2009

Reducing Bias in Cross-Cultural Factor Analysis through a Statistical Technique for Metric Adjustment: Factor Solutions for Quintets and Quartets of Countries

Kostas Mylonas
The University of Athens, kmylonas@psych.uoa.gr

Reducing Bias in Cross-Cultural Factor Analysis through a Statistical Technique for Metric Adjustment: Factor Solutions for Quintets and Quartets of Countries

Kostas Mylonas, *Department of Psychology, The University of Athens, Greece,* kmylonas@psych.uoa.gr

Differential item functioning or item bias is a usual threat in psychological research and many experts in the field such as Kline (1993), Nunnally and Bernstein (1994), and others have suggested various methods for its detection and removal. Item bias *in terms of culture* has been addressed by Poortinga, Van de Vijver, Leung, Muthén, and others, with most of the proposed methods attempting to minimize variance explained by culture itself by detecting and deleting culturally-biased items from the analysis. This is done through the detection of inequivalent –in terms of factor structure– items and their elimination before comparing for factor equivalence using "culture-free" correlation matrices. The currently proposed alternative statistical technique is to estimate indices of country eccentricity through individual differences Euclidean distance scaling models for the respective weirdness index to be computed and then proceed with variance and raw-scores adjustment, accounting for each item's variance caused in terms of culture only. Analyzing the before-adjustment and the after-adjustment correlation matrices through covariance structure analysis and Procrustean rotations, we might be able to evaluate the profits from the adjustment itself. The illustration data for this method were 20 measures from six countries (N=1,655) as presented in the Georgas et al. (2006) 30-nation study on family issues. Following the initial analysis of all six countries, all quintet and quartet combinations of countries were examined and evaluated for the before and after adjustment outcomes. For all 22 combinations, factor loadings after metric adjustment were slightly higher; also, for a large number of country sets, items not included in the factor structure before the adjustment were now entering it. Factor identities were not affected to a large extent, but the factor identities were somewhat altered, revealing interesting facets within each of the broader areas defined by the initial –before the raw score adjustment– factors.

Many types of bias such as method bias, content or construct bias, and item bias have been described in the methodology and psychometrics literature in respect to the extraneous variables causing them and in terms of detecting them (e.g., Anastasi, 1990; Christensen, 1988). Item bias or Differential Item Functioning (DIF) has been given wide attention, as in most cases it is the item parameters that differ across the subgroups in a sample (Nunnally & Bernstein, 1994). Methods and statistical techniques have also been proposed to neutralize or avoid such effects which silently and seriously threat internal validity (Christensen, 1988; Kline, 1993). For Kline (1993), evidence of test bias can be found through analysis of variance, internal consistency estimates, standard item analyses and Item Response Theory models, and also through exploratory factor analysis, with comparable solutions for each of the groups under consideration. Van de Vijver and Leung (1997) have focused on these sources of bias, seeking their detection and their possible avoidance in favor of "equivalence", a notion with large implications especially under the cross-cultural research scope.

One major cause of bias in cross-cultural research is culture itself. Poortinga (1989) has discussed several ways of dealing with the artefacts caused by "bias in terms of culture". Their elimination from cross-cultural comparisons has also been attempted through various ways proposed by Poortinga and other theorists. In a satisfactory cross-cultural study there is no variance left to be explained in terms of culture (Poortinga & Van de Vijver, 1987) and cultural variance should be reduced to zero to derive comparable measures and cross-culturally meaningful structures. Thus, inequivalence in cross-cultural studies can be mainly attributed to variance in terms of culture, which has to be reduced to null for the factor structures to be meaningful. Following these initial theses, and with respect to any quantification, differences in scores between cultural groups can reflect valid differences in the construct measured or they can result –at least partially– from measurement artefacts or bias.

