	3.265	2.846
Standard Deviation	.636	.714
Median	3.33	2.90
Range	3.75	3.90
Skewness	–.619	–.191
Kurtosis	.496	–.188
Kolmogorov-Smirnov z *	.074 (1,094), *p*<.001	.061 (1,094), *p*<.001
Shapiro-Wilk W	.976 (1,094), *p*<.001	.992 (1,094), *p*<.001
Cronbach's alpha	.80	.80
McDonald's omega	.81	.80

* Lilliefors significance correction

Table 5. Target rotated solution on the UK (Author2, 2006) factor structure

		SP	*SF*
1SP	A friend betrays me	**.43**	.02
2SP	I am the subject of a practical joke	*.34*	.01
3SP	People I live with show a lack of consideration	**.45**	.02
4SP	Someone steals from me	**.44**	.02
5SF	I feel frustrated	-.02	**.51**
6SP	Someone insults me	**.63**	.02
7SF	I have academic problems	-.02	**.55**
8SF	I experience family dispute	-.02	**.43**
9SF	I feel hot and crowded	-.02	**.49**
10SF	Someone ignores me	-.02	**.54**
11SP	Someone behaves in an inconsiderate manner towards me	**.60**	.02
12SF	I am in pain	-.02	**.48**
13SP	I am goaded or provoked by someone	**.52**	.02
14SF	I have been let down by someone	-.02	**.51**
15SF	I feel stressed	-.02	**.51**
16SP	Someone is drunk and behaves inconsiderately towards me	*.31*	.01
17SF	I hear a noise I cannot control	-.02	**.40**
18SF	I am frustrated with services	-.02	**.44**
19SP	Other around me are becoming aggressive	**.45**	.02
20SP	Someone makes offensive remarks to me	**.65**	.03
21SP	Another driver commits a traffic violation	*.31*	.01
22SP	I argue with a friend	**.45**	.02

Proportionality coefficient for SP and SF factors: .90 and .91, respectively.
Square root of the mean squared difference per factor, .19 and .18.

Note: Although the target loadings are the same with the CFA ones, through this approach the error loadings are now also available (full factor matrix) for across-solutions comparisons.

Table 6. Tucker's *φ* coefficients across a span of solutions

Compared solutions		Author2 (2006)	
		SP	SF
2017 study	SP	**.91**	.32
(Greek data only)	SF	.37	*.89*
2017 study	SP	**.92**	.32
(All countries)	SF	.38	**.90**
2017 study	SP	**.91**	.31
(Cluster 1 countries)	SF	.38	**.90**
2017 study	SP	**.92**	.33
(Cluster 2 countries)	SF	.38	*.89*
Current study	SP	**.91**	.32
(Model c, Tau-b)	SF	.37	**.90**
Current study	SP	**.90**	.33
Target-Rotated Solution	SF	.36	**.91**

Table 7. Summary of Psychometric indices for the two STAR^{GR} dimensions

	SP	SF		SP	SF
Mean	3.265	2.846	Standard Deviation	.636	.714
5% trimmed mean	3.290	2.853	Inter-Quartile Range (observed)	.83	.90
Huber's M-estimator	3.322	2.877	Range (observed)	3.75	3.90
Median	3.33	2.90	Range (max-scale, theoretical)	4	4
Mode	3.50	3.10	Standard error of the Mean	.019	.022
min. max.	1.00-4.75	1.00-4.90	Low Confid. Limit mean (*a*=.05)	3.228	2.415
N	1,094	1,094	High Confid. Limit mean (*a*=.05)	3.302	3.278
Cronbach's alpha	.80	.80	Standard error of the estimate	.284	.319
McDonald's omega	.81	.80	Estimated Range (score)	1.12	1.26

Table 8. STAR^{GR} Norms Table (*for distribution extremes*) **(*N* = 1,094)**

SP (raw scores range: 12 to 60 [or 1.00 to 5.00])			**SF** (raw scores range: 10 to 50 [or 1.00 to 5.00])		
T-score	Raw score (simple sums)	Raw score (averaged aggregate)	**T-score**	Raw score (simple sums)	Raw score (averaged aggregate)
20	12 to 16	1.000 to 1.333	**20**	-	-
21	17	1.417	**21**	-	-
22	18	1.500	**22**	-	-
23	-	-	**23**	-	-
24	19	1.583	**24**	10	1.00
25	20	1.667	**25**	-	-
26	21	1.750	**26**	11	1.10
27	-	-	**27**	12	1.20
28	22	1.833	**28**	13	1.30
29	23	1.917	**29**	-	-
30	24	2.000	**30**	14	1.40
31	25	2.083	**31**	15	1.50
32	-	-	**32**	-	-
33	26	2.167	**33**	16	1.60
34	27	2.250	**34**	17	1.70
35	28	2.333	**35**	18	1.80
36	-	-	**36**	-	-
	T-scores 37 to 49 are omitted			*T-scores 37 to 49 are omitted*	
50	39	3.250	**50**	29	2.90
	T-scores 51 to 63 are omitted			*T-scores 51 to 63 are omitted*	
64	50	4.167	**64**	-	-
65	51	4.250	**65**	39	3.90
66	-	-	**66**	40	4.00
67	52	4.333	**67**	-	-
68	53	4.417	**68**	41	4.10
69	54	4.500	**69**	42	4.20
70	-	-	**70**	43	4.30
71	55	4.583	**71**	-	-
72	56	4.667	**72**	44	4.40
73	57	4.750	**73**	45	4.50
74	-	-	**74**	-	-
75	58	4.833	**75**	46	4.60
76	59	4.917	**76**	47	4.70
77	60	5.000	**77**	48	4.80
78	-	-	**78**	-	-
79	-	-	**79**	49	4.90
80	-	-	**80**	50	5.00

Note: *For further information on acquiring the Greek items and on how to employ the STAR^{GR} scale and its Norms, please contact the first author's e-mail.*