



# Assessing factorial invariance of two-way rating designs using three-way methods

Pieter M. Kroonenberg\*

Department of Child and Family Studies, Institute of Education and Child Studies, Leiden University, Leiden, Netherlands

## Edited by:

Alain De Beuckelaer, Radboud University Nijmegen, Netherlands

## Reviewed by:

Heungsun Hwang, McGill University, Canada  
Pietro Cipresso, IRCCS Istituto Auxologico Italiano, Italy  
Wim Krijnen, University of Groningen, Netherlands

## \*Correspondence:

Pieter M. Kroonenberg, Department of Child and Family Studies, Institute of Education and Child Studies, Leiden University, Wassenaarseweg 52, 2333 AK Leiden, Netherlands  
e-mail: kroonenb@fsw.leidenuniv.nl

Assessing the factorial invariance of two-way rating designs such as ratings of concepts on several scales by different groups can be carried out with three-way models such as the Parafac and Tucker models. By their definitions these models are double-metric factorially invariant. The differences between these models lie in their handling of the links between the concept and scale spaces. These links may consist of unrestricted linking (Tucker2 model), invariant component covariances but variable variances per group and per component (Parafac model), zero covariances and variances different per group but not per component (Replicated Tucker3 model) and strict invariance (Component analysis on the average matrix). This hierarchy of invariant models, and the procedures by which to evaluate the models against each other, is illustrated in some detail with an international data set from attachment theory.

**Keywords:** stimulus-response data, Tucker3 model, Tucker2 model, Parafac model, three-mode analysis, rating scales, semantic differentials, Ainsworth strange situation

## 1. INTRODUCTION

Two-way rating designs may consist of, for instance, ratings of concepts on several rating scales. In this paper we tackle the problem of the invariance of the factorial structure of data arising from such designs when the data have been collected from several groups. In particular we will show that three-mode component models are ideally suited to assess factorial invariance for such designs. We will specify a hierarchy of models with increasing restrictions on the parameters resulting in more and more invariant factorial structures across groups.

Because in this paper we are dealing with component models we will use the term “components” rather than “factors,” unless factors are explicitly indicated. However, to stay within the standard terminology we will use the term *factorial invariance*, rather than *subspace invariance* or *component invariance*. A detailed treatment of the differences between factor analysis and component analysis for two-way data can for instance be found in Widaman (2007).

### 1.1. FACTORIAL INVARIANCE IN TESTS

Most of the research on factorial invariance assumes that an investigator wants to evaluate whether a test with a particular dimensional structure operates in the same way for different groups, so that the test, or the factors underlying it, can be used for all kinds of groups; a detailed technical exposition of measurement invariance, factorial invariance and their relationship can be found in Meredith (1993). Factorial invariance is typically of interest, for instance, when intelligence tests have been translated into other languages and researchers want to establish whether the translated tests function in the same manner as the original. Alternatively, a researcher may want to know whether a test has

the same structure for different groups, say both for regular and for clinical samples.

In a literature survey Vandenberg and Lance (2000, pp. 12–13) synthesized common practices in a list of sequential tests to assess the extent of factorial invariance. The steps in their hierarchy of hypotheses are listed below, but we have listed their first step as the final one, because it is the most restrictive of all invariance schemes, i.e., there is no intergroup variability. Here we present a compact version of their descriptions. Finally, we have added a new first step: Lack of factorial invariance. We need this step later on as a reference point or baseline for our analyses. Note that each next step introduces *additional restrictions* on the parameters of the models.

1. *Lack of invariance*: All groups have different factor patterns.
2. *Configural invariance*: Invariant patterns of factor loadings across groups.
3. *Metric invariance*: Invariant values of factor loadings for like items across groups.
  - \*a *Scalar invariance*: Invariant intercepts of like items regressions on the factor.
  - \*b *Unique variances invariance*: Invariant unique variances of like items across groups.
    - c *Invariant factor variances*: Invariant factor variances across groups.
4. *Invariant factor covariance matrices*: Invariant factor covariance matrices across groups.
  - \*d *Invariant factor means*: Invariant factor means across groups.
5. *Strict invariance*: Invariant factor means and covariance matrices across groups.

The hierarchy is primarily based on investigations using factor analysis within the context of structural equation modeling with and without estimation of the factor means. This means that it contains concepts and parameters characteristic of such models, such as unique variances, factor means and intercepts of regressions of items on factors. In this paper such concepts do not play a role, because our proposals are based on component analysis. In the sequel, the starred steps are therefore excluded for the following reasons: (\*a, \*d) all scales will be centered across concepts for each group (see below), so that means and factor means do not enter into the models; (\*b) the concept of unique variances does not play a role in component analysis. Note that when referring to Step 5, “Strict invariance,” we will assume only that the covariance matrices are equal across groups, again because the means have already been removed by centering.

The major analytical techniques for establishing the increasingly stricter types of invariance have primarily been structural equation modeling and item response theory as is evident in this special issue. In the hierarchy of hypotheses about factorial invariance it is implied that the models are nested, so that they can be evaluated, or in the context of structural equation models, tested against each other. This means that an a priori choice has to be made about the factor model itself: How many factors and which items are to be regressed on which factors. Therefore, a two-factor model may be invariant in a different way than a three-factor model for the same data. In this paper we will concentrate on series of both two-factor and three-factor models, but we will not attempt to make detailed comparisons between the two series.

Regarding the component models in this paper, comparisons between models are primarily based on the error sums of squares in relation to their degrees of freedom. These degrees of freedom are calculated as the number of data points minus the number of parameters to be estimated ( $N_{\text{parm}}$ ) where the means subtracted during the centering of the data are also counted as parameters. Details and formulas for calculating the degrees of freedom for three-way models can for instance be found in Kroonenberg (2008, Section 8.4, p. 177ff).

## 1.2. TWO-WAY RATING DESIGNS

In psychology a specific kind of measurement design is commonly used, i.e., a *two-way rating design* in which concepts are judged on scales by a number of judges such as in Osgood’s classical semantic differential design (Osgood et al., 1957). Alternative two-way rating designs generate stimulus-response data or situation-scale data. Characteristic for the designs is that a subject has to judge to what extent a particular scale or variable pertains to a particular concept or situation. For instance, in a study by Murakami and Kroonenberg (2003), a student had to judge the characteristics of the 24 preludes of Chopin on a number of scales. As example, the student had to indicate whether a prelude of Chopin (concept) is tempestuous or tranquil (scale). Another example, which will be our guiding explanatory case, is the two-way design in which a person with a multiple personality in each personality was asked to judge on a number of scales to what extent a number of concepts pertained to her personal situation. For instance, to what extent she considered her doctor to be good or bad (Osgood

and Luria, 1954). The aim in their study was to see whether each personality (Eve White, Eve Black and Jane; each measured twice) used the scales in the same way to rate the concepts.

Yet another kind of two-way rating data results from a design in which for several situations the mean characteristics of groups rather than of individual subjects are described by means of a number of variables. For our detailed example we analyzed a collection of two-way data sets consisting of episodes by variables obtained from several different countries. The data were collected using the Strange Situation, a procedure within the attachment theory paradigm (Ainsworth et al., 1978) (see Section 3).

A two-way rating design seems comparable to multitrait-multimethod (MTMM) designs where the traits and the methods mostly form a fully-crossed design for the response variables. An important difference with the MTMM design is that the two-way rating design is more like a two-way (concept  $\times$  scale) analysis-of-variance design with the intensity or strength of the judgment by a personality as the response variable.

## 1.3. TWO-WAY RATING DESIGNS AND THREE-WAY DATA

Two-way rating designs produce three-way data because they consist of three ways, i.e., concepts, scales and groups or individuals. For a more detailed discussion of such *three-way rating data* arising from two-way rating designs see Kroonenberg (2008, Chapter 14). As far as we have been able to trace, there is no or hardly no explicit literature on the topic of factorial invariance for two-way rating designs, and with this paper we aim to fill this gap. In particular, our aim is to look for both a consensus structure about the relations between the concepts and scales (i.e., invariance over groups) and for group differences, i.e., deviations from invariance. Even though we will primarily focus on the situation with a limited number of groups or individuals, also larger numbers can be analyzed. The emphasis in the present paper is an exploratory one, even though the comparative evaluation of different aspects of factorial invariance using fit measures is a central concern. However, the sizes and relevance of these differences have to be evaluated subjectively both by comparing fit/degrees of freedom ratios and by looking at substantive relevance and interpretability. Formal statistical testing is not part of the procedure.

