Original Article

Check for updates

A Modified Tucker's Congruence Coefficient for Factor Matching

Anikó Lovik [a] , Vahid Nassiri [ab], Geert Verbeke [ac] , Geert Molenberghs [ac]

[a] *I-BioStat, KU Leuven - University of Leuven, Leuven, Belgium.* **[b]** *Open Analytics, Antwerp, Belgium.* **[c]** *I-BioStat, Hasselt University, Hasselt, Belgium.*

Corresponding Author: Anikó Lovik, I-BioStat, KU Leuven - University of Leuven, Kapucijnenvoer 35 blok d, 3000 Leuven, Belgium. E-mail: aniko.lovik@kuleuven.be

Supplementary Materials: Data, Materials [see Index of Supplementary Materials]

Abstract

Since factor analysis is one of the most often used techniques in psychometrics, comparing or combining solutions from different factor analyses is often needed. Several measures to compare factors exist, one of the best known is Tucker's congruence coefficient, which is enjoying newly found popularity thanks to the recent work of Lorenzo-Seva and ten Berge (2006), who established cut-off values for factor congruence. While this coefficient is in most cases very good in comparing factors in general, it also has some disadvantages, which can cause trouble when one needs to compare or combine many analyses. In this paper, we propose a modified Tucker's congruence coefficient to address these issues.

Factor analysis is one of the most often applied methods of data analysis in psychology where survey methods are extensively used. Obviously, sometimes analyses across countries or different study populations need to be compared. Burt introduced in 1948 an "unadjusted correlation coefficient" to assess factor similarity, which later became known as "Tucker's congruence coefficient" (TCC; Tucker, 1951). It should be noted that this measure assumes a simple factor structure (each and every item loads highly onto one and only one factor). The TCC has been in use ever since but its use rocketed since the publication of a paper in *Methodology* by Lorenzo-Seva and ten Berge (2006), which established cut-off values to accept factor equivalence. The cut-off values were based on the comparison of expert opinion and calculated congruence coefficients from different

sets of factor scores. Naturally, with establishing cut-off values the congruence coefficient became decidedly more accessible to any researcher dealing with factor analyses since it allows easy decision making.

Factorial invariance, and therefore, measures such as congruence coefficients, are relevant to and need to be assessed in a wide range of fields, ranging from cross-cultural personality assessment studies (e.g., De Raad et al., 2010) to hydrological connectivity in ecology (Larsen, Choi, Nungesser, & Harvey, 2012) and from exploring bias in the evaluation of forensic evidence in criminology (Smith & Bull, 2012) to comparing event-related potential patterns in neuroscience (Barry, De Blasio, Fogarty, & Karamacoska, 2016) and the longitudinal assessment of dietary patterns (Batis et al., 2014).

Undoubtedly, the TCC is a useful measure of factor similarity, but the measure also has some disadvantages that are closely related to the dichotomy of congruence-incongruence for which we propose a solution by slightly modifying the congruence coefficient.

The paper is organised as follows. First, Tucker's congruence coefficient is presented together with its advantages and disadvantages. Then, the modified congruence coefficient is introduced and the difference between the two measures is presented in a real-life example and a small simulation study. Finally, all results are discussed.

The Original and the Modified Tucker's Congruence Coefficient

Tucker's congruence coefficient (TCC) was first proposed by Cyril Burt but then became popular based on Ledyard Tucker's (1951) work and is still in use as can be seen, among others, from the number of references to Lorenzo-Seva and ten Berge's (2006) paper (377 until March 2017 and 655 until May 2019 according to Google Scholar). While it is often used in combination with procrustes rotations – which maximises the congruence coefficient (Korth & Tucker, 1976; Brokken, 1983) – to our knowledge it has not been used to match factors for reasons that will be discussed later.

The Tucker's congruence coefficient ($\varphi(x, y)$) itself is very simple to calculate:

$$\varphi(x, y) = \frac{\sum x_i y_i}{\sqrt{\sum x_i{}^2 \sum y_i{}^2}} \tag{1}$$

where x_i and y_i are the factor loadings of variable (item) i for factors x and y, respectively. As it is obvious from its formula, the idea behind this coefficient is very simple and not dissimilar from the Pearson's correlation coefficient. They can both be interpreted as the cosine of the angle between two vectors, but the Pearson correlation is about the cosine of the angle of two *centered* vectors. Therefore, it is not surprising they are both used to assess the potential association between variables. They both range from -1 to

+1. However, while negative correlations are quite useful (and quite problematic when one wants to use the correlation of the loadings to assess invariance, see Pinneau & Newhouse, 1964), this is not really true for TCCs. Very negative TCCs are not terribly useful when assessing factor similarity where only the dichotomy between very high positive values and the rest is of interest. The interpretation is quite arbitrary and several cut-of values between 0.80 and 0.95 have been proposed (for details, see Lorenzo-Seva & ten Berge, 2006). These are mostly rules of thumb based on experience. While hypothesis tests have been developed (for example, Bentler & Bonett, 1980; Korth & Tucker, 1975) these are of moderate use as often values below 0.80, the minimum conceptual bound, are deemed relevant by providing a significant result, even when the factors are known to differ (Davenport, 1990). This is mainly due to the difficulty to establish a meaningful goodness-of-fit index that is capable of capturing the minimal difference that can discriminate between two different models. Bentler and Bonett (1980) argue that any index below 0.90 is too low. Two empirical studies found cut-off values of 0.85 (Haven & ten Berge, 1977 cited by Brokken, 1983 and Lorenzo-Seva & ten Berge, 2006) and 0.95 (Lorenzo-Seva & ten Berge, 2006). In summary, we can say that there is general agreement that anything below 0.80 means the factors are incongruent but there is a huge gray area between 0.80 and 1 (Barrett, 1986).