It has been supported that removal of item bias does not necessarily lead to scalar equivalence and that bias in general, cannot be merely reduced to item bias, and that a biased item can be treated as a disturbance at the item level that has to be removed (Van de Vijver & Leung, 1997). However, removal of items can easily affect the validity levels of a scale (i.e., if too many items are removed, even content validity can be threatened). In order to circumvent such a problem, a number of statistical methods have focused on bias detection and in some cases on bias elimination in order to achieve invariant scales across cultures. Psychometric-statistical methods may be used, e.g., merging the confounding variables in the design of the study (Poortinga & Van de Vijver, 1987) followed by a covariance or hierarchical regression analysis. Valencia, Rankin, and Livingston (1995) tried to account for cultural variance by controlling for age, gender and ability for an intelligence test through partial correlation coefficients; they found more than 50% of the items to be biased. Such an attempt though simply detects the source of bias but does not *per se* remedy for it. A methodological approach is to circumvent the bias effect by controlling for external criteria, such as gender, age, ability and other confounding variables, specific in each culture. Other approaches have focused on actually reducing bias by aggregating countries in terms of their common characteristics, such as ecosocial indices (Georgas & Berry, 1995), or of information contained in items of correlate measures collected for the same samples (Gari, Panagiotopoulou, & Mylonas, 2009; Mylonas, Gari, Panagiotopoulou, Georgiadi, Valchev, & Papazoglou, in review).

Yet, if a factor equivalent structure for a set of countries is the target, and if some biased items may be threatening this equivalence, then –reversing the argument– the items to be factor analyzed could also be themselves a possible source of bias estimation. That is, for a set of items across countries, we would ideally compute an index which might contain information about the variance explained only in terms of culture. Thus, such an approach would account for a specific variance component by estimating –for a set of items– the amount of variance caused by "culture" from the information contained in these same items and not by using external measures (such as control variables).

Along these lines of using item-inherent information, certain techniques may be applied which can multivariately gather variance-component information from each item in a scale. Such statistical techniques –amongst others– could be factor analysis and/or multidimensional scaling. Factor analysis in cross-cultural psychology is directly linked to data equivalence which is the key-issue for this analysis in the sense that construct bias may confound factor structures, as computed for more than one cultures. Kline (1993, p. 167) has argued that one criterion for the absence of differential item functioning is that "In factor analyses of the items the factor loadings of the items should be the same for both groups, within the limits of the standard errors." Van de Vijver and Poortinga (2002) extended the Covariance Structure Analysis proposed by Muthén (1994; 2000) into factor analysis in order to compute factor equivalent solutions –if possible. Although this method is effective in terms of acceptable equivalence levels, this is accomplished most of the times in the expense of erasing non-equivalent or biased items from the final structure.

Multidimensional scaling on the other hand has been widely used outside the cross-cultural research borders, but has also been employed to model cross-cultural similarities and differences; one striking example is the Schwartz studies (e.g., 1992; 1994; 2006; 2007) on values for the last 15 years or so. Even more, sophisticated ways to model similarities and differences simultaneously are available, such as the individual differences Euclidean distance model, through which we can compute as many as possible underlying dimensions for a set of countries and at the same time compute the relative importance of these dimensions for each country, in terms of dimension weights. A "weirdness index" computed for each country is also available, which corresponds to the proportionality of the individual dimension weights to the overall average weights, thus depicting the eccentricity of each country's similarity matrix in respect to the overall dimensions in the data. However, the *ALSCAL* algorithms (e.g., MacCallum, 1981) usually employed in this procedure are calculated on proximities (Everitt, 1996), which reflect similarities or dissimilarities among measures and not correlations, thus the solutions derived cannot really satisfy the construct equivalence criterion (Kline, 1993; Van de Vijver & Leung, 1997).

Any of these two statistical techniques would describe the multivariate properties for a number of scale items from two different statistical perspectives, but this was not the aim in the current attempt; we were in the specific need of finding some way of estimating the bias-in-terms-of-culture component in a dataset of two or more country samples. If we accept bias in terms of culture as "unwanted" variance, since it leads to inequivalence, the first thing that comes to mind is the usual approach in overriding unwanted variance; that is, partial correlation or similar-derivative statistical computations, such as covariance analysis or variations of structural equation modeling or multiple hierarchical regression modeling. So the basic aim would be to assess this bias in the simplest way possible and remove it from the original raw measures.