## 1.4. INVARIANCE IN TWO-WAY RATING DESIGNS

A problem for the invariance analysis of two-way rating designs is that there are often only a limited amount of judges or groups rather than large samples from a population so that there is no clear stochastic element in the data. The judges or groups need to be treated as another fixed factor in the analysis-of-variance sense, so that we really have a three-way design of concepts  $\times$  scales  $\times$  groups or concepts  $\times$  scales  $\times$  individuals. Even apart from the extremely small samples, this lack of stochastics in two-way rating designs makes using confirmatory factor analysis for testing invariance within the standard structural equation modeling context virtually impossible. Therefore, we propose to seek recourse to variants of component analysis, but it should be noted that the procedures discussed in this paper can handle large random samples as well.

Factorial invariance for two-way rating designs is cast here in a non-stochastic component framework in which we have

separate component spaces for the scales and the concepts. This has a disadvantage because components are generally not in themselves meaningful quantities but only maximum variance directions in the component space. What are invariant are the subspaces spanned by the components, rather than the components themselves. Therefore, we cannot automatically assume that the components themselves have intrinsic meaning like factors in confirmatory common factor analysis.

Only in some very specific models, such as the Parafac models which have unique solutions (see below), the components can validly be said to have intrinsic meaning. This will limit the kinds of invariances we can consider. Thus, generally we will have to discuss the invariance of subspaces across groups rather than the invariance of the components themselves. As already indicated in the introduction rather than refer to *subspace invariance* or *componential invariance*, we will use the standard term *factorial invariance*.

The two central questions in two-way rating designs are (1) how to define factorial invariance and (2) how to evaluate it. In contrast with the standard situation of assessing whether factorial invariance exists for a particular test across groups, in a two-way design one has to deal with the fact that *groups* or *individuals* use the rating scales to judge concepts. A definition of factorial invariance in this case must include three aspects of the data: (1) the component space or structure of the scales; (2) the component space or structure of the concepts and (3) the way the concepts (or the concept components) and the scales (or scale components) are linked for each group. The consideration of three different aspects of factorial invariance makes the situation for two-way rating designs fundamentally different from the standard situation. Both because of the design and the fact that we are dealing with component spaces rather than factors, makes that the Vandenberg and Lance steps have to be reformulated.

#### 1.4.1. Preprocessing

Variances of components in standard component analysis are represented by the eigenvalues. Whether they are actual variances or merely corrected or uncorrected sums of squares depends on the preprocessing, i.e., centering and normalization of the data. Standardization is more or less automatically carried out in regular component analysis but in two-way rating designs there are several options for preprocessing. Each option has different consequences for the data to be assessed for invariance, because it influences which part of the data is analyzed (see e.g., Kroonenberg, 2008, Chapter 6). To avoid such complications we will ignore the influence of preprocessing in this paper, and we will use the terms sums-of-squares and variances indiscriminately.

### 1.5. INVARIANCE HIERARCHY

When adapting the steps in the invariance hierarchy for two-way rating designs, we will assume from the start that we are attempting to approximate the centered data with lower-rank component spaces for the concepts and for the scales. This is in contrast with confirmatory factor analysis where covariance matrices are approximated.

Given the definition of a component, i.e., a linear combination of the original variables, any component is always present in a data set with the same variables given its coefficients; a property called *perfect congruence*; for a detailed discussion of this property see Ten Berge (1986a,b). What is generally different in different data sets with the same variables is the amount of variance explained by the components in each group. When it is not the full component space that is under consideration but only a limited number of (maximum variance) components, these group component spaces can be spanned by different linear combinations of the variables, so that component spaces of different groups may even be orthogonal to each other. The maximum variance components of one group, may account for very little variability in another group.

#### 1.5.1. Step 1. Lack of invariance

The most extreme form of lack of invariance is that each group has its own low-dimensional subspace. For two-way designs we take as our starting point the separate analyses of the group data without imposing any restrictions on the component subspaces other than considering a limited number of components, the same number for each group. The fitted sum of squares of the groups together, the *combined fit*, is calculated by summing their individual fitted sums of squares.

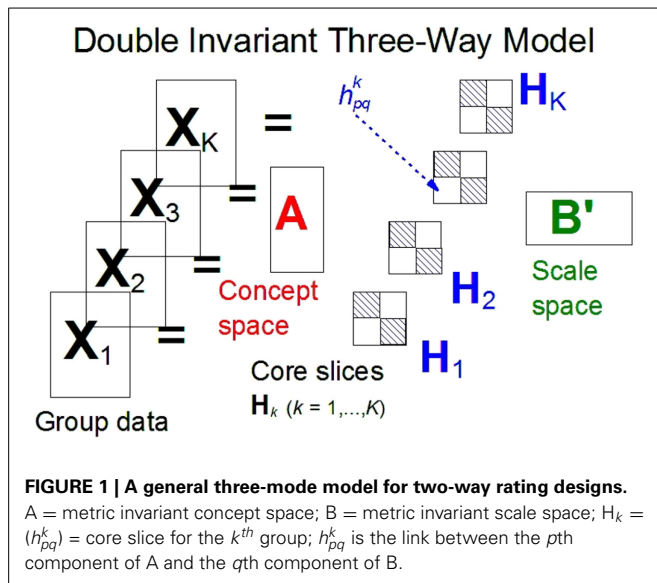
#### 1.5.2. Step 2. Configural invariance

Because every component returns in each data set with the same variables, i.e., components are always perfectly congruent across groups, configural invariance is not a limiting restriction in component analysis and is automatically true. Thus, it cannot be used as a limiting concept in a hierarchy of models, even though in different groups the same components may account for different amounts of variance and have different correlations.

#### 1.5.3. Step 3. Metric invariance

Of the models used to inspect factorial invariance, metric invariance is part of their definition. Thus, the component spaces (for the concepts and scales) specified in the models are such that the component coefficients are identical across groups. Three models can be used to investigate metric invariance. They have either (3a) an invariant concept component space, (3b) an invariant scale space or (3c) both. Metric invariance can be compared with a total lack of invariance by comparing the metric-invariant model fit with the *combined fit*. In addition, the metric-invariant space can be compared with the separate spaces of the groups, for instance via Procrustes techniques (see, for instance, Gower and Dijksterhuis, 2004); see also Section 4.

For the component models under consideration we will use the terms *links* and *interactions* to indicate the parameters which link the concepts and scales components. The links are contained in a so-called *core array*  $\mathbf{H}$  (see **Figure 1**). For each group this array contains a slice,  $\mathbf{H}_k$ , with the group's links between the components of the scales and the concepts. If both the concept and the scale space are orthogonal, the sizes of these links are the square roots of variation accounted for by the components. The invariance of the factor covariance matrices across groups translates into the equality of the core slices  $\mathbf{H}_k$  for  $k = 1, \dots, K$ .



#### 1.5.4. Step 4. Invariant component covariance matrices or core slices

As no common three-way models have restrictions on the variances without restrictions on the covariances, such models will not be discussed here; see Harshman and Lundy (1984) for detailed considerations about this issue. We will, however, consider (4a) models with invariant covariances (off-diagonal elements of the core slices) for all groups but with different variances (diagonal elements of the core slices). Even more restricted are models in which (4b) the invariant scale and/or concept components are uncorrelated in all groups.

#### 1.5.5. Step 5. (Weighted) strict invariance

The equality of the covariance matrices in Vandenberg and Lance's Step 5 translates into the equality of the centered data matrices of the groups. Such an equality implies equality of random errors which is of course nonsensical. However, a further tightening of the invariance in Step 4 is achieved in Step (5a) by restricting the slices of the core array to be identical, apart from a size coefficient (in the following referred to as a *weight*). Finally, the strictest factorial invariance situation is created in Step (5b) by specifying that also the weights are invariant across groups. In that case the structure of the scales and the concepts, as well as their linkages, are identical in all groups.

### 1.6. RELATED RESEARCH

Thus, for the two-way rating design the investigation of invariance is concentrated on the linkages between the invariant components for all groups. The discussion of the hierarchy of increasingly invariant three-mode models in this paper is strongly related to the hierarchy of three-mode models for fully-crossed raw data (Kiers, 1991). In addition, a similar hierarchy can be found in connection with simultaneous component analysis of covariance and correlation matrices (Timmerman and Kiers, 2003). However, in those papers the concept of factorial invariance is

not the focus of the investigation nor is the emphasis on two-way rating data.

## 2. MODELING FACTORIAL INVARIANCE

This section deals with three-way models for analysing data two-way rating designs. These models have as a common characteristic that the scale space and the concept space are invariant for all groups. However, they differ in the nature of the linkages between concept and space components. The models in Step 3a and 3b have *metric invariance* in one mode and all other models are characterized by *double-metric invariance*.