Putting aside the problem of the subjectivity of choosing a cut-off, there are several advantages of this measure, which are summarised by Lorenzo-Seva and ten Berge (2006) in four points:

1. $\varphi(x, y)$ is insensitive to scalar multiplication of x and y. Therefore, the factor similarity is measured independently of the absolute size of the loadings, meaning that factor pairs with high loadings can have low similarity and vice versa low loadings can result in high coefficients of similarity.
2. $\varphi(x, y)$ is sensitive to additive constants so two factors of similar pattern but with loadings very different in magnitude can still be incongruent.
3. $\varphi(x, y)$ is insensitive to a change in the sign of any pair (x_i, y_i), which reflects a change in the sign of variable i.
4. $\varphi(x, y)$ is mathematically attractive as it is a continuous function of x_i and y_i.

One problem with the TCC is that it is sensitive to changes in sign of individual loadings because, when only x or y change in sign then the sign of the product changes. This problem was already mentioned in Pinneau and Newhouse (1964) and in Barrett (1986) as the fact that $\varphi(x, y)$ overestimates congruence when the signs of the variable pairs (x_i, y_i) is predominantly the same since then the product is positive, hence the numerator's value increases with nearly each term while in the case of a difference in signs the product is negative and the total decreases. Vice versa, the similarity is underestimated if the signs are predominantly different. Pinneau and Newhouse (1964) give an illustrative example of a vector with only positive elements where the lowest possible TTC (when all

possible permutations are compared) is 0.67. Although this seems a minor disadvantage when one compares two factor analyses with very clear factor structures, in the presense of high cross-loadings it quickly becomes worrisome and seriously limits the use of TCCs for small samples, samples with missing values or analyses where sampling from the original samples is used.

Pinneau and Newhouse (1964) also point out that this overestimation problem means that the first factor for any centroid factor analysis will result in a high TCC as the factor loadings are predominantly positive. In fact, as Davenport (1990) remarks, factor loadings after rotation tend to be either positive or close to nil (in which case matching factors would be easy). This makes it easier to match the first factor but it is uncertain whether it affects the matching of the later factors. This is problematic because it is certainly possible that different samples will result in the same factors but that these factors are ordered differently. A very typical example would be matching the factor analytic after multiple imputation where it is a reasonable assumption - given that usually a high percentage of the data is shared across the imputations - that most complete sets result in the same factors albeit not necessarily in the same order (an example of such an analysis can be found in Lovik, Nassiri, Verbeke, & Molenberghs, 2018).

A second, more practical, problem arises if the survey on which the factor analysis is performed contains negatively framed items. Normally, these items are reversed before analysis. If one performs only one factor analysis without reversing these items, since the factor loadings are calculated based on the correlation/covariance matrix of all items, the factor loadings for these items will not change in magnitude, but the sign will reverse. This may or may not result in a TCC indicating incongruence erroneously and one needs to pay attention when interpreting such analyses. However, when such an analysis is compared to one where the negatively framed items were reversed before the analysis, the result is an extremely low TCC indicating incongruence while the modified coefficient will be high, indicating congruence. Similarly, an erroneously coded dataset where all items have been reversed may result in a TCC close to -1. Very negative TCCs can also be caused by high cross loadings in one of the factor analyses, for example when the analysis was performed on a small sample. Regardless of the chosen cut-off point, in these situations, the TCC incorrectly declares the factors to be incongruent.

The modification we propose, which solves both problems in a factor analytical context, is to use the absolute value of the products in the numerator:

$$\psi(x, y) = \frac{\sum \left| x_i y_i \right|}{\sqrt{\sum x_i^2 \sum y_i^2}} \tag{2}$$

This results in losing the nice geometric interpretation of the index but has several advantages. First, all advantages of the TCC are preserved: $\psi(x, y)$ is insensitive to scalar multiplication of x and y, and to change in the sign of any pair (x, y) but sensitive

to additive constants. Furthermore, it is also still a continuous function of x_i and y_i. Obviously, $\psi(x, y) \geq \varphi(x, y)$, $0 \leq \psi(x, y) \leq 1$ and there is no direct link between φ and ψ. As the TCC, this coefficient is also only valid in the context of a simple structure factor analysis.

A TCC of -1 becomes 1, which is advantageous since in practice this may only occur in case of erroneous coding. Similarly, very low values (between -1 and -0.90) normally happen when two factors are similar (in interpretation) but many signs are reversed for one factor compared to the other. In such cases, the Tucker's congruence coefficient would erroneously reject the possibility that the two are equal, while the modified coefficient results in a value above 0.90. This also may happen because of the many nil loadings mentioned earlier (Davenport, 1990) where the sign can change quite easily through the different analyses. Examples highlighting these effects can be found in the Supplementary Materials (SM) A of this paper.