A rather simple method for partialling out error variance in any correlational study is the standard deviation adjustment method for a variable *y*. The method is a bivariate one with *y* and *x* measures assessed at least at the interval level of measurement. For these measures, the correlation coefficient (Pearson r) can be computed and then squared to satisfy formula (1).

$$s = \sqrt{s^2 \quad s^2 r^2} \tag{1}$$

The above rationale (Paraskevopoulos, 1990, p. 172) can be derived directly from the bivariate normal distribution model (Winer, 1971, p. 765) with its properties (mean and standard deviation, respectively) calculated through (2) and (3) for the distribution of *y*.

$$_{y_j|x} = \ _{y_j} + \ (X \quad _{x_j}) \tag{2}$$

$$^{2}_{y_j|x} = \ ^{2}_{y}(1 \quad ^{2}) \tag{3}$$

Through formula (1), an adjusted standard deviation for the target measure can be computed and then employed to compute adjusted raw scores, through formulae (4) and (5).

$$z = \frac{X \quad \overline{X}}{s} \tag{4}$$

$$X = zs + \overline{X} \tag{5}$$

The adjusted target measure retains its original mean but its variance has been "relieved" from the *x* variable effect.

Earlier cultural and cross-cultural studies conducted by the author and his colleagues (Gari, Panagiotopoulou, & Mylonas, 2009; Sidiripoulou, Mylonas & Argyropoulou, 2008; Veligekas, Mylonas, & Zervas, 2007), have incorporated multidimensional scaling in its traditional form to model measurement scales by employing a two-dimension trigonometric transformation proposed by the author (Mylonas, 2009). An extended method has been employed in the current study, on a set of six countries and a scale of 20 items assessing father's roles in the family.[1] This extended method has been specifically introduced only in the 2009 article, discussing a part only of the overall outcomes to be presented hereon (as an illustration of the proposed method and as applied to the six-country set only). The overall method proposed is a combination of a) factor analysis, b) multidimensional scaling in its traditional *ALSCAL* form but also employing the individual differences Euclidean distance model and the resulting "weirdness" index, and c) covariance

[1] I would like to thank Professor Emeritus James Georgas of the University of Athens and his colleagues for the kind permission they gave me to analyze parts of the 30-nation family-study data set (for technical reasons three countries could not be analyzed). The entire project has been authored by Georgas, Berry, Van de Vijver, Kagitcibasi, & Poortinga (2006); this volume was awarded the 2007 *Ursula Gielen Global Psychology Book Award*, American Psychological Association, Division 52.

structure analysis as extended to factor analysis by Van de Vijver and Poortinga (2002). The combination of individual differences scaling solution indices with factor analytic techniques should first address the similarity/dissimilarity matrix across the variables of interest taking into account the country identity for each of the cases. If eccentricity in the similarity matrices computed for each of the countries separately could be depicted in terms of variance, then we could correct for this "noise" and then compare the factor analysis outcomes before and after this correction. The eccentricity in this case is represented by the weirdness index, to be used as a multiple r^2 index and the adjustment is the standard deviation one, in order to arrive into adjusted raw scores within each country and for each case separately. For each stage (before and after adjustment) the Muthén method, as extended to factor analysis by Van de Vijver and Poortinga, should be employed to arrive at the best possible "universal" factor solutions across countries. The factor solutions across the two stages should then be compared. In the current study, this procedure was conducted for six countries (the procedure itself, its statistical rationale and its outcomes have been extensively presented in Mylonas, 2009) and then to test for the method's potential in reducing bias in terms of culture, the procedure was generalized to all quintet and quartet combinations of these countries.

To briefly summarize, in the present study we focused on bias in terms of culture by employing two statistical techniques: factor analysis and multidimensional scaling for a scale of 20 items. These items assessed father's family roles for a set of six countries, and the statistical procedures on bias detection and removal were conducted for this six-country set, and for all its subsets of five countries followed by all its subsets of four countries.

Measures and procedure

The measures used in the current illustration of the proposed method are 20 items regarding paternal roles in the family such as provision of emotional support, keeping the family united, protecting the family, resolving disputes, managing the finances, etc. (see also Table 1). The responses were provided by university students in respect to their own families. The measurement scale was a 6-point one with 6 meaning that father undertakes the role within the family.