### 2.1. MODELS FOR TWO-WAY RATING DESIGNS

**Table 1** provides an overview of appropriate models, together with listing the nature of their invariances. To discuss these models in some detail we need some notation.  $A$  and  $B$  indicate the  $I \times P$  invariant concept space and the  $J \times Q$  invariant scale space, with  $P$  and  $Q$  the number of components, respectively. A subscript  $k$  indicates that a particular matrix belongs to the  $k$ th of  $K$  groups or levels of the third way; for instance,  $X_k$  is the concept  $\times$  scale data matrix of the  $k$ th group.  $H_k = (h^k_{ss})$  is the linkage matrix for the concept and the scale components for the  $k$ th group,  $D_k$  is a diagonal matrix of links used in the SVD as well as in the Parafac model. In the next section we will discuss these models in detail and indicate how they embody factorial invariance. As indicated in **Table 1** the Tucker2 model in principle allows for different numbers of components for the scales and the concepts, but as it is the only three-way model in **Table 1** for which this is the case, we will assume in the following that  $S = P = Q$ , i.e., that the numbers of components for the two spaces are the same throughout, so that  $A$  has size  $I \times S$  and  $B$  has size  $J \times S$ .

### 2.2. STEP 1: SINGULAR VALUE DECOMPOSITION PER GROUP

The singular value decomposition (SVD) is the motor of many multivariate techniques. For any  $X_k$  it may be written as:

$$X_k = A_k D_k B'_k + E_k = \hat{X}_k + E_k \quad k = 1, \dots, K \quad (1)$$

where for the SVD to have the form in Equation (1), the concept spaces  $A_k$  and scale spaces  $B_k$  have to have orthogonal components and the linkage matrices  $D_k$  have to be diagonal. The  $E_k$  contain the errors of approximation.  $\hat{X}_k = A_k D_k B'_k$ , and  $E_k = 0$  if all components are used. We will refer to the collection of independent analyses for each group as the *separate-analyses model* with abbreviations SVD\_2 and SVD\_3 for the two- and three-component models, respectively.

Thus, each data matrix  $X_k$  has its own decomposition as in Equation (1), and this decomposition is unrelated to that of any of the other data matrices. The total variance of a group  $k$  is equal to the sum of the squares of the singular values  $d^k_{ss}$  that make up the diagonal of  $D_k$  in the full decomposition, i.e.,  $SS(\text{Total})_k = \sum_k d^k_{ss}$ . Adding the  $SS(\text{Total})_k$  of the groups gives the total amount of variance of the groups indicated by  $SS(\text{Total})$ . In general, we will use only a limited number of components, here either 2 or 3. The components (columns) of  $A_k$  and  $B_k$  successively account for the largest amount of variance so that, given the dimensionality, the components for the concepts and those



**Table 1 | Models for two-way rating designs and their invariance.**

Model	Concepts	Scale	$P = Q?$	Interaction	Abbreviation
<b>STEP 1: LACK OF INVARIANCE</b>					
SVD per group	-	-	yes	no explicit invariance restrictions	SVD <sub>s</sub>
<b>STEP 3: METRIC INVARIANCE</b>					
Tucker1 - concepts invariant	x	-	no	concept space invariant; <i>single metric invariance</i>	T1A <sub>s</sub>
Tucker1 - scales invariant	-	x	no	scale space invariant <i>single metric invariance</i>	T1B <sub>s</sub>
Tucker2	x	x	no	concept and scale spaces invariant; <i>double-metric invariance</i>	T2 <sub>ss</sub>
<b>STEP 4: INVARIANT COMPONENT COVARIANCES</b>					
Parafac	x	x	yes	+ component covariances invariant; variances free	PFs
Parafac - Orthogonal	x	x	yes	+ component covariances invariant; variances free; components orthogonal for one or both ways	PFs_Orth
<b>STEP 5: (WEIGHTED) STRICT INVARIANCE</b>					
Tucker3 - Free	x	x	yes	metric invariance of orthogonal components variances invariant; group weights unrestricted	T3 <sub>ss1</sub>
Tucker3 - Fixed	x	x	yes	+ group weights fixed and constant	T3 <sub>ss1</sub> Fixed

$x =$  invariant;  $S-S =$  not invariant; SVD, Singular Value Decomposition;  $P =$  number of concept components;  $Q =$  number of scale components;  $s = 2$  or  $3$  number of components;  $ss1 =$  the first two ways have  $s$  components, the third way 1 component.

for the scales span the subspaces with the highest variance. Thus, we can use this variance accounted for,  $SS(\text{Fit})_{\text{separate}}$ , as an upper bound for the variance accounted for from any other model given the number of components. If the  $SS(\text{Fit})$  is the fit for a common model for all  $K$  groups, then if  $SS(\text{Fit})_{\text{model}} \simeq SS(\text{Fit})_{\text{separate}}$  the component space(s) are invariant. However, if there is a sizeable difference, the invariance restrictions on the common model are in doubt. We may also investigate *group invariance* by comparing the fitted variance of a particular group  $SS(\text{Fit})_k$  with the similar quantity calculated via the parameter estimates from one of the fitted models. Given the number of components, this will provide information on which groups fit well and which groups do not and are thus not invariant with respect to the other groups.

### 2.3. STEP 3A AND STEP 3B: SINGLE METRIC INVARIANCE - TUCKER1 MODELS

The first step into imposing restrictions on the solutions to investigate possible invariance is to demand that either the concept spaces can be properly represented by a single space (i.e., for all  $k$  the concept spaces are equal:  $\mathbf{A}_k = \mathbf{A}$ ), or that for all  $k$  the scale spaces are equal:  $\mathbf{B}_k = \mathbf{B}$  if there are  $s$  components. This can be investigated with the Tucker1 model, here referred to as Tucker1A (or T1A<sub>s</sub>) for concept space equality and Tucker1B (T1B<sub>s</sub>) for scale space equality. Metric invariance exists for the concepts if

$$\mathbf{X}_k = \mathbf{A}\mathbf{D}\mathbf{B}'_k + \mathbf{E}_k \quad k = 1, \dots, K. \quad (2)$$

Thus, there is a single orthogonal concept space for all  $k$  and separate scale spaces for each group. Metric invariance exists for the scales if

$$\mathbf{X}_k = \mathbf{A}_k\check{\mathbf{D}}\mathbf{B}' + \mathbf{E}_k \quad k = 1, \dots, K. \quad (3)$$

Thus, there is a single orthogonal scale space for all  $k$  and separate concept spaces for each group.

To compute the parameters, the three-way array is first converted to a two-way matrix of (Groups  $\times$  Scales) by Concepts or (Groups  $\times$  Concepts) by Scales, and these matrices are then subjected to a SVD. Note that the resulting  $\mathbf{A}_k$  and  $\mathbf{B}_k$  are no longer orthogonal because they are parts of a single orthogonal matrix of left and right singular vectors, respectively. We may compare the fitted variance of these models  $SS(\text{Fit})_{\text{model}}$  with the combined results of the separate SVDs,  $SS(\text{Fit})_{\text{separate}}$ , to investigate the metric invariance of either the concept or the scale spaces. However, it seems a bit odd to have an invariant concept space without having an invariant scale space, so we will not include the Tucker1A model further in our deliberations.

### 2.4. STEP 3C: DOUBLE-METRIC INVARIANCE - TUCKER2 MODEL

The next step in imposing invariance is to require *double-metric invariance*, i.e., for all  $k$  and given a number of components  $s$  both  $\mathbf{A}_k = \mathbf{A}$  and  $\mathbf{B}_k = \mathbf{B}$ , where both matrices orthogonal. Furthermore, the group linkage matrices  $\mathbf{H}_k$  are unrestricted and thus in general not diagonal. The model equation for the Tucker2 model (Tucker, 1972), as the model is commonly known (see Kroonenberg, 2008, Section 4.5.2) becomes

$$\mathbf{X}_k = \mathbf{A}\mathbf{H}_k\mathbf{B}' + \mathbf{E}_k \quad k = 1, \dots, K. \quad (4)$$

In other words, the metric invariance is present on both the concept space and the scale space, and the only differences between the groups can occur in the  $K$  interaction or linkage matrices,  $\mathbf{H}_k$ . The linkages matrices  $\mathbf{H}_k$  have sizes  $S \times S$ , where  $S$  is the number of components for both the scale and the concept spaces. An element  $h_{pq}^k$  of  $\mathbf{H}_k$  represents the link between the  $p$ th component of the concepts and the  $q$ th component of the scales for the  $k$ th group. So apart from their error terms, the variability between the groups lies in the strengths of their links between the concept and scale components or the sizes of the  $h_{pq}^k$ .

We can again compare the fitted variance of these models  $SS(\text{Fit})_{\text{model}}$  with the combined results of the separate SVDs,  $SS(\text{Fit})_{\text{separate}}$ , to investigate the double-metric invariance. Similarly we can make comparisons at group level.