Example of the Use of the Two Congruence Coefficients for Comparing Factor Analyses in Real Data

In this section, the original (TCC) and the modified (mTCC) congruence coefficients are compared in questionnaire data. The analyses were performed on the Big Five Inventory (BFI; Denissen, Geenen, van Aken, Gosling, & Potter, 2008; John & Srivastava, 1999) based on data from 7533 individuals from 4460 families who participated in the Divorce in Flanders study (see details in Mortelmans et al., 2011). The BFI measures the Big Five personality factors (Neuroticism, Extraversion, Openness to Experience, Conscientiousness and Agreeableness) and the analysis shows that the five independent factors emerge clearly from the data (Lovik, Nassiri, Verbeke, Molenberghs, & Sodermans, 2017). Unfortunately, factor analysis should not be used for clustered data and we decided to use multiple outputation (Follmann, Proschan, & Leifer, 2003) to deal with clustering. We generated 1000 random subsets from the original dataset using simple random sampling, thus one individual from each family was selected (n = 4460). There are 641 complete families with three participating family members, 1791 families with two participants each and 2028 families with one family member taking part in the study. The last group is included in each sample and thus the samples can be assumed to be fairly similar.

On each of these 1000 datasets, factor analysis with principal component extraction was performed and the results were rotated using direct oblimin (quartimin) rotation. It should be noted, that the choice of rotation method does not influence the results in this case (see Lovik, Nassiri, Verbeke, & Molenberghs, 2018; Lovik et al., 2017). Considering the Big Five factors are present in the full data, it seems reasonable to expect that a five factor structure will suit the subsets as well. If this is indeed the case, the five factors found in one subset could be matched to the factors of the other subsets. However, one

cannot assume that the factors will be ordered in the same way for all analyses and indeed, 449 analyses had a different factor ordering when compared to the first set. Considering the large sample, the low amount of missingness and low cluster size, this clearly shows high occurrence of differently ordered factor analyses and the the need for accessible factor marching procedures.

TCCs and mTCCs were calculated for each possible pair of factors (5×5) in all pairs of subsets and with the goal to find the five congruent factor pairs from the different analyses based on TCCs and mTCCs, which could be a step forward to combine factor analyses; after the matching factors are found based on congruence and factors are correctly ordered in the two (or more) analyses, the results of the different analyses could be merged. We assume that the highest congruence coefficient (CC) is when the factors are congruent. However, matching based on the TCC suffers from the previously described problem of not recognizing high negative loadings as potential matches, albeit in a very small amount of cases. For this reason, a scatterplot of all TCCs and mTCCs is presented in Figure 1.
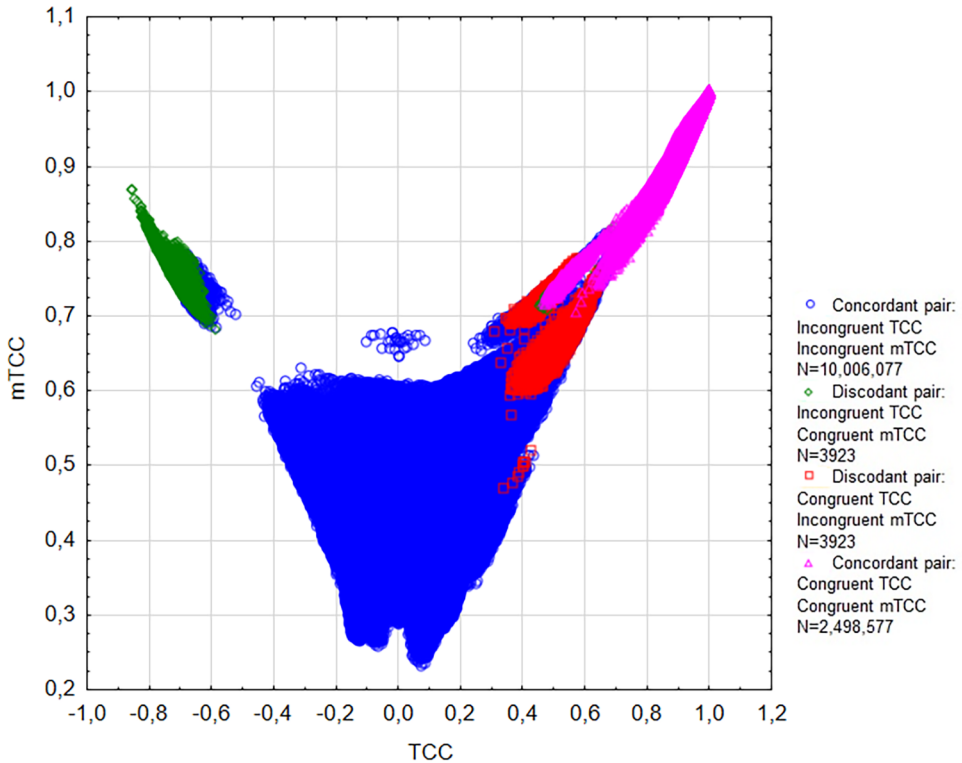
The left hand side of Figure 2 (2A) shows the TCC-mTCC pairs that are discordant (one CC assumes the factors to be congruent while the other states they are incongruent). This latter figure represents the potential gain and risks due to the use of mTCCs over TCC; potential good matches, not found by TCC, could be found by using the mTCC, but it can also lead to false positives: finding matches where there are none. The right hand side of Figure 2 (2B), shows the concordant TCC-mTCC pairs where the decision of both TCC and mTCC would result in a correct match.