Overall sample and sample combinations

We initially selected six countries out of the 27 available ones on the basis of their geographic location. First we selected six such geographic locations, representing different and distant parts of the world. We randomly selected one country from the specific countries composing each of these six initially selected globe regions. Greece was an exception, as it was selected for the illustration in a non-random way. The final six countries selected were Greece, Georgia, USA, Germany, Indonesia and Pakistan, with a total number of respondents of 1,655. Affluent and non-affluent countries are included in this set; a number of religious denominations are also present; finally, "western" and "eastern" countries are also represented in the set. Apart for representing basic socio-economic and geographic properties in the set, the main aim in selecting these countries was to employ as diverse countries as possible for the illustration, either at the unit level, or at the subset level (i.e., the USA & Germany subset can be considered diverse to the Georgia, Pakistan –and possibly Indonesia– subset). The countries were first analyzed as a set of six and then, for the second stage of the analysis, all "quintets" and "quartets" of countries were employed and tested for the proposed method.

Statistical procedures, Results and Discussion

The basic input for the analyses were the actual individual level country data, as follows: Greece N=350, Georgia N=200, USA N=263, Germany N=153, Indonesia N=239, and Pakistan N=450. The countries were not weighted for sample fluctuations because the overall multidimensional scaling solution was of no interest. However, one could of course employ country weights to remedy for possible sampling over- / under-representations (in terms of N).

The first step in the analysis was to compute the means for each of the 20 father's roles for each country and plot them in order to be certain that at least some variability existed in the data (Figure 1). That is, if selection of countries was to satisfy as much stringent testing as possible, then there should be considerable range for the means across countries for each item. The country profiles showed that there is large variability for most of these roles and little for a few of them. The patterns for those means are similar for most countries participating in the overall 30-country study. As the procedure described hereon has been applied to the six-country data set and separately to each "quintet" and "quartet" of countries, the outcomes of the analysis were 22 recalculations for all combinations of countries (six-country set inclusive). One example from the "quintet" combinations and one example from the "quartet" combinations will be outlined, along with the six-country set outcome, but the summary of all combinations' outcomes will be, of course, outlined and discussed later.

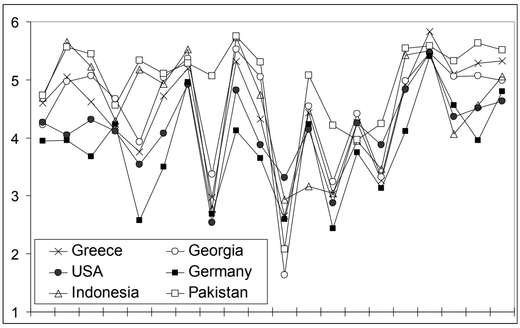

Figure 1. Means across countries.

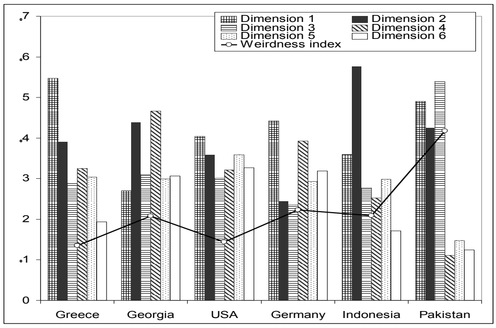

Figure 2. Country dimension weights (6-dim. MDS solution) and weirdness index levels for each country.

The next step was to factor analyze the data, through the Muthén method (1994, 2000) as extended to exploratory factor analysis by Van de Vijver and Poortinga. The pooled within and the between groups correlations matrices produced by the Muthén algorithm were factor analyzed and a Procrustean rotation followed. A 3-factor structure, produced without any adjustment of the raw data, was retained (Table 1) for further comparisons with the second stage factor structure, which would follow after the intervention-adjustment would have taken place; thus 22 factor structures, one for each combination (six-country set, quintets, quartets) of countries were available *before* adjustment.

The intermediate stage in the analysis was to compute an individual differences Euclidean distance scaling solution for each of the 22 combinations (six-country set, quintets, quartets) and calculate the respective weirdness indices based on six dimensions in all recalculations. Plots and tables in the same fashion with Figures 1 and 2 and with Table 1, were produced during each recalculation for each of the combinations of countries but obviously are not reported here.