## 2.5. STEP 4: DOUBLE-METRIC INVARIANCE WITH INVARIANT CORRELATIONS - PARAFAC MODEL

By requiring  $\mathbf{H}_k = \mathbf{C}_k$ , where the latter are diagonal matrices, and dropping the orthogonality restriction on the component spaces, we get the standard Parafac model with  $s$  components (PFs) which is a double-metric invariant model with as its model equation

$$\mathbf{X}_k = \mathbf{A}\mathbf{C}_k\mathbf{B}' + \mathbf{E}_k \quad k = 1, \dots, K. \quad (5)$$

The model can also be written by filling the rows of a  $K \times S$  matrix  $\tilde{\mathbf{C}}$  with the diagonals of the  $\mathbf{C}_k$ , i.e.,  $\tilde{c}_{ks} = c_{ss}^k$   $k = 1, \dots, K$ . In that case  $\tilde{\mathbf{C}}$  is considered a component matrix and is normalized like  $\mathbf{A}$  and  $\mathbf{B}$ , i.e., the lengths of the components in all three matrices are equal to one. The sizes of the  $S$  components are then contained in a diagonal matrix  $\mathbf{D} = (d_{ss})$ . However, for this paper we will stick with the  $\mathbf{C}_k$ .

Harshman (1970) that has shown this model implies that the groups have the same correlations between the components, which is a further imposition of factorial invariance. When at least one of the component matrices is orthogonal the  $d_{ss}^2$  are the variances of the  $S$  components.

One can even impose further restrictions on the components and so make the invariance even stricter by reintroducing orthonormality, non-negativity, or unimodality on one or both component matrices (see, e.g., Bro and Sidiropoulos, 1998).

Compared to other three-way models, Parafac models have a special characteristic in that their parameters are uniquely determined under rather mild conditions. This implies that the parameters in Equation (5) cannot be altered, for instance by rotation, without lowering the fit. The consequence is that the model has the *parallel proportional profile* property; (see Cattell and Cattell, 1955; Harshman, 1970; Harshman and Lundy, 1984). The only lack of invariance in these models consists of different strengths of the links between the concepts and scales, i.e., the  $c_{ss}^k$  vary between the groups. From the parallel proportional profile property and the uniqueness of the models it is the components themselves, not only the subspaces they span which are invariant; see Harshman (1970) or Harshman and Lundy (1984).

## 2.6. STEP 5: STRICTLY INVARIANT MODELS - TUCKER3 MODELS

To study factorial invariance with even more restrictions, we can demand that for each  $k$   $c_{ss}^k = c^k d_{ss}$ . In other words the weights for the components are invariant across groups apart from a group weight  $c^k$ .

$$\mathbf{X}_k = \mathbf{A}(c^k\mathbf{D})\mathbf{B}' + \mathbf{E}_k = c^k(\mathbf{A}\mathbf{D}\mathbf{B}') + \mathbf{E}_k \quad k = 1, \dots, K. \quad (6)$$

This model equals a simplified version of the full Tucker3 model (Tucker, 1966), and has been referred to as the *Replicated PCA* model by Van IJzendoorn and Kroonenberg (1990) and *Weighted PCA* by Krijnen and Kiers (1995). The only variable parts are the weights  $c^k$  for the group applicable to both components, and the

error terms  $\mathbf{E}_k$ . In other words, all groups have the same concept and scale spaces and the orthogonal components of each way are linked such that each concept component is linked exclusively to a particular scale component. The part between brackets has the form of a SVD valid for all groups. The only differences between the groups are their weights,  $c^k$ . This is in contrast with the Parafac model where each group has different link weights for the concept and scales component combinations, i.e., the  $c_{ss}^k$  are different for each group  $k$  and each pair of components  $s$ .

The ultimate invariant model is that in which we assume that all  $c_k$  are all equal with weight  $\bar{c} = \sqrt{1/K}$ , which is computationally equivalent to first averaging over groups and then carrying out a SVD on the average data matrix  $\bar{\mathbf{X}}$ , i.e.,

$$\mathbf{X}_k = \bar{c}(\mathbf{A}\mathbf{D}\mathbf{B}') + \mathbf{E}_k \quad k = 1, \dots, K. \quad (7)$$

Thus, in this case the only variable parts are the error terms and we may speak of *strict invariance*. We could reduce even further the number of parameters by specifying further restrictions on the concept and scale component spaces (see Takane et al., 1995), but this will not be considered here.

## 2.7. SUMMARY EVALUATING INVARIANCE

The conclusion from the above subsections is that one can define a hierarchy of models with an ever increasing number of parameters which are invariant over groups. By comparing the models with each other and with the combined separate analyses, it becomes possible to evaluate which models still provide an adequate fit to the data compared to separate analyses, and hence which type of invariance can be safely adopted. The two leading types of information for this purpose are the overall fitted variance and the fitted variance of each group.

In order to carry out model comparisons the number of parameters estimated for each of the models is determined. The models are compared by constructing a variant of the three-mode scree plot, in which the fitted sum of squares are plotted against the number of parameters estimated (see Section 3.3). Details on how to calculate the number of parameters can be found in Kroonenberg (2008, Section 8.4).

## 3. EXAMPLE: THE STRANGE SITUATION ACROSS THE WORLD

### 3.1. RESEARCH DESIGN

Attachment between adults, especially mothers, and infants is a lively research area—(see Cassidy and Shaver, 1999, 2008). Three types of bonds between adults and infants are generally considered: Avoidant attached, Securely-attached, and Resistant/Ambivalent attached, indicated by the letters A, B, and C, respectively. Here we will only look at attachment bonds with mothers, but those with other adults, especially other caregivers, have also been investigated (see, e.g., Sagi et al., 1985). The measurement procedure consists of a series of episodes of approximately 3 min, during each of which the infant is in a standardized room together with the mother (M), the stranger (S), both (MS), or alone (A); the episodes are the following: M1, MS2, S3, M4, A5, S6, M7. The idea is to increase the stress on the infant, especially by introducing the stranger and leaving the child alone, so that

the attachment relationship between mother and infant is put to the test. During the episodes, except when the infant is alone (A5), five core variables of an infant's reaction to an adult are measured: Proximity seeking, Contact maintaining, Avoidance, Resistance, and Distance interaction.

### 3.1.1. Strange situation data set

The data set under consideration consists of 11 samples: US-Belsky (USBel), US-Thompson (USTho), Germany-Berlin (GerBe), Germany-Bielefeld (GerBi), Israel-Kibbutz (IsrKi), Israel-City (IsrCi), Japan-Miyake (JapMi), Japan-Takahashi (JapTa), Netherlands-Younger infants (NLYng), Netherlands-Older infants (NLOld) and Sweden (Swed). The data set was put together by Sagi and Lewkowicz, and in their publication (Sagi and Lewkowicz, 1987) they supply full details of the origins of the different samples. For each of the samples the original investigators independently determined the infants' type of attachment. Earlier analyses can be found in Sagi and Lewkowicz (1987) and Kroonenberg and Van IJzendoorn (1987).

### 3.1.2. Invariance

The research question for this paper is whether the structure of the scales and that of the episodes, as well as the way these components are linked, are invariant across samples. The more parameters in the models are invariant, the more evidence this presents that the Strange Situation is a valid procedure across countries and researchers. For this example we only examine the average scores of the samples securely attached infants (B). These samples were chosen because each contained a sufficient number of B infants to make the average scores reliable. Thus, the two-way rating design consists of 7 episodes by 5 scales for 11 samples. This three-way data set was subjected to the models described above and their fit measures were compared.

## 3.2. RESULTS: THREE-WAY ANALYSIS OF VARIANCE

To acquire an initial perspective on the differences between samples, we carried out a three-way analysis of variance of the Strange Situation data. For this analysis the response variable was considered to be intensity of a reaction, and the Three-Ways were conceived as fixed factors in the ANOVA sense. This view is feasible because the samples are not exchangeable or drawn from a population. Moreover, it is the individual differences between the samples which are the focus of the analysis. Furthermore, the scales all had the same range from 1 to 7, so that averaging across scales is feasible and interpretable.

**Table 2** shows that the largest variability is between scales, indicating that the scale scores of the infant-mother dyads are effective in differentiating between behaviors across samples and episodes. On the other hand, the sample variability is comparatively very small (2.2% of the total), indicating that the investigating factorial invariance is a worthwhile exercise. This is confirmed by the size of the episode  $\times$  scale interaction compared to the interactions involving samples. Finally, the residuals (or the three-way interaction) only take up 7.5% of the total variability.

Parallel with standard component analysis, before the three-way analyses the data were centered but not normalized.