As it can be seen in Figure 1, although the five factor structure is quite clear in most of the datasets, the congruence coefficients show quite a bit of variability. Out of 12,512,500 (TCC, mTCC) pairs, 5000 are (1, 1) due to factors being compared with themselves. Excluding these perfect matches, the TCCs range from -0.8786 to 0.9997, while the mTCCs are between 0.2310 and 0.9997. The equality of the maximum is not surprising, the higher the TCC the smaller the difference with the associated mTCC. In case all factor loadings are positive the two coefficients are, of course, equal. This happens in 2.9% of the cases in our example (apart from the perfect matches).

Figures 1 and 2 show that there are 3923 cases where the TCCs and mTCCs are discordant, thus one would reach a different conclusion when using the mTCC instead of the TCC. In all cases but one, this means that the TCC would pair the factors incorrectly, meaning that two factors from analysis 1 would be assigned to the same factor in analysis 2 while one factor of analysis 2 would not have a match in analysis 1, so that in only one case the mTCC is responsible for mismatching factors. This suggests, that the gain due to the mTCC is far greater than the risk (see Figure 2). Furthermore, all discordant pairs occur in case of TCCs ≤ 0.80 and the matching mTCCs are always ≤ 0.90, however, the concordant pairs range from 0.4671 to 0.9997 and 0.7061 to 0.9997 for the TCCs and mTCCs, respectively. The right hand side of Figure 2 (2b) shows that many of

Figure 1

Scatterplot of TCCs and mTCCs in the DiF Study - for Each Possible Pair
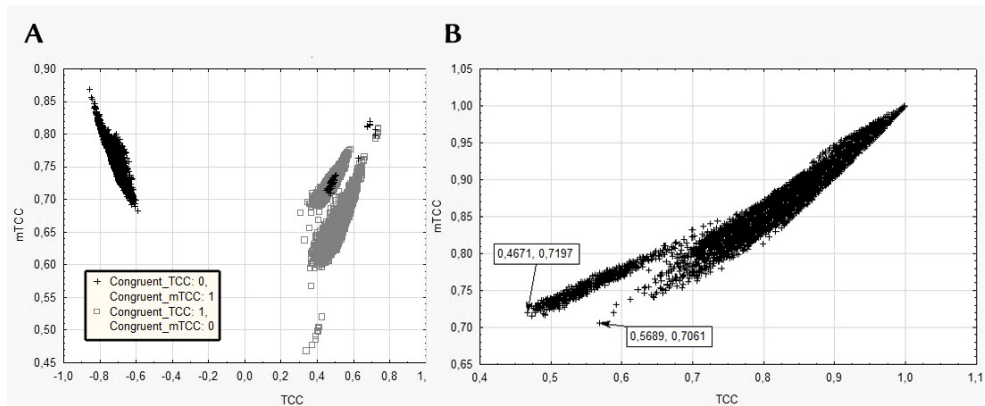

the concordant pairs would not be matched if one used a cut-off of 0.80 or 0.95. It should also be noted that the reason behind the mismatches and low congruent TCCs is partly due to the high number of negative loadings. In our example, there are on average 18.1 (*SD* = 0.41) negative loadings per factor, which means around 40% of the loadings are negative. The number of negative loadings is much lower for the first factor (13.65, *SD* = 0.85) than for the later factors, ranging from 18.13 (third factor before matching, *SD* = 3.45) to 20.5 (second factor before matching, *SD* = 3.58), which is as expected.

The overlap in Figure 1 shows that in this example it is impossible to set a cut-off point to accept congruence that will work in all situations. Having said this, Figure 3 shows that, in general, both TCC and mTCC separate congruent and incongruent factors well. Clearly, the concordant incongruent-incongruent pairs on the left hand side of the figure have much lower mTCC values than the concordant congruent-congruent