The adjustment stage was based on the computation of *z*-scores for the raw data and the recalculation of raw scores based on adjusted standard deviations. The adjustment for standard deviations was carried out through formulae (1), (4), and (5). In formula (1) r^2 is replaced by the weirdness index; this is reasonable, since this index can be considered as variance explained by the eccentricity of the similarity matrix per country. Thus, its effect can be removed from the original raw scores by adjusting for the standard deviation of each item within each country. However, we must note here that the statistical assumptions underlying this substitution of the r^2 statistic have not yet been defined. The means of the original raw scores were of course not affected by the intervention, but, as intended, the standard deviations were altered, producing adjusted correlation coefficients and therefore, an adjusted factor structure was expected for the same pool of items. The standard deviations for the six countries after the adjustment, in respect to their difference from the non-adjusted ones, showed that for some countries and some items, the adjustment was rather large, although in other instances these adjustments were trivial. This indication implied that for the second stage factor analysis, the adjustments of the correlation coefficients in the matrices employed could vary in magnitude.

Table 1. Unadjusted and adjusted raw score factor solutions for the six-country set

Paternal role items. *Father…*	F_1	F_2	F_3	F_1'	F_2'	F_3'
provides emotional support	**.68**	.19	.14	**.69**	.16	.20
keeps the family united	**.77**	.31	.00	**.77**	.32	–.01
keeps a pleasant environment	**.77**	.25	.14	**.78**	.28	.10
conveys traditions to children	**.57**	.01	.38	**.57**	.04	.33
conveys religion to children	**.65**	.14	.13	**.66**	.22	.02
preserves family relations	**.69**	.26	.15	**.69**	.31	.11
supports grandparents when in need	**.53**	.26	–.05	**.55**	.23	.02
takes care of grandparents (cooking, shopping)	.27	.27	.14	.32	**.44**	.05
protects the family	**.63**	**.47**	–.04	**.66**	***.44***	–.03
resolves disputes	**.50**	**.40**	.04	***.49***	**.48**	–.04
does housework	.01	–.15	**.55**	.03	–.31	**.61**
does the shopping, pays bills, etc.	–.03	**.55**	**.55**	–.05	**.61**	***.51***
takes children to school	.22	.22	**.60**	.22	.33	***.47***
plays with children	**.39**	.04	**.57**	**.41**	.01	**.62**
helps children with homework	.36	.14	**.65**	.36	.17	***.62***
teaches manners to children	**.64**	.35	.20	**.66**	.33	.22
contributes financially	16	**.65**	.03	.19	***.52***	.16
manages finances	.03	**.77**	.21	.03	**.77**	.22
gives pocket money to children	.29	**.67**	.18	.29	**.67**	.19
supports career of children	.36	**.53**	.13	.36	***.50***	.19

Key: $F_{1\ to\ 3}$: factor structure before metric adjustment. F_1', F_2', F_3': factor structure after metric adjustment. All underlined loadings are higher in the after-adjustment solution. All italicized loadings (in the after-adjustment solution) are lower than the respective ones before adjustment.

The six-country factor structure after the adjustment was similar to the first stage structure *but for* two main differences. First, the item that did not load on any of the factors before the adjustment, was now a part of the second factor (shaded loading, Table 1). Another difference was that a large number of the loadings originally observed in the before adjustment factor structure were now higher for the adjusted factor structure (underlined in Table 1). Although the profit seemed small, it indicated that the factor structure, after the standard deviation adjustment contained clearer information than the one computed for the unadjusted data. The generalization of computations followed (21 recalculations for the 15 quartets and the six quintets).

One example of the “quartets” of countries is Greece, Germany, Georgia and Pakistan. For the before and after adjustment solutions, Tucker’s φ indices (Tucker, 1951) were calculated to assess the levels of factor agreement and three pairs of identical factors emerged. However, there seem to be quite a few changes before and after the adjustment (Table 2) mostly in the factor identities which might better clarify the final structure. For example, there is stronger emphasis in family cohesion and protection of grandparents and children in the second factor, whereas this was not so clear before the metric adjustment.