**Table 2 | Three-Way analysis of variance (with a single observation per cell).**

Source	SS	% SS(Total)	df	MS	F
<b>MAIN EFFECTS</b>					
Episodes	79.3	16.4%	6	13.2	88.1
Scales	211.4	43.6%	4	52.8	324.0
Samples	10.7	2.2%	10	1.1	6.5
<b>TWO-WAY INTERACTIONS</b>					
Episodes $\times$ Scales	105.2	21.7%	24	4.4	26.9
Episodes $\times$ Samples	9.9	2.0%	60	0.2	1.0
Scales $\times$ Samples	29.5	6.1%	40	0.7	4.5
<b>THREE-WAY INTERACTION</b>					
Residuals	39.1	8.1%	240	0.2	
<b>TOTAL</b>	<b>485.1</b>				

*df* = degrees of freedom; *MS* = Mean Sum of squares.

Normalization was not deemed necessary because all the scales had the same range. Moreover, scales with more variability should be allowed to have more influence on the analysis than scales with little variability.

With respect to centering, the common type of centering for three-way rating scale data (averaging across the concepts) was used, i.e.,  $\tilde{x}_{ijk} = (x_{ijk} - \bar{x}_{.jk})$ . In other words, the scale means for each sample  $k$  were removed. In general, centering across samples is undesirable because it will eliminate the consensus configuration of the scales and concepts from the three-way analysis. Thus, due to this type of centering the means of the scales for each of the samples were not included in the invariance analysis, but depending on the purpose of a study, these means can be analyzed for invariance separately.

## 3.3. RESULTS: INVESTIGATING TYPE OF INVARIANCE VIA MODEL FIT

Because the procedure outlined for assessing factorial invariance for two-way rating designs is an exploratory one, deciding on the degree of invariance is a substantive and subjective matter, of course based on numerical information. **Table 3** provides the information on the series of more and more restricted, and hence more invariant, models. Any additional restriction on the parameters is going to incur a certain amount of additional loss compared to the separate analyses. However, the question is whether the decrease in fit can be acceptable, given that by restricting the number of parameters interpretability is enhanced. It is less useful to compare the two-component models with the three-component models, because they have different starting points, i.e., different separate solutions. Therefore, it seems best to first decide on the number of components one wants to use to model the data, and only after that to investigate the invariance. This is incidentally also the standard practice in structural equation modeling. Of course, one may come to the conclusion that a two-component model is more, or less, invariant than a three-component model and vice versa.

In **Table 3** we see that the most restrictive models are the Tucker3 models with a constant component for the samples (T3-221Fixed and T3-331Fixed), i.e., the strictly invariant models. At the other extreme the individual three-component SVDs are not much use in terms of data reduction, because the model for

each sample has only three degrees of freedom, and the rank of the centered data matrices is at most four. From a data-analytic point of view, it is doubtful whether a model with unrestricted three-component solutions for the separate samples is really useful because the three components fit about 97% of the total variability.

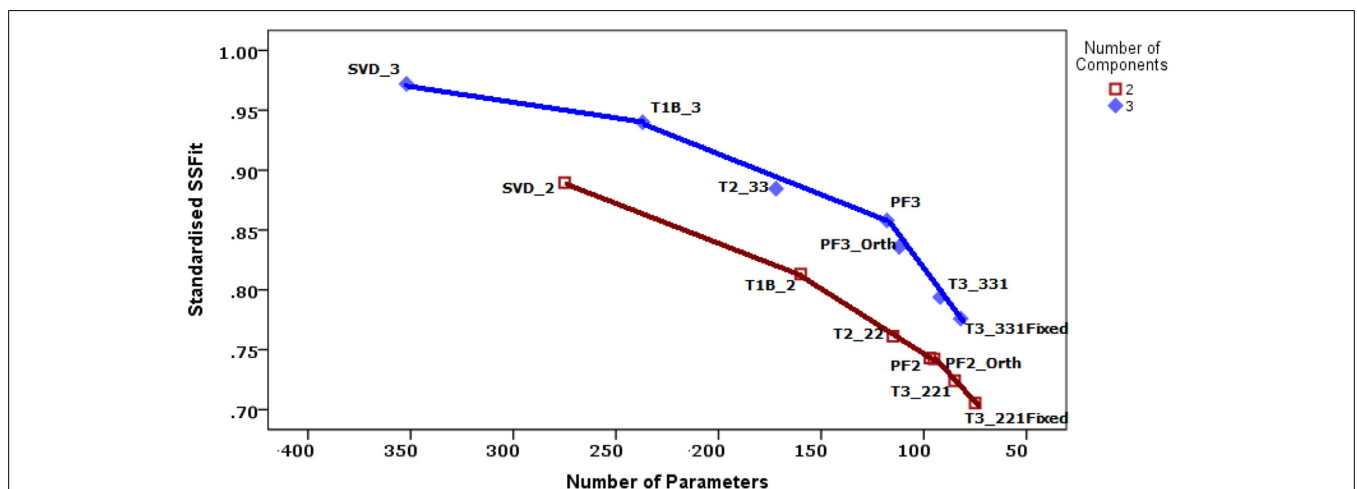
To decide upon the most appropriate model for these data, and thus on the extent of the invariance, it is useful to construct a variant of the three-mode deviance plot of the fitted sums of squares vs. the number of parameters (Figure 2); see Kroonenberg, 2008,

Section 8.5. The models with two components and those with three have been connected by part of a convex hull. Models on a convex hull are generally preferred to the models inside such a hull because of their more favorable  $SS(\text{Fit})/N_{\text{Params}}$  ratios. It is preferable to consider only models on or very close to the convex hull; the PF3-Orth model is less attractive because there are models with more favorable ratios (PF3 and T3-331) in the neighborhood. The more horizontal a hull, the more a model on the right is a good alternative for the models to the left on the hull, because the decrease in the number of parameters (i.e., increase

**Table 3 | Overall sums-of-squares for the Strange Situation data.**

Model	Abbreviation	SS(Fit)	SSS(Fit)	df	$N_{\text{Params}}$
<b>TWO-COMPONENT SOLUTIONS</b>					
Step 1	SVD per group (SVD_2)	205.78	0.89	110	275
Step 3	Tucker1 - Scales invariant (T1B_2)	187.94	0.81	227	158
	Tucker2 (T2_22)	177.78	0.76	270	115
Step 4	Parafac (PF2)	173.54	0.74	288	97
	Parafac - Orthogonal scale components (PF2_Orth)	173.36	0.74	290	95
Step 5	Tucker3 + Variable weights (T3_221)	169.18	0.72	302	83
	Tucker3 + Fixed weights (T3_221Fixed)	164.77	0.71	312	73
<b>THREE-COMPONENT SOLUTIONS</b>					
Step 1	SVD per group (SVD_3)	227.05	0.97	33	352
Step 3	Tucker1 - Scales invariant (T1B_3)	219.53	0.94	154	231
	Tucker2 (T2_33)	206.57	0.88	213	172
Step 4	Parafac (PF3)	200.40	0.86	267	118
	Parafac - Orthogonal scale components (PF3_Orth)	195.26	0.84	273	112
Step 5	Tucker3 + Variable weights (T3_331)	185.43	0.79	299	86
	Tucker3 + Fixed weights (T3_331Fixed)	181.23	0.78	309	76

$N_{\text{Params}}$  = Number of parameters (includes 55 removed means due to centering); The Total Sum of Squares of the centered data:  $SS(\text{Tot}) = 233$ ;  $SSS(\text{Fit}) = SS(\text{Fit})/SS(\text{Tot})$ ;  $df$  = degrees of freedom = number of data points ( $I \times J \times K = 385$ ) -  $N_{\text{Params}}$ .



**FIGURE 2 | Model comparisons.** The two-component and three-component models are connected by separate convex hulls. The horizontal axis is reversed because the investigation starts with the individual models. Legend: SVD\_s = separate SVDs with s components; T1B\_s = Tucker1 model with s components; T2\_ss = Tucker2 model

with s components for way 1 and 2; PFs = Parafac model with s components; PFs\_Orth = Parafac model with s orthogonal scale components; T3\_ss1 = Tucker3 model with s components for way 1 and 2 and 1 component for way 3; T3\_ss1Fixed = T3\_ss1 with a fixed value for way 3.



in the  $df$ ) does not seriously decrease the fitted sum of squares. In contrast, the steeper the hull turns downward for the next model to the right, the less attractive the model, because there is a large loss in fitted sum-of-squares for only a limited decrease in parameters. Note that a smaller number of parameters increases power and potentially simplifies interpretability.