Figure 2

Scatterplot of Discordant (A) and Concordant (B) TCCs and mTCCs
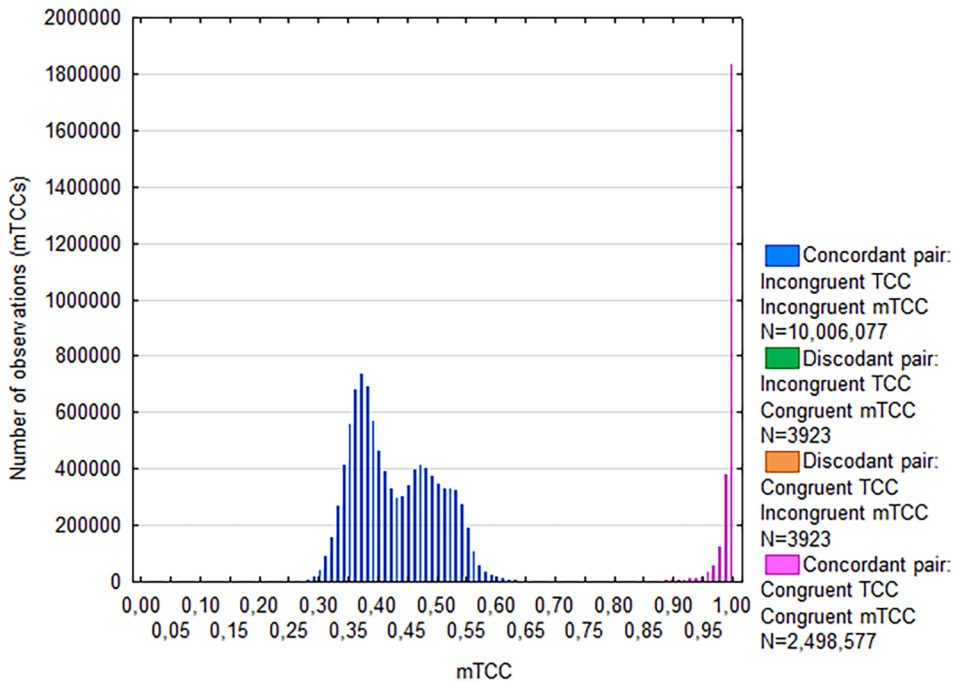

pairs (peak at the right). Thus, at least in similar settings, both congruence coefficients could be used to combine factors/factor analyses. Note that due to the low number of discordant pairs, the remaining categories cannot be distinguished, however, the values can be deduced from Figure 1 and these values fall mostly between the two concordant categories. The full analysis with regard to matching the sets based on congruence coefficients can be found in (Lovik, Nassiri, Verbeke, & Molenberghs, 2018).

Simulation Study Comparing the Factor Matching Performance of TCCs and mTCCs

A small simulation study to show the advantage of the mTCC, when combining a high amount of FAs, was performed. Four settings were considered: 2 and 5 factors with 2 or 10 items each. Because the congruence coefficient is calculated directly from the factor loadings, instead of generating datasets, factor analyses were generated where the primary loadings were chosen arbitrarily (means and *SD*s shown in Table 1), while the cross-loadings were drawn from a normal distribution with mean 0.10 and standard deviations (*SD* in the rest of the paper refers to these standard deviations) ranging from 0.025 to 0.40 truncated at -1 and 1, if necessary. It should be noted, however, that the choice (exact size) of primary loadings does not affect the conclusions (unless primary loadings are too low in comparison with the intended cross-loadings to achieve a clear factor structure at which point the the quest for a simplex EFA/PCA becomes pointless). To ensure this, different sets of primary loadings (*high* = 0.80; *moderate* = 0.60 and *low*

Figure 3

Histogram of mTCCs Categorised by Concordance in the DiF Study

= 0.40) were tested and no notable differences were found (see SM E), except when the assumption of a clear, simple factor structure was violated (only in the low primary loadings setting: SM E.3). The same is true for the mean of the cross-loadings. 0.10 has been selected because, in general, cross-loadings tend to be positive (e.g., Davenport, 1990) – due to the rotation applied to the FAs. Similar results have been achieved with different means (0 and 0.20) for cross-loadings (see SM G), with higher means showing a greater advantage of the mTCC over the TCC. Each replication compared 100 factor analyses and was repeated 100 times, thus for each *SD* 10,000 FAs were generated in a specific setting. Further details on the simulation, together with the R code can be found in SM B. SM D shows that the fixed random seeds have no effect on the results. Comparison of TCC and mTCC to the Pearson correlation coefficient can be found in SM H.

Figure 4 shows the percentage of mismatches for TCCs, mTCCs and also the percentage of shared mismatches (in case both TCC and mTCC fail to match congruent factors). The results, independently from the setting, clearly show that there is no difference