Another example, one of “quintets” of countries (involving Greece, Germany, United States, Indonesia and Pakistan) resulted into Tucker’s φ coefficients indicating factor equivalence for two out of three factors for the before-after adjustment solutions; again, there are quite a few changes in the factor identities, and this time they do not only clarify but they also alter the interpretation of the final structure. Thus, the “contributes financially” item is now a part of the non-traditional roles factor, and such a shift is allowing for an alternative interpretation of the specific factor. Also, the “conveys religious tradition to children” item is now a part of the first traditional roles factor only; other changes also exist between the two stages of analysis for this example.

Table 2. Unadjusted and adjusted raw score factor solutions: Quartet example (countries participating: Greece, Germany, Georgia, Pakistan)

Paternal role items. *Father…*	F_1	F_2	F_3	F_1'	F_2'	F_3'
provides emotional support	**.58**	.22	.18	**.40**	**.45**	.29
keeps the family united	**.71**	.31	.21	**.69**	**.42**	.19
keeps a pleasant environment	**.72**	.23	.31	**.75**	.31	.22
conveys traditions to children	**.44**	–.03	**.48**	**.39**	.03	**.45**
conveys religion to children	**.56**	.18	.38	**.79**	.08	.24
preserves family relations	**.63**	.23	.30	**.62**	.37	.22
supports grandparents when in need	**.48**	.28	–.08	.26	**.48**	.06
takes care of grandparents (cooking, shopping)	**.53**	.10	.18	**.71**	–.01	.17
protects the family	**.59**	**.48**	.06	**.68**	**.46**	.02
resolves disputes	**.55**	.31	.15	**.77**	.15	.01
does housework	–.24	.10	**.46**	**–.39**	.19	**.62**
does the shopping, pays bills, etc.	–.04	**.50**	**.51**	.29	.24	**.47**
takes children to school	.13	.20	**.68**	**.54**	–.02	**.48**
plays with children	.22	.08	**.58**	.08	.30	**.57**
helps children with homework	.22	.16	**.70**	**.46**	.02	**.61**
teaches manners to children	**.57**	.38	.30	**.68**	.35	.23
contributes financially	.04	**.73**	–.11	–.03	**.79**	–.03
manages finances	.07	**.72**	.14	.31	**.56**	.07
gives pocket money to children	.29	**.66**	.16	**.52**	**.55**	.10
supports career of children	.31	**.53**	.18	.35	**.54**	.17

Key: $F_{1\text{ to }3}$: factor structure before metric adjustment. F_1', F_2', F_3': factor structure after metric adjustment. All underlined loadings are higher in the after-adjustment solution. Shaded cells denote structural differences between the two solutions.

Table 3. Unadjusted and adjusted raw score factor solutions: Quintet example (countries participating: Greece, Germany, USA, Indonesia, Pakistan)

Paternal role items. *Father…*	F_1	F_2	F_3	F_1'	F_2'	F_3'
provides emotional support	**.68**	.22	.14	**.74**	.16	.26
keeps the family united	**.76**	.31	.01	**.81**	.34	–.04
keeps a pleasant environment	**.77**	.25	.16	**.81**	.34	.05
conveys traditions to children	**.56**	–.02	**.41**	**.51**	.12	.30
conveys religion to children	**.67**	.19	.10	**.71**	.38	–.15
preserves family relations	**.70**	.25	.17	**.72**	.34	.11
supports grandparents when in need	**.52**	.25	–.04	**.58**	.09	.16
takes care of grandparents (cooking, shopping)	.36	.32	.14	.31	**.69**	–.07
protects the family	**.64**	**.48**	–.02	**.75**	**.37**	.05
resolves disputes	**.51**	**.42**	.07	**.56**	**.56**	–.09
does housework	.01	–.18	**.56**	.04	**–.46**	**.57**
does the shopping, pays bills, etc.	–.05	**.54**	**.58**	–.10	**.69**	**.49**
takes children to school	.22	.23	**.58**	.28	**.58**	.19
plays with children	.37	.02	**.60**	.35	–.09	**.71**
helps children with homework	.37	.16	**.62**	.34	.32	**.45**
teaches manners to children	**.65**	.36	.18	**.72**	.33	.17
contributes financially	.14	**.63**	.01	.20	.25	**.41**
manages finances	.01	**.74**	.26	.02	**.71**	.36
gives pocket money to children	.27	**.67**	.19	.32	**.68**	.24
supports career of children	.35	**.55**	.14	**.39**	**.49**	.25

Key: $F_{1\text{ to }3}$: factor structure before metric adjustment. F_1', F_2', F_3': factor structure after metric adjustment. All underlined loadings are higher in the after-adjustment solution. Shaded cells denote structural differences between the two solutions.