For the Strange Situation data we see in **Figure 2** that for the three-component models the convex hull declines slowly at first, and a steeper downturn is observed only for the Tucker3 models, so that the Parafac model with three components seems a good choice. The choice for a two-component model is less clear. The relationship between the  $SS(\text{Fit})$  and the number of parameters is nearly linear. Again the Parafac models (PF2 and PF2-Orth) seem to be the best choice, and even though the orthogonal variant is marginally better, we decided to opt for the regular Parafac model. With respect to factorial invariance, the Parafac models incorporate invariant concept and scale spaces, and the correlations between scale components are constant over samples. The appropriateness of the Parafac models suggests that there is a considerable double-metric factorial invariance across the samples, only the size of the variances is different.

### 3.4. RESULTS: NON-INVARIANT SAMPLES

For three-way models with double-metric invariance which are not necessarily invariant with respect to their links, we can compute the model fit for each sample. These fit measures can then be compared with the separate-analyses model to determine whether overall lack of interaction invariance is due to specific samples or whether differences are present between all samples.

#### 3.4.1. Differences in proportional fit of samples.

For selected two-component models we calculated the proportional residual sums-of-squares  $\text{PrSS}(\text{Res}_k)$  for each sample and connected these values per model in **Figure 3**. In the figure we have arranged the samples such that the lack of fit is increasing for the two-component Parafac model.

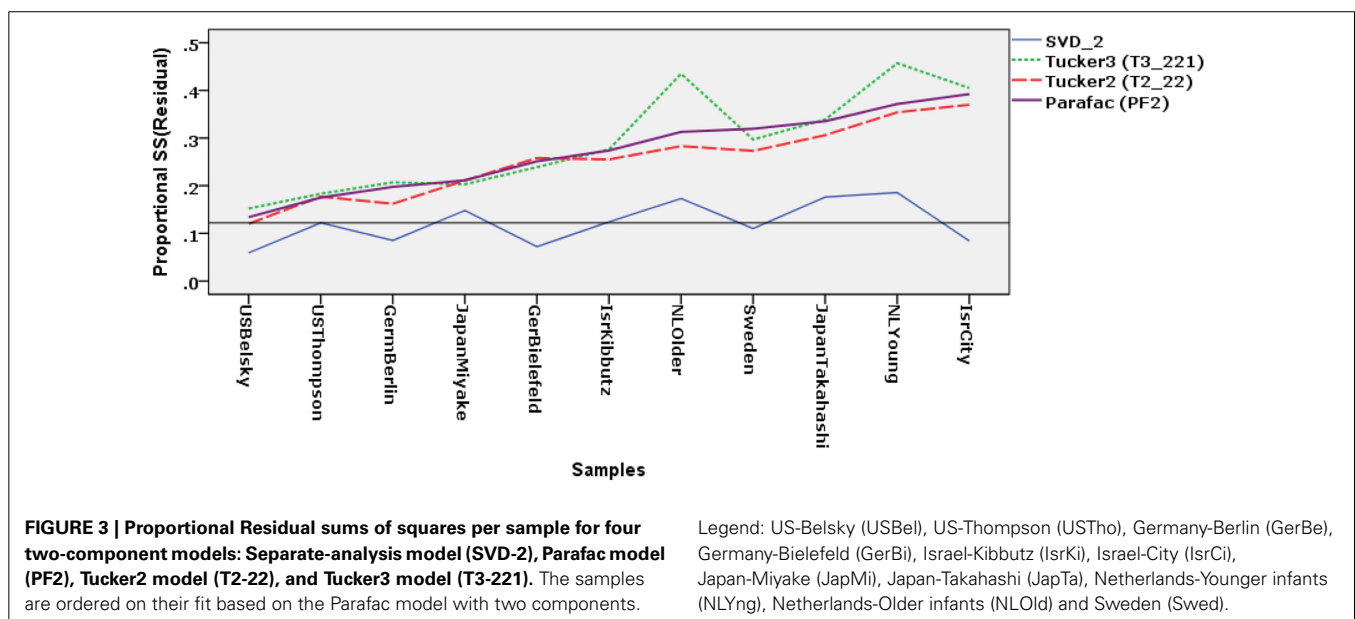
The solid line for the  $\text{PrSS}(\text{Res})$  represent SVDs of the separate samples. We see that their  $\text{PrSS}(\text{Res})$  fluctuate around the average value drawn as a horizontal line. In other words, a two-component SVD have about the same fitted sums of squares in all samples, but their concept spaces and their scale spaces are not necessarily equal.

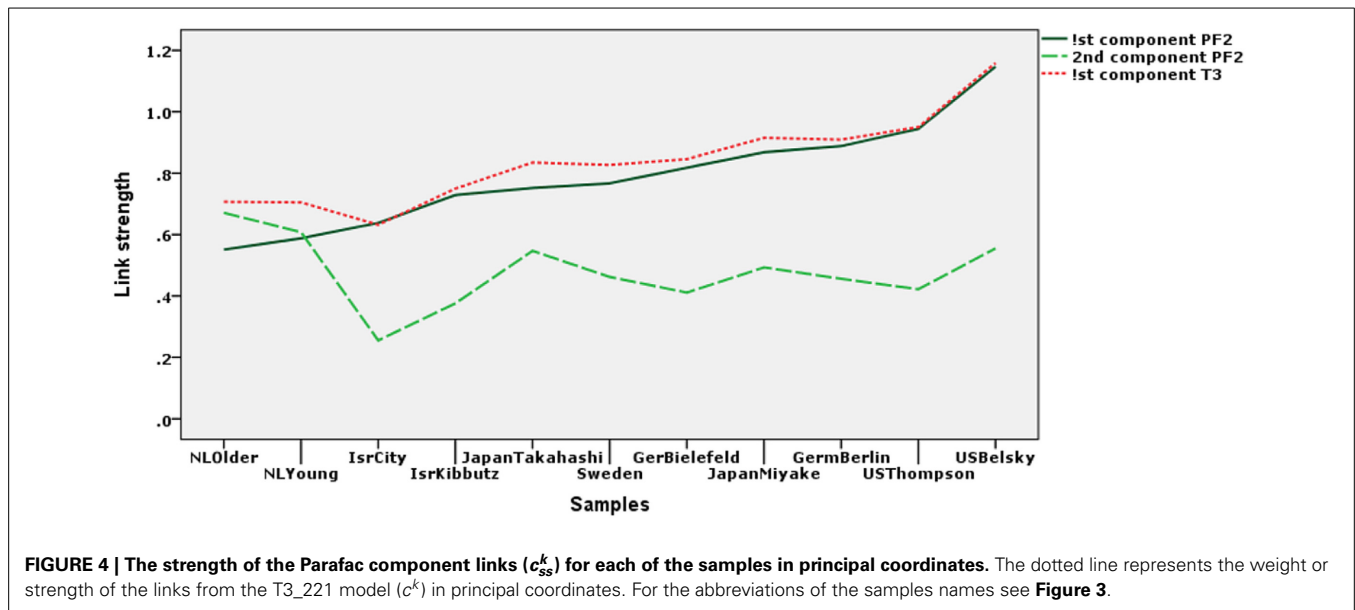
In the case of strict model invariance all lines would be more or less horizontal because the lack of fit would be equal for all samples. This is not the case here. The relative difference in fit varies between the solutions for the separate samples and those of the models displayed in the figure. Thus, for the US samples on the left-hand side of the figure the metric invariant subspaces for the concepts and the scales are more alike to their own separate spaces than to the subspaces for the younger Dutch sample and Israel-City sample on the right-hand side. In particular, the  $\text{PrSS}(\text{Res})$  for the two US samples is around 0.10 while it is around 0.30 for the younger Dutch and the Israel City sample.

All three metric invariant models displayed in **Figure 3** show more or less the same pattern with an increasing loss of fit from left to right. Given that the models are more or less equivalent, we may choose to interpret the most restricted and thus most invariant model, i.e., the T3-221 or PF2 models. **Figure 3** shows that the most right-hand samples fit marginally better, which is consistent with our earlier choice for this model. The Parafac model allows the components per sample to have a common oblique orientation with separate weights ( $c_{ss}^k$ ) for the links between these common components. In this data set the younger Dutch sample and the Israel City sample need further investigation, because it is their configurations that are deviating most from the common pattern.

#### 3.4.2. Differences in strengths of links between concept and scale spaces across samples

In **Figure 4** we have plotted the link strengths  $c_{ss}^k$  between the concept and scale components from the Parafac model with two components. The solid line represents the strengths of the





links between the first components,  $c^k_{11}$  and the dotted line the strength of the links for the second components,  $c^k_{22}$ . To provide a proper comparison these parameters have been depicted in principal coordinates. The third, dashed, line represents the weight parameter for each group according to the T3-221 model,  $c^k$ ; also in principal coordinates. The samples have been ordered so that the values for the first components,  $c^k_{11}$ , are increasing monotonically. The figure shows that  $c^k_{11}$  and the  $c^k$  are almost equal, but that there is a small compensation of the  $c^k$  for the absence of the links for the second components  $c^k_{22}$ . Thus, the choice between the models should take into account whether the fluctuations of the  $c^k_{22}$  are interpretable. At the same time the differences in the  $c^k_{22}$  point to where we should look for lack of invariance.