Table 1

Means and SDs of Selected Primary Loadings of the Simulation Study

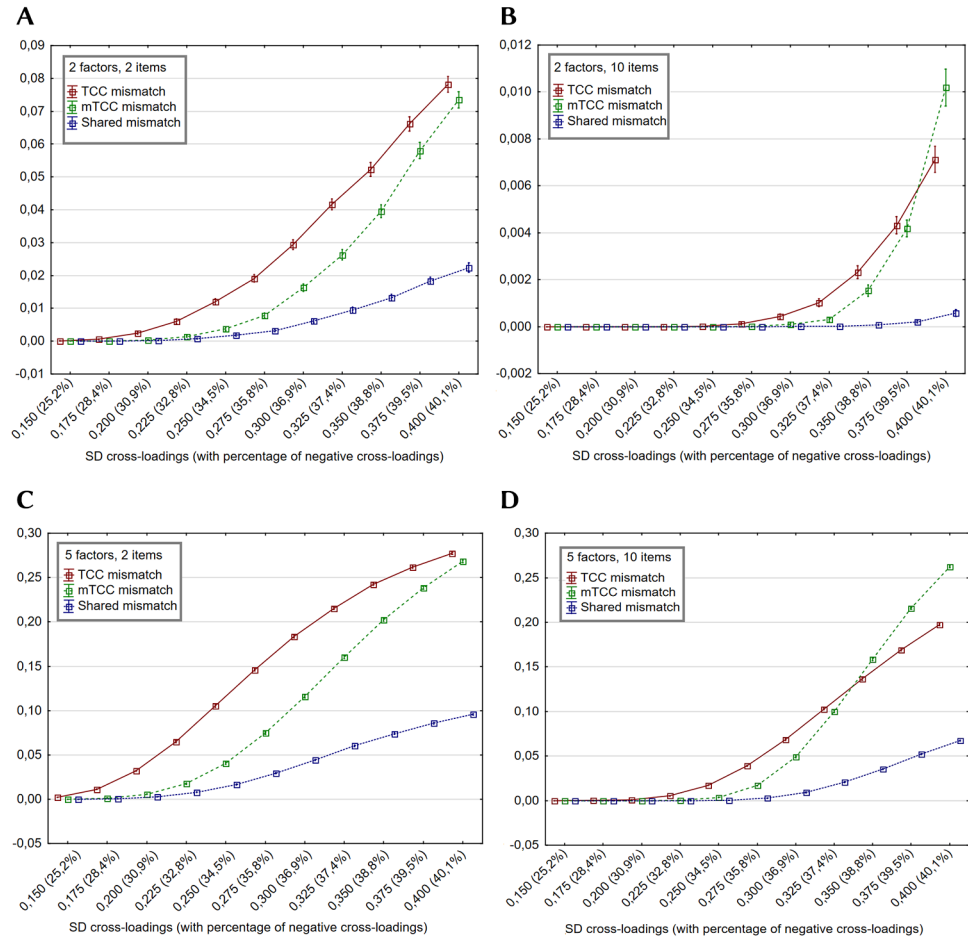| Factor	M (SD)				Range
	F2I2	F2I10	F5I2	F5I10	
Factor 1	0.75 (0.07)	0.68 (0.08)	0.85 (0.07)	0.68 (0.08)	0.6 − 0.9
Factor 2	0.65 (0.07)	0.58 (0.08)	0.75 (0.07)	0.59 (0.10)	0.5 − 0.8
Factor 3	-	-	0.70 (0.00)	0.55 (0.08)	0.4 − 0.7
Factor 4	-	-	0.65 (0.07)	0.51 (0.09)	0.4 − 0.7
Factor 5	-	-	0.60 (0.00)	0.49 (0.09)	0.4 − 0.6

Note. F = factor; I = item.

between TCC and mTCC in performance when the cross-loadings are negligible or low compared to the primary loadings – as can be seen on each of the four plots of Figure 4 on the left hand side (low *SD*s). "Low" obviously depending on the number of factors, number of items per factor and primary loadings. In the simulation study, in Figure 4A (2 factors, 2 items) there are no mismatches until *SD* = 0.175, while in Figure 4B (2 factors, 10 items) the first mismatches occur after *SD* = 0.25. Correspondingly, with an increased number of factors, the first mismatch occurs a bit earlier than for the simulations with the same number of items but less factors (first mismatch at 0.15 (Figure 4C) compared to 0.175 (Figure 4A); 0.175 (Figure 4D) compared to 0.275 (Figure 4A)). Fixing the number of factors, mismatches occur much later with 10 items/factor than for 2 items/factor. These patterns are present for the other simulations presented in SM E-G, but the exact number of mismatches and *SD*'s where the first mismatches occur are changeable. Both coefficients match the FAs without errors, in our simulation study there are absolutely no mismatches under *SD* = 0.15, hence these values were removed from Figure 4, SM C provides Figures 4A-4D with the full range of tested *SD*s.

Similarly, when cross-loadings become higher (approaching or exceeding primary loadings in magnitude), none of the coefficients do really well (mismatches ranging from 7.5% in Figure 4A to 27.5% in Figure 4C where *SD* = 0.40), although the TCC has fewer mismatches. However, this happens in situations where many cross-loadings are very high compared to the primary loadings, which does not happen often in data with a simple factor structure, since it means a clear factor structure is absent (or the factor structure is complex). Neither the TCC nor the mTCC are valid in these situations. For example, at *SD* = 0.40 about 15% of the cross-loadings fall between 0.50 and 0.90 which is the regular range of primary-loadings, resulting in a high number of mismatches. Especially the ten-item analyses are very sensitive to this effect as with the higher number of cross-loadings the probability of an extremely high cross-loading is higher and the lower primary loadings (see Table 1) for the later factors cannot compensate for

PsychOpen GOLD

Figure 4

Results of the Simulation Study

Note. The abscissa starts at *SD* = 0.15, since there are no mismatches below this value for any of the settings.

these. In a simulation study where all primary loadings were set to 0.80 this crossing of the TCC and mTCC lines did not occur (see SM E.1), while for lower primary loadings such as 0.60 or 0.40 (see SM E.2 and E.3, respectively) it happens earlier (meaning lower *SD* of the cross-loadings).

The interesting part, from the point of view of the mTCC, is when cross-loadings are not nil or close to nil but are, in general, lower than the lowest primary loadings, in other words, when ostensible primary and cross-loadings are clearly distinguishable from each other, with a certain number of negative cross-loadings. Obviously, the number of negative cross-loadings increases with the rise of the *SD* of the cross-loadings. In such situations the mTCC clearly outperforms the TCC regardless from the number of factors and items. Also, cross-loadings become more problematic in factor matching when the number of items per factor is low, in which case the difference in performance between TCC and mTCC is more evident. Likewise, when the primary loadings are lower (e.g., for later factors, here factors 4 and 5 of the five factor setting), the percentage of mismatches is consistently higher (close to 30% for 5 factors and *SD* = 0.40).