The main outcomes for all 21 quintet and quartet combinations of countries stemming from the comparisons of the before and after metric adjustment factor structures can be summarized as follows: **a**) the mean difference in number of items participating in each structure (before and after the adjustment) was 1.57, that is, one to two additional items were on average part of the structure after the adjustment, whereas they were not present in the structure before the adjustment. In more detail, for 29% of the before-after pairs of factor structures, up to two items were "lost on the way", but in most cases (52% of the factor structure pairs) there was a "gain" of at least one and up to six items. **b**) The mean number of items with higher loadings in the "after" solution was 8.19, whereas the mean number of items with lower loadings in the "after" solution was only 3.05. **c**) Tucker's φ indices as computed for each before-after pair separately, indicate a 68% overall factor identity, a 21% factor similarity and a 11% factor dissimilarity. **d**) Finally, the mean loadings of items in each of the factors were significantly higher (as tested through Student's *t* criterion for related samples), especially for the first and strongest factor in each of the solutions.

Conclusion

Our conclusions drawn from this combinatorial "exercise" are promising, as the loadings for the items participating in the factor structures increase in all combined country sets after the raw score adjustment. This may initially sound like a statistical artefact, but we are able to refute this possibility in two ways: **a**) for a rather large number of items not loading on the factors before the adjustment, the loadings decrease after the adjustment, whereas, if there was a general artefact inflating the loadings, those loadings would have increased too, and **b**) since the interaction among the item error-terms is reduced *due to the adjustment* and *not due to random or extraneous reasons*, our outcomes are not a type I error, but on the contrary, power has been enhanced in the analysis by taking out as much "noise" as possible.

Another interesting conclusion is that the "Takes care of grandparents (cooking, shopping)" item loads on one of the factors after metric adjustment for eight country combinations and this is true for the six countries initial set as well. Thus, we might support that if we gathered data on these 20 paternal roles from five or four coutries 100 times, for 52 of these times we would "profit" with at least one item in the structure after having adjusted our standard deviations and correlations analyzed through the weirdness indices.

A final conclusion is that the meaning of the "after" factors is not the same in most country combinations. Thus, for the "before" structures the main factor identities are: **a**) Father's traditional family roles, **b**) Traditional financial responsibilities for the family, and **c**) Non-traditional roles. For the "after" structures the main factor identities are: **a**) Traditional roles along with his emotional involvement with children, **b**) Financial obligations along with father's care for the previous generation (grandparents) and for the next one (offspring), and **c**) Non-traditional roles, including the item "father contributes financially". Although the latter sounds like an *oxymoron*, is it not, as this paternal role of being solely responsible for the family finances may not be considered a "traditional" one anymore for many cultures. In support of this, and according to an untested author's opinion, "Contributes" may have been misinterpreted by some or many participants; the initial meaning assigned to it implied that father attends to *a part* of the family finances, but it looks as if this was mistakenly interpreted by the participants as "father *takes full care* of the family finances", at least in some of the countries. This misinterpretation may be "traced" in the before the adjustment factor structures, as it partly affects those and places the item alongside the traditional paternal roles. Happily and quite interestingly, this method-metric "mistake" has also become apparent and corrected for in the after the adjustment factor structures for all combinations of countries in our illustration.

References

Anastasi, A. (1990). *Psychological testing*. New York: MacMillan.

Christensen, L.B. (1988). *Experimental methodology (4th Edition)*. Boston: Allyn & Bacon.

Everitt, B.S. (1996). *Making sense out of statistics in Psychology: A second-level course*. New York: Oxford University Press.

Gari, A., Panagiotopoulou, P., & Mylonas, K. (2009). Social axioms in Greece: Etic and emic dimensions and their relationships with locus of control. In K. Leung & M. H. Bond (Eds.), *Psychological Aspects of Social Axioms. Understanding global belief systems* (pp. 197-216). New York: Springer Science.