If we want to find out what exactly are the differences between the samples, we have to explicitly compare the invariant concept and scale spaces with the separate sample spaces. Thus, this analysis could be extended to find the causes of the differences by examining the Tucker1 model for scales (T1B), and possibly the Tucker1 model for concepts (T1A), to assess whether it is the scale space or the concept space which is not invariant. We will not pursue this here. The procedure described above should primarily be seen as a proof of concept, rather than a detailed analysis of a particular case (see, however, the Appendix for a more substantive interpretation).

#### 4. RESULTS: AN ADDITIONAL APPROACH TOWARD ASSESSING INVARIANCE

In a paper comparing Japanese and Australian children in the way they show respect to adults, Kroonenberg and Kashima (1997) tackled assessing invariance in a different way, even if they did not explicitly refer to factorial invariance. The children were given a questionnaire in which they had to indicate both to what extent they did *show* a number of respectful behaviors (greet, help, stick up for, etc.) toward a number of adults (father, mother, teacher, etc.), and to what extent they felt they *should* do so. This resulted

in a 5 (adults)  $\times$  7 (behaviors)  $\times$  4 (groups; Australian do, Australian should, Japanese do, Japanese should) three-way data set. Apart from a complete three-way analysis, the invariance was also assessed by first carrying out separate analyses for each of the four groups, and then using the adult space and/or the behavior space of one group as a restriction for the solution of another group. Essentially, of course, this is a cross-validating procedure, checking to what extent the parameter estimates in one group can also explain the variability in another group, or to what extent the two groups had invariant subspaces. However, one may equally see this as a procedure for establishing invariance. This procedure was referred to as *external analysis* by Van der Kloot and Kroonenberg (1985), because externally determined values for the parameters were used in fitting a particular data set.

For the Strange Situation data, this procedure could be used to investigate to what extent the separate solution of a sample is similar to that of another sample. In particular, the nature of the difference of the Dutch sample with respect to the other samples could be a focus of further analysis.

#### 5. CONCLUSION

In this paper we have presented an approach toward assessing factorial invariance in two-way rating designs such as stimulus-response and semantic differential designs. Such designs generate fully-crossed three-way data which can be analyzed by three-way component models. True three-way models like the Parafac and Tucker models and their variants already incorporate various aspects of factorial invariance, in particular the double-metric invariance of the concept and scale spaces. The models vary in how they treat the relationships or links between the components. A hierarchy of models with increasing factorial invariance is outlined, running from no invariance for separate SVDs for each group, via single metric invariance for Tucker1 models, double-metric invariance for Tucker2 models, double-metric invariance and correlational invariance of Parafac models, to strict invariance for a very restricted Tucker3 model.

These models, and hence the nature of the invariance, can be assessed and compared via deviance plots showing the sum of squares of fit against the degrees of freedom. By connecting the relevant models by convex hulls in the plot, a comparative evaluation can be made and an appropriate model can be selected. Moreover, information supplied by the three-way analysis can be used to assess which group is more deviant from the invariant solution, and what the nature of such differences are.

The descriptive approach toward model selection, rather than using a formal testing paradigm, has been shown to work well for the example presented here. Data from a multinational collection of Strange Situation sessions (Sagi and Lewkowicz, 1987) were analyzed to demonstrate the effectiveness and usefulness of the model hierarchy for two-way rating data.

By investigating data from two-way rating designs we have extended the concept of factorial invariances beyond its standard definition. The future will have to show to what extent this extension is going to make an impact on the research on factorial invariance. For the present it seems that using the conceptualization presented here and the proposed hierarchy of three-way models, can shed light on differences and similarities between the invariance in two-way rating designs.

## ACKNOWLEDGMENTS

Our thanks go to Avi Sagi, Kathleen Lewkowicz, and Marinus van IJzendoorn for making the international Strange Situation data available for reanalysis.

## REFERENCES

- Ainsworth, M. D. S., Blehar, M. C., Waters, E., and Wall, S. (1978). *Patterns of Attachment. A Psychological Study of the Strange Situation*. Hillsdale, NJ: Erlbaum.
- Bro, R., and Sidiropoulos, N. D. (1998). Least squares algorithms under unimodality and non-negativity constraints. *J. Chemometr.* 12, 223–247. doi: 10.1002/(SICI)1099-128X(199807/08)12:4<223::AID-CEM511>3.0.CO;2-2
- Cassidy, J., and Shaver, P. R., (eds.). (1999). *Handbook of Attachment. Theory, Research, and Clinical Applications*. New York, NY: The Guilford Press.
- Cassidy, J., and Shaver, P. R., (eds.). (2008). *Handbook of Attachment, 2nd Edn. Theory, Research, and Clinical Applications*. New York, NY: The Guilford Press.
- Cattell, R. B., and Cattell, A. K. S. (1955). Factor rotation for proportional profiles: analytical solution and an example. *Br. J. Stat. Psychol.* 8, 83–92. doi: 10.1111/j.2044-8317.1955.tb00323.x
- Gower, J. C., and Dijksterhuis, G. B. (2004). *Procrustes Problems*. New York, NY: Oxford University Press. doi: 10.1093/acprof:oso/9780198510581.001.0001
- Harshman, R. A. (1970). Foundations of the PARAFAC procedure: models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Work. Pap. Phon.* 16, 1–84.
- Harshman, R. A., and Lundy, M. E. (1984). “The PARAFAC model for three-way factor analysis and multidimensional scaling,” in *Research Methods for Multimode Data Analysis*, eds H. G. Law, C. W. Snyder Jr., J. A. Hattie, and R. P. McDonald (New York, NY: Praeger), 122–215.
- Kiers, H. A. L. (1991). Hierarchical relations among three-way methods. *Psychometrika* 56, 449–470. doi: 10.1007/BF02294485
- Krijnen, W. P., and Kiers, H. A. L. (1995). An efficient algorithm for weighted PCA. *Comput. Stat.* 10, 299–306.
- Kroonenberg, P. M. (2008). *Applied Multiway Data Analysis*. Hoboken NJ: Wiley. doi: 10.1002/9780470238004
- Kroonenberg, P. M., and Kashima, Y. (1997). Rules in context. A three-mode principal component analysis of Mann et al.’s data on cross-cultural differences in respect for others. *J. Cross Cult. Psychol.* 28, 463–480. doi: 10.1177/0022022197284005
- Kroonenberg, P. M., and Van IJzendoorn, M. H. (1987). “Exploring children’s behavior in the Strange Situation,” in *Attachment in Social Networks. Contributions to the Bowlby-Ainsworth Attachment Theory*, eds L. W. C. Tavecchio and M. H. Van IJzendoorn (Amsterdam: Elsevier Science), 379–425. doi: 10.1016/S0166-4115(08)61080-8
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58, 525–543. doi: 10.1007/BF02294825
- Murakami, T., and Kroonenberg, P. M. (2003). Individual differences in semantic differential studies and their analysis by three-mode models. *Multivariate Behav. Res.* 38, 87–96. doi: 10.1207/S15327906MBR3802\_5
- Osgood, C. E., and Luria, Z. (1954). A blind analysis of a case of multiple personality. *J. Abnorm. Soc. Psychol.* 49, 579–591. doi: 10.1037/h0054362
- Osgood, C. E., Suci, G., and Tannenbaum, P. (1957). *The Measurement Of Meaning*. Urbana, IL: University of Illinois Press.
- Sagi, A., Lamb, M., Lewkowicz, K., Shoham, R., Dvir, R., and Estes, D. (1985). Security of infant-mother, -father and -metapelet attachments among kibbutz-reared Israeli children. *Monogr. Soc. Res. Child Dev.* 50, 257–275. doi: 10.2307/3333837
- Sagi, A., and Lewkowicz, K. S. (1987). “A cross-cultural evaluation of attachment research,” in *Attachment in Social Networks. Contributions to the Bowlby-Ainsworth Attachment Theory*, eds L. W. C. Tavecchio and M. H. van IJzendoorn (Amsterdam: Elsevier Science), 427–459. doi: 10.1016/S0166-4115(08)61081-X
- Takane, Y., Kiers, H. A. L., and De Leeuw, J. (1995). Component analysis with different sets of constraints on different dimensions. *Psychometrika* 60, 259–280. doi: 10.1007/BF02301416
- Ten Berge, J. M. F. (1986a). Some relationships between descriptive comparisons of components from different studies. *Multivariate Behav. Res.* 21, 29–40. doi: 10.1207/s15327906mbr2101\_2
- Ten Berge, J. M. F. (1986b). Rotation to perfect congruence and the cross-validation of component weights across populations. *Multivariate Behav. Res.* 21, 41–64. doi: 10.1207/s15327906mbr2101\_3
- Timmerman, M., and Kiers, H. A. L. (2003). Four simultaneous component models for the analysis of multivariate time series from more than one subject to model intraindividual and interindividual differences. *Psychometrika* 68, 105–121. doi: 10.1007/BF02296656
- Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika* 31, 279–311. doi: 10.1007/BF02289464
- Tucker, L. R. (1972). Relations between multidimensional scaling and three-mode factor analysis. *Psychometrika* 37, 3–27. doi: 10.1007/BF02291410
- Van der Kloot, W. A., and Kroonenberg, P. M. (1985). External analysis with three-mode principal component models. *Psychometrika* 50, 479–494. doi: 10.1007/BF02296265
- Van IJzendoorn, M. H., and Kroonenberg, P. M. (1990). Cross-cultural consistency of coding the Strange Situation. *Infant Behav. Dev.* 13, 469–485. doi: 10.1016/0163-6383(90)90017-3
- Vandenberg, R. J., and Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. *Organ. Res. Methods* 3, 4–70. doi: 10.1177/109442810031002
- Widaman, K. F. (2007). Common factors versus components, principals and principles, errors and misconceptions, in *Factor Analysis at 100. Historical Developments and Future Directions* eds R. Cudeck and R. C. MacCallum (Mahwah, NJ: Lawrence Erlbaum), 177–203.