Discussion

In this paper, we presented an alternative to the Tucker's congruence coefficient using the absolute values of the products in the numerator. This method results in much higher values for incongruent factor pairs but for pairs with a TCC ≥ 0.95 the change is minimal. We suggest the cut-off value 0.95 (as established for TCCs by Lorenzo-Seva & ten Berge, 2006) to be adopted for factor equivalence, above which two factors can be considered to be equal, if any such cut-off value is really needed. Our example suggests that lower values may not separate congruent from incongruent factor pairs and further research is needed to prove that our suggestion can be used in practice. Furthermore, even in a setting where the factor structure is known and there is a generous overlap in the datasets, the congruence coefficients vary greatly (for congruent and incongruent factors both), implying caution is needed when such coefficients are used.

Nice potential lies in the combined use of the two coefficients where we suggest two options: an "inclusive" and an "exclusive" one. The exclusive combination only accepts congruency if both TCC and mTCC indicate factor congruence while the inclusive option would indicate congruency if at least one of the two coefficients signal factor congruence. Since the mismatches come in pairs, and either TCC or mTCC are correct while the other is incorrect, the use of either of these combinations comes with the cost of lower sensitivity or specificity in matching the factors. The exclusive combination tries to avoid including false positives but at the same time may miss out on true positives. Meanwhile, the inclusive combination may gain true positives at the cost of the inclusion of false positives in an increased number. Either of these combinations could be useful depending on the context.

Another potentially useful candidate could be the Pearson correlation coefficient (PCC). While our case study does not endorse this coefficient, the number of mismatches being 99, 97 and 1, for the TCC, PCC and mTCC, respectively, both the toy examples and the simulations including the PCC in SM H show settings where the PCC can

PsychOpen GOLD

be advantageous. Which coefficient works best, depends on many components, and therefore it has merit to add this coefficient as a reference point, especially if the number of mismatches is high with the other coefficients. Both the PCC's role in factor matching and identifying which coefficient is best for which setting should be explored more thoroughly in the future, but this investigation is out of scope of this paper.

As a reviewer pointed out, the mainstream approach to TCCs for factor matching, based on optimally reflected and ordered pattern matrices, should be considered. The main idea behind this approach is to rotate one factor analysis to a target matrix using a predefined criterion, for example by maximising the TCC (e.g., Cliff, 1966; Korth & Tucker, 1976), after which the factors are re-ordered and matched. Needless to say, there are different ways to achieve this. In a recent paper, Myers, Ahn, Lu, Celimli, and Zopluoglu (2017) described their freely available R package (REREFACT) which allows the user to match factor analyses of simulation studies. However, by relying on TCCs, equivalent sign and order pattern forms have to be evaluated. In case of 3 factors, 48 possible patterns exist when matching two sets, many of which are equivalent apart from sign and order (see Myers et al., 2017, Table 3). Removing the problem of signs by using the mTCC, the number of patterns to evaluate would be 6.

Because of the lack of clear geometric interpretation and well-established cut-off values, the new coefficient cannot simply replace the TCC. However, the mTCC may be more useful to combine factor analyses of a simple factor structure or to deal with a high amount of factor analyses where it is not feasible to assess each pair manually. In a small simulation study we have shown that with the increase of the cross-loadings both in range and magnitude (especially if some of the loadings are negative) the mTCC outperforms the TCC in correctly matching factors. A high amount of mismatches, or more mismatches for mTCC than TCC, may indicate the presence of too many (too high) cross-loadings compared to the primary loadings or even the lack of a simple factor structure.

Of course, even when factors are considered to be equivalent based on congruence this does not necessarily mean the interpretation of the factor is the same. However, while no congruence coefficient could possibly make sure that two factors have the same meaning, it allows to assess whether they are likely to measure the same concept. Also, our example shows that it is quite problematic to set cut-off values that are valid in every situation. Therefore, TCCs and mTCCs should be interpreted and used with caution.

Data Availability: The real data from the Divorce in Flanders study is available upon request. The simulation data for this article is freely available (see the Supplementary Materials section).

Supplementary Materials

For this article the following supplementary materials are available (for access options see Index of Supplementary Materials below):

- A – Toy examples of TCC, mTCC and Pearson CC
- B – Details on the simulations and R code
- C – Alternative Figure 4 of the main body of the paper
- D – Simulations with di_erent random seeds
- E – Simulations with primary loadings 0.80 (E.1), 0.60 (E.2) and 0.40 (E.3)
- F – Simulations with 2, 4 (F.1), 10 and 16 (F.2) items for 2 factors
- G – Simulations with mean cross-loading of 0 (G.1) and 0.20 (G.2)
- H – Simulation results comparing the TCC-mTCC-PCC triad
- I – Histograms of concordance of TCCs and mTCCs in the simulations

Index of Supplementary Materials

Lovik, A., Nassiri, V., Verbeke, G., & Molenberghs, G. (2020). *Supplementary materials to: A modified Tucker's congruence coefficient for factor matching*. PsychOpen. https://doi.org/10.23668/psycharchives.2786