Georgas, J., & Berry, J.W. (1995). An ecocultural taxonomy for Cross-Cultural Psychology. *Cross-Cultural Research, 29*, 121-157.

Georgas, J., Berry, J.W., Van de Vijver, F.J.R., Kagitcibasi, C., & Poortinga, Y.H. (Eds.) (2006). *Families across cultures: A 30-nation psychological study*. New York: Cambridge University Press.

Kline, P. (1993). *The Handbook of psychological testing*. London: Routledge.

MacCallum, R. (1981). Evaluating goodness of fit in nonmetric multidimensional scaling by ALSCAL. *Applied Psychological Measurement, 5*, 377-382.

Muthén, B.O. (1994). Multilevel covariance structure analysis. *Sociological Methods and Research, 22*, 376-398.

Muthén, B.O. (2000). Within and between sample information summaries. Software in FORTRAN. Retrieved in Spring 2002, from http://www.gseis.ucla.edu/faculty

Mylonas, K. (2009). Statistical analysis techniques based on Cross-Cultural research methods: cross-cultural paradigms and intra-country comparisons. To appear in *Psychology, The Journal of the Hellenic Psychological Society;* Special Issue: *Cross-Cultural Psychology.*

Mylonas, K. Gari, A., Panagiotopoulou, P., Georgiadi, E., Valchev, V., & Papazoglou, S. (in review). On Bias in terms of culture: Work Values country clustering for 33 European countries and Person-Job Fit factor equivalence testing for four clustered European countries. Chapter submitted for publication in the *Book of Selected Manuscripts of the 19th International Congress of the International Association for Cross-Cultural Psychology*, Bremen, Germany, July 2008.

Nunnally, J.C., & Bernstein, I.H. (1994). *Psychometric theory (3rd Edition)*. New York: McGraw-Hill.

Paraskevopoulos, J.N. (1990). *Statistike efarmosmeni stes epistemes tes simperiforas*. [Statistics, applied to behavioral science]. Athens.

Poortinga, Y.H. (1989). Equivalence of Cross-Cultural data: an overview of basic issues. *International Journal of Psychology, 24*, 737-756.

Poortinga, Y.H., & van de Vijver, F.J.R. (1987). Explaining cross-cultural differences: bias analysis and beyond. *Journal of Cross-Cultural Psychology, 18*, 259-282.

Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology (Vol. 25)* (pp. 1-65). New York: Academic Press.

Schwartz, S. H. (1994). Are there universal aspects in the content and structure of values? *Journal of Social Issues, 50*, 19-45.

Schwartz, S. H. (2006). A theory of cultural value orientations: Explication and applications. *Comparative Sociology, 5,* 136-182.

Schwartz, S. H. (2007). Value orientations: Measurement, antecedents and consequences across nations. In R. Jowell, C. Roberts, R. Fitzgerald & G. Eva (Eds.), *Measuring attitudes cross-nationally: Lessons from the European Social Survey* (pp. 161-193). London: Sage.

Sidiropoulou-Dimakakou, D., Mylonas, K., & Argyropoulou, K. (2008). Holland's hexagonal personality model for a sample of Greek university students. *International Journal of Educational and Vocational Guidance, 8*, 111-125.

Tucker, L.R. (1951). *A method for synthesis of factor analysis studies* (Personnel Research Section Report No. 984). Washington, DC: Department of the Army.

Valencia, R.R., Rankin, R.J., & Livingston, R. (1995). K-ABC content bias: Comparisons between Mexican American and White children. *Psychology in the Schools, 32*, 153-169.

Van de Vijver, F.J.R., & Leung, K. (1997). *Methods and data analysis in Cross-Cultural research*. London: Sage.

Van de Vijver, F.J.R., & Poortinga, Y.H. (2002). Structural equivalence in multilevel research. *Journal of Cross-Cultural Psychology, 33*, 141-156.

Veligekas, P., Mylonas, K., & Zervas, Y. (2007). Goal orientation and beliefs about the causes of success among Greek track and field athletes. *Perceptual and Motor Skills, 105*, 927-938.
Winer, B.J. (1971). *Statistical principles in experimental design (2nd edition)*. New York: McGraw-Hill.