**Conflict of Interest Statement:** The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 28 February 2014; accepted: 04 December 2014; published online: 08 January 2015.

Citation: Kroonenberg PM (2015) Assessing factorial invariance of two-way rating designs using three-way methods. *Front. Psychol.* 5:1495. doi: 10.3389/fpsyg.2014.01495

This article was submitted to *Quantitative Psychology and Measurement*, a section of the journal *Frontiers in Psychology*.

Copyright © 2015 Kroonenberg. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

**APPENDIX**

This appendix is presented here to offer some idea of the substantive outcomes of the invariance analysis of the Strange Situation.

For the Parafac model with two components, **Figure A1** shows the normalized components of the three modes in two panels, both for the first and the second component as well as the strength of the links between them.

If there was complete strict invariance, the samples would have been superimposed in both panels at the value  $(\bar{c}d_{ss})$  (the T3-221Fixed model). If there would have been weighted strict invariance, the rank order and spacing of the samples for each of the components would have been equal i.e., at the values  $c^k d_{ss}$  (the T3-221 model). As the figure shows, neither of these options was realized in the present data set, so that we must conclude that a double-metric invariant model (PF2) is the most restricted or invariant model that can be obtained.

The variances (or link strengths) of the components are  $d_{11} = 11.5$  and  $d_{22} = 6.4$ , respectively (see Equation 5), so that the ratio of their importance in reconstructing the model is 1.8. Thus, the differences between the samples with respect to link strengths of the first components are about twice as large as those for the second components.

**A.1. FIRST COMPONENT**

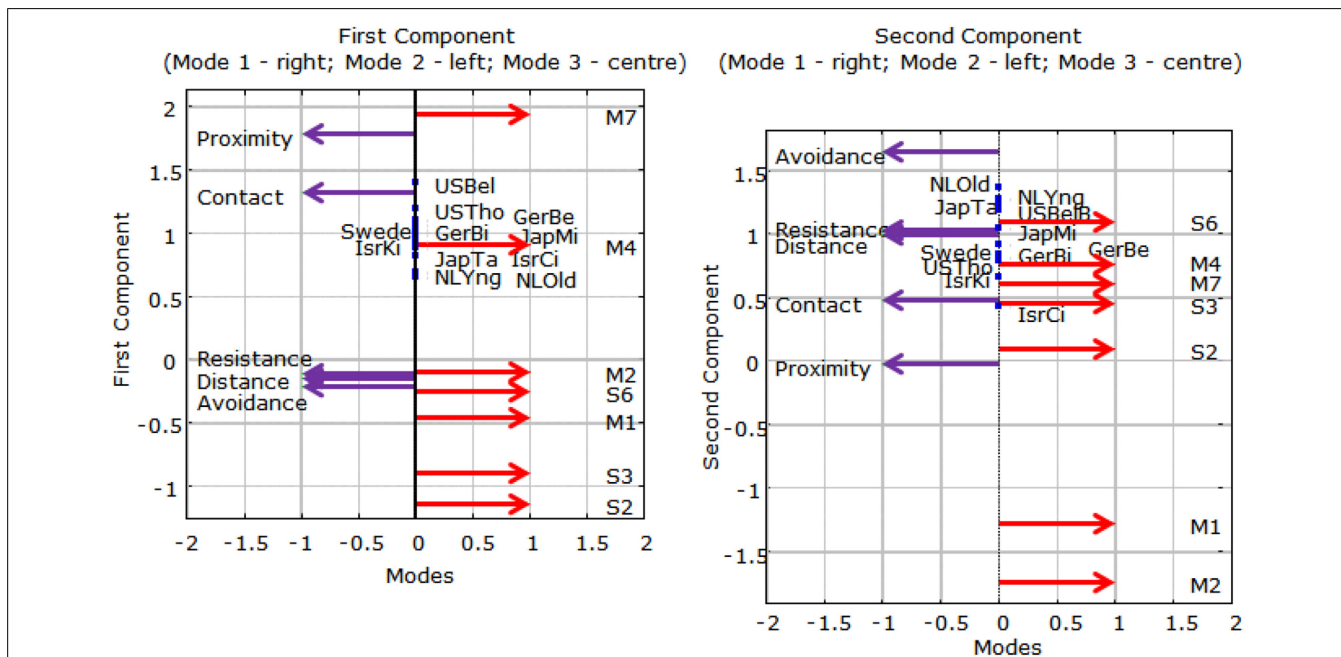
The left-hand panel of **Figure A1** shows that securely attached (B) children show increasing Proximity seeking and Contact maintaining during the Mother episodes of the Strange Situation, as is evident from the increasingly higher coefficients on the first component. Seeking closeness to the mother is indicative of increasing stress during the procedure, which the B children try to alleviate by showing more and more proximity to the mother,

i.e., showing a stronger secure attachment behavior. Treating the stranger with suspicion by staying at a distance is evident in the Stranger episodes; the coefficients remain negative but less so in S6 than in S3. Children’s suspicion is decreasing slightly during the procedure but it is never absent. The other three scales all hover around zero, indicating that these behaviors of the children are not related to the two behaviors mentioned first. The US securely-attached children show the described patterns to the largest extent and the Dutch children the least. It is interesting to see that samples from the same country are generally close together with the largest difference between the two Israeli samples.

**A.2. SECOND COMPONENT**

The second components in the right-hand panel of **Figure A1** describe mainly the avoidance, resistance and distance interaction behaviors toward the stranger, or stranger wariness. Such behavior is not typically present in the first two episodes but is present to a limited extent in the other episodes except for Episode 6, when it is the Stranger who returns rather than the Mother after the child has been alone in the fifth episode. With respect to the mother the situation is more complicated. There is a clear contrast between the earlier and later episodes, in that negative behavior toward the mother is not present in the beginning, but the children show a certain reserve when mother and child are reunited after the child has been alone with the stranger (Episodes M4 and M7).

These patterns are strongest in the Dutch and Japanese samples, as well as Belsky’s US sample. Again, samples from the same country are generally close together except for the US samples.



**FIGURE A1 | Two-component Parafac analysis: Invariant components plus link strength of samples.** *Mi* = Mother episode *i*; *Si* = Stranger episode *i*. For sample abbreviations, see **Figure 3**.



### A.3. INVARIANCE CONCLUSION WITH RESPECT TO THE B-CHILDREN

The original research question was whether the structure behind the data from the two-way rating design for the secure B-children was the same across samples. The double-metric invariance embodied in the well-fitting two-component Parafac model indicates that this is indeed the case. However, the samples show a lack of invariance with respect to importance of the linkage between the components of the episodes and the scales. This difference is primarily a matter of relative importance of the two components. The first component embodies the increasingly secure

attachment behaviors (Proximity seeking and Contact maintaining) over episodes, which is stronger in the US samples, especially in contrast with the Dutch samples. The second component represents 'stranger wariness' (Avoidance, Resistance and Distance interaction) which is especially strong in Episode 6, when the stranger rather than the mother returns, after the child as been alone in the fifth episode. To a lesser degree it also represents reservation toward the mother in the reunion episodes. These patterns are especially strong in the Dutch and Japanese samples and Belsky's US one.