References

Barrett, P. (1986). Factor comparison: An examination of three methods. *Personality and Individual Differences, 7*(3), 327-340. https://doi.org/10.1016/0191-8869(86)90008-5

Barry, R. J., De Blasio, F. M., Fogarty, J. S., & Karamacoska, D. (2016). ERP Go/NoGo condition effects are better detected with separate PCAs. *International Journal of Psychophysiology, 106*, 50-64. https://doi.org/10.1016/j.ijpsycho.2016.06.003

Batis, C., Sotres-Alvarez, D., Gordon-Larsen, P., Mendez, M. A., Adair, L., & Popkin, B. (2014). Longitudinal analysis of dietary patterns in Chinese adults from 1991 to 2009. *British Journal of Nutrition, 111*(8), 1441-1451. https://doi.org/10.1017/S0007114513003917

PsychOpen GOLD

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588-606. https://doi.org/10.1037/0033-2909.88.3.588

Brokken, F. B. (1983). Orthogonal procrustes rotation maximizing congruence. *Psychometrika, 48*(3), 343-352. https://doi.org/10.1007/BF02293679

Cliff, N. (1966). Orthogonal rotation to congruence. *Psychometrika, 31*(1), 33-42. https://doi.org/10.1007/BF02289455

Davenport, E. C., Jr. (1990). Significance testing of congruence coefficients: A good idea? *Educational and Psychological Measurement, 50*(2), 289-296. https://doi.org/10.1177/0013164490502007

Denissen, J. J. A., Geenen, R., van Aken, M. A. G., Gosling, S. D., & Potter, J. (2008). Development and validation of a Dutch translation of the Big Five Inventory (BFI). *Journal of Personality Assessment, 90*(2), 152-157. https://doi.org/10.1080/00223890701845229

De Raad, B., Barelds, D. P., Levert, E., Ostendorf, F., Mlačić, B., Blas, L. D., . . . Church, A. T. (2010). Only three factors of personality description are fully replicable across languages: A comparison of 14 trait taxonomies. *Journal of Personality and Social Psychology, 98*(1), 160-173. https://doi.org/10.1037/a0017184

Follmann, D., Proschan, M., & Leifer, E. (2003). Multiple outputation: Inference for complex clustered data by averaging analyses from independent data. *Biometrics, 59*(2), 420-429. https://doi.org/10.1111/1541-0420.00049

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (pp. 102–138). New York, NY, USA: Guilford.

Korth, B., & Tucker, L. R. (1975). The distribution of chance congruence coefficients from simulated data. *Psychometrika, 40*(3), 361-372. https://doi.org/10.1007/BF02291763

Korth, B., & Tucker, L. R. (1976). Procrustes matching by congruence coefficients. *Psychometrika, 41*(4), 531-535. https://doi.org/10.1007/BF02296973

Larsen, L. G., Choi, J., Nungesser, M. K., & Harvey, J. W. (2012). Directional connectivity in hydrology and ecology. *Ecological Applications, 22*(8), 2204-2220. https://doi.org/10.1890/11-1948.1

Lorenzo-Seva, U., & ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology, 2*(2), 57-64. https://doi.org/10.1027/1614-2241.2.2.57

Lovik, A., Nassiri, V., Verbeke, G., & Molenberghs, G. (2018) Combining factors from different factor analyses based on factor congruence. In M. Wiberg, S. Culpepper, R. Janssen, J. González, & D. Molenaar (Eds.), *Quantitative Psychology. IMPS 2017. Springer Proceedings in Mathematics & Statistics* (Vol. 233, pp. 211–219). Cham, Switzerland: Springer.

Lovik, A., Nassiri, V., Verbeke, G., Molenberghs, G., & Sodermans, A. K. (2017). Psychometric properties and comparison of different techniques for factor analysis on the Big Five Inventory from a Flemish sample. *Personality and Individual Differences, 117*, 122-129. https://doi.org/10.1016/j.paid.2017.05.048

PsychOpen GOLD

Mortelmans, D., Pasteels, I., Bracke, P., Matthijs, K., Van Bavel, J., & Van Peer, C. (2011). *Scheiding in Vlaanderen* (Divorce in Flanders). Leuven, Belgium: Acco.

Myers, N. D., Ahn, S., Lu, M., Celimli, S., & Zopluoglu, C. (2017). Reordering and reflecting factors for simulation studies with exploratory factor analysis. *Structural Equation Modeling, 24*(1), 112-128. https://doi.org/10.1080/10705511.2016.1230721

Pinneau, S. R., & Newhouse, A. (1964). Measures of invariance and comparability in factor analysis for fixed variables. *Psychometrika, 29*(3), 271-281. https://doi.org/10.1007/BF02289724

Smith, L. L., & Bull, R. (2012). Identifying and measuring juror pre-trial bias for forensic evidence: Development and validation of the Forensic Evidence Evaluation Bias Scale. *Psychology, Crime & Law, 18*(9), 797-815. https://doi.org/10.1080/1068316X.2011.561800

Tucker, L. R. (1951). *A method for synthesis of factor analysis studies* (No. PRS-984). Princeton, NJ, USA: Educational Testing Service.

PsychOpen GOLD